

Implementing and innovating:
Local governments in the development of China's New
Cooperative Medical Scheme

Lewis Husain

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Modern Languages and Cultures

October 2012

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Lewis Husain to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© 2012 The University of Leeds and Lewis Husain

Acknowledgements

This research has benefited from enormous support from a great number of people and institutions. Most important were the interviewees in counties and provinces in which I carried out fieldwork. Their generosity in giving up their time to explain their work and the functioning of their local health systems to an outsider was what made this research possible.

I wish to thank the Central Compilation and Translation Bureau in Beijing, particularly Professor He Zengke, Yan Jian, Li Yitian and Long Ningli for great help and support. My thanks also go to Yang Xuedong, Ding Kaijie, Lai Hairong and Chen Xuelian for valuable discussions. Thanks also go to many people who were generous with their time, thoughts and help: Professor Chen, Pei Xiaomei, Shi Guang, Fang Lijie, Cao Zhenghan, and participants at a 2011 workshop on rural health reforms at the Chinese Academy of Social Sciences. Thanks to Li Ling, Li Jinshan and David Goodman, who all provided encouragement at important moments. Particular thanks go to Gerald Bloom for thoughts, guidance and help along the way. Many thanks to Chen Lin at Beijing Normal University for alerting me to the resources available in the National Library. In Leeds, my thanks go to Jenni Rauch for continual help and constant professionalism. Many thanks to my supervisors: to Flemming Christiansen, for critical engagement and support, and to Heather Zhang.

This work would not have been possible without continual support of friends and family. In Leeds, huge thanks to Caroline Gunn (née Fielder), Jane Caple and Chris Bond for friendship, support, coffee and long walks. Finally, immeasurable thanks to my wife Vivian for almost endless patience, to my parents and sister, and to my Chinese family.

This research was made possible by financial support from the Economic and Social Research Council. All errors, omissions and inaccuracies are entirely my own.

Abstract

From 2002-2003, development of China's New Cooperative Medical Scheme (NCMS), a rural health insurance programme, was initiated based on an 'experimental' process of piloting and roll-out. The NCMS has developed rapidly since then: from limited coverage and funding, it has been extended nationally and funding has increased dramatically. Most analyses of the scheme focus on its impact on users' health seeking and spending, paying less attention to the policy and its development.

Much literature on central-local relations in China foregrounds questions of power and the centre's ability to enforce sub-national policy implementation. The NCMS, however, shows a policy principally run at the county level, under which counties have a responsibility both to implement the scheme and to develop workable local policy within a loose overall national policy framework. This gives a degree of freedom, or discretion, in operation of the scheme.

This study argues against seeing localities as simple implementers of pre-cast central policy, and argues for supplementing this with an understanding of the role of counties as frontline interpreters and developers of policy, and as innovators within supra-local policy frameworks. It examines the structuring of scheme implementation alongside ways counties operate within the overall NCMS framework, the degree of discretion they have, and the possibility and importance of local generation of policy, policy mechanisms and models. This gives a view of local practices and production of institutions on the periphery of the state policy making apparatus, where local diversity and implementation often run ahead of central policy. Based on county- and province-level fieldwork, this study examines the origins and systemic basis of selected county reforms, their systemic relevance and impact, and shows local practices as part of a loosely structured 'conversation' in which multiple levels of government play differentiated roles in a complex and ongoing process of reform.

Table of contents

Acknowledgements	3
Abstract	4
Table of contents	5
Acronyms and note on language conventions	14
Chapter One. Understanding local government innovation in China's healthcare reform: The case of the NCMS	15
1.1 Outline and approach	17
1.2 Framing of the study and research questions	20
1.3 Contributions to existing research	22
1.4 Structure of the thesis	23
1.5 Background to the NCMS	25
Chapter Two. Control and discretion in China's central-local relations and policy making	30
2.1 Introduction to theoretical chapters	30
2.2 Introduction to literature review	31
2.3 Approaches to central-local relations in the reform era	32
2.4 Pole 1: Declining central control	36
2.5 Pole 2: Decline in central control, but integrative mechanisms remain	40
2.6 Pole 3: More than zero sum and compliance	42
2.6.1 Pole 3a: Non-zero-sum approaches	43
2.6.2 Pole 3b: Discretion and differential implementation	46

2.6.3 Pole 3c: Experimental policy development and policy innovation	51
2.6.4 Pole 3d: New departures	59
2.7. Attempting synthesis	61
2.8 Looking forward	65
Chapter Three. Methodology: Looking for innovation	71
3.1 Introduction	71
3.2 Looking for innovation	72
3.2.1 Selecting cases: Looking for 'innovation'	72
3.2.2 Policy discourse and case selection	73
3.2.3 Profiling	74
3.2.4 Iterative research process	75
3.3 Implications for fieldwork	76
3.3.1 Implications for fieldwork: Practical	76
3.3.2 Implications for fieldwork: Conceptual	78
3.3.3 Are these cases outliers?	80
3.4 Hosting, interviews and documents	82
3.4.1 Access to research sites and interviewees	82
3.4.2 Interviewing	82
3.4.3 Trust, interviewing and completeness of data	83
3.4.4 Research ethics	84
3.4.5 Documentary sources	85

Chapter Four. Counties: Implementers and innovators in national policy schemes, part one	87
4.1 Introduction	87
4.2 Implementation and innovation	88
4.2.1 Compliance and 'syllogistic implementation'	88
4.2.2 Rhetorical attachment to decentralised initiative	89
4.2.3 Understandings and misunderstandings of innovation	92
4.2.4 The importance of local initiative: Two readings	93
4.3 Innovation within established policy frameworks	94
4.3.1 Freedom by design	96
4.3.2 Approaches to the cadre responsibility system	96
4.3.3 Approaches to policy texts and language	100
4.3.4 Target-setting and policy: Interim synthesis	101
4.4 Decentred practices and the necessity of tolerance	102
4.4.1 Models of the periphery: Expect a range of outcomes!	102
4.4.2 Tolerance and legitimacy of local practices	104
4.5 Interim summary and looking forward	104
4.5.1 Interim summary	104
4.5.2 Looking forward	106
Chapter Five. The NCMS: Between central and local policy	107
5.1 NCMS as central policy: Outline of the scheme	108
5.2 NCMS as central policy	110

5.2.1 NCMS as central policy: Implementing the scheme	110
5.2.2 Implementing the NCMS: Outline of the national policy framework	110
5.2.3 Target-driven development: NCMS as pressurised system	112
5.3 NCMS as central policy: NCMS as financing mechanism	114
5.3.1 NCMS funds: 'Fund risk' as a category of analysis	114
5.3.2 NCMS funds: Guiding principles and the fund transfer system	116
5.3.3 'Closed accounts': The local fund management system	119
5.3.4 Local scheme balancing: Enforcing a cellular structure	123
5.3.5 Interim summary	126
5.4 Local actors within the central policy framework	127
5.4.1 Counties as scheme managers	127
5.4.2 Scheme enrolment: Central targets and local responsiveness	128
5.4.3 Central targets and local responsiveness: NCMS as a 'new mode of governance'?	130
5.4.4 Tight budgets: Background to local scheme design	131
5.4.5 Local scheme design: Local setting of reimbursement plans	134
5.4.6 Interim summary	137
5.5 The developmental role of counties: Mechanisms and the importance of local initiative	138
5.5.1 NCMS as 'pilot'; policy principles and local practice	138
5.5.2 Central and local roles	139
5.5.3 The importance of policy 'mechanisms'	140

5.5.4 Pressurised system and local discretion	142
5.6 Caveats and complexity: Difficulties of treating the NCMS as a national system	143
5.7 The NCMS as an ongoing project	147
5.8 Interim summary	149
Chapter Six. Risk and innovation: Taoshan’s NCMS Oversight Bureau	152
6.1 Introduction	152
6.2 Background	153
6.2.1 NCMS management and oversight	153
6.2.2 Funding constraints and limits to work of the NCMS Office	154
6.2.3 The NCMS Oversight Committee	158
6.2.4 Other local oversight: Users and local Audit Bureaus	160
6.2.5 Ineffective oversight: Resulting problems	161
6.3 Taoshan: Establishment of an ‘Independent Oversight Bureau’	162
6.3.1 What Taoshan did	162
6.3.2 The need for increased oversight	164
6.3.3 A guardian angel for the NCMS: Remit of the Oversight Bureau	165
6.3.4 Functioning of the Bureau: Oversight	166
6.3.5 Functioning of the Bureau: Cases	168
6.3.6 Functioning of the Bureau: Problem solving	170
6.3.7 Management	171
6.4 Significance of Taoshan’s ‘Independent Oversight Bureau’	172

6.4.1 Effectiveness	172
6.4.2 How independent is Taoshan's Independent Oversight Bureau?	173
6.4.3 Shared histories; shared norms	173
6.4.4 Risk: Standing on the edge of the volcano	176
6.4.5 Risk and NCMS structure: Making good deficient central policy	178
6.4.6 Risk and NCMS structure: Ubiquity of risk framings	179
6.4.7 Risk: Perceptions and power	182
6.5 Taoshan's Oversight Bureau, the NCMS and policy change	184
6.5.1 Taoshan's initiative in the context of calls for structural change in local oversight	184
6.5.2 Will Taoshan's Bureau be sustained?	186
6.5.3 Supra-local influence of Taoshan's Oversight Bureau	187
6.5.4 Linking Taoshan's Oversight Bureau and policy change	188
6.6 Interim conclusions and looking forwards	192
Chapter Seven. Implementing and developing the NCMS: Institutional logic of Feitian's cost control reform	194
7.1 Introduction and background	194
7.2 Background to Feitian's innovation	195
7.2.1 Cost growth and misaligned incentives	195
7.2.2 NCMS and cost growth	198
7.2.3 NCMS and cost growth: Framings and legitimate scope of action	198
7.2.4 Feitian's cost control reform	200

7.3 NCMS as pressurised system: The view from Feitian	202
7.3.1 Implementing the NCMS: National targets and local policy	203
7.3.2 Implementing the NCMS: Targets, funds and plans	204
7.3.3 Interim summary	208
7.4 Implementing the NCMS: The logic of fund setting	209
7.4.1 Feitian: The 2007 report	210
7.4.2 Feitian: Cost growth and the logic of cost control	212
7.4.3 Other views of fund risk	215
7.5 Feitian: Cost control and the functioning of 'Controlling Cost Growth'	216
7.5.1 Understanding Feitian's cost control reform	216
7.5.2 The logic of Feitian's cost control reform	219
7.5.3 Functioning of CCG	222
7.5.4 Functioning of CCG: Feitian Chinese Medical Hospital	224
7.5.5 Interim summary: CCG	226
7.6 Feitian's reform and policy change: Risk, cost control and payment reform	227
7.6.1 Risk and the logic of provider payment reform	227
7.6.2 From cost control to provider payment reform	230
7.6.3 Implementing and interpreting policy: The logic of 'results orientation'	232
7.6.4 CCG, sub-national models and changing repertoires	234
7.6.5 Interim summary	238
7.7 Interim conclusions and looking forwards	239

Chapter Eight. Transitional institutions and fill-in development: The Second Round Reimbursement policy	241
8.1 Introduction	241
8.2 NCMS reimbursement, local capacity and rationale of the SRR	243
8.2.1 NCMS reimbursement schemes: Local design; low payouts	243
8.2.2 Capacity deficits, surpluses and conservatism	246
8.2.3 Rationale, functioning and effectiveness of SRR	249
8.3 Local roots and uptake: Provinces, the centre and transition to urban insurance	252
8.3.1 County-level innovation and sub-national spread	252
8.3.2 SRR on the central agenda	253
8.3.3 Standardising and regulating SRR	254
8.3.4 Interim summary	256
8.4 Capacity constraints, appropriate innovation and importance of context	256
8.4.1 Innovation	257
8.4.2 Innovation on the periphery: Appropriateness, context, metis, reified knowledge	259
8.4.3 Transitional institutions: Two conceptions of efficiency	263
8.5 Concluding comments	266
Chapter Nine. Counties: Implementers and innovators in national policy schemes, part two	268
9.1 Introduction	268
9.2 County roles: Risk, tolerance and legitimate action	269

9.2.1 Policy development and risk	269
9.2.2 Legitimate roles and judgements of legitimacy	272
9.3 Policy: Implementation, interpretation, and conversations of reform	273
9.3.1 Textual specificity: Policy discourse and the role of interpretation	274
9.3.2 Legitimate readings: Some readings are better than others	276
9.3.3 Repertoires, models and conversations	278
9.3.4 Synthesis	283
Chapter Ten. Conclusions: New ideas that work?	285
10.1 Adaptive development or making good central policy?	285
10.2 Differentiated responsibilities?	288
10.3 Learning from China?	291
10.4 Where have all the innovations gone?	292
10.5 Concluding comments	295
Appendix. List of interviews	297
List of references	301
1. Chinese policy documents	301
2. Other references	310

Acronyms

CCTB	Central Compilation and Translation Bureau of the Chinese Communist Party
CMS	Cooperative Medical Scheme
CRS	Cadre Responsibility System
HRS	Household Responsibility System
MOF (BOF)	Ministry of Finance (Bureau of Finance) ¹
MOH (BOH)	Ministry of Health (Bureau of Health) ¹
NCMS	New Cooperative Medical Scheme
NDRC	National Development and Reform Commission
RMB	<i>Renminbi</i>
THC	Township Health Centre
WHO	World Health Organisation
ZYB	<i>Zhuanyuanban</i> ²

Notes:

(1) 'Bureau' is used to translate both *Ju* (county and city level) and *Ting* (province). In all cases, which is referred to is clear from context.

(2) The *Zhuanyuanban* is discussed in detail in Chapter Four.

Note on language

Unless otherwise specified, all translations from Chinese are by the author. All transliterations of Chinese are given in *pinyin*, except where sources which use alternative transliteration systems are cited directly.

Chapter One

Understanding local government innovation in China's healthcare reform: The case of the NCMS

From 2002-2003, development of China's New Cooperative Medical Scheme (NCMS), a rural health insurance programme, was initiated based on a process of county piloting and subsequent roll out. The NCMS has developed rapidly: from initially covering a small number of counties and from very limited funding, the scheme has been extended nationally, funding has increased dramatically, and policy has been progressively refined and codified. The majority of existing English language analyses focus on impact on users' health seeking and spending, paying less attention to the policy, its management and development. The scheme is principally run at the county level, and my research examines implementation of the scheme, local 'innovation' and production of local reforms under this policy umbrella.¹

The significance of this is twofold. First, understanding functioning of the NCMS is important in its own right. The NCMS is a key national policy, affecting access to healthcare for hundreds of millions of rural Chinese. Its expressed aim is to provide health insurance to the rural population which, by the early 2000s, was largely not covered by health insurance of any kind. The policy is a cornerstone of reforms to the rural health system which started in the early 2000s but which have quickly increased in scope and pace since then. It is also fundamental to the broader reform agenda of the Hu Jintao – Wen Jiabao leadership, which has included extensive restructuring of rural government and institutions, and an aim to narrow the rural-urban divide under a broad policy framework of 'scientific development', 'taking people as the base' (*yi ren wei ben*) and construction of the New Socialist Countryside (Fewsmith 2004; Christiansen and Zhang 2009; Heberer and Schubert 2012).

Second, this study contributes to an understanding of the role of local government in

¹ In this thesis, 'local', and 'locality' refer principally to the county level.

China in policy implementation and reform processes – here, the development of a key health insurance programme. Much literature on central-local relations and local government in China foregrounds questions of power and the centre’s ability to enforce sub-national policy implementation and local compliance with central policy mandates. A restrictive focus on counties as implementers of central policy, however, tends to obscure their role as developers of policy in their own right, first in policy areas in which the centre plays a limited role, but also in areas in which the centre sets overall directions or parameters of reform, within which sub-national governments have a degree of freedom and are expected to have responsibility for development of workable policy.

This is the case of the NCMS, in which the centre has consistently defined policy principles, but sub-national governments, principally counties (the scheme is described as ‘*xian wei danwei*’, or ‘managed at the county level’, loosely translated, in much policy discourse), are given a degree of freedom in both operation and development of the scheme in the hope that this will allow, first, tailoring of policy to local circumstances and, second, development of innovative management mechanisms that can be fed into overall scheme development. In other words, a degree of policy innovation, at least, is presupposed. Given this, analysis of local governments cannot be confined to understanding their role as (frequently imperfect) implementers of policy – the role they are expected to play is larger and more complex than straightforward implementation of ‘pre-cast’ central policy. This framing attributes greater significance to the role of local governments in policy development and highlights their role as policy innovators. It also argues that local policy development and innovation should not, for the most part, be thought of as deviant and systemically-challenging behaviour, but as a component part of a loosely integrated, interactive mode of central-local policy development. In other words, local government innovation is, at least to a degree, systemically expected and tolerated as a constituent part of policy development.

In the case of the NCMS, it has been observed that the scheme has been ‘experimentally’ developed (Wagstaff, Lindelow et al. 2009a; Wang 2009b), but little

attention has been paid to actual scheme development or to understanding processes of local management, experimentation and innovation in relation to development of policy at larger scales within government. Understanding how counties innovate in policy development within overall national policy frameworks is of crucial importance to understanding not just development of the NCMS, but Chinese reform processes more generally.

1.1 Outline and approach

This study examines the role of local governments in policy implementation, specifically implementation of the NCMS, and re-frames this from a strict question of implementation to a question of both implementation and local government policy innovation. Study of local government in China has tended to analyse this as one part of a dualistic relationship with the 'centre', in which the core analytical question is the extent to which, and the mechanisms through which, the central government can exact 'compliance' with its policies. This framing of central and local roles underlies very much theorising on Chinese government and the role of local governments, both where analysis deliberately questions the extent to which the centre can control local agents (e.g. Wong 1987; Wang 1995; Lampton 1987a; Oi 1995; O'Brien and Li 1999), or analysis of the mechanisms through which control (normally limited control) can be exerted (e.g. Naughton and Yang 2004; Huang 1999; Edin 2003b).

Overall, in the literature on local government in China and on central-local relations, little room is given to description or understanding of the potential for local government to play a positive role in relation to the centre or superior levels of government. Where agency is attributed to local governments, with some exceptions, this tends to be unproductive agency – the agency to subvert the centre's plans (e.g. O'Brien and Li 1999). Those analyses that attempt to explain the very great complexity of local government roles and develop analyses supplementary to this dominant 'compliance' framing have tended to develop theories of 'non-zero-sum' relations (Li 1998b; Li 2005; Li 2006a), or to reframe questions of compliance in terms of local

'discretion' (Chung 2000; Chung 2011) or 'differential implementation' of central policy (Göbel 2011). A more recent analytical focus, largely pioneered by Heilmann, has emphasised the extent to which the Chinese state conducts 'experimental' policy making (Heilmann 2008a; Heilmann 2008b; Heilmann 2008c; Heilmann 2009; Heilmann and Perry 2011).

Analyses of the NCMS, and of health system reform in China more generally, have, with the exception of a small number of studies (e.g. Klotzbücher, Lässig et al. 2010; Wang 2011) been carried out by health systems analysts and health economists. There has been little analysis of the role of government, particularly local government, in implementation and development of policy such as the NCMS.

All the analyses briefly introduced here are limited in their ability to explain what is empirically visible in the case of the NCMS. Chapter Two reviews this literature in detail as a precursor to presentation of my analysis in Chapters Four and Nine.

In outline, reviewing existing literature, I argue the following. The NCMS, as framed and discussed in central policy and speeches, cannot be confined to a straightforward question of 'implementation' of pre-cast central policy. Rather, sub-national governments (principally counties) are expected to play a developmental role in refining the scheme and developing applicable policy mechanisms (*jizhi*), even under centrally-set policy frameworks such as this. While elements of 'compliance' analysis are useful for understanding the ways in which implementation of the NCMS is ensured, this fails to explain the ways in which local governments implement policy in different ways under a broad national policy framework and develop local policy solutions.

Existing theoretical approaches to 'differential implementation', local government 'discretion' and policy 'experimentation', as mentioned above, provide useful but incomplete explanations of the role of local governments in the implementation and development of the NCMS. With few exceptions (e.g. Li 2005; Li 2006a), analysts from this 'tradition' frame their understanding of policy and its implementation in centrist terms. What is visible is the *fact* of differential implementation or experimentation, but

not the understandings and framings of actors (particularly local government) engaged in this process.

Little scholarship has attempted to understand processes of bottom-up policy development, or innovation, in China, with the exception of a handful of studies of development of the 'household responsibility system' (e.g. Unger 1985; Watson 1987), village elections (O'Brien and Li 2000), and various changes to selection of officials (e.g. Fewsmith 2006; Fewsmith 2008b; Fewsmith 2008a; Saich and Yang 2003). Where local policy innovations are discussed, these are overwhelmingly framed as systemically challenging, risky and rare. There is very little analysis of 'marginal' (Bessant 2005), or systemically non-challenging, policy innovation.

The above two limitations of existing analysis, lead to a distorted understanding of both the degree of local policy development under central policy schemes, as well as the systemic significance of this. A focus on policy 'innovation' is a means to understand this. In Chinese policy discourse, local government 'innovation', local 'experimentation', and similar terms (see p. 73 for discussion) occupy a significant place as exhortations to creative local implementation, adaptation and interpretation of policy. 'Innovation' should be understood as both process and result: as *process*, innovation corresponds to what Schumpeter describes as operating "outside the boundary of routine"; as *result*, it corresponds to the creation of new 'combinations' (Schumpeter 1934, p. 84).

Seeing counties as innovators in policy processes does not mean that everything counties do should be seen as an 'innovation', if what we mean by innovation is, broadly speaking, something 'good', or corresponding to 'best practice'. It is a way of re-framing county roles (see p. 76). The importance of an innovation framing is to allow understanding of the mechanism by which Chinese government policy implementation/innovation mechanisms first create, then attempt to capture, the gains of decentred policy formulation and to recycle sub-nationally-generated policy solutions both across sub-national policy jurisdictions and in central policy formulation. This both corresponds to a long history of thinking about the role of central and sub-national governments in China (Mao 1977; Schurmann 1966), and to the

developmental imperative of polities, and systems in general, to adapt to changing circumstances (North 2005).

There is very little analysis of ways in which local government policy innovations are taken up by superior levels of government, become part of policy at supra-local scales, or the mechanisms underlying this (one exception is Li 2006b). There is, to the best of my knowledge, no analysis which questions the *form* of policy solutions, or innovations, developed by local governments.

1.2 Framing of the study and research questions

Counties play the main role in management of the NCMS, and in this respect the scheme is similar to many social programmes. My interest is in both explaining the origins and systemic basis of county-level practices and understanding and charting their systemic relevance or usefulness. In doing this, I am mainly interested in the county-province interface, and in charting (mainly) provincial understandings, evaluations and adoption of county practices. Understanding how counties run the scheme within a broad central framework, variation in management practices, and local production of institutional innovations, is, I believe, as important as understanding the scheme as expressed in overarching national policy. The policy processes visible here are relevant for an understanding of adaptive policy making in China more generally, and are not confined to the NCMS.

Understanding county processes of implementation and innovation is an empirical question, one which focuses on specific county practices and reforms and examines systemic judgements of their legitimacy and their relation to supra-local models, reforms and discourses, in an attempt to understand how localities use policy. My study does not focus on formal, centrally- or provincially-mandated experimentation, but rather on the way that counties operate within the overall structure of the NCMS, the degree of discretion they have, and the possibility and/or importance of local generation of policy, policy mechanisms and models, or 'locally generated' (*zifa de*, *zisheng de*) reforms. This affords a view of local practices and reforms on the

periphery of the state policy making apparatus, where local diversity and implementation frequently run ahead of central policy.

This study examines the administrative practice² of local governments (specifically NCMS management), and attempts to understand ways in which situated actors work within constraints (and make use of freedoms) under which they operate, the multiple ways in which they interpret policy and make judgements as to legitimacy of specific policy choices. This is an alternative strategy to one which takes as a strict focus 'policy' in order to question the fidelity of implementation to supposed textual authority. Paraphrasing Jiang, I believe it is more fruitful to ask how policy is 'practised' than to ask what a given policy 'means' (Jiang 2010). This requires contextualizing counties' policy implementation/innovation and attempting to understand "the policy process and the formal and informal rules that structure it", recognising that the Chinese policy system is "composed not only of laws and regulations but also of conventions, understandings, habits, and practices" (O'Brien 2010, p. 80). The perceptions and framings of actors in the policy process, their judgements regarding specific policy solutions and possible reforms, the legitimacy of these, and the ways these are linked with policies, discourses, models and reforms in other places and at other scales in the Chinese polity are important components of understanding policy implementation and innovation.

² I use this term very simply, taking 'practice' as simply "the actions and voices of people in history" (Gudeman and Rivera 1990, p. 189). This is an attempt to focus on what people and institutions do and say rather than on ideal-typical representations of their roles or functions. My use of this term comes close to, for example, Long, who argues that research should "focus upon intervention [implementation] practices as shaped by the interaction among the various participants, rather than simply upon [...] ideal typical representations" (Long 2001, p. 26). I do not engage in debates around 'practice' as a way out of structure/agency debates, as in e.g. Bourdieu (1992).

Accordingly, this study asks the following questions:

- 1) Given a clear and specific central rhetorical attachment to local government policy innovation in development of the NCMS, is there evidence that this exists?
- 2) If local government policy innovations are in evidence, what motivates innovation? What kinds of innovations are produced?
- 3) Given that the majority of analysis of local government policy behaviour is analysis of compliance and mechanisms used to ensure this, what mechanisms explain local government policy innovation in an existing central government policy area such as the NCMS?
- 4) How do local policy innovations relate to policy, reforms and discourses at larger scales in the Chinese policy system, including having the potential to contribute to policy at supra-local scales, including central policy?

Research is based on county-level fieldwork, mainly in two eastern provinces, but with some exposure to central/western provinces. Counties were chosen to allow examination of specific local 'innovations', and show different levels of development. In addition, I interviewed provincial officials about reform priorities and the role of counties in scheme management and development. Chinese documentary sources are used extensively. Details of research methodology, fieldwork, and use of documentary sources can be found in Chapter Three.

1.3 Contributions to existing research

This study contributes to existing research in the following ways.

- First, it presents a large amount of empirical material relating to development of the NCMS, both at the national and sub-national levels (policy documents, reports and academic sources which have received little, if any, treatment in English) to analyse the NCMS as national policy, and roles of national and sub-national government within this.

- Second, it presents two main case studies based on fieldwork in three principal counties in which I interviewed NCMS managers and policy makers, as well as one based on document analysis. (A full list of interviews is given in the Appendix.)
- Third, this is the first study of local government innovation in China at the time it is taking place (or thereabouts), and using fieldwork to detail perceptions and motivations of actors involved. This provides a very different view from the centrist perspective of Heilmann and others and points to processes of policy innovation in China being more common and less voluntaristic than normally recognised, at least in this area and as regards 'marginal', rather than systemically-challenging and 'disruptive', innovation (terms are from Bessant 2005).
- Fourth, the study presents an explanation of the way that counties innovate within this specific policy area. County policy innovations presented in Chapters Six and Seven, as well as a locally-generated policy innovation, which made the transition from sub-national to national policy (Chapter Eight), are presented as part of one case (cf. Heimer 2006) with discernable common characteristics, related to the 'peripheral' character of the counties and examples involved.
- Fifth, the study provides an account of systemic mechanisms underlying policy innovation, and how these are linked to mechanisms analysed in existing literature as underpinning policy implementation (specifically the cadre responsibility system and policy texts), and stresses the inevitability of policy interpretation, the low risk involved in this, and the dispersed nature of decisions as to the legitimacy of variant policy solutions and local innovations.
- Sixth, little has been written about policy innovation within health system reform in China. Equally, analyses of variation in the NCMS to date have largely focused on technical scheme variables (e.g. funding levels, co-payments, thresholds); my focus on local development of innovative policy solutions is an addition to this.

1.4 Structure of the thesis

The thesis is structured as follows. Chapters Two and Four review existing literature and start to develop an analytical framework. In Chapter Two, I review existing literature on central-local relations, policy implementation and experimentation in China and argue that this is inadequate to understand processes of development of policy such as the NCMS. Chapter Three describes research methodology. In Chapter Four, I set out the beginnings of an analytical framework for understanding how local governments in China are both under pressure to implement policy, and simultaneously have a degree of freedom as to how they do this. Chapters Five to Seven discuss the NCMS as both national framework and locally implemented and developed policy. Chapter Five discusses the NCMS as both national and sub-national policy: as a central policy in which the central government has consistently set the overall parameters of reform and principles (*yuanze*) of policy, but in which local governments play the main role in management and develop implementable (*caozuoxing de*) policy. I show how implementation of the NCMS is broadly ensured through vertical mechanisms, and how this is combined with a hope that localities will develop policy mechanisms of systemic significance. This is an analytical background to subsequent empirical cases.

Chapter Six examines the first of two local innovations discussed in depth, an NCMS oversight agency set up by the Bureau of Health in an eastern county. I examine functioning of this agency and argue that its creation was a response to perceived systemic risk in NCMS implementation on the part of the BOH. Chapter Seven examines a second local innovation, a cost control reform implemented by a second eastern county. This shows clearly how scheme implementation targets translate into local management but leave a large degree of discretion to local NCMS managers for unscripted policy development. In both cases, I examine local reforms in the context of changing national policy, both in terms of how these reforms respond to supra-local policy and discourses and the extent to which they are having, or could have, supra-local applicability and/or impact. Chapter Eight provides a detailed analysis of

the transition of an innovative local NCMS reimbursement mechanism to national policy. Following presentation of empirical material, Chapter Nine develops the analysis started in Chapter Four to provide a refined understanding of processes of local innovation within the NCMS. Conclusions are presented in Chapter Ten. Interviews are listed in the Appendix.

As background, the next section gives a brief sketch of the context of China's health reforms.

1.5 Background to the NCMS

Here, I give a brief background to introduction of the NCMS. This description largely deals with the period preceding recent reforms to the rural health system; current change is very rapid and of great significance.³

In rural areas, transition from a collective agricultural economy to household-based farming, from the late 1970s and early 1980s, was accompanied in most places by the dissolution of the Cooperative Medical Scheme, China's much-discussed pre-reform rural health insurance system.⁴ Rural health insurance coverage fell dramatically, to around 10% by the mid-1980s (Yip and Hsiao 2009) and to around 7% by 1993 (Wagstaff, Lindelow et al. 2009a). In 2003, with the initiation of the 'New' CMS, around 80% of rural residents lacked any form of insurance whatsoever (Wagstaff and Lindelow 2008a), and individual ('out-of-pocket') payments accounted for around 60%

³ This can only be brief, given the very broad range of issues and substantial relevant literature. See Wagstaff, Lindelow et al. (2009c) and Wagstaff, Yip et al. (2009) for recent overviews.

⁴ Most accounts present collapse as an unforeseen by-product of rural reforms (Blumenthal and Hsiao 2005; Wagstaff, Yip et al. 2009) and/or link its demise to technical explanations, including the impact of decollectivisation on the organisational and fiscal base of local health provision, restrictive risk pooling, poor administration, and a general movement away from collective structures with the introduction of market reforms (e.g. Meessen and Bloom 2007). In contrast, Duckett has recently described this as a "thoroughgoing form of programmatic retrenchment" ultimately deriving from a reversal in MOH policy, driven by ideological and political changes largely unchallenged by the bureaucracy, local governments or other stakeholders (Duckett 2011, p. 21 and Chapter Four).

of total health spending, up from around 20% in 1978 (Hu, Tang et al. 2008; Wagstaff, Lindelow et al. 2009c, p. 19), with government contributing an ever-decreasing share of spending over the reform period, though government spending grew overall, at just under growth in GDP (Wagstaff, Yip et al. 2009). This has led to a decrease in risk sharing and an increase in individualisation of the burden of medical spending.

Post-1978, management of the health system was significantly decentralised, and funding responsibility was largely devolved to local levels, resulting in a decrease in public funding for *de jure* public providers, most of which received around 10-15% of total revenue from public funds by the 1990s (Bloom and Gu 1997). Some providers receive less support than this, and village clinics are mostly formally financially independent. The collapse of public and cooperative funding, and subsequent devolution of financing responsibility (Tang and Bloom 2000), combined with decreases in fiscal capacity in many places and a 'broken' transfer system (Wong 2009) undermined provision, especially at village and township levels (Wang, Gu et al. 2008). Provision has become highly segmented, with low quality village provision set against investment-starved township health centres, and county hospitals with greater ability to invest in buildings and equipment and to attract quality staff. In many cases, county and city-level facilities are preferred by those with the means to pay. Inequalities between localities are very great and segmentation has also led to competition between providers for a limited pool of users and revenue, and a breakdown in referral mechanisms between providers (Bloom, Han et al. 2000).⁵

There is general consensus regarding the results of these changes. Government price controls, aiming to set prices of essential services and drugs below cost to ensure accessibility, simultaneously allowed mark-ups on higher tech services and drugs as a form of cross-subsidy, and this combined with increasing autonomy of providers to create incentives to invest in high-tech equipment and over-supply profit-making

⁵ On the inequitable and pro-rich nature of much health spending, and disparities in health see Tang, Meng et al. (2008), Wagstaff, Lindelow et al. (2009b) and Liu, Rao et al. (2008).

services at the expense of unprofitable treatments and drugs and public health work (Blumenthal and Hsiao 2005; Eggleston, Li et al. 2008; Yip and Hsiao 2008; Wagstaff, Lindelow et al. 2009c). Spending on healthcare increased dramatically, rising by around 16% annually since the early 1990s, faster than growth in GDP and incomes (Wagstaff, Lindelow et al. 2009c, p. 19; Yip and Hsiao 2009), and faster than price growth generally (Bloom and Fang 2003, p. 3), while quality, responsiveness to users and efficiency of services for the most part remain low (Eggleston, Li et al. 2008).

While it is hard to disentangle effects of market forces, low efficiency, and perverse system incentives in pushing up costs from changes in the overall burden of disease and changing health seeking behaviour, and to determine whether costs are rising 'too fast' (Eggleston, Li et al. 2008), it is recognised that restructuring of the rural health system and profit-orientation of providers has contributed significantly to increases.⁶ As Hsiao argues, changing incentive structures have resulted in "new and unfettered opportunities for hospitals and physicians to obtain higher incomes [which] have caused financial pursuits to triumph over professional responsibility and ethics for most physicians" (Hsiao 2008, p. 949). Information asymmetry within healthcare means that 'consumers' are rarely well-placed to make informed choices (e.g. Bloom, Han et al. 2000 citing Arrow 1963), and low levels of public funding translate into limited ability of local governments to supervise providers, exacerbated by an absence of alternative mechanisms to control providers' day-to-day behaviour, such as an insurance scheme acting as purchaser of medical services on behalf of rural populations.

Figures abound for the negative consequences of this. A recent World Bank report puts the out-of-pocket expense of a single inpatient episode at around 60% of per capita consumption (Wagstaff, Lindelow et al. 2009b, p. 14), while another study estimates this as "almost equivalent to China's annual income per head and [...] twice the average annual income of the lowest 20% of the population (Xu 2008, cited in Hu, Tang et al. 2008, p. 1846). More dramatically, one recent study, using data from China's National Health Survey, found that "out-of-pocket expenses associated with a

⁶ I return to this issue in more detail in Chapter Six.

single inpatient admission increased from 70 to 80% of per capita income in 1993 to more than 200% in 2003” (Yip 2009, cited in Wagstaff, Yip et al. 2009, p. S11). Data from the same survey show 50% of respondents who had been ill in the previous two weeks to have not sought treatment (an increase from 36% in 1993), and that 30% of survey respondents “said they had not been hospitalized despite having been told they needed to be [...] among those who did go to hospital, nearly half discharged themselves against their doctor’s advice. The largest barrier in all cases, but especially among the poorest, was cost of care” (Wagstaff and Lindelow 2008b, p. 266).

Where people do seek care despite high costs, this can result in impoverishment and very high levels of out-of-pocket spending. One recent study found ‘catastrophic’ health spending to be higher in China than in other Asian countries and, diverging from experience elsewhere in Asia, higher among the poor than among non-poor (2000 data) (van Doorslaer, O’Donnell et al. 2007), while another recent study found that in 2003 catastrophic spending affected “about 184 million Chinese people – a record that puts China’s health-system performance for financial risk protection among the poorest in the world” (Liu, Rao et al. 2008, p. 1922).⁷ Health spending has substantially increased China’s poverty headcount (van Doorslaer, O’Donnell et al. 2006), and illness and health-related expenditure have been estimated to be the main cause of poverty in between 30 and 50% of cases (Bekedam 2006).

While there is discrepancy in calculations and estimates arrived at by different analysts, the overall picture is not significantly contested: absence of health insurance and rapidly increasing costs have led to missed or inadequate care, neglect of preventive services and, frequently, impoverishment where expensive treatment is sought. This is an important social issue, with healthcare access frequently cited as a major worry in opinion surveys.

Since the 1990s, China has been experimenting with new ways of providing healthcare,

⁷ Measures of ‘catastrophic’ spending vary. Van Doorslaer et al. use measures of greater than 25% or 40% of non-food consumption; Liu et al. use a WHO standard of 30% of a household’s ‘capacity to pay’. See individual studies for actual definitions.

including changing incentives for providers, changing hospital funding arrangements and revamping urban health insurance. The 1990s also saw the beginning of attempts to resurrect the CMS, though with little success (see Jackson, Sleight et al. 2005, p. 141; Wagstaff, Lindelow et al. 2009a, p. 2). Since the early 2000s, there has been a reorientation in central policy towards rural issues, including reform of the rural healthcare system. Overall, this is a process of re-regulation, attempting to reverse much of the marketisation of the reform period to date, and the NCMS is a core component of this. From 2002-2003, development of a rural health insurance framework was begun, starting with a small number of counties and small amounts of money and with the expressed aim of reducing impoverishment resulting from health seeking. Since then, this system has evolved very rapidly and has been rolled out nationwide. It remains imperfect, however, providing only low levels of protection against catastrophic spending, and, structurally speaking, it has had a limited impact on providers and cost escalation. These issues, and detailed description of the scheme and its development, are taken up in subsequent chapters.

Chapter Two

Control and discretion in China's central-local relations and policy making

2.1 Introduction to theoretical chapters

This chapter is intended to be read in conjunction with Chapters Four and Nine which, combined, present the theoretical backbone of this thesis. Overall, I argue that analysis of central-local relations and policy implementation in China has changed greatly over the reform period, supplementing 'compliance-based' analyses with non-zero-sum and/or positive-sum analyses of central-local relations which attribute productive agency to localities in policy implementation and/or formulation. My analysis supplements this with a practice-oriented view of local behaviour, linked to policy and discourses at supra-local scales. This shows a need to understand localities as both implementers of policy and as innovators within policy frameworks.

Systemic policy and target-setting mechanisms have both controlling and generative functions, whose aim is to ensure both implementation of policy through the 'pressurised system' (*yalixing tizhi*) (see p. 73) of implementation targets and localisation of risk with implementing units, and local policy interpretation – in a quest for adaptive local policy setting and development of systemically useful policy solutions. Contrary to many analyses, I stress systemic tolerance of variant local practices, and show local policy innovation to be, at least in this case, relatively low risk. I analyse how local policy making exists in relation to national policy and to a range of sub-national models and practices in a loosely-articulated 'conversation' of reform (Gudeman and Rivera 1990), in which judgements as to the legitimacy of policy solutions are made even at low levels in the system. I question the types of policy solutions this system produces and their importance and usefulness in China's reforms and more generally.

This chapter critiques existing approaches. In Chapter Four, I provide an outline

analytical framework for understanding national policy and local action (implementation/innovation) in the case of the NCMS. Chapter Nine formalises this, following presentation of empirical material in Chapters Five to Eight.

2.2 Introduction to literature review

This chapter reviews literature on central-local relations and policy making/implementation in China. This analysis is selective and cannot be a comprehensive account of the functioning of local government or central-local relations. In outline, this chapter argues that, post-Reform and Opening, much study of central-local relations and policy implementation has been framed in terms of *control*, specifically the possible loss of central control as sub-national governments have benefited from economic development and greater responsibility for specific policy areas. This is valuable, but sidelines the question of local freedom and how this is used by policy implementers. Freedom, or 'discretion' (Chung 2001), is by no means absent from this literature, but many analyses showing sub-national discretion remain within an overall control framing. More relevant are non-zero-sum and positive-sum theories of central-local relations and theories of differential local policy implementation.

Post-reform, government of China has been significantly decentralised, and much has been written about transfer of power and responsibility downwards to spatial units (provinces, counties, etc.), conceived of as reducing the power of the centre vis-à-vis sub-national government. Here, I examine analytical approaches to central-local relations and policy implementation. I do not, more than superficially, deal with decentralisation as such.⁸

To a large extent, analyses that find decentralisation post-reform do so against the background of an assumed highly centralised and controlling government/Party in the

⁸ The relevant literature is substantial. For recent outlines and/or recapitulations, see e.g. Lieberthal (2004), Zheng (2007), Landry (2008) and Chung (2000). On devolution of authority in various policy domains, see e.g. Lampton (1987b) and Göbel (2007). Wong (2009) and Landry (ibid.) provide recent reviews of the fiscal picture.

pre-reform period, especially prior to the Cultural Revolution. Li, reviewing this period, for example, describes contemporary analyses as 'totalitarian', arguing, with Schurmann and Barnett, that sub-national units of government, particularly provinces, should be seen solely as agents of the centre in a system in which the centre had a monopoly on power through organisational and ideological control (Li 1998b, pp. 18-19), and in which sub-national government enjoyed "no inviolable autonomy" vis-à-vis the centre (Barnett, quoted in Li 1998b, p. 18). Similarly, for Chung, prior to the reform era, despite periodic central rhetorical attachment to decentralisation, "rigid ideological control and political propaganda dominated the policy process, depriving local leaders of incentives to risk their political fate for parochial local interests" (Chung 2000, p. 13): norms of obedience to the centre during the Mao era meant that local compliance with central mandates was "rigid beyond casual observation" (Chung 2000, p. 40). Elsewhere, Chung describes 'discretion' in policy implementation in the Mao era as "ideologically problematic and politically unsafe" (Chung 2001, p. 66). In contrast, Lieberthal, for example, believes the degree of central control is overstated, as "broad pronouncements from the Centre left considerable room for local adaptation [despite an] impression of disciplined conformity" (Lieberthal 2004, p. 320).

Whatever our views of the degree of centralisation/decentralisation pre-reform – and I return to the question of continuity in *commitment* to decentralised initiative below – there is widespread agreement that government in China in the reform era has been highly decentralised.

2.3 Approaches to central-local relations in the reform era

The study of central-local relations and policy implementation in the reform era is huge and complex. Here, I describe three 'poles' within this literature as an aid to analysis. The first revolves around the loss of central control linked to strengthening of sub-national government. The second acknowledges a dilution of control, but concludes that control remains at least functional and posits mechanisms preserving (a degree of) central-local integration. The third is less clear-cut and encompasses

attempts to re-define central-local relations as non-zero-sum, to show local variation and discretion in implementation in a positive light, and recent work on experimentation in Chinese policy making. I highlight, especially in relation to poles one and two, an underlying analytical thread in this literature: an excessive focus on control and on implementation conceived of as 'compliance' with central policy. Before setting out this analysis in detail, I sketch the context of these analyses. The range of issues and the timeframe involved mean that this can only be a brief introduction.

China's development in the reform era has had very great implications for central-local relations and policy implementation. Analyses have evolved over time, both as access to sources and possibilities of fieldwork have increased, and as the object of study has changed as reforms have progressed. The earliest reforms were rural, most significantly the shift from commune-based management of agriculture to household responsibility for farming and the very great increases in agricultural production that followed this, which had a great impact on rural incomes (at least initially) and the rapid development of rural enterprise ('township and village enterprises', TVEs) in the 1980s and 1990s, providing the motor for rural economic growth and national growth more generally.

Developments have been underpinned by changes to China's fiscal system in the late 1980s and 1990s, the effect of which was to make sub-national governments increasingly responsible for their own spending. According to Saich, the fiscal contracting system instituted in the 1980s transformed what had been a "province-collecting, centre-spending fiscal regime to an essentially self-financing regime for both centre and provinces" (Saich 2011, p. 301). The advantage of such a system was to increase sub-national enthusiasm for economic development, but at the expense of both declining central revenues and increasing inter-jurisdictional inequality. Where the previous system had been highly redistributive, the new system was inequitable and broke the link between revenue sharing and spending requirements. By 1993, almost 80% of total fiscal revenue was sub-national, and the central share of spending had decreased to around 28%, from around 47% at the beginning of the

reform period. State capacity, measured by control of revenue, had not declined, but had rather been realigned to the benefit of sub-national governments. Subsequent tax reforms in 1993-1994 changed this system somewhat, increasing the central share of tax revenues, but allowed richer provinces to retain more revenues, further increasing sub-national disparities (Saich 2011, Chapter 7). The transfer system which should offset these imbalances functions poorly (e.g. Wong 2009) and is insufficient to offset disparities in revenues. Sub-national enthusiasm for economic development, TVEs, off-budget funds and illegitimate extractions, as well as elimination of services in poorer areas, all, broadly speaking, derive from this conjuncture.

Sub-national economic development and greater local control of economic resources and initiative, as well as creating economic growth, have allowed sub-national governments a greater degree of freedom from central control in decision making, and created potentially divergent central-local interests, as well as sub-national solidarities with the possibility of weakening local adherence to central policy:

“while these economic reforms led to China’s accelerating economic growth, they also led to increasing political and economic decentralisation as local governments made economic decisions, used tax revenue for local projects, and received less financial support from higher levels. Furthermore, as the local governments facilitated the money-making capacities of the collective-private enterprises, they formed alliances with these enterprises that benefited both sides materially. It was often in the economic interests of local authorities to disregard central government injunctions [...]” (Goldman and MacFarquhar 1999, p. 8).

This dynamic, in one form or another, is the starting point for many of the analyses here and the analytical importance accorded to central control (mostly loss of control) of sub-national governments. As Baum and Shevchenko note, “rising provincial autonomy and incipient rebelliousness achieved a certain prominence at the end of the 1980s”, as a group of provincial governors resisted changes to the fiscal contracting system that would undermine their revenues in the centre’s favour. The same era also

saw a degree of loss of control by the centre of sub-national economic behaviour, as localities “ignored central exhortations to curtail expansion of local credit, investment and construction” and engaged in local protectionism Baum and Shevchenko (1999, p. 335). Various analyses examined below that address central control of the provinces are to a large extent products of this era and this conjuncture: Li’s (1998b) questioning of why the centre seems to complain about the provinces but do little to constrain their behaviour, Wang’s (1995) spectre of the ‘rise of the regions’ and Wong’s (1987) demonstration of breakdown in economic coordination as market reforms empowered sub-national governments are all examples.

At a sub-provincial level, this emphasis on local self-reliance (Saich 2011, p. 301) has certainly stimulated economic development, giving rise to new relationships between government and enterprise and to a number of typologies of the local state (Baum and Shevchenko 1999), but it has also led to very high, and frequently unachievable, local spending commitments, local extractions and illegitimate levies on rural residents, frequently described as the ‘peasant burden’.⁹ These are as much consequences of this mode of development as is sub-national economic dynamism. One of the main casualties of this mode of development has been rural social provision, especially in less economically developed areas. Local governments, often counties, are required to implement social policy – sub-national government spending as a proportion of total spending has been consistently very high, at around 70%, and almost 80% in 2008 (Saich 2011, p. 195) – but funds to meet central obligations must, for the most part, be found locally. This has been a driver of sub-national economic development, a cause of increasing inequalities in sub-national provision and persistent implementation gaps, and of extractions from the peasantry:

“financial pressure has meant that one emerging general imperative shared by economically developed and more resource-constrained localities is the

⁹ ‘Peasant burden’ is used to describe the, often huge, extractions – both legitimate and illegitimate – that rural residents have been frequently subjected to. In this thesis, ‘peasant’ is used as a translation of the standard Chinese term ‘*nongmin*’.

increasingly acute need to derive one's own sources of revenue to cover centrally mandated obligations. The resultant fiscal inequalities that arise from this system are a major cause of the significant variation in the provision of public goods and services [...] The concern over revenue generation is exacerbated by the fact that despite fiscal decentralisation, the central government has retained control over the policy agenda. The centre sets many tasks that must be implemented by local governments, and most of these are unfunded mandates" (Saich 2011, p. 200).

It has only been in the 2000s that central government has made a coordinated attempt to deal with rural development and social policy, though piecemeal reforms took place or were trialled prior to this. Recent or ongoing reforms are very many in number. Recent rural reforms fall under a development strategy aiming to narrow gaps between winners and losers from the reforms to date and to develop or re-regulate rural social services. As above, this is a brief sketch only, intended to provide a degree of context to the analysis that follows.

2.4 Pole 1: Declining central control

As Lieberthal neatly phrases it, in the reform era, the '*kuai*' have gained at the expense of the '*tiao*' in the government of China: territorial power has been strengthened, while vertical mechanisms of control have been weakened, at least in comparison with the pre-reform era (Lieberthal 2004, p. 316). In China, interest in territorial politics is, for historical reasons, animated by a concern for the integrity of the country/state in addition to a more general concern for local implementation of policy formulated at the centre – 'the implementation gap' – common in analysis elsewhere (Li 1998b, p. 1).

Naughton and Yang give a flavour of debates in the 1980s, when observers confronted a China in which central control over the economy, the political system and society was declining rapidly and visibly. They find that decentralisation of control over economic resources and decision-making "empowered local governments and enterprises at the expense of the national government" and, while decentralisation of economic decision

making undoubtedly spurred local initiative and growth, the centre appeared ill-equipped to deal with local and regional interests and its coordinating role was called into question by many analysts. Following 1989, a Chinese 'collapse' seemed plausible to many (Naughton and Yang 2004, pp. 1-2). Landry, in a study of cadre promotions, phrases the issue as one of fiscal decentralisation. In revenue/spending terms, China is extremely decentralised, and an already high level of fiscal decentralisation was increased in the reform era with the aim of stimulating economic growth (Landry 2008, pp. 3-4). Early analyses of the reform era, for him, pointed to declining central extractive capacity, arising from devolution of economic management to localities, as a key indicator of loss of control by the centre. While he cautions that fiscal indicators alone do not necessarily show how spending is ultimately determined (Landry 2008, pp. 12-13), this leads to an overall framing of the question of decentralisation and central-local relations in terms of control.

The number of such early analyses is large. Of the more commonly cited today, Wong notes in a similar vein that early market reforms increased local economic dynamism, but that local interests meant that local growth was at the expense of the state sector and overall economic efficiency and coordination. She concludes that attempts at control through administrative means were largely ineffective (Wong 1987). Wang notably discussed the 'rise of the regions' with economic development, a process in which the decline in central extractive capacity and regional control over economic resources would lead to competition between the regions and the centre and possible national fragmentation (Wang 1995). Other analyses point to cases of provincial defiance of the centre as evidence of impending fragmentation (Chung 1992; see also Goodman 1994 and Baum and Shevchenko 1999, pp. 334-336 for reviews).

Fragmentation is a specific case of a general concern with central-local relations and the 'implementation gap' – lack of congruence between centrally formulated policy and local implementation. One early edited volume (Lampton 1987a) presented multiple case studies showing the complexity of reform era policy implementation. Analyses showed a range of implementation deficits, from 'implementation bias' through which

implementers skewed policy towards their own interests (Naughton 1987), to the paradox of a Leninist system that was simultaneously a powerful tool for reform and suffused with patrimonial networks creating a brake on reform (McCormick 1987), to the watering down of policies once they transitioned from an experimental phase to full roll out (Bachman 1987).

At the local level, that of county and township, Baum and Shevchenko, reviewing the 'state of the state', divide characterisations of the local state into four types: entrepreneurial, developmental, clientelist and predatory, depending on the role of the local state in the economy (Baum and Shevchenko 1999, pp. 344-346). Many such analyses are phrased in terms of control: Oi, for example, while noting the vibrancy of 'local state corporatism', shows how greater control over resources gave localities greater freedom to both implement policy in line with local interests *and* to deliberately pursue strategies at odds with those of the centre (Oi 1999, especially Chapters 5 and 6). For Oi, "A decentralized strategy does not come without costs [...] there is a tension between the need to decentralize and the deteriorating effect this has on the need for a strong central state [...] The question is whether the localities can continue to grow in power without fatally damaging the strength of the Centre" (Oi 1995, p. 1148).

If Oi shows the generally benign face of the local state, others have clearly shown local governments riding roughshod over the interests of citizens. For Lü, the spiralling and crippling 'peasant burden' in the reform era is a result of a 'decentralised predatory state', resulting from a weak centre, unable to control its own agents (Lü 1997). Equally, Bernstein and Lü show that while the centre disapproves of excessive extractions and the peasant burden, these largely derived from unfunded central mandates, and the centre's inability to eradicate these problems stems from reliance on ineffective top-down control instruments (and an unwillingness to countenance genuine local supervision) (Bernstein and Lü 2000).¹⁰ Cases of mis-implementation of policy, corruption and thuggish local behaviour have been extensively catalogued (for a

¹⁰ For a recent treatment of fiscal disparities, unfunded mandates and the importance of changes in the tax system, see, for example, Wong (2009).

prominent example, see Chen and Wu 2006).

One highly influential thesis holds that, in the absence of genuine horizontal accountability, implementation is highly variable by policy type and the way that cadre incentives are structured. In this view, local cadres 'selectively implement' policy, both 'working hard' and 'shirking ably', depending on the case:

"On readily measurable policies the center has established effective controls that lead implementors [sic] to define their tasks as policymakers wish. Enough feedback reaches higher levels, and well-designed inducements and sanctions encourage most ground-level officials to execute even remarkably unpopular measures. On other policies, for which success or failure cannot be assessed without increased popular input, top-down controls have been largely ineffective, and grass-roots officials have easily frustrated efforts to monitor them. In these cases, the center has not been able to prevent implementors [sic] from ignoring or even sabotaging central plans" (O'Brien and Li 1999, p. 181).

In this analysis, 'success' or 'failure' in implementation is largely a question of the way higher levels structure targets through the cadre management system supplemented by *ad hoc* efforts to curb mis-implementation through (inadequate) local supervision and tolerance of 'rightful resistance' by those whose interests are infringed. Implementation is largely conceived of as compliance, and discretion in policy implementation is almost wholly seen as negative.

Similarly, Zhong asks why and how counties and townships, seen as policy implementers with little genuine autonomy, comply with directives from above. He concludes that a decline in ideological controls in the reform era and the inability of cadre management and oversight systems to enforce implementation of multiple and routine policies – as opposed to high priority issues and issues of immediate and pressing concern – lead to resistance, distortion and policy perversions by local cadres, though outright defiance is rare (Zhong 2003).

This chimes with Maria Edin's analysis of the functioning of the cadre responsibility system (CRS) which finds that the central state can be only 'selectively effective' in its control of localities – the system cannot handle too great a number of targets without overloading local governments. Edin concludes that,

“the Chinese party-state has the capacity to be selectively effective, that is, to implement its priority policies, and control its key local leaders and strategically important areas. This strategy no doubt leaves large discretion to local agents over implementation of non-priority policies, and little control over areas which are strategically less important” (Edin 2003b, p. 52).

Similarly, Whiting finds that the CRS provides an incentive to create growth, but also to distort central policies, and a corollary weakening of central state control in specific areas. Overall, she sees the incentives created by this system as 'highly dysfunctional' (Whiting 2001, p. 118).

2.5 Pole 2: Decline in central control, but integrative mechanisms remain

China has not fragmented, and much analysis of central-local relations is an attempt to explain what continues to hold the system together. Much of this is a continuation of the same, control-centred, analysis examined above. The most common approach is to locate control in the political system and to argue that this is enough to offset decentralised control of resources and administration and the fragmentation and/or implementation gap that should, it is hypothesised, result.

Landry's study of cadre promotions is an example. Landry shows huge fiscal decentralisation in the reform era, and points to the paradox of a fiscally decentralised authoritarian regime (Landry 2008, Chapter 1). The integrative mechanism he provides to explain this is the organisation and personnel system. Landry's study unpicks career structures of local cadres, arguing that local-level competition creates a “large pool of local officials [that] operates as a competitive feeding mechanism for higher office” (Landry 2008, p. 105). In this way, the incentive structure of local cadres is focused upwards, and the Party is, and is recognized to be, the determinant of formalized

career progression. This allows decentralisation while avoiding localism and maintaining a “surprisingly high degree of local compliance with key central policies” (Landry 2008, p. 259); for Landry, local deviance exists – the implementation gap is not fully closed, and corruption remains – but for the most part, the most egregious cases remain in check.

Another very influential thesis concerns investment control. For Huang Yasheng, this is an exemplary public goods problem: localities seek growth, but the centre must maintain control of inflation. In the reform era, local governments can no longer be considered simply an extension of the centre, at least in the economic sphere, and a decrease in the centre’s control of resources has led to a belief that the centre is losing control. This should lead to problems of control and coordination but, Huang argues, ‘large-scale coordination problems’ have largely been absent in inflation/investment control. Huang examines a number of integrative mechanisms, but lays most importance on the personnel system, which, he says, serves to align the interests of local officials with the centre. Where there is firm central commitment, compliance happens “once the central government demands such compliance” (Huang 1999, p. 314).

Naughton, Yang and co-authors in *Holding China together* also dispute the disintegration thesis. Various analyses argue that China is not fragmenting and that political institutions, most importantly the personnel system, are “the most important institution reinforcing national unity”, helping align local and national interests. One should not see a diversity of local policy outcomes as an indication of fragmentation (Naughton and Yang 2004, p. 9). Key chapters show systemic changes and their integrative value: Yang argues that reforms in the finance and banking systems, as well as regulatory reform, increase cohesion, and that the aim of the centre is to allow space for local initiative while ensuring that central policy is implemented (Yang 2004a), while Li shows how the centre has reformed management of leading provincial cadres, tying these in closely to national careers (Li 2004). At the local level, Whiting argues that the CRS enforces minimum acceptable levels of performance, but that caution is

needed in understanding the behaviour of local cadres, who are embedded in horizontal, as well as vertical, networks (Whiting 2004).

Yongnian Zheng's '*de facto* federalism' thesis criticises existing approaches to central-local relations, which he describes as 'structural', 'procedural' or 'cultural'. Within the 'structural' approach, Zheng critiques both (in his terms) bottom-up and top-down approaches. The former equate declining (central) state capacity with falling extractive capacity (e.g. Wang, 1995, cited above), whereas the latter, while acknowledging the impact of decentralisation on central-local relations, try to show how the centre has elicited provincial compliance through various political and economic strategies (e.g. Huang Yasheng). The procedural approach, focusing on 'actual policy processes' and represented by Lampton, Lieberthal and Oksenberg, and Shirk, he sees as limited by a rationality assumption and too great a focus on certain specific policy/issue areas and power politics (Zheng 2007, pp. 22-24). Zheng's solution is an analytical framework explaining distribution of power between centre and provinces in terms of coercion, bargaining and an explicitly behavioural/cultural component – norms of reciprocity. For Zheng, '*de facto* federalism' is a way to explain a system which he sees as functioning *like* federalism – in which bargains between, and the roles of, centre and provinces, are 'relatively institutionalised', and relatively distinct and stable central/provincial domains of decision making exist – but which is nonetheless a unitary system lacking constitutional checks on central power (Zheng 2007, pp. 39-41).

2.6 Pole 3: More than zero sum and compliance

Approaches examined above are variations on a theme of (mostly) binary opposition of national and sub-national governments. Analyses question state capacity, and the consequences of strengthened sub-national governments for continued central control/power. These analyses are derived from a principal-agent framework, and both the fragmentation thesis and problematising of implementation (understood as an implementation gap) are expressions of this. The common refrain that 'above there is

policy, below there is counter-policy' (*shang you zhengce, xia you duice*) tends to be the immediate (and often uncritical) riposte to the announcement of many central initiatives. Set against this, there is a separate analytical tradition which gives more scope to non-zero-sum analyses.

2.6.1 Pole 3a: Non-zero-sum approaches

Li has argued for a specific theory of non-zero-sum central-local relations. Her early work asks an explicit question about power in central-provincial relations and control of investment: why, if the centre is frequently heard to complain about the provinces, does it do little to seriously constrain their behaviour? Li argues that the position of the province as intermediary within China's 'continental' structure necessarily confers power in dealings with the centre. In this analysis, both centre and provinces have a degree of irreducible power. There is, inevitably she thinks, conflict, but no total victory is possible, leading to a politics of compromise. She argues that dominant analytical conceptions of power confuse the 'bases' of power (resources) with the operation of power and fail to see the importance of discretionary behaviour – the essence of a game structured by power, rather than outright domination. In this view, provinces have considerable leeway, actors' perceptions are important, and distinct patterns of behaviour, including bargaining, flexible implementation, feigned compliance and voluntaristic action, are visible. From this, Li derives a non-zero-sum account of central-local relations. This goes some way to specifying the inevitable differentiation of central and provincial roles, though her argument is somewhat Pascalian: behaviour of (specific) provinces must be tolerated because the province, as a unit of government, must exist due to China's size. The main criticism of this work is that it falls into the trap of criticising a dominant theory of power, only to introduce a new one in its place. It is not, to my mind, accidental that her phrasing of power distribution is 'non-zero-sum': specifically cooperative solutions fit uneasily in this framework (Li 1998b). Li's later work marks a significant divergence from this.

Elsewhere, Li is very persuasive in her analysis of local, particularly county-level,

institutional development and in questioning usefulness of 'compliance' as a framework for analysis of central-local relations and policy implementation. Remaining with the provincial level, Li believes that it is too simple to divide behaviour of provincial officials into compliance and non-compliance: provincial officials, she says, strategise about goals and means in the same way the centre does, with both parties engaged in a competition to define policy 'orthodoxy'. She attributes a focus on non-compliance to an analytical flaw, and an undue dominance of 'centralist thinking', according to which "differences in choices and judgements [are] treated as if they were indicators of non-compliance". Li's approach allows for much more pluralism in legitimate interest articulation within a basically coherent framework: the centre and provinces are independent, at least to a degree, and, as she states it, conflict and cooperation are normal occurrences between independent actors – the two things are not mutually exclusive as in 'unidimensional compliance analysis' (Li 1998a, pp. 178-179).

In several other studies, Li expands on this analysis at the county level. One study, of financial reforms in Yichang County, shows local reforms – technically detailed, speedily executed and in advance of national legislation – not to be classifiable as straightforward 'implementation', despite congruence with national policy. Li argues that, here at least, analysis derived from principal-agent models and focusing on monitoring and compliance fails to explain what she sees: complex patterns of local behaviour, including "partial resistance and partial implementation [...] innovation and activism" (Li 2005, p. 88). She sees local cadres as motivated by a desire for autonomy vis-à-vis superior levels, leading to pursuit of economic growth, and by domestic concerns, leading them to proactively regulate subordinate entities to prevent problems and forestall the possibility of intervention by higher levels.

A study of tax-for-fee reforms comes to similar conclusions: Li finds policy making and implementation to be 'non-dualistic', and describes tax-for-fee reform as a central-local 'joint-venture', in which "[l]ocal processes in reform implementation went far beyond the 'shirking' behaviour depicted in the principal-agent framework. In the case of agricultural taxes, local actions did not sabotage the central objective of reform, but

supplemented, and even led, central policy in its fine print" (Li 2006a, p. 172). This leads Li to differentiate roles of national and local actors. Local actors carry out "decision-making *qua* implementation", in which "local actions fill in details of policy during reform implementation, as diverse local situations logically dictate. The local adaptations buttress national policy and are very much expected in any account of policy making", and this "may develop into a second strand – local role *qua* decision-makers, when the local content illuminates major weaknesses in the national policy, or when feedback from local experimentation suggests a more desirable, and alternative, direction of change nationally". The centre, on the other hand, has

"a responsibility in (re)defining the major parameters of behaviour, and specifying the new rules of the game [...] It is insufficient for the central actors to announce the reform objectives, only to leave the specific mechanisms as to how to get there entirely to the provincial and local actors, on the pretext that all details are 'routine' implementation matters" (Li 2006a, p. 173).

There is potential here, at least, for actors at different levels to play complementary roles. What is crucial is *appropriate* division of responsibilities between the centre and sub-national government.

Similarly, Goodman, following a review of fears of fragmentation in the 1980s-1990s, makes a plea not to see central-provincial relations in China as necessarily zero sum. An excessive conceptual focus on 'compliance' is, he says, a late development in China's history compared to a longer – and analytically more useful – tradition of 'unity based on diversity'. Here and elsewhere, Goodman recognises clearly that variation is necessary within the scope of broad national policy frameworks, whose goals are outlined by the centre but detailed development of which is "the preserve of local cadres". For Goodman, central-provincial relations function on an understanding that there are inevitable limits to the possibility of central control (Goodman 1994). Similarly, Dali Yang believes that "centre and localities are not necessarily engaged in a zero sum game" and that indicators such as central/local distribution of fiscal resources are crude ways of assessing central/local power (Yang 1994, p. 89). There is a degree of

continuity here: Goodman doesn't ask whether China will fragment, but rather what the role of the centre (and by implication, other systemic actors) will be at time $t+1$ in the reform process.

2.6.2 Pole 3b: Discretion and differential implementation

Christian Göbel and Jae Ho Chung have, in different ways, taken up the issue of differential local implementation of policy. Göbel's concern is to explain why one (central) policy may 'work' in one place but not in another, which he sees as unexplainable by 'dualistic' and 'non-dualistic' approaches dominating the literature on central-local relations and implementation (Göbel 2011). For him, both these approaches aim to explain how bottom-up change can occur in China: dualistic approaches (e.g. O'Brien and Li, above) explain this as the result of local resistance to top-down policy, while non-dualistic approaches (e.g. Li, Zheng, Heilmann – see below) explain this in terms of cooperation. Neither approach, for him, can explain differential implementation. Göbel's interest is principally the sub-provincial level and tax-for-fee reforms. His analysis aims to show how local cadre incentives are structured by both *hierarchy* and *competition*. Local pioneers, spurred by the promise of promotion and other rewards, engage in competition to find novel policy solutions within the framework defined by central policies and in doing so provide a reservoir of policy solutions that can be drawn on in policy development at higher levels. Laggards, meanwhile, can be strong-armed into implementing policy through use of hierarchical control mechanisms.

Of most interest in Göbel's analysis is the way he discusses structuring of the policy development and implementation process. A number of points are of interest: tax-for-fee reform was developed as a combination of a 'main reform' (*zhuti gaige*) and 'complementary reforms' (*peitao gaige*); the reform set out to combine 'unified national policy' with 'dispersed policy making' to ensure local flexibility in implementation; policies stated broad aims, but there were no clear rules as to how aims were to be achieved – policy was characterised by both 'high expectations' and 'vague

instructions'. Within this framework, reform goals were passed down the system and "the task of specifying workable reform measures eventually fell to the county governments" (Göbel 2011, p. 68). What resulted from this conjuncture was a range of local outcomes, in which some localities pioneered, developing useful policy solutions, many were punished, and various kinds of local resistance to policy occurred; national policy developed in an incremental and responsive way.

Göbel's explanation is substantively correct. All the above factors are relevant in the development of the NCMS. Criticisms of Göbel's approach are that it lays undue stress on the voluntarism and rationality of local cadres in developing policy: for Göbel, cadres innovate for the sake of their careers and material rewards – localities choose, clearly and deliberately, to cooperate or to deviate. While I do not question this analysis overall, this phenomenon is less voluntaristic, and local judgements are less clear cut, than in Göbel's analysis. Second, while Göbel acknowledges local generation of policy solutions, in his analysis we see little of these, their formation and their congruence (or lack of) with higher level policy. Third, Göbel's analysis remains *scalar* in its understanding of implementation: progress of tax-for-fee reform is shown in comparative numerical terms across counties (see Göbel 2011, pp. 65-66). This focus on numerical outputs downplays the importance of locally-developed policy solutions, or institutional *form*. I return to this in Chapter Four. These are minor criticisms: Göbel's analysis is very useful and shows important components of local implementation.

Chung is interested principally in local discretion and adaptive (*yin di zhi yi*) policy implementation, to which, as he shows, the centre has had a long-standing rhetorical attachment. In a study of decollectivisation, Chung develops a typology of local responses to policy, characterising sub-national governments as 'pioneering', 'bandwagoning' and 'resisting'. In practice, in any given case, a mixture of responses is likely. The objective of decentralisation for Chung is to allow localities to be responsive to local conditions, rather than functioning as exclusive agents of the centre: he explicitly sees local discretion as having the potential to supplement central policy. He sees the centre in the reform era as trying to combat entrenched 'centralisation norms'

– local unwillingness to make use of available policy space due to a perception of risk, fear of reprisals for deviant behaviour, possible subsequent recentralisation of policy, and similar:

“In state-socialist systems like China, centralised political control may remain relatively intact despite the radical measures of economic decentralisation. Even locally initiated pilot programmes generally require strong support from the centre in order for them to survive and be later popularized as national policies [...]” (Chung 2000, p. 175).

Norms of decentralisation have been encouraged through persistent emphasis on local adaptation of policy, less emphasis on ideology and long-standing attempts to ‘liberate thinking’ (*jiefang sixiang*) of cadres, as well as a dilution in ‘statutory precision’, shown in use of fewer, and less minutely specified, policy documents, and the granting of specific implementation choices to lower levels in top-down reforms.

In encouraging adaptive implementation, however, the centre has opened up a “Pandora’s box full of potential deviations”. Discussing a later period in the reform era, Chung sees multiple patterns of central-local relations, including “coercion, collision, collusion, co-ordination, and competition”. In his reading, the centre is bound to accept divergent local outcomes for the sake of ‘system-level gains’ (in terms borrowed from Kohli (1990), the centre accepts a reduction in its ‘control power’ in exchange for an increase in ‘developmental power’ derived from the enthusiasm of sub-national governments for economic development) and because overly active intervention in local policy development risks resetting norms of decentralisation the centre is trying hard to cultivate (Chung 2000, pp. 173-174). This is the ‘centralising paradox’: that the centre has two incompatible preferences, to “continue to tolerate local variation for the sake of decentralisation, or to assert central control at the expense of decentralisation” (Chung 2000, p. 11); tolerance is necessary when adaptive implementation is required.

Chung (2001) gives a ‘balance sheet’ of discretion in the reform era. He argues that local discretion was not achieved during the Mao era, but that the reform era has seen

a large increase in local discretion arising from increased decentralisation. The centre has been forced to admit that local under-achievement (mis-implementation and differential implementation of policy) may not be an act of willful non-compliance, and, as such, should not automatically be stigmatized. The balance sheet shows central-local relations as an extreme example of agency problems in which, though, systemic tolerance has increased where system-level gains result. The degree of tolerance can be quite high, especially in economic matters, though less in political matters. In the reform era, Chung argues, “uniform compliance has largely become a bygone concept”, with more locally-tailored policy, more bottom-heavy policy and greater regional variation. The space for local action in the reform era is increasing and the increase is durable: space is consistently greater at time $t+1$ than at time t . (Though, commenting on Zheng’s thesis, Chung states clearly that China cannot be seen as federal, as the centre retains the authority and power to unilaterally impose changes on the terms of central-local relations should it so wish.)

More recently, Chung discusses central-local relations in the context of experimental policy development in China. Here, he stresses the importance of both time and a genuine commitment on the part of the centre as conditions for change in behavioural norms, and reiterates his view that reform has been successful in bringing this about: “it took some time for the *yindi zhiyi* principle to take effect in the actual implementation of central policy”. The result is a complex picture of selective control: localities tend to comply in areas which the centre really cares about, though there may be foot-dragging in areas where localities have genuine economic interests at stake and hope to negotiate a concession, and on issues where certain localities have “crucial interests at stake”, Beijing may not require compliance (Chung 2011, pp. 303-304). Chung cautions, in various places, that decentralisation is issue-specific and that analysis must therefore also be issue-specific. Most studies of decentralisation and central-local relations, he says, have dealt with resource-intensive sectors and fiscal issues and these may not provide a basis for generalization. It is hard, but also very fruitful, he thinks, to attempt to identify and explain local policy variation (Chung 1998).

Differential implementation, as an idea, in fact has a long history in studies of the reform era. As above, implementation deficits were identified early on in the reform era, but not all this work showed divergence in implementation in an entirely negative light. Manion, for example, showed how cadre retirement policies were reshaped at the local level as 'middlemen' sought to appease those affected by the policies, even though this might mean deviating from central policy (Manion 1992). Manion hypothesised that middlemen must judge costs and benefits of policy and that "deviation that takes place in the course of policy implementation may be a useful supplement to an otherwise crude information [monitoring/feedback] system" (Manion 1991, p. 275). Solinger also addresses differential implementation, showing how economic and employment policy in three different cities during the Asian Financial Crisis demonstrated not just compliance with central policy during a period of stress and potentially divergent central-local interests, but also a shared understanding of the goals of policy combined with a tailoring of actual policy solutions to specific local conditions (Solinger 2004). In many ways, there is continuity between this and Solinger's earlier work, which showed divergent sub-national outcomes as depending on *both* the centre and sub-national governments (Solinger 1995).

A related, and extremely influential approach to policy and decision making in China is the 'fragmented authoritarianism' model. Lieberthal and Oksenberg (1988) describe China's political system as characterised by fragmentation of authority, in which resources and authority are structurally distributed across individual bureaucratic units, each with its own legitimate remit and authority, and each embedded in specific chains of bureaucratic authority. Divergence in the remits of bureaucratic units ('where you stand is where you sit'), coupled with dispersal of authority, lead to a process of bargaining and accommodation over how to achieve tasks requiring collaboration. It is clear that fragmented authoritarianism is not applicable to all levels of the Chinese system (note the incompatibility with compliance-focused approaches, as in Edin) and all policy areas, and this was a question specifically raised by Lieberthal (1992). It should also be clear that smooth 'implementation' of policy in a fragmented system is a

logical inconsistency.

2.6.3 Pole 3c: *Experimental policy development and policy innovation*

Ideas of discretion and differential implementation are attempts to account for ways in which policies produce variant local outcomes from within a non-zero-sum (e.g. Li), or positive sum (e.g. Chung), view of central-local relations. As above, Göbel mentions the possibility of local development of policy tools later taken up by higher levels in policy development, but this is not his main focus and he does little to concretely show this. Bottom-up 'innovation' and experimental policy development, however, have a certain pedigree in the China literature.

How widespread local innovation is in China is a question for debate. One prominent Chinese academic who works on China's health reforms and who I contacted during my research stated baldly that there is no local innovation in the NCMS (personal communication, 2009). In contrast, Sebastian Heilmann, whose work I deal with in detail below, believes that in the reform era, experimental policy making has been the dominant mode of policy making in China, and that localities play an important role in this. Equally, one researcher from the Central Compilation and Translation Bureau (CCTB), has described the period from the end of the 1990s to the present as a golden age of local government innovation (Gao 2008b), and the CCTB runs a *Local Government Innovation Prize (Difang Zhengfu Chuangxin Jiang)* programme to showcase and reward innovators in local government (see e.g. Chen and Yang 2009). Goodman, very early on, noted that sub-national experimentation was a major part of policy development (Goodman 1986; Goodman 1994). Concrete reform era examples are numerous: a case can be made for the household responsibility system having a component, at least, of local innovation (e.g. Unger 1985; Tsou 1986; Watson 1987), though this is contested (Hartford 1985, p. 43). According to one analysis, village elections started as a local innovation in Guangxi (O'Brien and Li 2000). Case studies of current and recent examples exist within both the English and Chinese language literatures: Fewsmith, for example, has profiled various local innovations which have

gained a degree of recognition, mostly in selection of officials (see Fewsmith 2006; Fewsmith 2008b; Fewsmith 2008a), while Saich and Yang have examined development of township elections in Buyun (Saich and Yang 2003), and Foster has shown city-level public service innovation in Yantai (Foster 2005).

Among early reform era studies, and departing from studies broadly within the fragmented authoritarianism framework, an insightful study by Paine shows how local education policy makers find their way through 'groping' (she takes the term from Behn (1988), to render the Chinese term '*mosuo*') in a loosely-structured national policy space. Paine describes curriculum development as an issue "looking for a policy", showing how broad Ministry of Education plans were carried out locally, with "local institutions [having] to decide what [plans] meant in curricular terms" (Paine 1992, p. 188). She describes this policy process as characterised by "local interpretation, mutual adaptation, and policy fluidity"; policy documents tend to be consultative (*cankaoxing de*) and "act as guidelines rather than regulations. Without enforcing power, the guidelines allow for some measure of autonomy [...]". Overall, discretion and fluidity allow continual recasting of policy in an interactive process:

"under the guidelines of these vague policy discussions, this reform policy has had fluidity that allowed it to change over time. Typical of the groping process, the current standards represent an evolutionary compromise between the broad objectives of the central bureaucracy [...] and specific experiments of local experience. Policy is recast by those carrying it out [but] those involved are relatively weak actors with limited power to revise policy. Groping is characterised by responsiveness, as implementation proceeds alongside the process of continual formulation. This policy process is iterative" (Paine 1992, p. 193).

Heilmann has brought a lot of attention to policy experimentation and innovation, conceiving of China as a "learning authoritarian state" in which experimentation in policy development underlies 'adaptive efficiency' (the term is from North; see below). For him, this is the most important mode of reform era policy development:

“In most domains of China’s economic reform since 1979, policy changes were produced by a process in which the central policymakers encouraged local experimentation to generate novel policy options that could then be fed back into national policymaking. Experimentation served as a crucial means for avoiding policy deadlock and reducing the frictions and delays that are characteristic of top-level consensus-building and inter-agency accommodation. It helped to reduce risks in policymaking, stimulate policy entrepreneurship, and contributed to a fine-tuning of policy content and implementation” (Heilmann 2008a, p. 2).

A number of points in Heilmann’s account are crucial. Policy development occurs in what, in policy cycle models (see Hill and Hupe 2009), can heuristically be considered the implementation phase. In contrast to models derived from developed country experience, there is little *ex ante* modeling or legislation: policy development occurs during implementation; codification and legislation occur late in the policy cycle. Objectives of reform are the business of the centre; sub-national government is the locus of actual experimentation and innovation, exploring and developing policy instruments and tools, as “policy adaptation is built into administrative practice and made a permanent enterprise. It is based on the administrative discretion and entrepreneurship exercised by local officials” (Heilmann 2008a, p. 4).

One consequence is that local policy development frequently occurs in a grey zone of dubious legitimacy. For Heilmann, the authoritarian nature of the Chinese state (and, implicitly, the lack of *de jure* sub-national bases of authority) requires specific systemic mechanisms, including experimental regulation, experimental points and zones, and patronage, to legitimise local activity which might otherwise be seen as deviant. Experimental regulations function to encourage local innovation, and titles of policies make reference to their provisional nature through use of terms such as ‘trial’ (*shixing*), etc. As with Paine, regulation tends to be consultative, rather than providing concrete stipulations. ‘Experimental points’ (*shidian*) are used to carry out experiments, with sub-national units selected for controlled trialling of specific measures aimed at

rigorous testing of policy tools from a range of reform measures and assessing their usefulness for wider-scale application. Experimental zones allow normal rules to be suspended for the purposes of policy experimentation. According to Heilmann, specific permission is required for experiments (*shidian*), and authority to grant this is situated in dedicated offices at the provincial level, where it was moved in 1998 from a centralised national office in order to prevent this office being swamped by the increase in volume of experimental work (Heilmann 2008c). In addition, local innovators require high-level backers to hedge against the risk inherent in innovating, and to help disseminate local innovations (Heilmann 2009).

There is a tension in Heilmann that derives from his not fully separating two closely related concepts, 'experimentation' and 'innovation', though he discusses both. Early Heilmann sees experimental policy development as a highly controlled, rationally conducted and a specifically *experimental* process of comparative trialling of policy options:

"Experimentation [...] implies a policy process in which experimenting units try out a variety of methods and processes to find imaginative solutions to predefined tasks or to new challenges that emerge during experimental activity. Policy experimentation is not equivalent to freewheeling trial and error or spontaneous policy diffusion. It is a purposeful and coordinated activity geared to producing novel policy options that are injected into official policymaking and then replicated on a larger scale, or even formally incorporated into national law" (Heilmann 2008c, p. 3).

For Heilmann, this derives from a conception of a "well-conceived process of practical experimentation" in policy development (Heilmann 2008b, p. 19), and he points to opposition to this mode of development by (rationalistic) 'neoclassically oriented Chinese economists' (Heilmann 2008c). In later work, Heilmann's approach softens, and he discusses policy experimentation as less deliberate and purposive. Citing Taleb (2007), Heilmann hypothesises that adaptability depends on how well systems allow decentralised 'tinkering', understood less as purposive, than as unprogrammed,

exploratory, activity. Allowing tinkering allows the system to capitalise on chance. Tinkering, as an axis of analysis, is, he thinks, more important than more common distinctions between plan and market in thinking about system adaptability (Heilmann 2009).

This duality leads to confusion as regards the role of higher levels of government: as instigators and guides of reform, versus filters for useful and generalisable experience. Locally-generated policy solutions are replicated, says Heilmann, in a way described as 'point-to-surface' (*cong dian dao mian*), and this overall mode of policy development has changed little since the early days of the PRC and before. Heilmann and Perry describe this as a distinctive Chinese mode of 'guerrilla policy making' (Heilmann and Perry 2011). Policies are "made and re-made endlessly. Policy-making is a process of successive approximation to some desired objectives in which what is desired itself continues to change under reconsideration" (Heilmann 2009, p. 452, citing Lindblom 1959).

In line with Heilmann's later tinkering thesis, decentralised development is a strength, since "this method is decidedly open with regard to the means of reform, it naturally leads to decentralized initiatives that cannot be anticipated by the center" (Heilmann 2008c, p. 12), though 'best practice' solutions are rarely the result.

The two poles of Heilmann's thinking are reflected in his changing formulation for the mode of policy development he describes, from 'experimentation under hierarchy' to 'maximum tinkering under the shadow of hierarchy'. Both formulations attempt to provide a framework for action at sub-national levels while showing that the centre remains the guide to the direction of reform and legitimacy of specific policy solutions. The first emphasises scripted development; the second emphasises the law of large numbers and the benefits of a degree of productive systemic fragmentation in development of policy mechanisms.

Heilmann is at pains to say that he sees Chinese policy making as neither top-down nor bottom-up: "distinguishing between bottom-up ('spontaneous') and top-down

('mobilization-style') initiation of experimentation is nearly meaningless since there is a strong element of both, local initiative and central sponsorship, in the initial stages of major experimental efforts" (Heilmann 2008c). He is correct to be wary of top-down vs. bottom-up dichotomies, but this element of Heilmann's analysis requires refining. Practically speaking, within the NCMS, it is possible to see both highly managed experimental policy development (*nominally*, top-down 'experimentation') and a range of locally-generated policy solutions (*nominally*, much more bottom-up and 'innovative'). A second criticism is analytical. Heilmann's position is reminiscent of a statement by White to the effect that, as an analyst, understanding whether the intentions of policy formulators have been expressed in actually implemented policy is hard (White 1997, p. 167): the question is how one should judge something as intangible as intentions. The implication, in Heilmann, is that there is little point in trying. This is wrong.

First, Heilmann's discussion of, firstly, a structured experimental mode of policy development and, secondly, of capturing the gains arising from tinkering, represent two different phenomena in the harnessing of bottom-up activity or creativity. In the first case, specific jurisdictional lines are drawn to allow either controlled experimentation or rule breaking (supplemented by informal mechanisms of patronage, etc.). In the second, natural systemic fragmentation and, as I shall argue below, the structuring of control and communication in policy implementation, inevitably produce local variations and innovations which can be captured as 'system-level gains', in Chung's term. The mediating concept here is *discretion*. Heilmann sees the Chinese system as authoritarian and hence seeks to explain how rule breaking can occur without being punished. In other words, he sees innovation and deviation from the norm as risky. On the other side, Chung seeks to show that there is an increasing degree of *de jure* space for flexible implementation, and that norm change – actual acceptance of this *de jure* central commitment to local discretion – is the key to unlocking *de facto* local creativity. This parsing of risk reflects a dominant trend in the literature, which tends to

see deviation, including flexible implementation, as risky.¹¹

Second, framing experimentation/innovation as a bottom-up vs. top-down question, while the prevalent framing in the China literature, is a red herring. In this analysis, in refusing to state whether specific policy development dynamics are, schematically speaking, bottom-up (innovative) or top-down (experimental), Heilmann refuses to be bound by limited analytical concepts, but remains bound by the overall framing of the problem. A more productive framework for analysis is to ask whether local reforms can be seen as 'within the system' (*tizhi nei*) or 'outside the system' (*tizhi wai*). I discuss this in Chapter Nine.

Heilmann's work is an extremely valuable contribution to analysis of policy development in China. A few general points deserve mention, however. First, and overall, Heilmann's work is a very serious engagement with the development of institutions in the reform era and how this is mediated by state structure. This goes a long way towards filling a lacuna in the literature. Related to this, though, Heilmann does not seriously deal with the notion of *control*, though he posits hierarchy and 'shadow of hierarchy'¹² as the means ensuring implementation. In common with, for example, Göbel, local innovation in Heilmann is voluntaristic: deliberate

¹¹ Risk in policy implementation is not foregrounded by most analysts, but remains present within the literature. The clearest statement of this is, in many ways, Chung's discussion of local deviation in pre-reform China, when he discusses local 'boldness' and central 'vengefulness' and argues that local compliance with central mandates was "rigid beyond casual observation" (Chung 2000, p. 40). Similarly, Oi recognizes that strong local and central states are complementary, but that when it wants to be, the central state is "every bit as awesome as the totalitarian model would predict" (Oi 1989, p. 229). Both analyses carry over understandings of risk and local deviance from a previous era, but this framing is often retained in contemporary analyses. One recent article, for example, assumes risk in local government political innovation (He and Thøgersen 2010), though one should note that Zhang Jinming, behind one of the most systemically challenging recent local reforms, the introduction of direct township elections in Sichuan, later declared unconstitutional (Saich and Yang 2003), was subsequently promoted within the Party, to no less than head of the Ya'an Organization Department (He and Thøgersen 2010).

¹² 'Shadow of hierarchy' is used to describe the *possibility* of being observed and application of hierarchical sanctions should something go wrong. The term is from Scharpf and is introduced by Heilmann in *Maximum tinkering* (Heilmann 2009).

experimentation by local proxies of the centre aims to find solutions to policy problems. Heilmann spends little time on motivations, perceptions or conduct at the local level, or how experimentation fits with, for example, the CRS and target-setting by higher levels of government, being more concerned with system-level effects.¹³ Heilmann's analysis frequently sits alongside, rather than being integrated with, other literature on central-local relations and implementation. Secondly, Heilmann's discussion of experimentation, and risk, should be considered a *special case* of discretion in policy making. If one agrees with Lieberthal and Oksenberg that to a certain, though unspecifiable, extent, all 'nodes' within the Chinese system have a degree of autonomy, *by design*, then adaptation of a specific policy during implementation is inevitable. A third point is that Heilmann's analysis is retrospective – deliberately so. Heilmann and Perry and their co-authors in *Mao's invisible hand* start with "prominent features of China's contemporary political scene and work backward in search of their (often tortuous) historical origins". This allows them, they say, to sidestep teleological searches for, for example, "a 'real' market economy or 'real' democracy" and to seek to explain the actual genesis of the somewhat idiosyncratic institutions that China seems to produce (Heilmann and Perry 2011, pp. 9-10). While this is useful in making visible the tortuous paths to the present, analytically it necessarily privileges the winners in China's institutional development game. If you start from the end and work backwards, you risk missing the multiple dead ends, false starts and the like that are part of this process.

Wang Shaoguang's analysis of the NCMS suffers from this defect. Wang examines development of the Cooperative Medical Scheme from its inception to its

¹³ This does not imply complete absence of local analysis: Heilmann thinks, for example, that experimental policy making can be effective where it plays to the interests of local actors (one of his main foci is economic development), but is sceptical of its usefulness where it conflicts with entrenched interests, particularly in provision of social goods. Under this rubric, Heilmann discusses rural health reform, arguing that this was only really propelled onto the policy agenda by SARS, and that a previous failure to address this issue was due to the lack of benefit to be derived by local elites from experiments in this area (Heilmann 2008c, p. 19), and he cites the failure of 1990s CMS pilots as evidence of this. Note that this is taken up and contradicted by Wang (Wang 2011).

transformation in the early 2000s into the *New Cooperative Medical Scheme*. He does an excellent job of piecing together this story, but the conclusion is disappointing: “The NCMS is of course imperfect and fraught with problems. Therefore, additional experimentation has been conducted since its introduction [...] Thus after nearly sixty years of development with a number of twists and turns, a cooperative health-care system finally reached an all-time high” (Wang 2011, p. 125). The problem is methodological: seeing the NCMS as one national system, and its history as a process of development to its apogee, visible today, privileges a national and coherent evaluation and narrative. In doing this, Wang privileges the narrative over the thing. As Behn says, “When any management story is told, the emphasis is on premeditated and purposeful action rather than on any groping [...] The chronicler [...] is looking for interesting lessons – lessons that can be found in the manager’s intelligent and flawless (or misguided and inept) forecasts, decisions, or actions” (Behn 1988, p. 653). This is the pitfall to be avoided in retrospective analysis. A second problem is that Wang’s focus on *the* NCMS confuses an overall umbrella policy with actual policy *mechanisms*, the level at which experimentation and innovation is most important (see Chapter Five).

2.6.4 Pole 3d: *New departures*

Here, I highlight a number of recent discussions of ‘practice-oriented’ approaches to Chinese politics. In one, Jiang discusses what he calls China’s ‘effective constitution’ as a supplement to understanding the country’s ‘normative constitution’ (Jiang 2010). His analysis criticises Chinese constitutional studies for focusing excessively on the written constitution and texts. Jiang argues that a separate tradition of constitutional analysis, focusing on the ‘unwritten constitution’, the set of norms and assumptions within which the written constitution functions, is more appropriate and informative. As O’Brien comments, “Jiang’s article is first and foremost a critique of formalism. He adeptly draws readers away from the standard question ‘Is the state constitution ‘fake’ or ‘widely abused’?’ to the more pressing issue of what China’s constitution is” (O’Brien 2010, p. 82).

Similarly, Zhou examines local 'collusion', deviation in policy implementation and local policy interpretation. There is a tension in his analysis between collusion, proper, in which local levels come together to deceive higher levels to evade monitoring and similar, and local variation in policy implementation, seen as inevitable due to the centre's monopoly of policy formulation. This monopoly, he says, leads to broadly formulated national policy (inevitably, he says, as central policy must be applicable in widely different contexts), and, equally inevitably, to local interpretation. He calls this the 'paradox of uniformity in policy making and flexibility in implementation'. Flexibility, for Zhou, is a neutral term, and can be both good and bad, including "adaptive behaviors that lead to a better fit of policy goals and local conditions as well as those deviations from intended policy goals due to interest conflicts, political sabotage, or other circumstances" (Zhou 2010, p. 57). Flexibility in implementation is both spatial and temporal:

"Spatially, we find considerable variations across regions and arenas in the implementation of the same policy, and such differences are accepted by the supervising governments. Over time, we also observe significant, temporal variations in the effectiveness of policy implementation in the same locality" (Zhou 2010, p. 57).

Various discussion articles deal with themes from these two studies, of which several are particularly relevant. O'Brien draws out the salience of Jiang and Zhou's approaches: both examine "conventions, understandings, habits and practices" and, citing Zhou, "[i]nformal, but highly institutionalised practices" in the functioning of Chinese politics, in addition to laws, regulations, texts and the like (O'Brien 2010, p. 80). Of Jiang, he says that his "main interest is practice rather than textual exegesis" (O'Brien 2010, p. 82). Both studies take seriously an attempt to explain minutiae of administrative practice. Similarly, Huang discusses Jiang and Zhou's approaches in terms of a 'history of practice': most important is to focus "on operational realities in order to uncover the underlying logics of such" (Huang 2010b, p. 6). In a related article, Shue, rather flamboyantly, says that "in their day-to-day performance of the arts of

governance [...] state actors in the Chinese system, like actors in any good repertory company, have rarely if ever been able to adhere to some single, uniform, or preselected script for rule”, and that in China, “the idea of state authority is then, perhaps, always singular, transcendent, and universal; the realization of rule, however, is always plural, rooted, and particular” (Shue 2008, p. 141).

The relevance of this for my study is highlighted by O'Brien, who draws on Zhou to say that “weak policy implementation literature in China is misdirected”, and that the range of practices, formal and informal, evident in implementation are worth taking seriously as the ‘glue’ that makes the system work, rather than necessarily as evidence of deviance. Flexibility is ‘baked in’ to the way the system functions – indeed it is an ‘open secret’ of Chinese politics. For O'Brien this implies that the “notion of design and its purposefulness [...] should not be reified [...] it is nearly impossible to encourage only ‘good’ flexibility while eliminating ‘bad’ flexibility” (O'Brien 2010, p. 81).

2.7. Attempting synthesis

This chapter has reviewed approaches to central-local relations, local government, and policy implementation in China in the reform era. The weakness of an approach covering so many, and such varied analyses, is a lack of substantive analysis of the content of the reforms under discussion. The advantage is to show a basic analytical cleavage within this large body of literature and how this has been dealt with. My aim has been to show that analysis of central-local relations, local government, and policy implementation has tended to be dominated by analyses framed in terms of control, whether this is loss of control (Pole 1) or dilution of control, but maintenance of adequate systemic integration (Pole 2). A third pole tries to understand central-local relations in non-zero-sum, or positive-sum terms, centering on analyses of differential implementation, discretion and experimentation in policy making and implementation.

All of the analyses above are useful, but leave much unexplained. This is recognised in many places: the realisation, for example, that fragmented authoritarianism, which originated in analysis of resource-intensive sectors where bargaining was the norm,

might not hold for other policy areas; the caution against over-reliance on fiscal indicators as measures of central-local distribution of power; Chung's repeated insistence that decentralisation cannot be treated as a blanket phenomenon, and that different policy areas must show different dynamics; the different paths and outcomes of development and reform in different sub-national areas, whether reflected in varying provincial strategies or different local development models; the contradictions between, say, the fragmented authoritarianism model as seeming to describe upper levels of Chinese government, and the equally useful analysis of the CRS, which seems to show how compliance of local agents with policy is attempted; the apparent contradiction between Heilmann's picture of an innovative polity, in which cooperative sub-national actors develop useful policy solutions, and analyses which stress mis-implementation. All analyses discussed here are, inevitably, partial pictures – reflections of specific phenomena at specific times and places and in relation to specific policy areas or components of reform, but inevitably limited in their generalisability.

To recapitulate, in development of the NCMS, a range of sub-national practices can be observed, and there is clear central attachment to local government policy innovation and decentralised initiative (see Chapter Five). I ask the following questions:

- 1) Given a clear and specific central rhetorical attachment to local government policy innovation in development of the NCMS, is there evidence that this exists?
- 2) If local government policy innovations are in evidence, what motivates innovation? What kinds of innovations are produced?
- 3) Given that the majority of analysis of local government policy behaviour is analysis of compliance and mechanisms used to ensure this, what mechanisms explain local government policy innovation in an existing central government policy area such as the NCMS?
- 4) How do local policy innovations relate to policy, reforms and discourses at larger scales in the Chinese policy system, including having the potential to contribute to

policy at supra-local scales, including central policy?

For the purposes of this study, the body of literature analysed here has many strengths, but also weaknesses. Analyses that exclusively stress local governments' 'compliance' with central policy arise from a specific historical conjuncture. The overall picture visible to analysts of government, policy and policy implementation in China today shows much greater complexity than can be accounted for within this analytical framework. Elements of this framework remain highly useful, however: analyses of the cadre management system (e.g. Edin 2003) are important in understanding one of the main central mechanisms structuring policy implementation, and I return to this in subsequent analysis.

Overall, though, this framework is limited in the extent to which it can explain local variation under national policy schemes, or differential local implementation, as can be observed empirically in the case of the NCMS (and other policy spheres). It also fails to explain sub-national agency, except as a function of national policy implementation pressures. Where sub-national agency is allowed for, it is seen as negative and running counter to the interests of the centre (e.g. O'Brien and Li 1999). This results in an unhelpful polar opposition between 'good' national government and recalcitrant sub-national government.

For the purposes of the present study, analyses grouped under Pole Three offer greater explanatory power. Overall, these are all attempts to reframe debate and analytically deal with sub-national agency and show the existence of discretion or freedom at sub-national levels. These analyses are very helpful, but have limitations.

Heilmann has done more than anyone else to show experimental and policy making and policy innovation in China. This gives an overall framework for analysis of how the Chinese state carries out experimental policy development, showing that localities play an important role and that much policy development is bottom-up, though I disagree with Heilmann on certain points, especially his scripted, specifically experimental, work which fails to capture the vibrancy of what is visible empirically. Heilmann's

system-level focus also offers little understanding of how localities function within experimental-innovative policy frameworks.

Chung and Göbel, in different ways, go a long way towards explaining the existence of sub-national differential implementation of national policy. Common to all non-zero-sum analyses is attribution of a degree of agency and discretion or freedom to sub-national units. If localities have a degree of freedom, though, how do they use this? This is a problem, given that analyses of the local level have tended to focus on control and compliance.

Both Li and Göbel (Li especially), go some way towards showing the thinking of local cadres in the way they use, rather than simply implement, policy, but this picture is limited. Zhou and Jiang, and an emerging discussion of practice-oriented approaches to Chinese government (e.g. O'Brien 2010), provide an analytical approach privileging 'informal, but highly institutionalised practices' over formal descriptions of the way that policy/government work.

One critique is common to almost all analyses discussed here: with the exception of Li, all implicitly examine policy development from a central perspective. For example, Chung's analysis is phrased in terms of pioneers, bandwagoning and resisting (within central policy), while Göbel adopts a scalar measure of local burden reduction. There is an implicit rationality assumption underlying these analyses. In effect, they suppose that local government voluntarism and cooperation are sufficient to ensure good implementation outcomes, and that localities can successfully implement policy if they choose to; the problem is that they don't always choose to. These pay no attention to either (widely varying) levels of capacity or to local conceptualisations of appropriate and feasible government roles, or to the range of reforms that can be envisaged or imagined by local governments.

Missing from almost all studies examined here is any substantial discussion of the types of institutions that are developed through the processes described by these analysts. As above, Göbel's discussion of burden reduction is scalar. This is significant.

Göbel's analysis of the way in which implementation of the burden reduction policy was structured is very persuasive: he shows clearly the way in which incentives are structured at the sub-provincial level and the combination of both pressure to implement and uncertainty over the form that implementation should take – reforms, he says, are characterised by 'high expectations' and 'vague instructions'. In evaluating burden reduction in scalar terms, however, Göbel overlooks a large part of the significance of the structures he is observing: that in the gap between pressure to implement and the ontological indeterminacy of policy is a substantial engine for the production of a range of local practices. Heilmann comes closest to a consideration of the actual types of institutions produced in experimental reform processes when he comments that these are rarely 'best practice' solutions. As discussed above, however, a limitation of Heilmann's analysis is a lack of detailed attention to local reform processes.

A final comment: with Chung, it is important to be clear that decentralisation, central-local relations and policy development cannot be studied in the abstract. The ways in which county-level policy implementation and development are structured within the NCMS will inevitably differ from other areas of policy (though, intriguingly, Paine paints a picture of local reform processes very close to my own). It is inevitable that existing analyses will need to be tailored, at least, to fit the NCMS.

2.8 Looking forward

My theoretical analysis is set out over the following chapters. Chapter Four provides the building blocks of this, and this is completed in Chapter Nine. Intervening chapters first present an analysis of NCMS policy development, followed by three case studies: two county policy innovations, and one county reform which transitioned to national policy. As set out in Chapter Three, these should be considered one case; all three case studies form part of one overall argument. This conforms to what Heimer (2006) terms a 'one-case multi-field-site' approach.

Overall, I argue the following:

- 1) I argue that local 'compliance' with central policy is both a theoretical impossibility, and in fact should not be expected, given a clear central attachment to decentralised initiative. A focus on policy innovation is the expression of an older current of thinking on central and sub-national roles clearly visible in Mao (1977) and Schurmann (1966).
- 2) I argue that analyses of local government innovation in China have relied on very restrictive framings of 'innovation', almost exclusively seeing this as disruptive, radical, and system changing, at the expense of consideration of a much larger volume of non-disruptive, 'marginal' (Bessant 2005) innovation. This framing leads to an unwarranted analytical separation between implementation and innovation, which are much closer than is commonly accepted.
- 3) I argue that Chinese central government attachment to decentralised initiative can be understood by reference to Hayek (1945), Schumpeter (1934) and North (2005). Local governments, as frontline policy implementers, possess knowledge of local circumstances "not given to anyone in its totality" (Hayek 1945, p. 519). A central commitment to limited decentred decision-making is intended to capitalise on this. The fact that this should be considered a developmental function, and not simply one of local tailoring of policy (*yin di zhi yi*) is in the potential for implementers-innovators to develop new 'combinations' (Schumpeter 1934), or institutional form, in the process of policy implementation. The importance of these is as valuable lessons or 'experience' (*jingyan*) that can be recycled systemically in policy at supra-local scales. This is, ultimately, a developmental question and part of developing novel policy solutions and ensuring institutional adaptation (North 2005).
- 4) In contrast to the dominant analyses of Chinese policy and structuring of cadre management examined earlier, a separate analytical tradition exists, allowing us to see both of these as generative, not just restrictive, structures. The cadre

management system, and the textual construction of policy are, at least in the cases I examine, decidedly open as to the means of reform, if not the ends. This is a formalisation of my analysis of Göbel, above, in which I state that his description of the policy implementation structure in terms of 'high expectations' and 'vague instructions' is an engine for the production of differential institutional form by sub-national implementers (see also below on Eco 1989). As in point two, above, this points to the need to consider 'implementation' and 'innovation' side by side: this system enforces implementation, but leaves the form of that implementation unspecified. Innovation as *process* – localised production of (differential) practices – is forced by this system.

- 5) 'Innovation', aside from use of this term in Chinese policy discourse, which should be considered a special case, should be understood in two ways, as both process and result: as process, 'innovation' corresponds to what Schumpeter describes as operating "outside the boundary of routine"; as result, it corresponds to the creation of new 'combinations' (Schumpeter 1934, p. 84). Seeing counties as innovators in policy processes does not mean that everything counties do should be seen as an 'innovation', if what we mean by this is, broadly speaking, something 'good', or corresponding to 'best practice'. It is a way of re-framing county roles (see p. 76).
- 6) I argue that the structuring of implementation/innovation visible here is likely to create a range of local policy practices. This is an inevitable result of counties' working without an exact template for reform. Some resulting practices (innovation as *result*), will be useful and have the potential for application at a greater scale than that of the locality which developed them – *supra-local* applicability. The majority will not. Decontextualised judgements of 'innovativeness', or potential usefulness of policy solutions, rely on a form of technocratic thinking which is of less use, especially in low capacity, peripheral and changing contexts, than local and contextually-situated judgements. We clearly see such judgements being made in the cases examined in subsequent chapters.

Empirical chapters show the following main points:

- 1) An elaborate structure, straddling the health and finance systems (*xitong*) is in place to ensure implementation of the NCMS. I analyse this in terms of the 'pressurised system' theory (He 1997; Rong 1998; Rong 2009), which serves to make clear the degree to which implementation is structured through the systemic localising of risk with implementing units. Implementation, in and of itself, implies risk.
- 2) County reforms described in detail show several main points. First, these are part of a process of innovation to the extent that the counties in question have no template for action and are operating "outside the boundary of routine", carrying out exploratory local policy making. Second, neither of the principal counties examined deliberately set out to blaze a trail; rather, their innovation processes are part and parcel of the process of policy implementation. One implies the other. Third, in both county cases examined, there is a high degree of systemic (province and city) tolerance of county innovations. This county behaviour is expected and tolerated, and not seen as radical or system-changing. Fourth, in both county cases, there is a degree of supra-local impact and propagation. While this is limited, it shows a degree of systemic recycling of local practices deemed to be useful. This process is seen most clearly in Chapter Eight, in which I use a 'back to time line' (Li 2006b) analysis to show transition of a minor, though systemically significant, process innovation from county to national policy. Fifth, both county cases show innovations that arise in situations of low capacity and peripherality. The importance of this is developed fully in Chapter Eight, where I describe the Second Round Reimbursement policy as an 'appropriate' innovation and tie this to its development in a low capacity and peripheral environment.

Following presentation of empirical material, I further develop the theoretical argument.

Chapter Nine argues the following:

- 1) I argue that the empirical chapters show a high degree of consistency of

understanding of roles across levels of government I principally deal with (county, city, province), and that counties are systemically expected to be, and recognise themselves as being, in the position of developers of policy, not simple implementers of policy. The kind of marginal process innovation (Bessant 2005) shown in this study is low risk – especially when set in the context of a ‘pressurised’ policy implementation system which enforces implementation through the localisation of risk – and is systemically accepted and encouraged. This is a reversal of conventional judgements of the significance of government innovation in China.

- 2) Innovation as result is concerned with a question of form: useful institutional form, or “new ideas that work” (Mulgan and Albury 2003). The motor for the production of institutional form in my case is the under-specification of the form of implementation in a ‘policy of principles’ (*yuanze*). I formalise analysis of this, using Eco (1989) to show how Chinese policy texts are constructed as ‘open works’, which create an “expectation of the unpredictable”.
- 3) At any one time, any number of ideas regarding concrete means of developing and implementing policy may be in existence. These are not confined to content of top-down policy instructions; rather, they exist in a ‘conversation’ (Gudeman and Rivera 1990) of reform, which is comparatively open as to possible forms of implementation.
- 4) Implementers’ understanding of feasible ways to implement policy or carry out reform is not simply a question of rational calculation (as in analyses of policy implementation I critique above as relying on an implicit rationality assumption as to the possibility of successful implementation), but rather is conditioned by their starting point and their rationalisation of reform possibilities in terms of ‘repertoires’ (Behn 1988) of action (as shown in both county analyses), situated in relation to China’s ‘imaginative geography’ (Said 2003).

Chapter Ten concludes, and questions the effectiveness of this form of structuring of

policy implementation, and the extent to which local governments are able to rise to the challenge of solving problems thrown up at the coal face during China's health system reform. Before this, the next chapter sets out research methodology.

Chapter Three

Methodology: Looking for innovation

3.1 Introduction

My starting point in researching the NCMS was to question how local governments innovate in social policy schemes and the importance of this for national reforms. I hoped to carry out broadly ethnographic research based on long-term immersion, interviewing and observation in one, or possibly two, counties, and a case study approach seemed appropriate, given my focus on local innovation processes. In making this choice, I was less concerned with generalisability of my findings than with generating a detailed understanding and rich description of local processes, an absence in existing literature, and relating this to theory. I discarded the idea of large-scale comparative approaches as impossible, for reasons of access and time, and as methodologically inappropriate: how would one collect comparative data on processes and on perceptions of local implementers and reformers? A survey approach might yield data useful for such a comparison, but would sacrifice depth – if it were possible. The best way to understand local innovation processes, I reasoned, was to examine a small number of cases in depth. On this basis I set out to look for innovative local practices.

Looking for 'innovation' in management of the NCMS was a specific (though misguided) methodological choice: I set out to look for what, at that time, I considered extreme cases (Bryman 2001, p. 50; Flyvbjerg 2006) which I hoped would illuminate a specific category of local government behaviour that the China-related literature terms 'innovation'. My reading had persuaded me that local governments had a degree of leeway within the NCMS, and I tested this proposition on a number of academics during early desk research. All English language literature, and the limited Chinese sources I had at that point been able to acquire gave only a very limited indication of what variation in local implementation might actually look like. This reading combined

with a preconception derived from existing literature on local government innovation in China as something largely radical or system changing and pushed me to look for outliers in scheme implementation and to the process of 'top-down' profiling of local practices described here. As I later came to understand, and as I have argued, the border between implementation and innovation is, in China at least, porous, and a more directly 'bottom-up' research strategy, had this been possible, would have likely been as fruitful (and possibly more fruitful) than the one employed here. It would certainly have been easier. I was extremely lucky that this research approach succeeded: the chances of getting access to places I had profiled in advance and targeted as interesting and significant seem now so vanishingly small as to make this a very risky strategy.

The next section discusses the process of selecting cases and describes the iterative nature of the research process. The following section discusses the implications of this, both practical and conceptual, for my research. The final section discusses access to fieldwork sites, interviewing and the use of documents.

3.2 Looking for innovation

3.2.1 Selecting cases: Looking for 'innovation'

My theoretical starting point conditioned my approach to finding local innovative practices. My profiling of local innovations was entirely top-down: I made very extensive use of Chinese sources, through which I identified localities deemed (by government, academics, etc.) to be significant or innovative. I spent a long time reading about local practices in policy documents, MOH briefs, news reports, and specialist websites, and trying to gauge their significance. This method of selecting cases was less rigorous than I would have hoped: in the absence of a developed understanding (especially in early stages of research) of the range of county practices, determining significance of specific reforms was hard. As I argue in Chapter Five, though, a comparative understanding is largely unachievable. In Scott's term, I lacked a 'native tracker' – someone familiar with the lie of the land who would point out the

most significant navigational features (Scott 1998). I had conversations with academics, and these were helpful in orienting my research, but I had little guidance on selection of specific local reforms. The reforms I ended up observing were targeted through this method of finding local practices accorded significance by authority, doing background research on these, and attempting to gain access to carry out fieldwork. The final criterion in all cases was whether I could visit the county to carry out fieldwork, and this is an inevitable factor in any study of this kind. In all, my research converged on what Heimer has termed a 'one-case multi-field-site' approach (Heimer 2006), in that, while it addresses several principal local cases, these are manifestations of one case – local government innovation within the NCMS.

3.2.2 Policy discourse and case selection

In adopting the approach described here, in looking for *chuangxin* (innovation), my case selection was inevitably mediated by Chinese policy vocabulary, and my profiling was complicated by the ubiquity of the term *chuangxin*/innovation (and near homologues) in Chinese policy discourse. It takes time to start understanding the significance (rather than putative definition) of terms such as this. Understanding significance requires understanding how frequently terms are used, what they are used in relation to, who uses them and why, whether they are intended to refer to something of significance or relatively minor, and so on.

Chuangxin is a key word in current Chinese policy discourse, and can mean many things. Used by the centre, it is mainly an invocation to localities to be creative in carrying out policy (similar to *jiji* ('enthusiastic') and *yin di zhi yi*) and a marker of practices deemed to be 'good'. Used by localities, it is a signaling device. *Chuangxin* is overused, and many local practices are labeled *chuangxin* either by the locality in question, in an effort at positioning, or by some higher level of government wanting to confer legitimacy on a specific local model. What, as an outsider – especially in the beginning stages of research – one reads about as a *chuangxin*, may or may not be of significance outside the locality whose practices are being described. This is hard to

know in advance. A wonderful indicator of how the word is overused came at a large dinner in one county where I carried out fieldwork: a senior member of the BOH satirically described *chuangxin* as making egg fried rice ... but having the egg on the side. In other words, the exhortation to *chuangxin* is overused and many things of limited significance or originality can, in fact, be labeled *chuangxin*. There is little rigour in the way the word is used. The same applies to the panoply of near cognates: overall, development of the NCMS is described as a *tansuo* ('experimentation' / 'investigation'), indicating, broadly, that it is uncertain how the process will unfold. Specific county practices, though, can be described as *tansuo*, with much the same meaning as if they were described as *chuangxin*. Seemingly more rigorous is *shuaixian* ('first', as in 'the first to'): a county can be described as the first in the country, province, etc., to develop a certain practice (e.g. '*quan guo shuaixian fazhan*'), for example, and it seems that this generally does connote a degree of 'real' newness.

3.2.3 Profiling

In profiling local reforms, I used a range of sources. The most important of these were:

- reports, news items, speeches, etc., sourced from MOH and sub-national BOH websites; Ministry and provincial yearbooks;
- MOH rural health policy 'briefs' (*Nongcun Weisheng Gongzuo Jianxun*): extremely useful near-monthly short reports on national policy and local reforms; this series was started in 2004, reaching its hundredth issue in May 2010 (*Weisheng Bu Nongcun Weisheng Guanli Si* 2010). This is replicated, though possibly not uniformly, at sub-national levels;
- academic sources, including Blue Books on social development, Green Books on health, as well as book-length academic studies; the National Library in Beijing, which houses many otherwise inaccessible sources; CNKI (China National Knowledge Infrastructure), an extremely useful source of background information and analysis on local reforms;

- national and sub-national media; specialist websites, such as China Health Reform (*Zhongguo Yigai Pinglun*), China Reform (*Zhongguo Gaige*), etc.;
- discussions with experts.

This approach generated a large range of potentially interesting sub-national practices. Profiling revolved around a number of things. First, I targeted local reforms of possible supra-local significance: reforms that appeared to deal with acknowledged weaknesses in the NCMS. In this, my thinking was guided by academic analyses, policy pronouncements and speeches. Second, I was biased towards pro-poor reforms, such as those in Yinchuan (see Chapter Five), though finally I was unable to study these. I spent a lot of time profiling sub-national reforms that are not included in this study, and the selection of cases could have been very different, reflecting the vast range of sub-national practices. The number of possible foci is huge, and if I had examined different local reforms I may well have come to different conclusions. At the same time as profiling local reforms, I sought advice from academics and experts on the significance of these. A number of Chinese academics were helpful, giving advice and affirmation regarding my approach, or the significance of specific local practices.

3.2.4 Iterative research process

A typical approach to researching a specific local reform involved a number of steps:

- profiling and background reading to determine whether a given reform merited investigation;
- tentative approach through academics and other contacts to determine whether access to the field was likely;
- in-depth desk research in preparation for fieldwork;
- fieldwork visit and interviewing;
- reviewing of interviews and documents collected;
- second round of desk research;

- return fieldwork visit;
- third round of desk research to fill in knowledge gaps and/or requests to interviewees for supplementary materials.

This approach had two main advantages: first, it was iterative, allowing results of reading and interviews to be fed into subsequent stages of fieldwork and planning. Attention to risk and local discretion in policy interpretation, for example, were themes that emerged during fieldwork and subsequent reading, rather than specific starting points for research. Second, the heavy reading requirement helped build a large degree of contextual understanding of local reforms other than those specifically studied. This process is extremely labour intensive, however: investigation of county reforms I examine in detail – and those that I don't – required examination of very large numbers of documents.

3.3 Implications for fieldwork

3.3.1 Implications for fieldwork: Practical

Searching for 'innovation', combined with the type of fieldwork I was able to carry out, has methodological implications. In early stages of study, I was hosted by the Central Compilation and Translation Bureau of the Chinese Communist Party, and it was through the CCTB that I first visited fieldwork sites (one county in Hubei, not profiled here). The visit was arranged through the CCTB and I was accompanied by a CCTB researcher. The visit was of a '*kaocha*' ('inspection' or 'investigation') type: it was short and involved concentrated meetings, interviews and site visits with the BOH, the NCMS Office, hospitals, etc. This clearly conforms to a certain paradigm of government inspection (and, frequently, academic research) in China.

My trips to other counties were similarly structured and were also of a *kaocha* type. In all the counties substantially discussed in this thesis, I visited twice and stayed for approximately a week each time. This was a disappointment: in my initial research proposal, I stated that I would attempt to stay in one county for an extended period of

time, which I saw as appropriate given the broadly ethnographic approach I hoped to use. The barrier to longer immersion was both practical and conceptual. First, I was attempting to interview local government officials who, almost exclusively, were busy and could only allocate me limited time. Most interviewees tended to approach interviews in a businesslike way, wanting to know clearly what I wanted to know, so they could answer my questions quickly and directly. Not all interviews or conversations were highly scripted and formal – there was often a significant degree of informal and social interaction also – but it was difficult to immerse myself in this environment.

The conceptual barrier lay in my interviewees' expectations of research. Hansen discusses how researchers in China must frequently follow 'in the footsteps of the Communist Party', by which she means that the existence of an established mode of fieldwork practice centred on short *kaocha*-type (my word) trips can hamper researchers (particularly anthropologists) who wish to work outside this tradition (Hansen 2006). This was visible in the way that my research experience was shaped by latent conceptions of my interviewees. In my understanding, most interviewees saw my research, and our interaction, in terms of a question and answer, or 'presentational', process, and this was compounded by my research profiling strategy which (unwittingly) conformed closely to an existing government/Party *kaocha* mode of research, ruling out the possibility of longer-term immersion. I attribute this to a clash between my hopes of ethnographic immersion and an expectation of presentational, *kaocha*-type fieldwork.

This had a number of repercussions. First, I did not achieve my goal of long-term immersion. Second, the *kaocha*-type approach combined with a focus on innovation to mean that as a researcher I quickly found out about specific county reforms, but that it took longer to develop a rounded understanding of the NCMS overall. This led to modification of my research. First, it led to a realisation that a large amount of documentary research was necessary as a complement to sub-national fieldwork. Documents such as the interim reports commissioned by the State Council and

academic sources became essential in understanding context. Second, it underlined the importance of triangulation: in this thesis, Meijiang functions more as a comparator to the two main cases than as a case in its own right.

Third, research became highly iterative. My first (brief) fieldwork trip was extremely exploratory. I found out about the county's 'innovation', but floundered in understanding NCMS functioning more broadly: the importance of computer systems, or fund management, for example, were not apparent at this stage. Such management issues receive little or no analysis in English language materials and it was only through the iterative process of fieldwork – reading – fieldwork – reading that I became aware of these. One indicator of this was my evolving familiarity with the vocabulary of NCMS managers. Initially, it was hard to find an adequate way to ask interviewees about the degree of latitude they had in scheme implementation. On several occasions, I asked people what could be considered 'the core' of the NCMS, but this tended to lead to formalised responses reflecting official formulations (*tifa*) that the NCMS was a policy 'benefiting the people' (*hui min zhengce*) and similar. On other occasions, I asked interviewees how much of NCMS policy was dictated by superior levels, and how much was up to them, and this proved more successful in understanding that counties perceived themselves as having a degree of autonomy. The most direct way of asking about this I found, however, was to ask whether specific initiatives were 'within the scope of decision making' (*juece fanwei nei*) of the BOH or the county, though it took some time to arrive at this phrasing.

3.3.2 Implications for fieldwork: Conceptual

Specifically searching for local innovation has conceptual, as well as practical, implications. First, this focus allowed me, to an extent, to sidestep an implementation and power framing of local government behaviour. As in Chapter Four, local innovation in China has largely been framed as being 'outside the system' (*tizhi wai*), dissociating it from questions of implementation, policy and the cadre responsibility system. In looking for innovation, I was not specifically questioning the ability of the centre to

enforce policy implementation – conceptually, the two things remained largely separate. What I found, though, was that looking for innovation led me to implementation. A second implication relates to use of *chuangxin*. As above, this is highly loaded and subject to multiple interpretations. On a number of occasions, framing of my research in terms of innovation in discussions/approaches to Chinese academics hindered communication (most often with health systems analysts and/or economists), though this framing never created a problem with government interviewees.

A third implication is that this approach leads to a focus on ongoing developments. This is both good and bad. One negative implication is that it is impossible to make definitive statements about the endpoints of the reforms I examine: all are in evolution and their significance (locally and outside the locality) is uncertain. This problem is replicated when one looks at the NCMS overall: there is (and can be) no resolution: as in Chapter Five, the scheme continues to evolve and change in many ways and in many places. The NCMS, in turn, is one strand of ongoing changes to rural (and national) healthcare provision. These, in turn, are components of larger institutional changes – urbanisation, migration, industrialisation, integration of local economies and communities, etc. – which dwarf the NCMS and healthcare reforms. Examination of processes of innovation as they happen is unavoidably messy. The objects under examination are, of necessity, half formed. Their future is, and has to be, left hanging.

The justification for this approach is that studying ongoing innovation processes shows things not visible in *post hoc* reconstructions. This is important to stress: the range of practices visible at the time of large scale reforms is very great, and this approach allows the researcher to see elements of ongoing processes likely to be subsequently swallowed up by larger reform currents. This is the case with all the county-level reforms and policy innovations addressed here, most obviously the Second Round Reimbursement policy (Chapter Eight), which was visible because of the timeframe within which I was carrying out research. I doubt the significance of this would have been so obvious in a *post hoc* analysis of the NCMS. In this, my research is methodologically distinct from the approach of Heilmann and Perry, despite superficial

similarities (Heilmann and Perry 2011). Seen 'from below' (Porter 1985),¹⁴ local government innovation in China seems less containable in an overall narrative structure than the picture given by Heilmann and Perry or by, for example, reverse-constructed accounts of the development of the NCMS (Wang 2011). Of necessity, studies that work backwards from currently-accepted institutional innovations – local innovations taken up as central policy – privilege the 'winners' in the competition for national influence and find greater clarity than was likely there at the time.

3.3.3 Are these cases outliers?

My profiling of local *chuangxin*, as extreme cases, leads to practices that are in some way judged 'good' or legitimate. The fact of having some kind of systemic designation of approval, imprecise though this is, does mean something. In Chung's terms, my fieldwork sites are all 'pioneers' (Chung 2000) in that they have gained a degree of recognition for doing something of significance in development of the NCMS, though this phrasing is too simple. There are legitimate methodological questions, however, with this approach, the main one being that of determining the significance of specific counties' practices in the absence of a basis for comparison. As above, my judgement of significance of local reforms, especially early on, was a poor guide to case selection, and this was compounded by non-transparent use of *chuangxin* and related terms and absence of a 'native tracker'. If the key attribute of counties one is attempting to profile corresponds to a highly loaded term in Chinese policy discourse, this is bound to create difficulties.

The obvious question that arises is to what extent these counties' practices reflect those of other counties – am I just seeing the best that China has to offer? Does other counties' experience lag behind this or is it somehow qualitatively different? Does the fact that all my fieldwork counties were in the first waves of counties to start

¹⁴ I borrow this phrasing from Roy Porter, who argued that medical history had been dominated by official views, by views 'from above', and argued for doing 'medical history from below' – from the patient's perspective.

implementing the NCMS mean I am seeing outliers? It is impossible to say with certainty, but I believe not. Two points are relevant. First, comparisons with other counties, where possible, indicate that my counties' experience is similar to other places. I have tried, when discussing reforms in both Taoshan and Feitian, to put these in the context of other counties' experience; frequently, this involves a degree of cross-comparison, showing similar framings of risk, using comparisons with Meijiang, or drawing in comparisons available from Chinese sources. My SRR chapter shows very similar features to my fieldwork-based chapters, but at a system level. The objective, in these cases, is both to situate my cases in a larger framework, and to claim a degree of commonality with other counties' experience. As in Chapter Five, though, the NCMS is not amenable to overall analysis and it is impossible to state with certainty the extent to which my cases reflect other counties' experience. Saying this, the counties in which I carried out fieldwork are not exceptional: they are relatively poor and peripheral (both geographically and intellectually/discursively). While we cannot know the degree to which they are representative of general local reform dynamics, there is no great reason to think of these counties as significant outliers. In most respects, they should be considered fairly *normal* places. Second, while the number of cases I examine is limited, the resulting analysis of policy implementation should be thought of as *structural*, and not confined to the cases examined. As above, the choice of a case study method was deliberately oriented to providing detailed analysis of local processes over generalisability.

Saying this, this research could have been different and could have turned up different findings. Research was determined by circumstances and luck (good and bad). An early decision to be hosted at the CCTB, for example, fostered a certain view of what kind of research was feasible, while access to academics and fieldwork sites determined the avenues I explored and what I discovered. This was inevitable, but should not undermine my conclusions: while a different study, focusing on different places or reforms, might have privileged different things, this is simply a reflection of carrying out fieldwork in a messy context in which idealised approaches are rarely

possible.

3.4 Hosting, interviews and documents

3.4.1 Access to research sites and interviewees

Early in my study, I was hosted by the CCTB as a visiting scholar, allowing me access to experts there and to collect materials and carry out research scoping in Beijing. During fieldwork, I was hosted by universities. Access to fieldwork sites was arranged through host institutions and academic contacts. Interviews were generally arranged for me by Bureaus of Health where I carried out fieldwork. In most cases, a member of the BOH would be assigned to help arrange interviews and would refer me to appropriate interviewees in line with my requests. This inevitably introduces an element of bias, but was unavoidable. Interviewees were almost exclusively staff of Bureaus of Health, of NCMS offices or of health providers. I had informal conversations with non-officials in many of my fieldwork sites, but did not significantly interview members of the public. In addition, in only one county did I interview members of the local Bureau of Finance or Development and Reform Commission. In no counties did I interview county leadership. These are all limitations of my research.

3.4.2 Interviewing

Interviews were mostly semi-structured or in depth interviews, though format depended to an extent on context, length of time available, number of interruptions, etc. A mixture of open-ended and closed questions was used, and I would ask for supplementary information where necessary and feasible. Within Bureaus of Health and NCMS Offices, I generally interviewed a range of people connected to the NCMS, though most interviewees were with middle-ranking members of staff who were less busy than people in leadership positions (and often best placed to answer questions on operational realities). In all fieldwork counties, I interviewed BOH heads. Most interviews lasted between forty-five and ninety minutes, and in most cases I could re-interview people to confirm details or ask supplementary questions. As well as

questions regarding specific reforms, interviews also enquired about NCMS management and operation. These questions inevitably became more detailed over time. This should be expected in an iterative, empirically-driven, research process, but inevitably reduces the basis for comparison between counties: elements of scheme functioning which entered my questioning relatively late were not explored in early interviews, for example. Interviews were conducted in Chinese, with the exception of a small number of interviews with English-speaking academics.

I visited five counties between May 2010 and July 2011: Taoshan County (X Province); Meijiang City (X Province); County P (Z Province); Feltian County (Y Province); Qianjiang District (Chongqing Municipality). Provinces are represented by letters: X and Y are eastern/coastal provinces which, while relatively rich, show provincial disparities in development levels; Z is a comparatively underdeveloped inland, southwestern province. I carried out around seventy interviews. Around fifty were with BOH staff, NCMS managers and other health system employees (hospital managers, administrators, doctors, etc.). A list of interviews (anonymised) is given in the Appendix.

3.4.3 Trust, interviewing and completeness of data

Successfully interviewing local NCMS managers required building rapport, and that they trust me. This was not a significant problem in any of the counties where I worked, but it has methodological implications: in all cases, I had to display trust in order to be trusted. The significance of this is in the questions I was able to ask, or the degree to which I was willing to push interviewees to discuss certain topics. The most prominent example of this was in Taoshan, where I asked interviewees whether the county's reform of oversight mechanisms was related to a specific local event and was told it was not. I subsequently discovered, through detailed reading about the county, that there had been some kind of crisis there in the years before the reform and that the two things were likely linked. I did not feel able to raise this with interviewees in Taoshan, though I did raise it with an interviewee elsewhere with a link to the county. He was

reluctant to discuss the event and I did not insist. This points to a limitation to research of this type and in this context: it is impolitic to attempt to discuss certain topics, at least in depth. This is unfortunate, but in the case cited, ultimately does little to change the overall conclusion. Also, this should be contrasted with the analysis of oversight given by interviewees in Taoshan, which I believe accurately reflects oversight problems in general, even though these were presented through a form of 'displacement'.

There is a related question as to how complete an analysis can be that relies on interviews and where there is no 'unfettered' access to people, documents and data. One of my fieldwork visits coincided with a visit to that county by a State Council NCMS audit team. The county had been audited once before and were unfazed by this. According to the NCMS Office, their being evaluated twice within a few years was entirely random. The audit team, however, were to stay for two weeks and go through the books meticulously. The realisation that came only subsequently was that a team of State Council auditors/accountants (presumably highly qualified and experienced) requires two weeks to properly audit the NCMS scheme in a moderately sized county. The implication of this is that achieving a 'full' understanding of a county scheme, making it fully 'legible' (Scott 1998) would require this kind of access and time.

3.4.4 Research ethics

This study was approved by the AREA Faculty Research Ethics Committee of the University of Leeds. I used a short research protocol in Chinese and English explaining the nature of my research, intended use of data, voluntary participation and protection of interviewees. Where possible, I gave this to interviewees to read. Where this was not possible, I explained the content orally and asked for consent. Interviewees were granted automatic anonymity. Signed consent was not required, but recorded approval was required if interviews were to be digitally recorded. I attempted this in one county in early stages of fieldwork, but discontinued it as I believed it to be a barrier to building rapport with interviewees. Handwritten notes were made during interviews; these are stored securely by the author. In-text references to interviews are by interview number

used in the Appendix. Interviewees and locations are given pseudonyms in the text or are referred to in non-specific terms.

3.4.5 Documentary sources

I use two main groups of Chinese language sources. Publicly available sources: national and sub-national policy documents, academic articles and books, media and specialist websites, as above. Second, locally-provided documents and reports form a major component of my documentary sources. These were supplied by interviewees to help explain a given policy or reform or to provide background on scheme operation. Mostly, such documents are used for the arguments they provide and the underlying framings of scheme management they show, rather than treated as necessarily factually accurate: the motivation for a particular local reform, as expressed in a report from a given county, for example, can be examined while sidestepping, at least to a degree, the question as to whether all figures provided in the report are exactly accurate. The appendices to the 2006 report to the State Council on functioning of the NCMS, for example, give reporting forms used in compiling of the report and contain the admonition that counties must report 'actual numbers' rather than those numbers used in local documents and policies ("*biaoge zhong suo tian de shuju wei shiji shuju, er fei wenjian guiding shuju*") (NCMS Pilot Evaluation Group 2006, p. 198), and my own research shows at least one case of a county strategically inflating figures in order to get support for a local reform.

This grey literature is cited in an indicative way only, and brief titles and sources are given in English only. Indicative references are not accurate translations of document names or sources, but do accurately reflect the significance of these. In some cases, due to the local and particular nature of reforms I discuss, this leads to the unfortunate need to anonymise references for publicly available materials. Where provincial and county-level documents are mutually referential or where published academic sources referring to a county initiative are used, these too are given indicative citations. In some cases, cited figures are marginally altered where exact figures might make it possible

to identify sites or people. In all cases, anonymising sources, people and places and any marginal changes to figures have no effect on meaning.

Gathering and use of documentary sources was an iterative process which started with review of English language materials and main Chinese policy documents, and proceeded from there as I tracked down cited sources and as these led to further sources. This approach is limited, however, due to the scope of the NCMS, the huge volume of materials published on this (4,774 studies according to one meta-study; see Chapter Five), and the fact that initial understanding is inevitably coloured by the sources used (and by the limited range of issues addressed in English language sources). As above, my fieldwork created new areas of *specific* interest within the scope of the NCMS (oversight mechanisms, payment reforms, etc.), requiring both referring back to existing sources and new searches for academic literature and location-specific materials. In a number of cases, 'forensic' reading was required to track down specific items of information, such as background to Taoshan's reforms, which was uncovered from government yearbooks. Documents were marked up by hand and annotations used evolved over time with my evolving understanding.

Chapter Four

Counties: Implementers and innovators in national policy schemes, part one

4.1 Introduction

The NCMS is frequently described in policy documents and by local governments as 'under county management' (a rough translation of *'xian wei danwei'*). We should take this phrasing seriously: what does it really mean to say that the NCMS is 'under county management'? What freedom and what constraints do counties have in running the NCMS? How do they act within this framework? Is there a range of different ways that counties act, or are county responses to policy quite standardised? What might either standardisation or different reactions mean, or even look like? Understanding the way counties operate within the NCMS is an empirical question: there is likely to be a difference in what implementation means in a 'hard' policy area, such as birth control or maintenance of order (*zhi'an*), and in a more loosely structured implementation environment, such as the NCMS.

The county is an important unit of analysis: counties manage most spending on social programmes, including education and health (Göbel 2007; Liu, Wang et al. 2009; Wong 2009, pp. 105-108). Most importantly, counties are responsible for most day-to-day government in rural China (Zhong 2003), and are the lowest relatively 'complete' level of government – below this are townships, which lack many of the offices and functions of the county level (Lam 2010), though their position on the interface with rural areas makes them a potentially important object of analysis (Hsing 2006).

Here, and building on Chapter Two, I set out an initial approach to analysis of the NCMS. I argue for the inadequacy of seeing counties as simply (imperfect) 'implementers' of policy and argue that we should take seriously their role in policy development, or innovation, and that we should attempt to understand the way they

use the systemic space they have.

Broadly, I argue several main things: first, local 'compliance' with central policy is an impossibility. Second, persistent central commitment to local initiative in China tends to indicate that compliance, strictly defined, is not the desired goal of policy implementation. Third, existing literature underestimates the volume of local innovation in China due to a restrictive focus. Fourth, local innovation is understood as having two parallel systemic functions: allowing adaptation to diverse circumstances and development of novel policy mechanisms. Fifth, counties should be seen as having a degree of freedom by design, and there exist analytical approaches to the cadre responsibility system and to policy discourse framed in this way. This chapter sets the scene for subsequent empirical chapters. Chapter Nine draws together empirical material with the analysis started here.

4.2 Implementation and innovation

4.2.1 Compliance and 'syllogistic implementation'

What we mean by implementation in China must be complex. The term 'compliance', as used by Li (Li 1998a), is useful in directing our attention to an underlying understanding of implementation in China-related literature, but it hinders as well as helps: Li wanted to show that multiple actors have power in the Chinese system and that implementation should not therefore be considered solely in terms of compliance. 'Compliance', though, is hard to pin down – compliance with what? Compliance draws implicitly on what Umberto Eco terms 'syllogistic logic' (Eco 1989): it assumes a possibility of congruence between intention and reception in communication; that, in policy terms, implementation can, in theory at least, be a question of 'carrying out policy' or 'putting policy into practice': it assumes a baseline, an ideal case in which implementation can proceed mechanically, in syllogistic fashion – if 'a', then 'b'. This is not just hard to believe in the case of China; it is necessarily an impossibility: the underlying logic of the principal-agent framework, in which compliance analyses are rooted, is the inability of the principal to both specify exactly the tasks the agent is to

complete and to monitor the accomplishment of these. The lack of specificity that *must* result from this necessarily gives agents discretion (Downs 1967, Chapter 11; Brehm and Gates 1997, p. 9). There is a limit to the ability of governments to carry out totalising planning, and space within plans and policy frameworks is *inevitable: metis*, or location- and situation-specific understanding of how to act, must be the complement to planning (Scott 1998, see also Hayek 1945).

Independent of this, many factors conspire to indicate that implementation in China should not, *even ideally*, be framed in terms of 'compliance', or smooth translation of policy intent into action, independent of the existence of implementation gaps, identified by many authors discussed above and whose existence I do not question. Implementation is a contested phenomenon – policies get transformed in the process of being actualised, and a general concern with this is neither new nor unique to China (for a recent review, see Hill and Hupe 2009). In China, specific factors complicate the issue: a degree of apparent autonomy of all nodes in the Chinese government system militates against any idea of syllogistic implementation, as do intersecting vertical (*tiao*) and horizontal (*kuai*) mandates within government and the sheer size of the government system. The literature on discretion and non-zero-sum approaches shows many of these complexities. Separate to this, consistent rhetorical attachment on the part of the centre to sub-national initiative, if we take this seriously, should make us wonder whether the hope of the centre is, in fact, implementation conceived of strictly in compliance/syllogistic terms. I start from the premise that not all exercise of discretion by local governments is necessarily deviant or oppositional, and that not all exercise of agency necessarily indicates non-compliance. Other forms of agency, other than stylized compliance/deviance extremes, exist and must be taken seriously.

4.2.2 Rhetorical attachment to decentralised initiative

An exclusive focus on control and implementation-as-compliance ignores a strong tradition of rhetorical attachment to decentralised initiative, stressed as having the potential to be complementary to, rather than in conflict with, the centre. Given this

continued central commitment to local initiative, it seems strange to focus purely on local *compliance* in policy implementation. Schurmann provides the classic pre-reform analysis of this: for him, there was a consistently high degree of attachment to 'democratic centralism', implying a commitment to local initiative in the implementation of policy: "Peking laid down the general policy guide lines, but the regions were allowed to develop specific policies to make sure to 'Get the Best Out of Each Area" (*yin di zhi yi*). For Schurmann, the aim was to promote 'democracy', by "arousing the 'initiative and creativity' of administrative units below the central level" (Schurmann 1966, pp. 86-87). In this view, centralism, the other component of this uneasy dyad, kept local agents within an overall line of shared goals: the 'unity of opposites', if it "works in practice [...] creates the possibility of dynamic politics within a framework of general agreement on ultimate ends and values" (Schurmann 1966, p. 56).

The classic statement of this is from Mao, in *On the ten major relationships* (Mao 1977). This clearly presents the harnessing of local initiative as vital to China's development, given the size of the country, its population and the complexity of the issues involved in building socialism. Local and central initiative, 'unity' and 'particularity', should be combined: it is essential "to bring the initiative of local authorities into full play and let each locality enjoy the particularity suited to its local conditions" – particularity, here, being "that [which] is necessary for the interest of the whole and for the strengthening of national unity". This document, first formulated in 1956, has been core reading for successive generations of Chinese leadership (Zheng 2007, pp. 359-363). As above, Chung emphasises the reform-era realisation that without sub-national policy space and a central commitment to local discretion, China's overall national development would be impossible (Chung 2011).

Recent examples of central enthusiasm for local initiative are not hard to find. A speech by Hu Jintao at the 2008 Party Organisation Work Conference, for example, called for all levels of the Party to strengthen their 'innovation ability' (*chuangxin nengli*), arguing that innovation is the responsibility of the whole Party at all levels, and that in building a

*xiaokang*¹⁵ society and bringing about socialism with Chinese characteristics, it is important to 'keep learning', 'keep putting reforms into practice' and to 'keep innovating' (Zhong 2008). 'Innovation', here, is clearly used to communicate enthusiasm for, and tolerance of, initiative at all levels. Specific high-level encouragement of local innovation is clearly visible in the case of the NCMS (see Chapter Five).

Experimentation in reform of China's health system is recognised. In the mid-1990s, China started pilots of a revamped Cooperative Medical Scheme (CMS) in 14 counties (Carrin, Rona et al. 1999; Jackson, Sleigh et al. 2005), and local initiative in reform of other elements of the healthcare system is also evident: Bloom, Han et al. (2000), for example, discuss reform of health worker incomes, during which the centre encouraged localities to find appropriate ways of working, some of which were taken up by other localities or made the basis of regulation. Wagstaff and Lindelow, among others, discuss such processes in 1990s urban health insurance reform. They describe design of the NCMS as based on the same principle:

"Counties are being given considerable discretion in the design of NCMS [...] One reason for this was simply an acknowledgement that local choice on design issues is an integral feature of China's highly decentralized health system. But there was also another reason – to ensure that lessons could be learnt from local experimentation, and that they could be fed into the scaling-up process" (Wagstaff, Lindelow et al. 2009a, p. 3).

Counties, in other words, should be expected to have a degree of freedom within the scope of the NCMS. Early analyses noted substantial local variation in insurance packages, local setting of reimbursement thresholds and similar. Overall, though, counties' degree of freedom is an empirical question. This is dealt with in detail in subsequent empirical chapters.

¹⁵ 'Moderately prosperous'; '*xiaokang*' is a key term of the Hu-Wen leadership.

4.2.3 Understandings and misunderstandings of innovation

The literature on local innovation in China tends to see this as risky, radical and system-changing. Analysis of the household responsibility system (HRS) early in the reform era is an example, as are recent analyses focusing on democratisation measures, local elections, and similar (Fewsmith 2006; Fewsmith 2008b; Fewsmith 2008a). There are two problems with this: first, this underestimates the quantity of local innovation through a restrictive focus; second, this conflates result and process.

First, there is a quantitative misunderstanding deriving from a restrictive conception of 'innovation'. There is debate within the literature on public policy innovation as to what the magnitude of a given reform must be in order to qualify for the title 'innovation'. On the one hand, if the bar is set too low, a wide range of practices of varying degrees of novelty can be described as innovation (Hartley 2006, p. 27); much literature attempts to classify practices according to their perceived degree of novelty (e.g. Roper 2010). Alternatively, innovation,

"must not simply be another name for change, or for improvement, or even for doing something new lest almost anything qualify as innovation. Innovation is properly defined as an original, disruptive, and fundamental transformation of an organisation's core tasks. Innovation changes deep structures and changes them permanently" (Lynn 1997, cited in Hartley 2006, p. 25).

Others promote attention to processes of 'continuous' or 'marginal' innovation (Bessant 2005). In the case of China, most (foreign) analysis has tended to think innovation must be disruptive, radical and system changing; such studies tend to examine innovations that are 'outside the system' (*tizhi wai, zhidu wai*), in that they fall significantly 'outside' existing policy parameters – the HRS, unsanctioned promotion of township elections, and similar. There are very few studies of local government innovation in low-key policy areas, such as housing for the elderly, or pension delivery, for example, though compilations published by the CCTB are a useful corrective (Chen and Yang 2009; Yu 2009). This leads to a focus on, first, innovation-as-result and,

second, elections and similar, perceived to be (potentially) systemically significant.

On the second point: a focus on innovation-as-result means that we tend to overlook the importance of innovation-as-process. A focus on reforms such as the HRS privileges a view that only things that come from 'outside', that change the overall paradigm, are significant: action is only innovation if it is disruptive, and non-disruptive activity is highly unlikely to be innovative. This reasoning creates a false opposition between ideas of *implementation* (understood as an idealised state of compliance or syllogistic implementation) and *innovation*. In the next chapter, I argue that counties are expected to both implement and actively develop policy, and this forms the basis of empirical cases examined in subsequent chapters.

4.2.4 The importance of local initiative: Two readings

Two different understandings underlie the preceding descriptions of local initiative in China's reforms. First, the logic of adaptive (*yin di zhi yi*) implementation is one of making use of dispersed information in policy development. This is Hayekian:

"The economic problem of society is thus not merely a problem of how to allocate 'given' resources – if 'given' is taken to mean given to a single mind which deliberately solves the problem set by these 'data'. It is rather a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only these individuals know. Or, to put it briefly, it is a problem of the utilization of knowledge not given to anyone in its totality" (Hayek 1945, p. 519).

Second, the importance of local initiative in generating novel policy solutions is more directly developmental: localities are expected to play a role in developing policy solutions which, it is hoped, will be of supra-local and/or systemic importance. This is more Schumpeterian than Hayekian: innovation (as process), in this reading, occurs where counties operate in a dynamic environment, beyond the boundaries of routine practice: "every step outside the boundary of routine has difficulties and involves a new element" – existing ways of thinking and working are of limited use. Innovation, as a

process, takes place where we lack “those data for [our] decisions and those rules of conduct which are usually very accurately known”: operating on the boundaries of routine experience, countries must find new ways of doing things, or new ‘combinations’ in Schumpeter’s term (Schumpeter 1934, p. 84).

I equate this to new ways of doing things, with “putting invention into practice” (Gudeman 1992),¹⁶ and with ‘new ideas that work’ in Mulgan and Albury’s rather blank, though influential, definition of government innovation (Mulgan and Albury 2003). In more directly institutional terms, I equate Schumpeter’s ‘combinations’ with North’s ‘artifactual structure’. According to North, we live in a ‘non-ergodic’ (non-repeating) world, where the future is never derivable from experience. The ability of societies, of polities – of systems in general – to adapt is crucial to their long-run survival. North’s thesis rests on this distinction between ‘allocative efficiency’ (the ability to ensure efficiency in allocation of goods in the here and now) and ‘adaptive efficiency’; for him, problem solving takes place through learning, rather than use of brute logic, and the key to long-term ability to adapt is a rich ‘artifactual structure’. Faced with uncertainty, says North, actors construct regularities (institutions, practices, mechanisms) to codify behaviour, though the resulting institutional forms “depend on how novel [situations] are and on the cultural heritage of the actors”.¹⁷ This ‘cultural heritage’, for North, is an artifactual structure, a composite of practices from the formal to the informal, encompassing “beliefs, institutions, tools, instruments [and] technology” necessary for ‘adaptive efficiency’ and which are amassed over time and differ from culture to culture (North 2005, p. 36). The richer the artifactual structure, the easier it is to convert the ‘novel’ into the simply ‘new’, to convert unknowable uncertainty into quantifiable risk, to

¹⁶ For Schumpeter, ‘innovation’ is distinct from ‘invention’. Innovators may invent/discover the ideas they implement, but they need not – new possibilities, “are always present, abundantly accumulated by all sorts of people. Often they are also generally known and being discussed by scientific or literary writers. In other cases, there is nothing to be discovered about them, because they are quite obvious” (Schumpeter 1934, p. 88). This is very visible in my empirical cases.

¹⁷ North distinguishes between ‘new’ and ‘novel’: many situations may be new, but novel situations are unprecedented: “we have no historical experience that prepares us to deal with [them]” (North 2005, p. 20).

adapt and survive (adaptive efficiency is discussed at length in Chapter Eight) (North 2005). I return to this in Chapter Nine with a discussion of the 'repertoires' employed by actors in locally developing policy innovations.

The significance of this process is that it is dynamic and developmental.¹⁸ Judging the functioning of such a system should be based not on how it 'administers existing structures', but rather how it creates and destroys them:

"A system – any system, economic or other – that at every given point in time fully utilizes its possibilities to the best advantage may yet in the long run be inferior to a system that does so at no given point in time, because the latter's failure to do so may be a condition for the level or speed of long-run performance" (Schumpeter 1976, p. 83).

In other words, the basis on which to judge such a system is its capacity for 'adaptive efficiency', rather than in terms of the efficiency with which it distributes resources ('allocative efficiency').

In the case of the NCMS, new combinations are institutional *form*: policy solutions, specific policy mechanisms, organisational and management technologies and similar. This should be understood to include a range of practices of varying degrees of (adjudged) usefulness: those practices which, through systemic approbation and concatenation of circumstances, come to be described as 'innovations' (*chuangxin*), local 'experience' (*jingyan*), 'experiments' (*tansuo*), 'models' (*moshi*), 'firsts' (*shuaixian*) and similar – all meaning something close to 'best practice' in Chinese policy discourse – as well as others which are either *de facto* less useful, or simply not taken up and given an explicit systemic stamp of approval. In subsequent empirical chapters, I

¹⁸ Schumpeter was concerned to show how capitalism functions as a dynamic system which "incessantly revolutionizes the economic structure from within", and his description of "capitalism as 'the perennial gale of creative destruction', [...] has become the centerpiece for modern thinking on how economies evolve" (Cox and Alm 2008). The core of this, the "fundamental phenomenon of economic development", is the creation of 'new combinations' of the means of production – new ways of doing things, new products, and the like.

examine local reforms in relation to both adaptive implementation and the developmental Schumpeterian function described here.

4.3 Innovation within established policy frameworks

4.3.1 Freedom by design

Analytically, we are accustomed to innovation occurring outside existing policy schemes: this is what we see in cases such as the HRS and township elections. How, though, to explain innovation within existing policy domains, where the default analytical approach is to look for *implementation* of policy? Overall, I argue that both the CRS and, to a large extent, policy, deal with 'function', not 'form'. They specify *what is to be done*, without specifying *how this is to be done* (this was raised in my discussion of Göbel, above, though my analysis goes further than his). In Schumpeterian terms, *this absence of a template for action is a disjuncture at the centre of the policy implementation process that forces innovation by policy implementers*. This is why we must consider implementation of the NCMS (and many other policies) in terms of both implementation *and* innovation. Understanding this requires that we reorient our understanding of both the CRS and policy discourse. Clear precedents for this exist and I draw on these here.

There is a tension between ideas of control and communication in analyses of the CRS and policy in China. Schematically, two poles of analysis are visible: first, analyses that examine policy, communication, and target-setting and lament their imprecision and/or ineffectiveness – their inability to achieve control; second, analyses that attempt to take seriously a lack of precision in targets, policy language and stipulations, and argue a systemic function for this. Both poles are important: how the CRS and policy enforce both implementation and innovation.

4.3.2 Approaches to the cadre responsibility system

The first approach, centred on control, is exemplified by Edin, whose analysis starts with a clear question: why, against expectations, did local cadres in China promote

growth in the early reform era? The question is significant as she judges that local cadres should, *a priori*, engage in predation, rent seeking and similar (Edin 2003a, p. 83). Whereas previous analyses had stressed fiscal explanations for local cadres' behaviour, Edin supplements these with a political explanation located in the contracting system and the way that targets of differing degrees of rigidity (normal, hard and veto) are set for a range of policy priorities. This is an understanding framed in terms of *control* and stems directly from Edin's initial question.

Other studies of the CRS work within a similar control framework: one recent study describes the system as a tool for the central state to achieve 'compliance' from its agents, but finds this to be flawed (Gao 2009). In Gao's analysis, the system enforces compliance, but does not necessarily ensure good outcomes; the use of vague targets makes measurement of achievement hard and leaves room for personalised evaluations and cronyism. Gao does not problematise 'compliance', but the meaning is familiar: on one hand, an excessive focus on economic growth targets, mandated by the CRS, creates growth at all costs, irrespective of environmental degradation and similar; on the other hand, while cadres must take orders from above seriously, "a substantial number of the indicators adopted to measure the achievements of local officials in meeting functional targets [...] are abstract, vague, poorly defined and hard to measure objectively" (Gao 2009, p. 29). This is a reflection of the argument of O'Brien and Li, discussed above.

Li has also examined functioning of the CRS, in her case in management and oversight of 'public service units' in Hubei.¹⁹ In this case, the CRS is seen as a means of communication and sending a message about work priorities, but one whose control

¹⁹ 'Public service units' is used to translate 'shiyue danwei'. The translation comes from Christine Wong: "the vast majority of public services in China are provided by publicly owned entities that are attached to government at various levels, known as public service units (shiyue danwei). They include schools, agro-technical extension stations, cultural centres, health clinics and hospitals" (Wong 2009). Elsewhere, shiyue danwei are described as 'public institutions'. The importance of this term here is that it denotes a para-governmental entity lacking the full regulatory authority of a government department.

function is relatively weak, especially as regards targets not amenable to quantified measurement, particularly 'soft' tasks and those requiring detailed local knowledge, or *metis* in Scott's (1998) term (teaching quality, for example). Li also notes that the CRS seems to follow an "implicitly 'negative' approach – whether rules were contravened, or some quantitative yardsticks had been met – rather than focusing directly on the quality of services". This, for Li, is where the system is weakest (Li 2008, p. 265).

A recent article by Heberer and Senz takes a different tack, interpreting the CRS as a flexible communication mechanism which can be tailored to local ends. These authors examine, as they put it, the 'creative space' allowed within this system. To this end, they show how some localities adjust targets within the CRS to prioritise environmentally sound development as part of a process of local innovation, development of local models, and inter-jurisdictional competition. In this process, the centre sets the agenda and localities interpret and modify policy in line with the "standards of the Centre [...] fixed in documents which state the intentions of the party-state and its priorities". These authors conclude that,

"evaluation processes in China show not only that the counties have room to manoeuvre in terms of policy priorities but also that these processes offer incentives for policy implementation, particularly for implementing environmental policies. In China, decentralisation is a gateway to improving policy implementation, even in the domain of environment. The incentive and communication system that has been described here [the CRS] is a relevant factor in the specific design of decentralisation; otherwise all benefits of decentralisation – not just those in the domain of environmental policies – would have come to naught" (Heberer and Senz 2011, p. 107).²⁰

A greater departure from this dominant control focus lies in questioning the *specificity* of contracts. Specificity is the underlying concern of analyses of Gao and Li, as well as

²⁰ This departure from a strict emphasis on control is welcome, but highlights something discussed by Edin in early work, that while core targets are nationally determined, exact content of contracts can be varied locally: "[e]ach unit and area may formulate its own evaluation methods and targets" (Edin 2000, p. 127).

O'Brien & Li (see above). For Gao, the failure of the CRS to specify targets exactly is a failing in comparison to measurement systems in OECD countries, where detailed specification is the norm. Specificity is at the centre of Li's analysis: her question is how to oversee public services which are frequently not easily quantifiable (though she recognises that this concern is not unique to China and cites a World Bank study to the effect that one hundred percent specificity in design of contracts is absurd, as demonstrated by "work-to-rule' strikes by public service workers in various countries") (Li 2008, p. 265).

An important and related point is made by Edin in her PhD thesis, but seemingly was never included in subsequent publications, possibly because Edin wanted to explain *why* cadres promote growth, rather than *how* they promote growth. Edin's analysis, prefiguring Göbel, discusses control of local agents in terms of both hierarchical and market mechanisms. The CRS allows setting of goals under hierarchical steering mechanisms; the contracting system, says Edin, following Rigby, is *goal rational*, rather than *procedural*: the CRS is a means of communicating goals without specifying how these are to be achieved (Edin 2000, p. 36). In doing this, it communicates *function*, but leaves *form* unspecified, introducing a disjuncture into policy implementation.

This disjuncture between form and function is recognised in a prominent Chinese analysis of local cadre management: the 'pressurised system' (*yalixing tizhi*) theory (He 1997; Rong 1998; Rong 2009), a consciously empirical and atheoretical description of local government functioning in the 1990s. The basis of this system is the setting of targets by higher levels of government, which are passed down through the government system with implementation indicators, deadlines, rewards and punishments attached. The system has both positive and negative aspects: it can effectively arouse local entrepreneurship and promoted development, but unfunded mandates, poor local accountability, plus local cadres' incentive to curry favour with their superiors, all lead to negative outcomes (poorly planned development, environmental degradation, peasant burden, etc.). In this, the 'pressurised system' largely mirrors Edin's or O'Brien and Li's analysis. Two things recommend this

description, however: first, analyses of the ‘pressurised system’ clearly state that this specifies tasks, but leaves the means for their achievement to implementers (“*bing bu guanxin ge xiangzhen, ge zhineng bumen yong shenme shouduan wancheng renwu zhibiao*”), and this can have serious negative consequences – excessive extractions, and the like (He 1997, p. 70). Second, the name of the theory – the *pressurised system* – derives from the vocabulary of local cadres, who complain of the pressure they face in carrying out their work. This vocabulary is picked up by Heberer and Senz, who quote a People’s Daily survey of “100 leading local grassroots cadres” of whom “over 30 per cent complained that [CRS] evaluations put enormous pressure on them” (Heberer and Senz 2011, p. 91). A vocabulary of risk (*fengxian*), pressure (*yali*) and a systemic focus on results, rather than means (*iieguo daoxiang*) is extremely evident in my fieldwork and gives a very different baseline for analysis from approaches which start by questioning control and implementation.

4.3.3 Approaches to policy texts and language

There is a strong tradition of attention to textual specificity (or its absence, *indeterminacy*) in analysis of Chinese policy discourse. The classic statement of this is from Schurmann, who sees indeterminacy of much Chinese policy discourse as allowing communication of objectives without tying down policy formulators in endless detail:

“categories and language must be precise enough to be applied practically, but also general enough to cover a wide range of different conditions. Policy orders are put in general terms, but they must be carried out under particular conditions. The local cadres must understand the intent of the central policy decision. If all key terms were exactly defined, it would make them too precise and so tie down the policy makers” (Schurmann 1966, p. 60).

This requires highly codified language, truly understandable only to insiders: “As long as the categories and language of communication are systematic, the receivers of communication will decode the messages in the way they were coded, thus assuring

congruence between the intent of the message and its interpretation". Much of cadres' understanding of what, concretely, is to be done derives from study of authoritative editorials, models held up for emulation and similar, informing their reading of texts: "over time, people learn the *real* meanings of [significant] terms" (Schurmann 1966, p. 60, emphasis mine). Schurmann's use of a decoding metaphor is unfortunate, as it suggests a hidden, privileged, meaning within texts; elsewhere, in a discussion of the importance of models and emulation, Schurmann's intention is, I think, different: low level cadres study models and "get the point" (Schurmann 1966, p. 64). 'Getting the point' is much closer to an idea of separation of *form* and *function* in policy discourse.

Schurmann is not alone in focusing on policy *texts*. Early work by Lieberthal claims there to be "considerable room for policy making in the process of policy implementation" (Lieberthal 1978, p. 15). Chung, Göbel and Manion all consider this: Chung's use of 'statutory precision' allows him to show a decrease in specificity in much policy, seen as a means of communication, in the reform era, with the aim of increasing local 'policy remake' (Chung 2000, p. 5). Something similar can be seen in Göbel, when he discusses the divergence between policy aims and the ways that localities are to achieve these. Manion's and Paine's conclusions regarding local adaptation are in a similar vein (Manion 1991; Paine 1992). Such studies show a more complex understanding of communication in the process of implementation than that found in, for example, Wedeman, who argues that progressive distortion of central messages is inevitable as they are passed down through multiple layers of government and that mis-implementation arises from 'incompetence, noise and fear' (Wedeman 2001).

4.3.4 Target-setting and policy: Interim synthesis

There are two understandings of the CRS and policy texts intertwined in the above analyses: target-setting and policy have, or can have, both restrictive (controlling) and generative (communicative) functions. In separating these two understandings, my aim is to show that both are important, and fundamentally linked, in local policy

implementation. The control function discussed by Edin and others clearly exists, and functions to a considerable degree. Conversely, the communicative function of these systems is also important, helping understand how target-setting and policy can play a generative role at the local level. I show local implementation and innovation-as-process as fundamentally linked, and show local innovation as ultimately *forced*, at least to a degree, by these systems. This goes further than Göbel, in decreasing the importance of local voluntarism in policy innovation, and differs from Heberer and Senz in pointing to the way that implementation of policy frequently *requires* innovation (rather than innovation, in their analysis, arising from strategic adjustment of the CRS). While my analysis departs from authors examined above, it should be understood as supplementing, rather than denying, these analyses. This analysis runs through subsequent empirical chapters and is theoretically developed in Chapter Nine.

4.4 Decentred practices and the necessity of tolerance

4.4.1 Models of the periphery: Expect a range of outcomes!

Seeing counties as systemically engaged in innovation as well as implementation must change the way we view what they actually do. To be clear: counties innovate in implementation of policy in that, and to the degree that, their activity is unscripted. Some policy areas are relatively scripted; others less so; the NCMS shows both characteristics and this changes over time. Seeing counties as innovators in policy processes shows, in the case of the NCMS, huge local variation in policy. This does not mean that everything counties do should be seen as an 'innovation', if what we mean by innovation is, broadly speaking, something 'good'. What it means is that counties are engaged in development of specific local practices, integrated to varying degrees with larger reform processes. In vocabulary borrowed from Gudeman and Rivera, wherever there is a margin, there will inevitably be 'models of the margin' (Gudeman and Rivera 1990, p. 188).

We should expect to see a range of outcomes of decentred innovation or reform: some

good, some less so. This is an inevitable outcome of the way counties respond to working without an exact template for action, and will be influenced by any number of factors: the level of implementers' commitment to the policy in question, their understanding and (technical) competence, the impetus and incentive they have to 'run with' a policy, ways a given policy can be integrated with other local priorities, possibilities for rent seeking, playing to specific constituencies, local connections and oversight, implementers' centrality or peripherality in debates on, for example, health reform and access to new information and ideas. We should also expect to see opportunism, voluntarism and currying favour with higher levels, pioneering, bandwagoning and resisting, as in other analyses. My cases highlight a number of these aspects, but the range of variables is so huge and actual outcomes so contingent that generalisation about the conditions of production of policy solutions is highly risky. Especially in an area as technical as the NCMS, counties are unlikely to stumble upon new 'best practice' policy solutions, if what we mean by this is models of national (or international) applicability. We are more likely to see a range of idiosyncratic practices emerge from this kind of decentred policy development.

I am most interested in locally-generated practices. In Chinese, and when interviewing local Bureaus of Health, such practices are frequently referred to as '*zisheng de*' or '*zifa de*', meaning 'self-generated' or 'self-developed'. This, by definition, shows localities operating at the margins of policy processes, and clearly flags local practices as peripheral. Some local practices, the most 'successful' from a certain point of view, gain supra-local recognition and may be studied and copied by other implementers or taken up and refined in policy at higher levels. If localities play the role of policy developers and innovators by virtue of their position within the policy system, 'form' (or 'combinations') is the mediating term that shows this to be not just a process of decentred local implementation, but rather something akin to a decentred innovation system: there is a clear hope and recognition among policy formulators (and implementers) that local policy solutions may have supra-local or systemic relevance.

Systemic, semi-regularised, channels for recycling local and peripheral practices exist,

as do multiple informal and non-regularised channels. Specific local reforms examined later clearly exist in 'conversation' (Gudeman and Rivera 1990) with central texts, officially-promoted models and other sub-national practices. The degree to which systemic channels are useful and/or promote local practices in a neutral and fair way (as opposed to being particular and used for self-promotion) is unclear, but must be evaluated empirically. Horizontal spread and dialogue between peripheral practices are more common, in the NCMS at least, than in the dominant picture derived from the China policy literature, which presents vertical adoption and promotion of local models as the main mechanism underlying diffusion (e.g. Saich and Yang 2003; Fewsmith 2006).²¹

4.4.2 Tolerance and legitimacy of local practices

Tolerance of variant local implementation(s) is a requirement for this mode of policy development. Tolerance has been discussed above; for Chung, it is a prerequisite for the development of norms of decentralisation and adaptive (*yin di zhi yi*) implementation. To the extent, in Chung, that this is about adapting policy to local conditions, this is a Hayekian argument, and the payoff for the centre of adaptive implementation is an increase in 'developmental power' (see above). My argument, in this case, is more explicitly Schumpeterian: tolerance is necessary for local development of a wide range of policy solutions through decentred innovation; some of these will be useful, some not. Some will be successful; some (maybe many) will fail. The logic is one of acceptance of reduced allocative efficiency as a trade off for increased adaptive efficiency. In my empirical chapters, I show a high degree of tolerance, and argue that 'innovation' is understood as relatively low risk, against the dominant picture painted in existing literature.

²¹ Vectors for horizontal transfer clearly exist, however: Oi (1995) and others show the importance of the networks through which local leaders acquire raw materials, negotiate contracts, etc, while Pieke (2007) has shown that study trips, cross-postings for the purposes of observation and learning (*guazhi*), as well as vertical secondments, etc, are common.

4.5. Interim summary and looking forward

4.5.1 Interim summary

Here, I give the beginnings of an analytical framework for understanding county policy implementation/innovation within the NCMS. I have argued several main things. First, that compliance, or syllogistic approaches to implementation are insufficient, especially in China. China's consistent central attachment to decentralised initiative should indicate that compliance is likely not the desired, ideal, outcome of policy implementation; this certainly applies in the case of the NCMS. Second, that approaches to innovation in China are restrictive, tending to focus on a small number of marginal outliers; this risks blinding us to a much larger volume of systemically legitimate local innovation and creates a false opposition between ideas of implementation and innovation. This does not mean that innovation is always 'good', it is simply a reorientation of understanding of local action.

Third, that (in this case at least) counties have freedom thrust upon them through the CRS and through indeterminate policy. These force action without specifying in detail means to achieve stated goals, requiring that implementers translate abstract policy goals into concrete management mechanisms. This introduces a disjuncture in the policy process, which must be considered in terms of both implementation and innovation. Fourth, that any number of variations in local conditions, aptitude and commitment will change the ways in which localities implement and interpret policy, leading to a diverse range of peripheral practices. These are, to a degree, in 'conversation' with dominant policy texts, officially-promoted models and other sub-national practices. Local and peripheral practices are not likely to be 'best practice' solutions, but are part of a process of dynamic policy development in which adaptive efficiency is privileged over allocative efficiency. Tolerance of variant local practices, and possibly implementation failure, is likely to be high.

4.5.2 Looking forward

The following four chapters present empirical cases building on the analysis started here. As in Chapter One, my cases focus on county administrative/management practices in an attempt to understand how counties both implement policy and innovate within this framework – how they *use* policy and the systemic space they have. My cases adopt a broadly ethnographic approach, based on interviews with county and provincial official, local policies and documents to examine how policy is ‘practised’, the perceptions and framings of actors in the policy process, their judgements regarding specific policy solutions and possible reforms, the legitimacy of these, and the ways these are linked with policies, discourses, models and reforms in other places and at other scales. This is an attempt to question what implementation and innovation actually *are* in the case of the NCMS and in China – to understand the “conventions, understandings, habits, and practices” (O’Brien 2010, p. 80) within which implementation/innovation take place in this area – over ideal representations of what implementation and innovation *should* be. Cases were selected through top-down profiling of county innovations, as explained in Appendix One.

In the next chapter, I examine structuring of the NCMS as a national policy and the importance of the county within this as the level at which the scheme is managed. I show how county management is ensured by the ‘pressurised system’, but that, following on from the analysis here, this is accompanied by a clearly expressed central desire for local innovation and creativity. Chapters Six and Seven discuss two local cases. On the surface, the cases presented are dissimilar, but both clearly show counties as *innovators within*, not just *implementers of*, central policy, and show a degree of consensus over county roles in the policy process as understood by counties themselves and by superior levels. Chapter Eight shows the transition of a local reimbursement mechanism to national policy and develops arguments started here about the Schumpeterian, developmental, nature of the NCMS policy process and the importance of local initiative. Chapter Nine brings together the analysis started here with material from subsequent chapters to provide a theoretical conclusion.

Chapter Five

The NCMS: Between central and local policy

In this chapter, I give an overview of the NCMS as both national and sub-national policy; this is intended as an analytical background to empirical cases discussed in subsequent chapters. It also complements existing English language analyses which tend to focus on the impact of the NCMS on users' health seeking and spending, paying less attention to the policy and its development.

The paradox of the NCMS is that it is a central policy, in which the central government has consistently set the overall parameters of reform and principles (*yuanze*) of policy, but in which local governments – principally counties – play the main role in management and develop implementable (*caozuoxing de*) local policy. Overall, this can be understood through the lens of the 'pressurised system' (*yalixing tizhi*): ever-increasing implementation targets are passed down and must be realised at the county level. Perhaps surprisingly, this implementation mechanism is combined with a clearly-expressed central hope that localities will develop usable policy solutions that can be fed into overall development of the scheme. The paradox lies in the coercive nature of the system which forces implementation but which frequently remains agnostic as to the *form* implementation takes.

The chapter is structured as follows: the first section gives an introduction to the NCMS as a national scheme. The second section analyses the NCMS as an example of functioning of the pressurised system, through target setting in the health system and through operation of the NCMS financing system. The third section examines the role of counties within the national framework in both implementing and, it is hoped, developing policy. The fourth section develops the theme of variation within the NCMS and the difficulty of understanding this as one, national, scheme. The fifth section concludes and sets the scene for subsequent chapters.

An initial and clear caveat is necessary: *it is impossible to adequately understand the*

NCMS as a national system. Central and provincial policy has changed over time, tending to regulate counties' freedom in implementing the scheme, while development of the scheme has led to ever-changing demands on counties. The distributed nature of policy making, at both provincial and county levels, means that counties' scope of action, and provincial and county policies, differ, possibly considerably, across sub-national jurisdictions, while implementation of many elements of policy remains inconsistent. I return to this theme in more depth below. As an outsider to the policy process, the amount of information available on scheme operation as a whole is limited. My analysis in this chapter relies for the most part on policy documents, ministry yearbooks, and academic sources, most usefully Chinese sources.²²

5.1 NCMS as central policy: Outline of the scheme

The NCMS is a central policy: the State Council and, principally, the Ministries of Finance and of Health set the overall parameters of the scheme and local governments, nominally at least, work within this framework. Development of the NCMS, according to the conventional narrative, starts in 2002/2003, with the issuing of overall framework documents, first by the Party Centre (Zhongfa 2002, No. 13) and then by the State Council (Guobanfa 2003, No. 3) (though, as above, various attempts were made in the 1990s to develop a new rural health insurance scheme). The 2003 State Council Opinions set the overall framework for development of the scheme and provide the basis on which subsequent central policy builds. Many non-Chinese analyses have described the NCMS as revolving around two core stipulations: that the scheme is voluntary and that it focuses principally on 'serious illnesses' (*da bing*), with the aim of reducing impoverishment resulting from 'catastrophic' health-related spending (e.g. Yip and Hsiao 2009). This is true to an extent: both of these stipulations have been core to

²² Chinese academic sources, while immensely valuable, suffer from a problem of huge supply: a meta-analysis carried out by Beijing Normal University in 2010-2011 found 4,774 studies of the NCMS published to date, the vast majority in Chinese (unpublished presentation, given at Chinese Academy of Social Sciences workshop, March 2011).

the NCMS since the beginning, but the reality is somewhat more complex,²³ and Chinese analyses have tended to focus more on the fact that for the first time the centre directly contributes a certain proportion of funds to a health insurance scheme for rural residents (at least in poorer places) (NCMS Pilot Evaluation Group 2006, p. 2; Gao 2008a), and have been much more concerned with actual management of the scheme than have non-Chinese analyses.

The NCMS pools funds from users, local governments and, in many places, central government, at the county level, normally under the management of the county Bureau of Health (though in some counties, and in various pilots, management may be placed under other departments) for the reimbursement of medical expenses incurred by scheme members, normally in designated health care providers (*dingdian yiliaojigou*; those providers included within the NCMS system). The scheme is 'pay-as-you-go': users contribute a premium in any given year for insurance coverage in that year. In many places, users who do not make use of the scheme in a given year are entitled to a free health check (*tijian*) as a form of reward or incentive. Nominally at least, the scheme is based on voluntary sign up; from the outset, concerns were articulated that this would lead to adverse selection – that the healthier people in any given county would not sign up and that the NCMS would be left to insure only the worst health risks, prejudicing sustainability of the scheme (e.g. WHO 2004). One result of this is the decision that families should sign up together: individuals are allowed to sign up only if their immediate family members also do so. Similarly, the scheme has been remarkably successful in covering the majority of the rural population, and maintaining coverage, somewhat surprisingly given the number of analyses that pointed out the difficulty of operating insurance schemes aiming for generalised population coverage on a voluntary basis, especially given the scheme's focus on serious illnesses which inevitably only affect a small number of people at any one time. (I return to the question

²³ Yip and Hsiao, for example, definitely go too far in claiming that the "NCMS incorporates two important policy features: voluntary enrollment and coverage of catastrophic illnesses [...] Apart from these two requirements, the design of the program is left to the local governments" (Yip and Hsiao 2009). See Section 5.5, below, as well as Chapters Six and Seven.

of how voluntary the scheme really is below.)

5.2 NCMS as central policy

5.2.1 NCMS as central policy: Implementing the scheme

In this section, I briefly outline the national NCMS policy making framework. I also show how the (national) setting of targets to be implemented, ultimately, at the county level – the pressurised system – combines with the operation of the NCMS funding system to localise responsibility and, at least as importantly, implementation *risk* at the county level. This serves a systemic function of risk containment; it also influences counties' implementation behaviour. The systemic role of implementing counties is addressed in the following section.

5.2.2 Implementing the NCMS: Outline of the national policy framework

Overall, central policy is passed down through multiple levels of government and is eventually implemented by counties, though provinces have an intermediate role in this system. Development of the scheme is led by an inter-ministerial coordination group under the State Council. Below the State Council, the majority of NCMS-related policy is issued by either the Ministry of Health or the Ministry of Finance (though the remit of a number of other ministries, most importantly the Ministry of Agriculture, the Ministry of Civil Affairs and the National Development and Reform Commission, also includes NCMS work²⁴). The Ministry of Health established an NCMS Technical Guidance Group (*jishu zhidao zu*) and a dedicated NCMS Research Centre early in operation of the scheme (Guobanfa 2004, No. 3; Weibannongweifa 2004, No. 46). The role of the Finance system (*xitong*) is discussed in detail below. Governments at the provincial level are charged with coordinating NCMS work among related line departments and with establishing provincial level leadership coordination groups, with an office in the provincial Bureau of Health (*Weisheng Ting*). This office is charged with

²⁴ The full list of ministries and departments whose remit includes NCMS work is longer than this (see Gao 2010, p. 42).

day-to-day NCMS work and reports to the provincial NCMS leadership group, which in turn should report to the Party Committee, the People's Congress and government at the same level. The province is also charged with establishing an NCMS Technical Guidance Group, with members drawn from universities and others familiar with rural health work. Since 2003, annual national NCMS work meetings have been convened, reinforcing the importance attached to the scheme by the centre.

All in all, provinces play a less important role in development of NCMS *policy* than do counties: to a large degree, provinces are charged with overseeing scheme operation in counties under their jurisdiction and their policy role tends to revolve around regulating counties' management of the scheme, though a number of areas of regulation within the NCMS have been developed at the provincial level in advance of national policy, including finance (*caiwu*) and accounting (*kuajiji*) systems, both of which were set on a provincial basis before becoming the object of national regulation in 2008 (see below). Other areas in which provinces regulate the scope of activity of counties are in setting lists of drugs and procedures that are reimbursable by the NCMS and in pushing counties to converge on a certain number of 'models' of NCMS operation (see below), and/or in attempting to harmonise county reimbursement plans. Provinces are also responsible for oversight of county-level NCMS reimbursement plans, carrying out training of county NCMS managers and ensuring county baseline assessments are completed; they are also charged with researching problems in NCMS implementation and developing solutions, including latterly, for example, the establishment of provider payment pilots, experiments with ways in which county NCMS schemes can interface with out-of-county hospitals (difficult, given China's hierarchically structured administration) and similar (e.g. Zhang 2011b, Chapter 3). Provinces also collate and report NCMS information from counties under their purview and report this to central ministries (see NCMS Pilot Evaluation Group 2006, p. 22; Gao 2010, pp. 41-44) and contribute funds to the scheme.

5.2.3 Target-driven development: NCMS as pressurised system

Development and roll out of the scheme has been consistently based on national targets; main targets are set nationally and must be reached by county schemes. This clearly conforms to the logic of the pressurised system (see Chapter Four). In 2003, *de jure* piloting of the scheme started, and provinces were asked to choose two to three counties as pilots (Guobanfa 2003, No. 3); indicative criteria for choosing pilot counties were set by the centre (Guobanfa 2004, No. 3). In 2005, provinces with comparatively better conditions were allowed to add new pilot counties (Weinongweifa 2005, No. 319) and in 2006 targets for stepped national roll out of the scheme were set – 40% (approximately) of counties in 2006, 60% in 2007 and basically full national coverage in 2008; by 2010, the scheme should ‘basically cover’ all rural residents (Weinongweifa 2006, No. 13). In 2007-2008, the centre released national regulations on a number of elements of scheme functioning, including changes to central NCMS transfers, financial management and accounting systems, use of check ups within the scheme and similar.

Amounts of money in the scheme have grown very rapidly over time: from a very low base of around 30 RMB per user in most places in 2003, government funding was increased in 2006-2007 (at least 40 RMB combined central and sub-national contributions) (Weinongweifa 2006, No. 13), in 2008 (80 RMB from government, to be accompanied by a doubling of user contributions to 20 RMB) (Weinongweifa 2008, No. 17), in 2010 (150 RMB) (Weinongweifa 2009, No. 68),²⁵ and in 2011 (government contribution to increase to 200 RMB, with user contributions to increase in ‘appropriate’ degree) (Guobanfa 2011, No. 8). Scheme funding levels are national targets and do not necessarily reflect exact amounts of money in any given county scheme: early analyses showed funds to be variable to a degree, tending to be higher in the east than in poorer central and western regions (NCMS Pilot Evaluation Group 2006; Gao and

²⁵ In 2009, a 120 RMB target was first announced, followed later in the year by the 150 RMB target given above (Zhonggongzhongyang Guowuyuan 2009; Guofa 2009, No. 12).

Han 2007). Generally, provinces are to determine the split in contributions among sub-national levels of government (Guobanfa 2003, No. 3). Richer counties and cities may well increase funding for the scheme significantly beyond the central requirement at any given time. One study, for example, reported combined contributions of 450 RMB per participant in Shanghai in 2007 – when national targets were just 50 RMB per person (Hu, Tang et al. 2008). One city in which I carried out fieldwork has been consistently ahead of national and provincial funding levels, to the extent that when the province released regulations requiring that scheme funding reach the equivalent of 3% of local incomes, this brought the provincial funding requirement more or less in line with the level at which the city was already funding the scheme (Interview no. 46). Conversely, poorer counties may be given extra time to reach targeted funding, in recognition of their circumstances (e.g. Weinongweifa 2006, No. 13).

From 2009, the centre set enrolment (as opposed to coverage) targets, with sign up of 90% of eligible users targeted over three years (Guofa 2009, No. 12).²⁶ From 2009, and the inclusion of NCMS targets under the umbrella of the New Health Reform (*Xin Yigai*), reimbursement rate targets have been included in national policy: a target of 60% inpatient reimbursement (*buchang bili*; see discussion of terms, below) and an increase in inpatient reimbursement ceilings to six times local average incomes was set in 2009 (Guobanhan 2010, No. 67). In 2011, an approximate (*zuo you*) target of 70% inpatient reimbursement was announced and the stipulation on inpatient reimbursement ceilings was strengthened, with the requirement that these normally be no less than 50,000 RMB (Guobanfa 2011, No. 8). In 2012, national targets stood at: 95% of eligible users signed up; approximately 300 RMB funding per user (rising to 360 RMB by 2015), of which government is to contribute 240 RMB; an inpatient reimbursement rate target of approximately 75%; an inpatient reimbursement ceiling of eight times local average incomes and not less than 60,000 RMB; generalised use of outpatient pooling (see p. 108), to replace family accounts, progressive increases in

²⁶ Though from the start, 80% sign up levels were required in order for counties to be eligible for matching funding (Brown and Theoharides 2009).

outpatient reimbursement rates and increases in reimbursement for certain specific high-cost outpatient items (anaemia, diabetes and the like) (Zhonghua Renmin Gongheguo Weisheng Bu 2012). Targets are not confined to increases in funding; other numerical targets have increased over time, including use of outpatient pooling, on-the-spot reimbursement, inclusion of specific categories of illnesses and similar.²⁷

5.3 NCMS as central policy: NCMS as financing mechanism

Development of the pressurised system theory was largely in the context of economic development in the 1990s. The NCMS shows a slightly different picture: national targets drive scheme development, as above, but actual functioning of the scheme revolves around the aggregation, transfer and spending of large amounts of money. The existing pressurised system theory requires supplementing with an understanding of the way that funds are managed within the scheme and 'fund risk' is managed, both systemically and locally.

5.3.1 NCMS funds: 'Fund risk' as a category of analysis

The NCMS is a fund aggregation and transfer system: it aggregates funds from multiple sources (principally government and users) at the county level, and transfers these to healthcare providers for reimbursement of healthcare spending. The majority of analyses available in English are evaluations of the effectiveness of the scheme in improving health outcomes and in reducing impoverishing individual health spending (e.g. Zhang, Cheng et al. 2010; Zhang, Yi et al. 2011). In contrast, the financing and management system underpinning operation of the NCMS – the process by which funds are aggregated, transferred and controlled – has received no analysis in English.

²⁷ It should be noticed that, while the NCMS is primarily driven by targets, these are output-based targets, relating to specific scheme variables, rather than to actual health outcomes (on outputs vs. outcomes, see Hill and Hupe 2009, p. 9). One provincial health manager I interviewed said he would be very much in favour of experimenting with setting targets related to health outcomes, and that he thought this could greatly change the ways counties implement the scheme. He thought this would be beneficial, but that such a system would need to be fair and impartial, and would need to be seen to be so, or it would be challenged by counties. The province, he said, lacked resources to implement such a scheme.

The analysis that follows revolves around a notion central to very many Chinese analyses of the NCMS, especially analyses of the NCMS financing system, and one which occupies a central place in the discourse of county NCMS managers whom I interviewed: 'fund risk' (*jijin fengxian*). The term exists in opposition to 'fund safety' (*jijin anquan*). 'Fund', here, refers to the funds for operation of the NCMS; use of this term is broadly constant across various types of analysis. The *risk* associated with the NCMS fund, however, depends to a certain degree on the position of the observer or analyst. In many academic and policy-related analyses, discussed below, 'fund risk' refers to the potential for misuse of NCMS funds through fraud, diversion of monies, improper implementation of policy and the like. Gao, an analyst of the NCMS financing system, divides fund risk into three broad areas within the operation of the NCMS overall: there is risk in getting the funds in place (through the transfer system and collection of user fees at the local level); there is risk in managing the funds at the county level (relating principally to the management of NCMS accounts and technicalities of payments); and there is risk in the fund use part of the system (poor setting of reimbursement plans and the possibility of counties overspending, improper use of funds and similar) (Gao 2008a). In other words, Gao's framework encompasses the entire chain of operation of the NCMS, from fund collection, to management, to final transfer to users and providers for the reimbursement of users' medical expenditure. County NCMS managers also talk in terms of fund risk, related to the pressure of implementing the NCMS and, principally, the possibility of overspending and/or going bankrupt, due to the very tight funding constraints they generally work under, the speed of change of the environment they are working in and, frequently, the difficulty of anticipating what results of a given intervention or change in NCMS operation will be. This is examined in detail in subsequent chapters.

NCMS funds are derived from multiple sources, but are pooled and managed at the county level. The county is relatively low as a level of pooling, and the small size of the pool may make local schemes vulnerable to sudden or large changes in local disease incidence, health spending or simple mismanagement (e.g. Yi, Zhang et al. 2009b). Of

most interest for my analysis are the ways that NCMS funding and the NCMS financing system serve to contain risk at the county level and minimize systemic risk in the operation of the scheme. Several main mechanisms are important. First, the two main components of the NCMS financing system – the fund transfer system and the system of ‘closed management’ of funds – are systemic attempts to contain fund risk at the county level and ensure that NCMS funds are used for their proper purpose. Second, various management mechanisms, such as county-level ‘risk funds’ (*fengxian jijin*) and retention of fund surpluses (*jieryu*) serve to buffer operation of the scheme locally. Finally, and most importantly, county schemes have, for the most part, limited funds and are subject to a hard budget constraint.

The result of these systems is that fund risk is largely, if not entirely, localized at the county level – counties face great risk in operating the NCMS on limited budgets and in an uncertain and rapidly changing environment. Fund risk combines with NCMS targets handed down to counties to create great pressure at the county level, especially in poorer counties. This is exacerbated by two related stipulations of NCMS operation: that counties provide a proportion of scheme funds, and that management expenses (which may be considerable) be provided out of county budgets and not be deducted from NCMS funds; more general capacity constraints are also a concern.

5.3.2 NCMS funds: Guiding principles and the fund transfer system

The MOF and MOH did not release substantive national regulations on NCMS financial management until 2008, though provincial regulations were released before this. The 2003 State Council Opinions on the establishment of the NCMS (Guobanfa 2003, No. 3) contained only an outline of NCMS financial management; of most interest are the following stipulations:

- NCMS funds are ‘tripartite’, in that they are composed of user fees, plus contributions from sub-national and central government (in less developed areas); user contributions are to be in place before sub-national and central matching funds can be applied for; user participation is, in theory, voluntary.

- The NCMS is a pay-as-you-go scheme (*yi shou ding zhi, shou-zhi pingheng*): with the exception of certain, specific, fund categories, schemes raise funds in a given year to be spent in that year and incomings and outgoings must be balanced at the county level.
- NCMS funds are earmarked funds (*zhuan kuan zhuan yong*): they may be used for NCMS reimbursement and health check ups (*tijian*) only.
- Funds are to be managed in dedicated NCMS accounts (*zhuan hu chu cun*) in state-owned banks, and the NCMS Office is responsible for actual management of this account.

The 2003 Opinions provide a large-scale framework only. Initially, counties were asked to manage NCMS funds according to these principles and to devise their own rules on financial management (Guobanfa 2003, No. 3) and provinces were required to develop provincial financial management and accounting systems from 2004 (Guobanfa 2004, No. 3). During the early stages of scheme development, in other words, no national regulations existed governing financial management of the scheme. Despite this, central financial support for localities implementing the NCMS was rolled out from 2003, at a rate of 10 RMB per enrolled user in counties (excluding cities and districts) in central and western regions. Central support was conditional upon local governments enrolling people in the scheme and providing matching funds. Applications for funds were to be made through the BOF and BOH at the provincial level and provinces were to report key indicators on county-level scheme functioning (financial support provided by different levels of government, numbers signed up and main scheme indicators, including categories of illnesses covered, amounts paid out, etc.) in order to secure central funding (Caishe 2003, No. 112). The 2003 document specifies that application for central funds is to be overseen by the *Zhuanyuanban* (ZYB), a dispatched organ of the MOF at the provincial level, responsible for oversight and audit on behalf of the central MOF.²⁸

²⁸ In the mid-1990s, due to changing needs for inspection in a more diversified economy and greater use of project-specific (*zhuan xiang*) transfers by central

In 2004, very detailed regulations were released setting out the work of the *Zhuanyuanban* in oversight of NCMS fund applications (Caijian 2004, No. 91), according to which all documents relating to provinces' applications to the centre must be checked by the ZYB prior to submission to the MOF. The review procedure is specified in great detail and the ZYB is specifically charged to make sure that local NCMS funds are in place and that numbers of scheme members reported by counties and provinces are accurate. This documentary oversight is to be complemented by detailed county inspections of numbers of users signed up, amounts of money paid by users, whether NCMS funds have been deposited in counties' designated NCMS accounts, whether the county has an effective fund management system in place, and the state of counties' fund management. According to these regulations, the ZYB is to carry out on-site inspections of at least one to three counties each year. In addition to acting as a gatekeeper for the MOF in checking provincial applications, the ZYB reports data and findings directly to the MOF, providing a separate reporting channel and central check on local NCMS data as well as the state of local financial management systems. In 2007, the timing and scope of the ZYB's work was amended slightly (Caishefa 2007, No. 5).

The *Zhuanyuanban* plays an extremely interesting role, at least in early stages of development of the NCMS, in that it allowed piloting of the scheme before a well-developed financial management system was in place to deal with this. The detailed nature of ZYB regulations, and the degree of specification of the role of the ZYB is in very great contrast to the level of detail in MOH documents of the same period. Academic analyses point to the fact that the ZYB played an important role during the early stages of the NCMS, providing feedback on implementation of the

government, the *Zhuanyuanban* system was created by dismantling and repurposing what had been the central SOE/factory inspection system, under which inspectors carried out embedded oversight within central state-owned industries (*Zhongyang Qiye Caizheng Zhuchangyuan Jigou*) (Jin 1995; Zeng 1999). The ZYB has seemingly been involved in most major reforms since the 1990s (China Financial Publishing House 2002, p. 112; China Financial Publishing House 2005, p. 134; China Financial Publishing House 2007, pp. 166-168).

financial system at the local level (Liu 2008, p. 8), and publicly available reports written by provincial ZYBs show a number of concerns during early stages of the NCMS, including improper management of NCMS accounts and funds at the county level, delays in central and provincial funding getting through to the local level and structurally-induced conflicts of interest in local NCMS management leading to poor operation of the scheme.²⁹ One recurring criticism is that NCMS accounts are frequently not properly managed, in particular that 'closed management' systems (see below) are not properly implemented (Hu and Lei 2004; Chen 2005). A 2008 Jiangxi ZYB report (Caizhuguanjian 2008, No. 39) found a number of problems in the province: provincial funds were not in place on time; in some counties, user contributions were paid by collectives in violation of NCMS rules; fund management did not conform to regulations (counties were found to be paying NCMS monies to providers as operating expenses, to be using NCMS funds to invest in government bonds and to be holding funds in (higher interest) term accounts, rather than instant access accounts); a number of counties were found to be guilty of reporting incorrect figures upwards.

5.3.3 'Closed accounts': The local fund management system

The 2003 Opinions fail to specify in detail how NCMS finances are to be managed locally, but do specify that the principle governing management is that funds be deposited in dedicated NCMS accounts in state-owned or state-controlled banks. Subsequent development of the financial system revolves around the notion of 'closed management' (*fengbi yunxing*) of funds, first stipulated in 2004 in instructions to provinces on developing the finance system (Guobanfa 2004, No. 3). The 2004 regulations set out principles of local fund management: finance departments are to set up NCMS accounts in local branches of national banks and management of funds is to be carried out by the NCMS Office at the county level. The guiding thinking of the 2004 document is that banks should manage funds, but not manage the NCMS account; that the NCMS Office should manage the account, but not manage funds;

²⁹ This structural conjuncture is discussed in the next chapter in the context of development of a county-level oversight innovation in Taoshan.

that receipt and payment of funds should be separate, as should management and use; and that funds should be managed according to a system of 'closed management' (Guobanfa 2004, No. 3, Article 13). In 2008, national fund management regulations were issued (Caishe 2008, No. 8), elaborating on the principles set out in the 2003 and 2004 documents and clarifying detailed management procedures; alongside these, national NCMS accounting regulations were released.

Actual implementation of closed management has two main variants: 'closed management' (*fengbi yunxing*³⁰) and 'separation of payments and receipts' (*shou-zhi liang tiao xian*) (Gao 2008a). The core of both variants is the same: that counties establish only one NCMS account (i.e. the BOH, NCMS Office, townships, etc., are all prohibited from opening separate accounts), that this is held in an appropriate local bank (specified by the province), that this account is under direct oversight of the county BOF, and that payments from the NCMS account can only be made with direct approval of the BOF. The main difference between the two systems concerns whether NCMS payments to providers are routed via the NCMS Office (separation of payments and receipts), or whether the BOF makes payments directly to providers (closed management), and a significant amount of Chinese analysis has been devoted to the merits of both systems (Ying 2006; Gao 2008a; Gao and Zhu 2009; Gao 2010). Both systems should be cashless, relying on use of official receipts for payment and/or direct transfers between the NCMS bank account and providers' bank accounts where conditions allow this.³¹

Oversight of the NCMS account is to be carried out by Bureaus of Finance and Health and local audit offices, and NCMS Offices are required to report regularly on fund levels and scheme management to Bureaus of Health and Finance, which report to local government and to government at the next level up. The core function of both

³⁰ Confusingly, the name of the management variant is the same as the term for the overall principle of fund management.

³¹ Over time, NCMS reimbursement has shifted in most places from a system in which users pay in advance and present receipts to the NCMS Office for reimbursement to one in which reimbursements are deducted from payments at point of treatment and providers are later compensated by the NCMS.

systems is the same: to insulate NCMS funds from possible interference to the greatest degree possible by placing them under the BOF at the county level. As above, in early stages of development of the NCMS at least, the state of local closed management systems and the effectiveness of fund management received close oversight by the ZYB. A State Council-mandated interim report on the state of the NCMS, carried out early in the development of the scheme (2006), found management systems, including financial management systems, to be generally in line with central requirements in counties examined, and that local NCMS managers believed that while fund risk existed in the NCMS, the fund management system was basically secure (NCMS Pilot Evaluation Group 2006, pp. 10, 42).³²

Saying this, Chinese academic analyses, including an analysis of NCMS fund risk conducted for the MOH in 2009 (Gao and Zhu 2009), have found application of closed management systems to be uneven. Even very late analyses, following roll out and release of standardised national NCMS finance regulations, say that this is not rigorously implemented: Gao, a member of the national NCMS technical guidance group and analyst of the NCMS finance system, concludes that while counties operate closed management, in fact “any number of different practices” exist and that fund problems are *increasing* (Gao 2010, pp. 135, 143), though it is hard to know to what extent her analysis reflects current management versus the state of management at earlier stages of operation of the scheme. However, the fact that very recent MOH policy continues to stress the need for localities to improve fund security – including implementation of separation of payments and receipts systems – tends to reinforce Gao's argument (Weibannongweifa 2010, No. 53; Weinongweifa 2011, No. 52).

Analysis of the management of fund transfers in China has found that local governments may divert earmarked funds from their intended use (Liu, Wang et al.

³² In 2005-2006, two large-scale interim evaluations of the NCMS were carried out on instructions from the State Council. They bring together a number of academic institutions and departments of the Ministry of Health. Together, they come to more than five hundred pages and represent a large investment in understanding the state of the scheme at that time. Full references are given in the text.

2009), and this seems to be visible in the case of the NCMS. Gao, for example, cites cases of counties diverting NCMS funds to other local spending (principally public health) and the continued purchasing of government bonds with NCMS funds (Gao 2010, pp. 139, 143), something specifically prohibited in the 2008 regulations (Caishe 2008, No. 8). Liu et al., in an examination of a huge county finance dataset find, however, that use of closed accounts and similar methods reduced amounts of funds diverted in their case. Local NCMS managers I interviewed stated that the NCMS Office, at least, sees no actual money (indicating that the closed management systems seem to be operating, at least in these counties). One county, Feitian, has used a system of closed management since they started implementing the NCMS, and the pre-2008 system does not seem to have been significantly different from the later national system. At the provincial level, one NCMS manager from a developed eastern province, estimated that around three to five percent of NCMS funds had been diverted (Interview no. 63), a figure close to figures given by Liu et al. for leakage in their county sample. Training materials prepared by Hunan provincial BOH on implementation of the finance system discuss fund risk, and argue that the greatest risk to funds is in deliberate diversion of these by county governments (Hunan Sheng Hezuo Yiliao Ban 2006), and that this risk exists even where closed management systems are in operation. In 2010, the National Audit Office audited NCMS operation in 45 counties in nine provinces and, publicly at least, reported finding only small problems (though they found diversion (*nuoyong*) of around 27 million RMB in five counties/cities audited) (Zhonghua Renmin Gongheguo Shenjishu Bangongting 2011, No. 7). Gao concludes, rather sanguinely, that “objectively [speaking], fund risk exists” (Gao 2008a, p. 129).

How should we see this system? Despite the above criticisms, we should interpret the above as a *relatively* robust fund delivery system: significant effort goes in to making sure that central funds are disbursed only when local user contributions and matching funds are in place, and the *Zhuanyuanban* provides oversight of the local level on the centre's behalf.³³ Once funds are aggregated at the county level, in theory they are

³³ The ZYB acts principally as an overseer of the centre's NCMS funds: its role is

under tight oversight of the county BOF and this should reduce, at least, the potential for local misuse: the greatest potential for large scale misuse is in local *government* diverting funds. In other words, closed management systems concentrate fund oversight within the 'safest' part of the county-level administration but, *in extremis*, they remain vulnerable if county governments are determined to misuse them. The provincial interviewee cited above was direct in his assessment of this system: I raised the fact that much Chinese language analysis laments the poor state of local oversight of the NCMS, and the impossibility of this under current management and capacity constraints, but that there does not seem to have been huge diversion of funds. The explanation given for this was that funds are contained in the closed management system at the county level. When asked the extent to which he thought funds had been diverted, this interviewee gave the 3-5% estimate cited above. Whatever one thinks of the magnitude of this figure (or the 11% figure given by Liu et al. (Liu, Wang et al. 2009) for diversion of health spending), structurally speaking, the NCMS financing system appears to function in a *relatively* robust manner in ensuring delivery of funds to the county level and minimising, though not eliminating, direct misuse.

5.3.4 Local scheme balancing: Enforcing a cellular structure

As above, the 2003 State Council Opinions on the establishment of the NCMS (Guobanfa 2003, No. 3) contained the provision that scheme funding must be balanced at the county (city) level, described as 'setting payments according to receipts and balancing revenue and expenditure' (*yi shou ding zhi, shou-zhi pingheng*). In health insurance terms, this is simply a statement that NCMS pooling takes place at the county level and, as above, one can debate whether pooling at this level is optimal for spreading disease-related risk. Seen from another angle, requiring local balancing of funds is a means of enforcing a cellular funding structure and containing fund risk.

specifically to oversee the process of local application for central funds, the veracity of relevant information, and to ensure that NCMS funds actually get through to the county level; it is not charged with oversight of actual fund use at the county level. Oversight of actual functioning of the NCMS is principally at the county level; see Chapter Six.

The strongest expression of this is that county schemes face something approaching a hard budget constraint. This is shown by the numbers of counties whose schemes have overspent and/or gone bust (*po pan*; '*pan*' refers to the '*jijin panzi*' – the pool of funds). Zhang Xiulan, from the School of Social Development and Public Policy at Beijing Normal University, gives the following figures: "In about 2,750 counties, 1,219 counties suffered fiscal deficit in social pooling, among which 122 counties were still in deficit after making use of previous revenue and risk fund, with an overspend of 970 billion [RMB] accounting for about 25% of their year revenue in 2009" (Zhang 2011a). In 2006, 109 counties of around 1,400 counties then implementing the NCMS 'overspent' (*chaozhi*) (Gao 2008a, p. 128). According to one provincial interviewee, approximately a third of counties in that province (a developed eastern province) went bust as late as 2010 (Interview no. 63). Another provincial NCMS manager explained the rationale for this: all funds are subject to closed management and this is audited every year by the province – counties must know, he said, that the province will not allow 'problems' to occur in management of these funds, and that if problems do occur, these are the responsibility of the counties ("*bu yongxu chu wenti – shei chu wenti, shei chengdan zeren!*") (Interview no. 48). This was mirrored by interviews in Feitian, one of my main fieldwork counties (Interview no. 57).

Provisions exist to give counties a buffer in the management of their schemes: they are permitted to retain a small percentage of scheme funds as 'surplus' (*jieyu*)³⁴ (capped at 15% of funds for a given year, or 25% of historically accumulated funds; see Chapter Eight) and in 2004, regulations on the NCMS 'risk fund' (*fengxian jijin*) were released, according to which counties should maintain a separate fund (of around 10% of NCMS funds) to be used in the case of 'abnormal' (*fei zhengchang de*) fluctuations in spending which threaten the viability of the scheme (Caishe 2004, No. 96); the importance of this risk fund has been reiterated over time.³⁵

³⁴ Note that in Professor Zhang's analysis, above, 'previous revenue' refers to 'surplus' as I use the term; my term conforms to normal Chinese usage.

³⁵ One recent analysis points out that 'risk fund'-type arrangements are common in pay-as-you-go schemes elsewhere, as can be easily understood (Zhang, Cheng et

The 2008 finance regulations specify what is to happen in the event that a county scheme spends all its funds: the county should first use its surplus, then apply (to the finance department, either at the county or provincial level – this varies locally) for money from the risk fund. If these funds are exhausted and the scheme still cannot meet its obligations, the BOH should approach the county government for top-up funds (Caishe 2008, No. 8). County NCMS managers questioned about this pointed out the risk inherent in this: there is no guarantee, even in comparatively wealthy areas, that the county (city) government will have funds to bail out the scheme at the end of the year – this point was even made to me by the NCMS director of a wealthy eastern city, who was, though, at pains to assure me that the city government would never let the scheme go bust. The provincial interviewee cited above stated that of the schemes that had gone bust in that province in 2010, none had been bailed out and that this was entirely the responsibility of the county governments concerned; the province would not intervene under any circumstances (Interview no. 63).

If counties do overspend, there are several main, though less sanctioned, options. One is to curtail scheme reimbursement, which involves manipulating scheme variables so as to reduce reimbursement levels in order to avoid going bust. Possible levers include manipulating drug/service lists (though this is likely becoming harder over time, see below) and/or changing the scope of reimbursement and/or reimbursement levels for specific items. *In extremis*, providers may be forced to internalise costs; this is probably most likely in cases where providers are seen as having contributed directly to overspending. While the interests of providers and Bureaus of Health diverge much of the time, they converge in the need to prevent the scheme going bust, which – as well as being a clear indicator of management failure by the BOH and county – would both damage hospitals' revenues and could prompt government intervention in their management (Interview no. 46).

al. 2010a). An interesting point to note is that use of risk funds in China seems to originate not in contemporary development of PAYG insurance, but in stabilising and managing grain supply (Guofa 1993, No. 70). In other words, they seem to be a part of the Chinese government repertoire for risk management and maintenance of a cellular risk structure.

Systemic mechanisms allowing counties to buffer their schemes – risk funds and maintenance of local fund surpluses – are clearly inadequate, at least to a degree, in preventing overspending. The fact that even late on in the operation of the scheme a significant number of counties, even in developed provinces where, all being equal, levels of management should be greatest, continue to overspend and/or go bust should be a cause for concern. The larger point, though, is that this relates to the setting of the county as the pooling unit or, in other words, the enforcement of a cellular scheme management structure, in which counties bear responsibility for fund risk. Chapter Eight examines sub-national development of an innovative mechanism for dealing with this risk and subsequent spread and incorporation into national policy.

5.3.5 Interim summary

In the sections above, I have described how the NCMS enforces a cellular structure under which county schemes (in the absence of an increase in the pooling level, which is not normal) are responsible for their own spending and budget management and operate under a hard budget constraint. The systems according to which funds are transferred to the county and managed at that level should be considered *relatively* robust (though by no means perfect) and serve to reinforce this cellular structure. In health insurance terms, this is a question of the level at which scheme pooling is set and, as above, pooling can be criticised for being excessively local. Chinese analyses of this system have predominantly been phrased in terms of ‘fund risk’, and the ways that these systems contain (or fail to contain) fund risk at the county level. This fits with, and complements, our understanding of the pressurised system (*yalixing tizhi*) and central target setting for implementation at the local level.

The enforcement of this cellular structure creates the overall environment within which management and development of the scheme takes place, and counties – and within counties, principally Bureaus of Health and NCMS Offices – must deal with ever-increasing targets and requirements handed down in policy from above. In addition, the fact that local governments must contribute to the scheme, and the

unfunded mandate of NCMS management, create a significant financing burden in many, especially poorer, counties.

5.4. Local actors within the central policy framework

Counties are charged with actual implementation and management of the NCMS and they are also asked to play a considerable role in policy development. This section makes three inter-related points: first, NCMS management at the county level is where the pressure of implementation is mainly felt; second, while much scheme design is local and schemes vary significantly within an overall national framework, as can be seen in the range of reimbursement models trialed and used within the scheme, counties have been hampered by very limited levels of funding; third, sub-national governments are called on to play a developmental role, coming up with innovative policy mechanisms within the scope of the scheme.

5.4.1 Counties as scheme managers

The last sections have shown how the NCMS aggregates funds and places these under the oversight of county Bureaus of Finance. In contrast to the above, the county is the level at which NCMS funds are actually spent – on reimbursement of users' medical expenses (either directly, or through payments to providers). A 2006 interim survey of functioning of the NCMS divides functions of the NCMS Office into, broadly: fund collection, reporting and analysis of fund use; publication of NCMS fund use; auditing local medical fees; oversight of services provided by 'designated providers' (*dingdian yiliao jigou*; those providers included within the NCMS system); checking and approval of providers for inclusion in the NCMS (NCMS Pilot Evaluation Group 2006). At the county level, main responsibilities are divided between government, responsible for fund collection (dealt with briefly below) and publicising the scheme (*xuanchuan*), and the NCMS Management Committee (*Guanli Weiyuanhui*) and NCMS Management Office (*He Guan Ban*, hereafter NCMS Office). (The BOF has been discussed above.) The NCMS Committee is under the dual leadership of the county government and the BOH at the superior level; the NCMS Office is charged with actual implementation and

management. These agencies may be replicated at the township level, but this is not universal, though township governments tend to have a responsibility for fund collection and publicising the scheme. Villages may also have responsibilities delegated to them, but, again, this appears not to be universal (Gao and Han 2007).³⁶

It is important to understand the significance of this structure, which *places the county-level NCMS management at the focal point of the scheme*: funds from multiple sources are channeled to the county and aggregated under county control; overseeing NCMS funds, budgeting, reimbursing providers and users and overseeing reimbursement procedures are the core of the work of the NCMS management (the NCMS Committee and Office under the BOH), which stands at the interface between the funding system, users and providers (hospitals, township health centres, village clinics), and it is at the county that targets for implementation of the NCMS are principally felt.

5.4.2 Scheme enrolment: Central targets and local responsiveness

In theory, the NCMS is based on 'voluntary' sign up of users to the scheme on a rolling annual basis. This decision was a result of debates in the 1990s and early 2000s about the 'peasant burden': a new rural insurance scheme, while necessary, could not be seen as a burden by rural populations (Wang 2008). The result has been that local governments must, to a degree at least, 'sell' the scheme to potential users and persuade them to sign up. Two main points deserve mention: first, while localities must clearly make their scheme attractive to local users, the NCMS is better thought of as *semi-voluntary* than *strictly* voluntary; extremely high membership rates are the result, in many cases, of a degree of hard selling at the local level (see below). Second, signing up a semi-willing, frequently dispersed, rural population for the most part not in formal employment is often hard, time consuming, and expensive; the costs of doing

³⁶ This structure should be considered *reasonably* uniform, though from the start of the scheme, some counties have contracted day-to-day scheme management to commercial insurance companies (e.g. Mao 2005) and experiments have been conducted placing NCMS management under the social insurance administration.

this fall to local levels.

While signing users up to the scheme has certainly become easier over time, as people personally benefit, or see friends and neighbours benefit from the scheme (e.g. Gu 2008), this has been, and remains, hard, time consuming and costly in many places. In the majority of cases, fee collection involves township and village officials going door-to-door and persuading people to sign up; this seems to be an inevitable consequence of running a system based on (semi-) voluntary sign up where the majority of potential participants lack a formal workplace and salary. Various estimates exist for the numbers of users who sign up 'fully' voluntarily: between 20 and 30% (NCMS Pilot Evaluation Group 2006, pp. 47, 61). Signing users up may well require multiple door-to-door visits – an average of three, and sometimes as many as nine in one early analysis (NCMS Pilot Evaluation Group 2006, p. 4). In one of my fieldwork counties, Taoshan, NCMS managers described the necessity of 'doing work' (*zuo gongzuo*) to sign users up to the scheme: this involves a lot of publicity and village cadres going door-to-door to persuade people. If potential users don't sign up straight away, cadres must go back again to cajole them and try to persuade them. While this is getting easier, as more and more people derive benefit from the NCMS, it remains hard, according to interviewees in Taoshan: without active intervention, they estimated that sign up rates would likely be around 50-60% (as opposed to the current figure of more than 90%). This is not specific to Taoshan; other counties where I have carried out fieldwork report that fund collection remains hard.

This is costly: various reports have estimated costs of collecting funds in the early stages of the scheme at between 12% and 17% of funds actually collected (estimates are for various counties; see Mao 2005; Ying 2009); one estimate, from Ningxia, was of 20-30% of funds collected (Yan, Raven et al. 2010). The time spent on fund collection may also be significant: from two to three months in one county in Ningxia (Yan, Raven et al. 2010) to an average of three months (and half a year in extreme cases) in another study (Gao, Zhao et al. 2006). Over the course of development of the NCMS, there have been innumerable calls for development of a stable, low-cost fund

collection system, including at national NCMS meetings (Wu 2007); to date, this problem remains unresolved (Zhang 2011b, p. 85).

5.4.3 Central targets and local responsiveness: NCMS as a 'new mode of governance'?

One 2010 study interprets the NCMS as a radical departure in central government control of local agents: by requiring that the scheme be voluntary, the centre binds the hands of local governments, which must make the scheme appropriate to their constituents or risk seeing it become unpopular and fail (Klotzbücher, Lässig et al. 2010). This reading is an attempt to escape from a target-driven view of policy implementation (cf. Edin, O'Brien and Li, etc.), and to locate NCMS implementation outside the pressurised system and its deficiencies. These authors argue that use of soft performance targets based on numbers of users voluntarily signed up requires localities to implement the scheme, but leaves operation – how to make this a success – to be determined locally. The result is that local governments become 'responsive' to users, modifying schemes to increase popularity. This argument hinges on the NCMS being voluntary, and the study found "no evidence for any systematic coerced participation" (Klotzbücher, Lässig et al. 2010, p. 49). This claim is in line with much English language analysis (e.g. Brown, de Brauw et al. 2009; Yip and Hsiao 2009), but contradicts most Chinese analysis. Since the start, success of local schemes has been judged, at least in part, on numbers of users (e.g. Gu 2008, p. 96; Brown and Theoharides 2009), taken as an indicator of political performance (*zhengji de tixian*) (NCMS Pilot Evaluation Group 2006, p. 46). One interim report described sign up as 'semi-coercive' (*ban qiangzhixing de*) (MOH Center for Statistics Information 2007, p. 134), while another argued that "under current social and economic conditions, 'full voluntary' sign up, in line with government regulations is very difficult" (NCMS Pilot Evaluation Group 2006, p. 46).

This has resulted in a wide range of sign up practices, from legitimate (advertising, use of flyers, etc.) to questionable and illegitimate practices: one report described many

local practices as 'lacking an institutional guarantee' (*quefa zhiduhua de baozhang*), a signal that these are less than legitimate (NCMS Pilot Evaluation Group 2006, p. 48), while in one nationally-publicised case, Guangnan County (Yunnan) enlisted teachers and other public employees to collect user fees, threatening to withhold salaries if quotas were not met. This was clearly illegitimate and the case was reported nationally (Weibanjifufa 2003, No. 147).

A strong indicator that the NCMS is not a radical departure from target-based policy implementation, but remains within the pressurised system, is the importance of risk funds. Above, I describe these as a systemic means allowing local governments to buffer their schemes, allowing maintenance of a cellular financing structure. Risk funds have a second important function: in setting specific fund quotas to be reserved for fund stabilisation, they were a mechanism used, not to force schemes to run a surplus (most schemes were running large surpluses for fear of overspending), but rather to force schemes to spend the funds they had collected (I deal with fund surpluses and a linked reimbursement mechanism in detail in Chapter Eight). In other words, localities initially hoarded funds for fear of overspending, and this derived from then dominant targets – *exclusive* use of sign up targets (Gu 2008; Zhang, Cheng et al. 2010a) – remarked on by Klotzbucher et al. Counties had to be *forced* to spend NCMS funds – through use of a top-down target. Saying this, schemes are responsive *to a degree*: NCMS managers in Taoshan confirmed the overall point that schemes must be popular in order to attract users (Interview no. 23), and a degree of local responsiveness is also visible in my analysis of Feitian (Chapter Seven).

5.4.4 Tight budgets: Background to local scheme design

Many analyses of spending on social policy programmes in China have pointed out that a very large proportion takes place at the sub-national level, that much funding is derived from sub-national levels of government, and that this is inequitable. In the reform era, this has been the case for much health-related spending: central government spending on health overall constitutes only a small part of total

government health spending (Ter-Minassian and Fedelino 2008); by far the largest component of health spending overall occurs at the county and township level, at around 44% in one analysis (Wong 2009). One result of this is very variable levels of spending by sub-national jurisdiction; overall, government health subsidies, both supply-side (subsidies to providers) and demand-side (subsidies to insurance schemes) tend to be pro-rich (Wagstaff, Yip et al. 2009), both across and within provinces, and sub-national inequalities are not adequately compensated for by central government transfer payments. According to Wong, “the Chinese system depends on local governments to implement social policies, but lacks mechanisms for ensuring that they have adequate financial resources to carry this out” (Wong 2009, pp. 942-943).

For most of the operation of the NCMS, amounts of money in the scheme have been extremely limited compared to overall health spending and need. Various figures exist for per capita rural health expenditure. One 2009 analysis estimated per capita health expenditure in the western/central region at around 150-180 RMB, compared to 120 RMB average scheme funding at that point (Yip and Hsiao 2009). Two World Bank analysts give an estimate of 240 RMB (Wagstaff and Lindelow 2008b, p. 273), while a recent OECD analysis reckoned that the scheme would require around 300 RMB “in order to stand a reasonable chance of markedly lowering poverty due to catastrophic illness” (Herd, Hu et al. 2010, p. 26).³⁷ Many, even very recent, analyses have shown how very limited scheme funding has resulted in schemes having very low payout rates, severely limiting usefulness. Barber and Yao cite MOH data that average *de facto* inpatient reimbursement rates under the scheme in 2009 were 41.5% and that “major catastrophic events are not fully covered under the current programs” (Barber and Yao 2011, p. 350).

A recent study based on nationally representative panel surveys (2004/2007) found reimbursement rates so low that “they only help [users] meet less than 15% of the total

³⁷ Various measures of ‘catastrophic’ spending exist; see e.g. Wagstaff, Yip et al. (2009), Zhang, Cheng et al. (2010b).

medical expenditures when they have catastrophic illnesses” and that users frequently borrow, draw down savings or sell assets to be able to afford treatment (Yi, Zhang et al. 2009a, p. S126), while a 2011 report based on some of the same data shows decreasing payout rates as expenditure increases (as the absolute sum an NCMS user spends on inpatient care increases, the rate of reimbursement falls), with obvious consequences for reimbursement of catastrophic spending (here 4,000 RMB and above), though the study found that reimbursement rates had increased between 2004 and 2007. Overall, the authors believe, users are not being reimbursed at the rate promised by the scheme – the *de facto* reimbursement rate (see below) falls (well) below the advertised rate, with consequences for users (Zhang, Yi et al. 2011). Failure to be hospitalised after referral remains very common, and the main reason for this remains economic (MOH data reported in Barber and Yao 2011).

As above, scheme funding is increasing rapidly and, at around 300 RMB per user, is now around the level estimated by most analysts to be necessary to enable proper scheme functioning. While increasing scheme funding is clearly a good thing, it has the potential to increase the burden on local governments: as one recent study points out, the core of the equation of funding the NCMS for local governments is the discrepancy between growth in spending on social insurance and growth in fiscal revenue. This study cites the head of a county BOF in Ningxia as saying that

“[i]n recent years, the growth of input to expand social insurance has been much faster than the growth of fiscal revenue. For example, compared to 2004, the input to social security in 2005 increased by 157% in our county. Meanwhile, the revenue only increased by 4%. In this situation it is a heavy burden for government to support NCMS ... As NCMS develops further, the burden will be heavier” (Zhang, Cheng et al. 2010a, p. 106).

As these authors point out, in the absence of changes to the overall structure of revenue sharing and/or redistribution, this problem is likely to become more, rather than less, serious over time.

Given the limitations of redistributive measures within NCMS funding,³⁸ and few attempts to increase the level of NCMS pooling, the pressure of NCMS financing is clearly felt most acutely in poorer counties.³⁹ In addition, the pressure on counties to find NCMS matching funds has certainly been exacerbated, at least, during certain periods of implementation of the scheme, by delays in the disbursement of central and provincial funds (Gao and Han 2007; Zhang, Cheng et al. 2010a), though central policies have mandated changes to funding schedules in an attempt to overcome this problem (Gao and Han 2007, p. 89; Wagstaff, Lindelow et al. 2009b, pp. 37-38).

5.4.5 Local scheme design: Local setting of reimbursement plans

Counties are expected to play a role in developing the NCMS; the aspect of this which has received most attention is the design of scheme reimbursement plans. These have received much more attention than have local management models, reflecting, at least to a degree, the fact that the majority of analyses of the NCMS in English have been written by health systems analysts and/or economists. A number of analyses of the early period of development of the scheme, for example, pointed out the high degree of heterogeneity in the design of reimbursement plans by counties. In the early stages of

³⁸ Indeed, the 2009 World Bank study, which benefited from very significant collaboration with the MOH, devotes a full four pages to a discussion of possible methods of increasing inter-jurisdictional sub-national solidarity in NCMS funding (Wagstaff, Lindelow et al. 2009b, pp. 91-94).

³⁹ While the NCMS is a departure from previous practice in committing central funds to rural health insurance, the overall burden of funding remains inequitable. Overall, as a recent World Bank study points out, “county contributions are regressive – [NCMS] contributions compose a larger share of a poor county’s income per capita than of a rich county’s” (Wagstaff, Lindelow et al. 2009b, p. 89). This derives from the structuring of county contributions and inadequate targeting of central and provincial transfers to counties to even out imbalances: “counties pay the same contribution irrespective of their capacity to pay, except insofar as those in the west and center receive central government assistance and provinces decide to target central government and provincial government assistance on poorer counties (few, in practice, do so)” (Wagstaff and Lindelow 2008b, p. 274). This tends to confirm early evidence regarding NCMS spending, which found that the overall cost of implementing the scheme (i.e. local government contributions, plus operating expenses) was around twice as high as a percentage of government finances in the west as in the east (Mao 2005).

operation of the scheme, counties were almost entirely free to design their reimbursement schemes as they saw fit, with the proviso that these focus mainly on 'serious illnesses' (*da bing*), and that they carry out a baseline assessment (Weibanjifufa 2003, No. 47). As Mao, in an early analysis, phrased it, "every county designs its own reimbursement scheme according to its understanding and conditions" (Mao 2005, p. 15).

There has been much discussion of various 'models' trialled under the umbrella of the NCMS, depending on the *type* of spending covered: inpatient only, or inpatient plus some form of outpatient spending (often 'catastrophic' – i.e. large expenses only), or inpatient spending plus coverage of outpatient spending through either 'family accounts' (*jiating zhanghu* – medical savings accounts in which a family's contributions are 'saved' for their exclusive use) or generalised pooling of outpatient spending (*menzhen tongchou*). 'Family accounts' have received a degree of attention: they are not optimally efficient in terms of pooling risk of outpatient spending (they function as a pool of just a few people), but functioned as a stopgap mechanism to help increase trust in the scheme in the early stages of its operation by showing that user contributions could not be appropriated by government or used by others (NCMS Pilot Evaluation Group 2006, p. 34; Bloom 2011). Over time, family accounts have been progressively replaced by outpatient pooling: Mao (2005) reports early MOH data as showing that 65% of pilot counties nationally opted to use family accounts; by 2009 this had fallen to around 47% (Barber and Yao 2010).

As a result of limited budgets, the majority of schemes have, from very early on, put in place various 'demand side' measures to control costs (e.g. Wagstaff and Lindelow 2008b; Brown, de Brauw et al. 2009), though overall, most schemes have instituted fewer 'supply side' measures to control costs through control of health care providers.⁴⁰ 'Demand side' restraints used by schemes are complicated and various: schemes set

⁴⁰ This is linked to a very frequent criticism of the NCMS: that it failed to institute provider payment mechanisms to control costs. While largely true, this misses many solutions of varying degrees of formality adopted by implementing counties (see Chapter Seven).

thresholds and ceilings for reimbursement, and within the range of reimbursable expenses, set a specific reimbursement rate (i.e. scheme users face a co-payment even within the band of reimbursable expenditure between the reimbursement threshold and ceiling). This is complicated even further by the high number of items excluded from reimbursement: provinces set lists of reimbursable drugs and procedures (Guobanfa 2003, No. 3; Guobanfa 2004, No. 3), and these may be modified by implementing counties.⁴¹ Many items are excluded, though these may well be expenses routinely incurred during treatment. The narrowness of drug lists has been a consistent complaint. Procedures for referral between providers have tended to be set at the county level. Schemes tend to set higher reimbursement rates (or lower thresholds) for providers at lower levels (and *vice versa*), in order to encourage users to seek treatment at cheaper providers. Many schemes do not reimburse medical expenses deriving from self harm, accidents due to drug use or drunkenness and the like.⁴² Analytically, the above gives rise to a distinction between advertised reimbursement rates (*buchang bili*) and *de facto* reimbursement rates (*shiji buchang bi*). *De facto* reimbursement rates may be much lower than promised rates (e.g. Zhang, Yi et al. 2011); this is obviously an important concern for scheme effectiveness in reducing impoverishment related to health spending.

Scheme variations, as above, result in mind-boggling variation and complexity. In 2005, the State Council commissioned two large-scale reports by academic institutions and the MOH on the development of the NCMS at that point (NCMS Pilot Evaluation Group 2006; MOH Center for Statistics Information 2007). Together, these reports come to

⁴¹ Under the aegis of the New Health Reform plan, China started developing and piloting a national essential drugs list. This is significantly beyond the scope of this study, though this clearly will interact with the NCMS in counties where it is trialled and/or implemented (see e.g. World Bank 2010; Zhang 2011b).

⁴² Feitian County, for example, lists the following exceptions, which are fairly typical of county NCMS plans: industrial and work-related injuries, traffic accidents, accidents in which there is third party liability, fighting, excessive drinking of alcohol (*xu jiu*), suicide, disability resulting from the user's own actions (*zi can*), drug taking "and other illegal behaviours which result in harm" (Feitian County Bureau of Health 2011b).

more than five hundred pages, much of which is a description of variation in scheme implementation, different operating models, different funding levels and origins of funds, different spending levels, variations in local management and oversight procedures and similar. Analysis of variation in scheme design by Brown et al., based on a large early dataset, found that “fundamental characteristics of the NCMS programme, such as participation rates, the minimum spending levels eligible for reimbursement and reimbursement rates, all vary dramatically from one county to the next” (Brown, de Brauw et al. 2009, p. 328), while Wagstaff, Lindelow et al. examined variation across a number of county models, concluding that the impact of the scheme, understandably enough, varied by county, but that “the policy of ‘letting a thousand flowers bloom’ in the piloting of NCMS has much to commend it in terms of encouraging innovation, but it makes pinpointing the secrets of success very hard” (Wagstaff, Lindelow et al. 2009a, p. 17).

5.4.6 Interim summary

Local implementation of the NCMS is driven to a large extent by national targets. For most of its duration, and in most places, the NCMS has been run on very tight budgets. Local contributions required by the matching funding principle can be a considerable burden, especially in poorer counties, and this burden is likely increasing over time as overall levels of scheme funding are increased in line with central targets. In management terms, low levels of scheme funding, able to cover only a small part of potential health spending, combine with the hard budget constraint enforced by the NCMS financing system to locate risk in implementing the NCMS with the county BOH and NCMS management. Over time, funding has increased, but, as above, implementation targets have also increased dramatically to include not just scheme coverage rates, but also reimbursement rates, ceilings and the like. In other words, the pressurised system of target setting is combined with a cellular funding and management structure; the result is that perceptions of risk, especially fund risk, are a major factor underlying counties’ behaviour, and this is shown in the cases presented in later chapters.

Above, I argued that the core of the work of the NCMS management at the county level is in budgeting, reimbursing users' expenses and in overseeing reimbursement procedures. I return to these functions, and how they are linked with the pressurised risk system outlined here, in detail in subsequent chapters. In the next section, I show that counties have been expected to contribute to scheme design and development – surprisingly, perhaps, given the analysis above.

5.5 The developmental role of counties: Mechanisms and the importance of local initiative

In English language literature on the NCMS, while a degree of variation in the scheme is recognised, little of this is captured beyond variation in elements of reimbursement package design, except for a few exceptions (e.g. Wagstaff, Lindelow et al. 2009b), and various early studies which discuss county-level variation in a limited number of management arrangements (e.g. Mao 2005; Brown, de Brauw et al. 2009). Overall, however, this literature underestimates the degree of variation in the scheme.

5.5.1 NCMS as 'pilot'; policy principles and local practice

The NCMS is referred to in Chinese policy discourse (in policy, academic analyses, official documents and by policy makers and implementers) as a 'pilot' (*shidian*), even in the post-roll out period, and there is a strong rhetorical commitment on the part of the centre to local initiative in scheme development. The vocabulary used here risks misinterpretation by non-Chinese speakers: early development of the NCMS started with 'pilots' in a defined number of counties, roughly corresponding to use of the term in English – experimentation with a certain form of design on a limited scale. Overall, though, the NCMS is also termed a 'pilot', and here the term diverges from its normal English meaning: we should understand this as signaling that the NCMS as policy remains in flux and denoting a degree of tolerance and desire, on the part of the centre, for local generation of policy solutions. This is extremely evident from speeches at yearly national NCMS meetings and the like, and has been reiterated many times. At the 2003 national NCMS meeting, for example, Vice Premier Wu Yi explained the

rationale for this approach as “departing from actual conditions, summarising experience through piloting, ceaselessly perfecting and progressively standardising [the NCMS]”. In developing the scheme, the role of central policy is to set principles according to which local policy is developed:

“Something that needs to be stated clearly is that the ‘Guiding opinions’ [Guobanfa 2003, No. 3] only state principles [for development of the scheme]; questions of actual implementation are left to each area to research and we hope that you can set concrete [management] regulations in line with your local conditions” (Wu 2004).

Elsewhere, this emphasis on local adaptation of the scheme is contrasted explicitly with the possibility of a ‘one size fits all’ approach to development:

“The reason the document forwarded by the Office of the State Council was titled ‘Guiding opinions’ is that the difference in conditions between different rural areas is very large, and it is impossible to adopt a one size fits all policy; the main thing is for each area to experiment in the process of implementation in line with local conditions” (Wu 2005).

Piloting is framed as a process of discovering and solving problems, whose occurrence is an inevitable part of development of the scheme:

“... a very important point is discovering problems and solving them. The problems discussed above [in Vice Premier Wu Yi’s address to the 2004 national NCMS meeting] could not have been discovered, and could not have been resolved if we hadn’t carried out a pilot. It is normal that these problems should occur; the key thing is that we correctly understand these, treat them seriously and effectively solve them” (Wu 2005).

5.5.2 Central and local roles

Core to these appeals are three distinct notions: first, the centre and localities have distinct roles – the centre in setting the principles of reform, and the localities in

developing actual working methods. Second, this stems from a recognition, or belief, that designing the reform in advance is impossible – that adaptation and problem solving are an inevitable part of reform. Third, tolerance of multiple means of achieving the centre's policy goals is necessary. In her address to the 2007 national NCMS meeting, Wu Yi stated this extremely clearly:

“We use a method in which the centre sets the principles of policy and localities concretely organise implementable policy; we encourage each locality to experiment and innovate” (Wu 2007).

In calling for localities to develop an appropriate fund collection (*chouzi*) mechanism – a longstanding problem in development of the scheme – Wu Yi's frankness with regard to tolerance of multiple local policy solutions is astonishing:

“In summary, *whatever the means used*, it is only necessary that peasants are willing and agree [to accept this fund collection method], that the method be effective and low-cost – then it can be promoted and continually developed as it is implemented” (Wu 2007, italics mine).

From the above, several points deserve highlighting. First, the vocabulary of ‘principles’ (*yuanze*) and concrete management ‘methods’ (*caozuo*) reflects at least two things: a central recognition of the differences between localities and the importance of local management in adapting policy locally, plus the belief that piloting or experimentation, whether explicitly guided or organic, will produce valuable ‘experience’ (*jingyan*) useful for reform. The level at which this is expected to work is that of policy ‘mechanisms’. Second, existence of a degree of local discretion should not mask the fact that policy implementation is obligatory and governed by targets.

5.5.3 The importance of policy ‘mechanisms’

Mechanisms (*jizhi*) can be understood in relation to key terms including ‘policy’ (*zhengce*) and ‘system’ (*zhidu*), in which it is the junior partner. ‘Mechanism’ should be understood as referring to a component and dependent element of a larger, and

separately instituted, policy.⁴³ Speeches at national NCMS meetings frequently make reference to locally-developed mechanisms which are perceived or profiled as solutions to specific problems in the development of the scheme. Wu Yi's 2004 speech, for example, mentions development of a closed fund management mechanism (*guanli jizhi*) in Hubei, a mechanism for separate management and operation of NCMS agencies and providers in Jiangxi and the use of 'single-disease payments', a form of provider payment mechanism, in Shaanxi (Wu 2005).⁴⁴ Such examples are very common. I return to this vocabulary of 'principles' and 'methods' in subsequent chapters, where I argue that this vocabulary and understanding of differentiation of roles is shared by policy implementers, at least to a reasonable approximation.

There is continual profiling of sub-national developments through media, and through national meetings. A near-monthly Rural Health Brief (*Nongcun Weisheng Gongzuo Jianxun*) issued by the MOH since 2004 is particularly interesting, setting out national policies and showcasing sub-national policy developments. Though the Brief relies mainly on localities reporting local practices, and is therefore a non-representative selection of policy activity going on at any given time, it shows a great range of work going on sub-nationally, in different ways and at different speeds in different places and at different levels within the system.⁴⁵ At various points, the MOH has circulated materials on approved local practices. In 2009, for example, the MOH published a

⁴³ I prefer the term 'mechanism' to the vocabulary of 'main' and 'complementary' reforms as used by Göbel. At least in the case of the NCMS, these terms tend to be used most frequently to refer to other reforms being carried out in parallel to the NCMS, but on which this relies, such as reform of township health centres, vertical integration of rural health care providers and similar, rather than to refer to development of components of NCMS policy.

⁴⁴ My term, 'single-disease payments' is a simplified translation of '*dan bing zhong ding'e fu fei*' and '*dan bing zhong xian'e fu fei*'. This practice is often conflated with diagnosis related groups (DRGs) in English. This is discussed in more detail in Section 7.6.2.

⁴⁵ In 2010, the Brief reached its hundredth issue and was prefaced by an editorial note stressing its importance in "publicising classic rural health practices and successful experience" and praising its role in "stimulating information exchange and arousing 'work thinking' [and] guiding [localities] in doing rural health work well" (Weisheng Bu Nongcun Weisheng Guanli Si 2010).

collection of studies run by its NCMS Research Centre in the previous years, though this is not publicly available (Anon 2010). Very many academic analyses profile local practices in order to ask if these are valuable models for more widespread consideration or emulation. This list could easily be extended: there are many vectors for profiling and circulation of locally-developed mechanisms and models.

5.5.4 Pressurised system and local discretion

A lack of specificity in policy should not mask the fact that localities are obliged to implement this. A 2007 report for the State Council on implementation of the NCMS contains a very illuminating analysis of NCMS policy and the roles of centre and localities:

“Because of China’s size and the difference between levels of social and economic development in different areas, much central policy cannot be very concrete; often it just sets large principles. These large principles are coercive [*qiangzhixing*], but they only set boundaries that behaviour cannot cross – they specify ‘what you can’t do’ and often do not set the scope of legitimate action – ‘what you can do’. The boundaries ... are flexible; under the proviso that regulations are not infringed and policy is not opposed, the space for ‘innovation’ is quite large” (MOH Center for Statistics Information 2007, p. 136).

In other words, policy principles are passed down through the government system before requiring, for the most part, concrete articulation at the county level, where implementation is enforced, though the form of that implementation is, at least to a degree, unspecified (Sun and Chai 2009, pp. 122-124). This corresponds clearly to the ‘pressurised system’ (*yalixing tizhi*).

Third, this framework allows all sub-national levels a possibly significant degree of freedom in developing mechanisms within the overall umbrella of the NCMS. This is called for and expected and requires a degree, at least, of *tolerance* of variant local practices as long as these further implementation and development of policy. It also

requires, as above, a degree of central recognition that problems will be encountered in the course of policy development and tolerance of these. The corollary of this, however, is that variation, lack of standardisation and frequent lack of specificity in policy may give localities significant space to deviate from policy and implement in illegitimate ways. While the 2007 report cited above acknowledges that this use of framework policy, to be developed at the local level, allows space for innovation, it also results in localities developing countermeasures (*duice*) to central policy and 'playing edge ball' on the boundaries of vague central policy, increasing the costs of central monitoring of local implementation (MOH Center for Statistics Information 2007, p. 136). Sun et al. criticise this mode of policy development as making standardisation of policy hard to achieve (Sun and Chai 2009), and I return to this theme in later chapters.

5.6 Caveats and complexity: Difficulties of treating the NCMS as a national system

The NCMS is too complex and too heterogeneous to be amenable to overall analysis: it is very hard to treat the NCMS as *a system* given both the huge change that has taken place over a very short time, and the huge range of variation in policies – and practices of various degrees of correspondence to official policy – at multiple levels and in different places. The range of variation 'within' the NCMS, both in space and time, is huge, though (probably) declining over time. This range of variation is, however, ultimately unknowable, though it is possible to be fairly confident of *trends* within scheme development, in that national and provincial policies point in certain directions, and these directions seem to be confirmed by publicly available data and analyses: county-level arrangements for fund management (*caiwu guanli*), for example, have ceded to first provincial and then, in theory at least, a more or less nationally-standardised system of fund management; use of family accounts is being phased out in favour of outpatient pooling; and provinces have over time taken an increasing role in streamlining county-level scheme operation, reducing the number of county-level reimbursement models in operation and in trying to ensure that inter-county variation in funding and reimbursement levels is not too great

(Weinongweifa 2005, No. 319; Weinongweifa 2007, No. 253), in setting up and standardising NCMS information systems (under way at least to a degree and in some places) (Zhang 2011b), in developing on-the-spot reimbursement systems (Weinongweifa 2009, No. 62) and similar. National standards for NCMS reporting were released in 2008 (Weinongweifa 2007, No. 304).

Despite this, a huge amount of sub-national variation clearly remains, and there is no way to get a clear overview of the range of practices in different places and the way that different local schemes function. As above, the State Council commissioned two large-scale reports on scheme implementation during early phases of its development. These provide a very useful milestone, showing variation at a certain point in scheme development. To the best of my knowledge, there has been no repeat of this exercise, and while one could expect the scheme to have become more standardised after national roll out, and especially with the *de jure* consolidation of various components of NCMS policy at the national level around 2008, whether this is really the case is ultimately an empirical question to which there is no satisfactory answer.⁴⁶

In looking at the scheme overall, we have to rely on partial indicators and analyses: a recent Chinese review of rural health system development, for example, profiles swathes of sub-national developments, including increasing use of outpatient pooling, use of on-the-spot reimbursement systems, ongoing capacity building and provider payment reforms at both county and provincial levels (Zhang 2011b, pp. 78, 89-91). Central priorities for development of the scheme change, to an extent, over time:

⁴⁶ As above, as an outsider to the policy process, the amount of information available on scheme operation as a whole is limited: the centre, for example, uses various channels to understand progress on implementation of specific components of the NCMS, in addition to normal reporting required of sub-national governments. These include the use of 'designated' (*ding dian*) counties and provinces in which scheme development is observed more closely than elsewhere, local observation visits (*kaocha*) and the like, as well as semi-controlled experiments and discrete research projects and reporting on specific components of scheme development. Results of such analyses are mostly not publicly available, though related academic analyses may be in the public domain. Analysis in this chapter relies mainly on policy documents sourced through relevant government websites, ministry yearbooks, and academic sources.

provider payment reform, for example, whose absence in the NCMS has long been recognised by analysts as a weakness, was the object of fully-fledged central policy opinions only in 2012 (Weinongweifa 2012, No. 28), though changing policy priorities coexist with areas of work, such as strengthening management of funds and improving local fund collection and management systems, whose importance has been reiterated many times since the start of the scheme.

In addition to the dominant narrative given here, it should be remembered that any number of separate modes of development of policy are taking place under the NCMS umbrella, including centrally and provincially managed experimentation, managed development of sub-national policy models and locally-initiated (*zifa de, zisheng de*) behaviour, in addition to 'straightforward' implementation of policy. Specific jurisdictions may, for example, run quite distinct NCMS 'models': Tianjin has largely included NCMS users in an catchall rural-urban scheme (e.g. Tianjin Shi Laodong he Shehui Baozhang Ju 2009), for example, while Yinchuan has worked on development of a potentially interesting pro-poor 'version' of the NCMS, under which treatment for many common illnesses, many common prescriptions and public health items are made available for a fixed charge of 1 RMB (Sun 2009). Such examples are common and frequently given in Chinese media, policy briefs and academic work. In addition, the centre also engages in specific experiments: in 2008, for example, the MOH announced experimentation with development of specific facets of the NCMS and specified counties in which pilots were to be conducted (Nongweisi 2008).⁴⁷

The above analytical complexity is complicated by the existence in many places of 'un-standardised' (*bu guifan*) practices. As well as multiple legitimate sub-national variations in policy and processes of distributed sub-national policy development, many local practices do not conform to the way the system should, in theory, work. The

⁴⁷ Areas of pilot work in this instance were: increasing the pooling level to the city (*di ji shi*), integrating serious illness coverage with outpatient pooling and integrating the NCMS with the urban resident basic insurance scheme. Around 10-15 counties/cities were specified for each pilot, alongside experts in charge of each pilot.

NCMS is poorly 'institutionalised', and variation occurs not just at the level of formal policy (variant management models and their differential implementation over time and place, gradual standardisation of policy, etc.), but also at the level of practices of differing degrees of formality which in many cases exist alongside prescribed systems. A recent book-length study by Sun and colleagues discusses standardisation of the NCMS and the need for legislation to help regularise scheme functioning. Based on fieldwork in ten counties in one province ('S') in 2006/2007, Sun et al. describe a large range of local practices in many areas of NCMS operation, including who at the county level sets NCMS reimbursement plans; who sets policy and how; who takes part in plan setting; whether users take part in scheme oversight; what agencies have oversight of reimbursement plans, and similar. To take one example: in several counties, the head of the BOH personally set NCMS reimbursement plans; in one county, the head of the BOH also served as part-time head of the county hospital, derived his salary from the hospital, but was given responsibility for setting the plan as "no-one [else] understood this work"; in another county, the NCMS Office set the plan in conjunction with township cadres and township health centre (THC) representatives; in one case, the county directly copied the reimbursement plan of another county as there was no-one available to do the work; in another, a retired hospital director set the plan (Sun and Chai 2009, pp. 98-100). The authors find that this level of variation at the margins of policy exists in many areas of local NCMS work. These findings, for Sun et al., are a case of insufficient 'institutionalisation' of the NCMS; in many ways, they reflect at least some of the findings of the 2006 State Council report, which shows a range of variation in very many areas of scheme management (NCMS Pilot Evaluation Group 2006), not simply in the use of one or another 'management model' to some degree prescribed under the NCMS umbrella, as in the analyses by, for example, Mao (2005), above.

All this creates huge complexity. By 2009, the NCMS was being implemented in more than 2,700 counties, and it is likely that the scheme differs, at least to some extent, and possibly to a considerable extent, in each county in which it is implemented,

irrespective of the large differences visible in the overall thrust of policy development. Most of this variation is unknowable. In addition, the NCMS does not exist in isolation: reforms being 'piloted' in parallel with the scheme – such as attempts to introduce a national basic drug list, to vertically integrate rural health care providers, or to reform public hospital management, or to develop a system of zero mark-up on drug sales in rural providers – are all likely to interact with the scheme in different ways in different places. These items are cited as examples only; the maelstrom of parallel reforms, especially since the launch of the New Health Reform (*Xin Yigai*) in 2009, is huge. The analysis in this section is unavoidably schematic; taken together, however, the above indicators point to a system which varies across time, across space and along multiple axes in numerous, and frequently unknowable and un-analysable, ways.

5.7 The NCMS as an ongoing project

It might be assumed that the NCMS, ten years into its development, could be considered a largely 'finished' project, and that remaining development would revolve around complex areas of policy peripheral to core functioning of the scheme. To an extent, this is likely true: provider payment mechanisms under the umbrella of the NCMS were quite late in getting serious national attention, for example; equally, integration of rural and urban insurance schemes, management of reimbursement across scheme boundaries (developing mechanisms to allow rural-urban migrants to be reimbursed in their place of temporary residence, for example), developing mechanisms allowing the NCMS to deal with providers outside the county, and increasing the pooling level from the county to the city (or possibly, eventually, higher) are all major horizons in the development of the scheme, though in all these areas there has been at least some sub-national experimentation and policy development (e.g. Gao 2010; Zhang 2011b).

Paradoxically, perhaps, given the rhetoric of piloting, trialling and roll out, the NCMS, as it was extended country-wide, was a very imperfect system. It is understandable that certain elements of scheme functioning, such as provider payment reform or

rural-urban integration, should require concrete experimentation, and should be developed over time. It is less clear, though, that other elements of scheme functioning should require such trialling or take so long to pin down. One Chinese academic described the development of the NCMS as based on three main phases: enrolling users, improving scheme management, and integration with urban insurance schemes (Interview no. 16). While this may help provide an overall view of the NCMS as a system, it is an idealised view of a situation characterised by enormous complexity and opacity. As above, it is hard to know the extent to which long-standing problems in scheme management have been resolved.

Since the start of the scheme, for example, the importance of developing an efficient and cost-effective NCMS fund collection mechanism (obviating the need for costly, multiple, door-to-door visits) has been stressed (Gao and Han 2007; Gu 2008; Zhang 2011b), but fund collection remains hard and time-consuming in many places (and in at least two of my fieldwork counties) (Interview nos. 57, 64). Equally, the need for a legal basis linking local government scheme contributions to user incomes or local government revenues has long been stressed – many Chinese analysts believe local allocation of funds to be excessively dependent on the voluntarism of local governments, and therefore vulnerable to a shift in overall policy priorities (Gao and Han 2007; Zhang 2011b, p. 85), though this has recently been included in legislation in one province (Jiangsu Sheng Renda Changweihui Gonggao 2011).

Even very late central policy calls for improvement of many elements of local scheme management and oversight (problems which have been the subject of central policy as well as sub-national experiments and policy development for some considerable time), while certain components of scheme functioning, such as the stipulation that users take part in scheme management and oversight, are in most places likely little more than formulaic (I return to both these themes in Chapter Six). Similarly, local capacity deficits have been a concern throughout the development of the scheme – one recent analysis concludes that “it is difficult to understand how any county could be thought to have the technical and financial capabilities to run a complicated health insurance

programme" (Zhang, Yi et al. 2011, p. 15). I return to this issue at a number of points in subsequent chapters.

As above, even very recent analyses of the impact of the NCMS on users' spending find this to be very limited, and it remains unclear whether the NCMS will be able to succeed in translating increasing amounts of funding into better health outcomes. The scheme remains very much in development. As various analyses point out, though, the NCMS is only a small part of rural health system restructuring and its effectiveness and impact are related to other reforms taking place in parallel; its ability, as a discrete rural health financing mechanism, to respond to the "perfect storm that ravages health care in China" (Tang, Meng et al. 2008, p. 1493) is limited. Saying this, in 2003 very few counties in China ran any kind of health insurance scheme or had any experience of this. Ten years on, through a somewhat ramshackle development process in which experimentation and piloting have certainly played a role – though this is hard to analytically separate from a great mass of unprogrammed activity and poorly institutionalised systems – China now has around 2,700 counties running a rural health insurance scheme, though with varying degrees of competence. This, in itself, is a huge achievement.

5.8 Interim summary

The argument presented here is unavoidably long and complex. I have described the NCMS as a central policy principally implemented at the county level. Functioning of the NCMS corresponds broadly to descriptions of the pressurised system in existing literature, in that ever-increasing central targets are passed down through multiple layers of government to the county level where they must be met. Understanding the NCMS requires, though, that we supplement this analysis of target-driven implementation with an understanding of the way the financing system functions to transfer funds to the county level and aggregate these under control of NCMS management and the BOH. Chinese policy and academic discourse presents this as a question of 'fund risk', and subsequent empirical chapters show how this concept of

risk is also used by county NCMS managers in the way they think about implementation.

The NCMS has an uneasy existence as both local and national policy: implementers must follow national targets, but they design their own reimbursement plans. Rhetorically, this has been very much based on the idea that schemes should be tailored to local conditions, but limited funding has limited schemes' ability to meet user needs, and this has been expressed in a range of demand-side measures to control costs. In addition to designing reimbursement plans, counties are called on to contribute to scheme design by developing concrete local policy and mechanisms (*jizhi*) within, and for the furtherance of, central policy. This is where the paradox of the NCMS is most apparent: it is a central policy, in which the centre sets targets and policy principles (*yuanze*), but in which local governments must develop implementable (*caozuoxing de*) local policy; the scheme is coercive, in that implementation is prescribed, but frequently agnostic as to the *form* implementation should take. Many factors combine to mean that it is very hard to view the NCMS as a unified scheme; and this creates a great analytical challenge to any overall presentation of the scheme.

As above, I am principally interested in the role of counties in the development of the NCMS – in the way counties *innovate*, as well as *implement*. County development of specific mechanisms – functional local policy solutions – under the umbrella of NCMS policy creates enormous complexity and analytical difficulty: specific mechanisms have their own genealogies; understanding these requires substantial examination of policy trajectories, national (and sub-national) debates and detailed process tracing. This is rather different to studies which attempt to show development of 'the NCMS' as though this were a more or less unitary object amenable to analysis (cf. Wang 2009b). Within a study of this scope, this kind of analysis is impossible for more than a very small number of cases. Subsequent empirical chapters follow on directly from this assertion, showing the development of specific county-level policy mechanisms. These are the product of specific places and specific *moments*, as well as falling within the overall

framework of the national scheme as described here. Chapter Six describes the development of a county-level NCMS oversight mechanism and shows how this makes good, at least in part, a deficit in the NCMS, while Chapter Seven describes one county's experiment with development of a local payment reform system ahead of the curve of national policy development. Chapter Eight is an example of detailed process tracing, showing the local genesis, spread and codification in national policy of a distinct NCMS reimbursement mechanism.

Chapter Six

Risk and innovation: Taoshan's NCMS Oversight Bureau

6.1 Introduction

This chapter examines the case of an NCMS oversight agency set up by the Bureau of Health in a county in an eastern province. A significant literature in Chinese has criticised the effectiveness of local oversight of the NCMS, resulting from local staffing and capacity deficits, structural ambiguities and problems inherent in China's political structure. In addition, genuine oversight by scheme users, though a requirement of the NCMS structure, is likely a formality in most cases. The promise, therefore, of a county experimenting on its own initiative with increased and, apparently, 'independent' (see below) oversight is attractive.

In 2008 Taoshan County,⁴⁸ a not very prosperous eastern county, set up what the county describes as an Independent Oversight Bureau (IOB) to oversee operation of the NCMS. This was a local innovation in scheme management, to the degree that this was a locally driven reform and not dependent on policy at superior levels. This was reported on favourably by the MOH, which described the county as 'exploring' (*tansuo*). Creation of the Bureau was a local initiative, but is understood to fit within, and be a legitimate development of, the NCMS overall. I examine functioning of the Bureau and argue that its creation was a response to perceived risk in scheme implementation on the part of the BOH. I examine Taoshan's reform within the context of changing NCMS policy, and ask to what extent this mechanism is having, or could have, supra-local applicability and/or impact.

I carried out fieldwork in Taoshan in 2010. The chapter is structured as follows. Section

⁴⁸ Names of places and people have been changed to preserve anonymity. This applies also to published materials where necessary. In some cases, figures and dates have been marginally altered to prevent identification of sites or people. In all cases, anonymising people and places and any other marginal changes have no effect on meaning. See Chapter Three.

two discusses county-level NCMS management and oversight and argues that oversight should generally be considered weak. Sections three and four show development of Taoshan's initiative and examine its functioning. Section five situates Taoshan's reform in the context of calls for change in NCMS oversight and developing national policy. Section six concludes.

6.2 Background

6.2.1 NCMS management and oversight

At the county level, and in the majority of cases, the NCMS is overseen by an NCMS Management Committee (*Guanli Weiyuan Hui*), with day-to-day management carried out by an NCMS Management Office (*He Guan Ban*, hereafter NCMS Office). This structure may be replicated at the town/township level, though this is not universal (Gao and Han 2007). The NCMS Committee falls under the dual leadership of the county government and BOH.

Broadly speaking, the NCMS Office is charged with day-to-day oversight and management of the scheme. A 2006 interim survey of functioning of the NCMS broadly divides functions of the NCMS Office into: fund collection (*chouzi*), reporting and analysis of fund use; publication of NCMS fund use; auditing local medical fees; oversight of services provided by designated providers; checking and approval of providers for inclusion in the NCMS (NCMS Pilot Evaluation Group 2006, p. 42). In addition, an important component of the role of the NCMS Office is the oversight of reimbursements and behaviour of providers: correct reimbursement procedures must be employed, the NCMS Office should normally approve all claims (or claims exceeding a certain threshold) against the scheme to ensure healthcare providers reimburse only sanctioned procedures, drugs, etc., before these are passed to the Bureau of Finance for payment. To a very large extent, *day-to-day* work revolves around overseeing fund use, overseeing and processing claims (approving claims/payments), and overseeing providers (Gao and Han 2007). The NCMS Office then reports back to the NCMS Committee, part of local government. To this extent, the

NCMS Office should perform a gatekeeping and oversight role within the county-level management of the scheme.

6.2.2 Funding constraints and limits to work of the NCMS Office

Two main problems exist in relation to the functioning of the NCMS Office in oversight of providers/services and the claims/reimbursement process: lack of funding, personnel and capacity; and lack of independence from the local health system. The capacity of NCMS Offices to carry out effective oversight and management in most places is limited.

As described in Chapter Five, NCMS funds are earmarked funds (*zhuan kuan zhuan yong*) and may only be used for NCMS reimbursement and health check ups (*tijian*). This stipulation, that NCMS funds are earmarked for the reimbursement of NCMS expenses, and not to be used for NCMS management and operation, derives from a stipulation in the 2002 Central Committee and State Council Decision on strengthening rural health work (Zhongfa 2002, No. 13) that users' NCMS contributions not be regarded as a 'burden' ("*bu neng shi wei zeng jia nongmin fudan*"). Following this, it was specified clearly in the 2003 State Council Opinions that NCMS management expenses not be taken from scheme funds, which are to be used exclusively for reimbursement: personnel and operating expenses must be found from county budgets (Guobanfa 2003, No. 3). This requirement has been reiterated many times, in speeches and policy, including in the 2008 NCMS financial management regulations, which clarify the issue even further, specifying that even bank charges (*yinhang shouxu fei*) related to running the scheme are not to be paid out of scheme funds (Caishe 2008, No. 8). As in Chapter Five, the NCMS is, in theory at least, voluntary. This stipulation is a relic of debates in the 1990s and 2000s on the reduction of the peasant burden: a new rural health insurance scheme could not be seen to increase the peasant burden at a time when government was working to eliminate this (Wang 2008). It seems likely that the initial framing of NCMS funding in terms of peasant burden underlay the decision to prohibit use of scheme funds for management and

operating expenses.

Early assessments of available management capacity and funding pointed out that policy specified in insufficient detail how management of the NCMS was to be paid for (e.g. WHO 2004). An article on the NCMS policy process written by Yang Tuan, from the Chinese Academy of Social Sciences, describes the decision to start NCMS pilots as hasty, with little attention to practicalities of management or funding. As Yang phrases it, the requirement that funds for management be found locally

“equates to requiring all county governments to set up new administrative organs for the management of rural health, [and that] county governments must bear the assorted costs of establishing organisations, hiring personnel and operating expenses, with the result that within a year, 310 new service organisations (*shiye jigou*) had been established around the country, and to date their personnel allocation (*renyuan bianzhi*) is not fully in place and their operating expenses are very far from sufficient” (Yang 2004b).⁴⁹

In the case of the NCMS, lack of provision for local management costs is, perversely, exacerbated by the early stipulation that the scheme be voluntary – the cost involved in running a scheme on these lines is large, and this cost falls to local levels (see Section 5.4.2).

An evaluation of local costs incurred in setting up and running the NCMS, carried out in six counties in Yunnan and Heilongjiang in 2004 (Gao, Zhao et al. 2006), found that counties spent on average around 200,000 RMB in setup costs, and that when initial management costs and ongoing office and personnel expenses were included, this figure rose to around 650,000 to 850,000 RMB (2004), equivalent to around 14.6% of NCMS funding at that point. This spending must come from local budgets. To put these costs in perspective, Gao et al. cite one county they evaluated whose *total* annual NCMS costs (maintenance and management, plus matching funds, based on

⁴⁹ More frequently called ‘*shiye danwei*’, or ‘public service units’. This is addressed in detail in Section 4.3.2.

estimated 90% sign up) were around 1.3 million RMB. This is a poor county: total incomes, from local sources and transfers (the majority of the budget), came to almost 135 million RMB. Existing essential county spending commitments for wages and continuing operational (*yunzhuān*) costs came to more than 155 million RMB. Gao et al. describe this county as having 'bread and butter finances' (*chīfān cǎizhēng*) only, and point out how common this kind of situation is in poor counties, especially in the west. It is easy to see how addition of NCMS funding is an additional burden for such counties.

Under such circumstances, the NCMS may be a significant burden for already tight county finances. One Anhui study found extremely low levels of funding for scheme management, with around a quarter of pilot counties operating on funds for management of just 0.09 RMB per signed up user, and a third operating on funds of 0.17-0.25 RMB. None exceeded 0.64 RMB. Average national funding was reported as 1.32 RMB per signed up user in the same study. This is around one fifth of per capita management spending on the urban employees' insurance scheme (*Chengzhen Zhigong Yiliao Baoxian*) (Wang et al., cited in Gao 2008a). The 2006 State Council interim report on the state of the NCMS came to similar conclusions, noting also that investment in computer systems, and the fact that NCMS work requires a car to be effective, all added to the difficulty of local financing of the scheme, and found funding to be clearly inadequate, with an impact on implementation and on oversight and management work (NCMS Pilot Evaluation Group 2006, pp. 8, 41). One recent study also found very variable, and in some cases very limited, funding arrangements in the early stages of scheme development in six counties in Ningxia and Shandong (Yan, Raven et al. 2010).

One result of this pressure on funding is a degree of 'cross subsidy' within counties: Gao et al. give an example of a county in Heilongjiang in which in 2004 county finances contributed only 17% of setup costs. Another 20% of costs were contributed by the local Agricultural Bank (where the NCMS account was held), and various healthcare providers (THCs and county-level institutions). The remaining 63% had to be found out of the budget of the BOH, already taken up with existing commitments for wages,

personnel costs, and the like (Gao, Zhao et al. 2006). This 'collateral' financing burden may extend in many cases to township and village cadres, who are required to enroll users in the scheme and to collect contributions and are unlikely to be compensated for this (e.g. Yan, Raven et al. 2010), and may well extend to village clinics (Babiarz, Miller et al. 2010; Zhang, Yi et al. 2011). Financial constraints lead, in most places, to Bureaus of Health and NCMS Offices 'requisitioning' staff from THC's to help implement the NCMS, creating conflicts of interest. Overall, the picture described here fits with the exporting of fiscal burdens from the county to subordinate levels of government described elsewhere (Liu, Wang et al. 2009).

The personnel (*bianzhi*) allocation of most NCMS Offices is set at 5-6 people, though this is frequently supplemented by staff effectively requisitioned from providers under the purview of the BOH. On average, county-level NCMS managers are each responsible for oversight of a population of around 50,000 people (more than twenty times the ratio for the urban employees' insurance scheme) and in addition, due to problems of personnel allocation and lack of funds, positions may not be filled: capacity may be available locally, but there may be administrative blocks to its employment in NCMS management (Gao 2008a, pp. 125-126).

Discussions of this arrangement with the head of one city NCMS Office, Director Chen, underlined the difficult position in which NCMS Offices find themselves: Chen was very clear about the challenges involved in doing his job: supervision and management (*jianguan*) of the scheme (including reimbursements and oversight of providers) is hard, and five staff are not enough to oversee all medical treatments. This is also exacerbated by generally low skill levels at township/THC level, meaning that the NCMS team frequently has to directly examine large quantities of medical files (*bingli*), and in many cases question the judgement of doctors. For this NCMS Office, as for many, oversight and management (*jianguan*) are the bulk of day-to-day work. During one visit to a township health centre (THC), Director Chen took away a stack of forty medical files to go over that evening, in addition to other work – while I had been conducting interviews, Chen had been looking over records in the THC office and had

found a number of files he suspected of containing problems. Problems of oversight become even more acute when supervising providers at higher levels: county-level facilities inevitably employ specialists, and it can be hard to question their judgement, according to Chen (Interview no. 58).

In addition to low staffing (*bianzhi*) levels and low levels of training and expertise at the township level, the potential for conflicts of interest exists at, and below, the county level. The 2006 interim report bemoans the lack of independence of the NCMS Office: in most cases, this is situated within the BOH and draws leadership and staff from within the health system (NCMS Pilot Evaluation Group 2006, p. 43).⁵⁰ The position of the NCMS Office within the local health system is deemed to cause difficulties for effective oversight of providers: in most cases, the BOH appoints the head of the NCMS Office, as well as having oversight of personnel, finances and material supplies (*wu*) of providers (NCMS Pilot Evaluation Group 2006, pp. 51-52).

In addition, due to personnel shortages, NCMS Offices mostly rely on assistance from employees of hospitals and THCs under their purview to help oversee NCMS operation (reimbursement, standards, etc.) in those same providers. In practice, this mostly means that each designated provider must allocate staff to NCMS duties, including overseeing reimbursements, reporting, etc. One review of scheme functioning found that half of township-level personnel were sent down by the county, with half locally seconded (NCMS Pilot Evaluation Group 2006, p. 40). Staff seconded from providers, however, remain on the payroll of the provider in question, and continue to be managed by their 'parent' provider, despite performing duties for the NCMS Office, creating a clear possibility of conflicts of interest.

6.2.3 The NCMS Oversight Committee

The 2003 regulations on the establishment of the NCMS allow localities to set up NCMS Oversight Committees (*Jiandu Weiyuanhui*) 'according to local conditions'

⁵⁰ Transferring NCMS management to departments other than the BOH is, at least in part, an attempt to resolve this.

(*genju bendi shiji*) (Guobanfa 2003, No. 3). By the time of the 2006 interim report, however, the Oversight Committee seems to be a universal structure at the county level. This same report outlines functions of the Oversight Committee as observed in pilot counties studied: checking whether funds are in place and are being spent correctly; checking whether funds are being managed correctly, according to a system of closed management, whether designated accounts are being used and whether balance is being maintained in fund use (between revenues and payments); checking on implementation of NCMS plans (*shishi fang'an*) and other work. In addition, the Committee has the responsibility to make sure funds are in place to pay contributions for poor groups,⁵¹ publicise information on the state of NCMS fund use and investigate incorrect and illegal behaviours (NCMS Pilot Evaluation Group 2006, p. 43).

In theory, then, there exists provision for supervision at the county level, independent of the NCMS Office, whose purpose is to oversee functioning of the scheme and report on this. A number of Chinese language analyses of county-level oversight arrangements, especially Oversight Committees, criticise the effectiveness of this, however. Two main criticisms stand out. The first is the lack of technical expertise available: members of Oversight Committees are unlikely to have medical training, making oversight of limited value. The second main criticism is derives from local institutional structuring: committees tend to be staffed by People's Congress and People's Consultative Conference members,⁵² and to have close links to local governments. In addition, it is thought that committee members are likely at the end of their careers and unlikely to be motivated to carry out effective oversight. Another State Council-supported interim NCMS report gives excerpts from interviews with members

⁵¹ A Medical Financial Assistance programme was started in conjunction with the NCMS in order to enable poor groups to join the scheme. I do not discuss this significantly. This is discussed in the 2006 report to the State Council (NCMS Pilot Evaluation Group 2006).

⁵² This is the dominant picture given in available literature; however, the 2006 interim evaluation of the scheme – seemingly the only source giving a statistical picture of Committee composition – listed various government departments (including agriculture and audit) as well as NCMS users as making up the bulk of membership of Oversight Committees (NCMS Pilot Evaluation Group 2006).

of Oversight Committees. Interviewees believe oversight to be formulaic; they point out that these committees, as above, are staffed by people at the ends of their careers who haven't managed to get promoted, and who have no further promotion prospects. If an Oversight Committee were to uncover a problem, says one interviewee, it would be unlikely to bring this to the attention of government leaders, as this would be unwelcome. A third interviewee maintains that the main purpose of Committees is to maintain the image of the local leadership, and that the composition – NCMS users, People's Congress members, and People's Consultative Conference members – shows how toothless and symbolic they really are (MOH Center for Statistics Information 2007, p. 128). In all, both interim reports on development of the NCMS believe that oversight is toothless, and can only be toothless, under the conditions described here. Interviews with BOH staff tend to confirm the view that local People's Congresses lack the ability to carry out effective oversight (Interview no. 20).⁵³

6.2.4 Other local oversight: Users and local Audit Bureaus

The 2003 regulations specify that local audit departments should carry out periodic audits of scheme funding (Guobanfa 2003, No. 3). In addition, the same regulations stipulate that NCMS Committees and NCMS Oversight Committees should include user representatives (*nongmin daibiao*). The 2006 interim evaluation report on the NCMS reports surprisingly high numbers of both committees to have user representatives: almost 65% of counties studied in the case of NCMS Committees, and 77% in the case of Oversight Committees (NCMS Pilot Evaluation Group 2006, pp. 42-43).

Saying this, most analyses point to a lack of real importance of this participation. According to the 2006 interim NCMS analysis, while users (*nongmin*) are generally concerned that the scheme should be run properly, they are reluctant to take part in managing it (NCMS Pilot Evaluation Group 2006, p. 52). This point is also made by

⁵³ In addition, other systemically less significant factors, such as the fact that Oversight Committees inevitably carry out after-the-fact oversight, rather than real-time oversight, are also likely to impair their effectiveness (Chen 2010).

Gao, who cites a paper given at a MOH NCMS Technical Guidance Group work meeting to this effect (Gao 2008a, p. 131). Other analyses argue that users have a low level of understanding of the scheme and that, even were institutional conditions for participation more favourable, this would be a block to effective participation in management (Chen 2010). A number of authors argue that effective user participation in oversight is important for healthy development of the scheme, and that this has received insufficient attention to date (Chang 2010; Zhu, Ye et al. 2010), though, seemingly at least some experiments aimed at increasing user participation have been taking place (Feng, Wang et al. 2009).

6.2.5 Ineffective oversight: Resulting problems

Various official and near-official sources give some understanding of problems thought to arise from insufficient oversight. The 2006 report specifically highlights problems relating to fundraising, publicising of the scheme, open publication of data on scheme use, etc. According to this report, local NCMS managers believe that financial management is unlikely to be a large problem (as in Chapter Five, due to existence of solid finance and accounting systems). Most managers, it is said, believe that problems are most likely to occur in the payment/reimbursement part of the system. Problems cited are lending/borrowing of NCMS cards to illegitimately claim against the scheme, illegitimately acquiring prescriptions for others, over-provision of treatment, doctors (and/or doctors in collaboration with patients) attempting to cheat the system, and similar. Lack of effective oversight of providers is said to be likely to lead to inflation of diagnoses, costs, and similar (NCMS Pilot Evaluation Group 2006, p. 10).

Evaluation by Gao broadly confirms this diagnosis: she believes localities pay more attention to collecting scheme contributions than to scheme management. Most problems, says Gao, are likely to occur in fund use (rather than collection). This author lists problems of over-spending (due to sudden increases in health seeking following scheme implementation, poor fund setting, etc.) as well as improper fund use (incorrectly allowing reimbursement of procedures, lending of NCMS cards, cheating of

the fund by scheme users), and believes that these contribute to fund risk (Gao 2008a).

Any number of other management problems are likely to exist, to different degrees, in different places. The number of possible 'holes' in the NCMS is large. Different reports present different worries, though there is a degree of consistency: lack of implementation of closed fund management and/or stipulated financial management systems, in contravention of rules; use of incorrect reimbursement practices, etc. Studies have even found certain counties to be using NCMS monies to pay for users to sign up to commercial insurance (NCMS Pilot Evaluation Group 2006, p. 10; Wu 2009). Summary results of a 2009 National Audit Office audit of the NCMS (forty-five counties in nine provinces) state that no large problems in breach of regulations were discovered, but that small problems were discovered in various counties, including misuse of funds (several counties were found to have routed NCMS funds into other social security programmes), incorrect reporting of numbers of scheme members with the intention of claiming higher levels of matching funds from the centre, insufficiently rigorous oversight resulting in expanding the scope of reimbursement (*buchang fanwei*), etc. (Zhonghua Renmin Gongheguo Shenjishu Bangongting 2011, No. 7).

My aim here is not to exhaustively present problems arising from lack of detailed scheme oversight: this is impossible. Rather, it is to give an idea of the *range* of problems presented in Chinese analyses. Below, I examine illegitimate behaviours listed as targets of supervision by Taoshan NCMS Office and by the Oversight Bureau.

In the above sections, I have argued that county-level NCMS monitoring and oversight is, in most places, to be considered weak. Capacity, personnel, and funding constraints combine with insufficient separation of management and supervision organs to undermine the possibility of effective scheme oversight. This is compounded by little *de facto* user participation in scheme governance in most places. A number of problems are said to result from this oversight deficit, discussed briefly above. In the next section I examine Taoshan's initiative: what the county did, the thinking behind this, and the possible usefulness of this initiative.

6.3 Taoshan: Establishment of an 'Independent Oversight Bureau'

6.3.1 What Taoshan did

In 2008, Taoshan piloted what the county describes as an independent NCMS Oversight Bureau, said by a local newspaper to be the first of its kind in the country and described in provincial materials as a local innovation. Taoshan's Oversight Bureau was envisaged and promoted by Director Wang of the BOH. From 2008 to 2009, the Oversight Bureau was developed locally, and initial trials were conducted in two townships, before being rolled out county-wide; detailed operating provisions were released subsequently. The townships chosen as pilots were said to represent different extremes of local development. The BOH approached county government for permission (and funding) to carry out trials, and this was given.

According to Wang, carrying out local trials before county-wide roll out enabled the BOH to minimise risk (if the initiative were to be unsuccessful, the impact would be limited) and to iron out problems during the development process: for Wang, this was about problem solving, about discovering problems (*"faxian wenti"*), and finding policy solutions (*"xiang banfa sheji fang'an"*). This was a practical solution: something you "can see clearly, something you can touch" (*"neng kan de jian, neng mo de dao"*). In addition, the county allocated staff and funding to the pilot, something Wang would have been unable to do without approval. Carrying out a pilot was also a way for the BOH to demonstrate to the local Party Secretary the potential of this initiative, and to show that it could reduce incorrect behaviours in the local health system, improve scheme operation and increase effectiveness of NCMS implementation agencies and users' satisfaction levels (Interview no. 20).

At the time of writing, Taoshan's initiative has received a degree of supra-local attention and has been studied by a number of other counties but has not achieved significant national fame.

6.3.2 The need for increased oversight

Taoshan's Oversight Bureau aims to separate scheme oversight and management: oversight is placed in the Oversight Bureau, not with the NCMS Management Office, and the Bureau is independent, financially and in management terms, from the NCMS Office and from providers. In articles written by BOH staff, the need for increased management and oversight of funds and fund use is presented as a natural consequence of increasing NCMS funds (Wang 2009a). Per capita NCMS funding increased from 30 RMB in 2003 to 150 RMB in 2010, and 98% of Taoshan's eligible population is signed up to the NCMS (Taoshan County Bureau of Health 2010). According to Wang's article, which is representative in its main points of other explanatory-analytical materials published by the BOH, in the period before the establishment of the Bureau, problems were encountered in oversight of NCMS users, providers and even some NCMS implementation agencies (*jingban jigou*). Problems listed include excessive drug prescription, inappropriate use of drugs (*chao fanwei yong yao*), inappropriate reimbursements (*chao fanwei baoxiao*), high costs and so forth. In the case of NCMS implementation agencies, problems cited include laxness and delays in getting work done. None of these problems should be seen as special nor unique to Taoshan; rather, they are common to hundreds of NCMS policy documents, and presumably to hundreds, if not thousands, of counties around China.

A report from Taoshan BOH (Taoshan County Bureau of Health 2010) gives main reasons for these problems: difficulties faced by the NCMS Committee in overseeing the NCMS; too few staff in the NCMS, and a faulty oversight model relying on staff drawn from THC's, creating a conflict of interest. Wang et al. state the problem this way:

"NCMS staff in designated providers come from within those work units [providers], and it is unavoidable that in overseeing treatment and reimbursement [carried out] by their work units, self-monitoring (*ziji jianguan ziji*) should be an issue. What's more, the majority of problems discovered are solved in house, and it's inevitable that providers will think of revenue

generation, their own survival, development, and social impact (*shehui yingxiang*), and it's hard to avoid this having a negative impact on policy implementation; in some cases, providers' internal oversight exists in name only" (Wang 2009a, p. 23).

This conflict of interest is compounded by lack of staff: Taoshan's NCMS Office oversees almost 400 providers and around 900,000 users (Chen 2009). Standard NCMS staffing levels are, quite obviously, inadequate. These explanations mirror those given by Oversight Bureau and BOH staff: problems in the NCMS, common, in fact, to many places, are a result of insufficient oversight, stemming from deficient 'internal management' (*neibu guanli*) on the part of providers, which requires supplementing with strengthened 'external management' (*waibu guanli*). This is the point of departure for establishment of the Oversight Bureau.

6.3.3 A guardian angel for the NCMS: Remit of the Oversight Bureau

Local materials describe the Oversight Bureau, variously, as a 'guardian angel' and/or 'firewall' for the NCMS. Structurally, the position of the Bureau in the local administrative hierarchy is that of a 'Sub-Bureau', with a personnel (*bianzhi*) allocation of 25 people (compared to a personnel allocation of six people for the NCMS Office). The Bureau reports directly to the NCMS Committee (not Office). Its offices are located in the BOH. Staff are recruited from within the local health system and are managed by the head of the Bureau, Director Wen. Staff are expected to have expertise in areas key to the work of the Bureau.

Functions of the Oversight Bureau are set out in two main documents: the Oversight Bureau Regulations and a document specifying operating rules (*chuli banfa*); supporting documents fill in this overall framework. The overall scope of the Bureau's work is given in the Regulations (Taoshan County Bureau of Health 2009b): to supervise and encourage township-level NCMS fundraising; to check on NCMS reimbursements, making sure that they are within the scope specified by the NCMS; to check that township-level NCMS Offices are effectively monitoring village-level NCMS

reimbursements; to check fund collection and use at village, township and county levels; to collect public complaints; to check injury claims;⁵⁴ to assist providers in carrying out internal auditing; to publicise the NCMS; to report on oversight work; to carry out door-to-door surveying.

Roles and responsibilities are specified for the Bureau and five Sub-Bureaus, each of which oversees a 'district'⁵⁵ within the county, and is staffed by a manager and three to four other staff. Bread and butter work is carried out by Sub-Bureaus, whose work is supervised by the county-level Bureau, though Sub-Bureaus retain a certain amount of decision-making authority and are headed by seemingly highly competent individuals.

Responsibilities of the heads of the Sub-Bureaus, and of investigators, are specified in some detail, while local implementation documents attempt to enforce standards of correct behaviour, specifying that staff should not carry out inspections in the place they originate from (an avoidance rule), should not accept gifts, should always operate in pairs, and similar. The Bureau's operating rules go so far as to specify rights of targets to request proof of the identity of inspectors and to request written proof of any decisions made relating to their case. (Almost obsessively clear guidance is given on the behaviour of inspectors and the various bureaus, covering protocol when dealing with the public, case recording, confidentiality, management of case files and similar.) Assessment (*kaohé*) procedures for staff are given, and results are to be linked to bonuses and other benefits (Taoshan County Bureau of Health 2009a).

6.3.4 Functioning of the Bureau: Oversight

Various types of oversight are carried out by the Bureau and Sub-Bureaus. The

⁵⁴ The NCMS frequently does not pay where there is third party liability, notably in injury cases. Determining whether a third party is liable frequently requires significant checking on the part of NCMS implementation organs. In one county where I carried out fieldwork, Meijiang, *all* injury claims are checked door-to-door by a dedicated team including members of the NCMS Office and doctors.

⁵⁵ Management is based, here, on a spatial unit below the county level, but not corresponding to the existing town/township administrative structure. According to Wang, this is used routinely in other administrative areas, but not normally in the health system.

majority is routine inspection, in which case targets are notified in advance of a Bureau visit; however, provisions exist for *ad hoc*, unannounced and targeted inspections where illegitimate behaviour is suspected (following a public complaint, for example), or where it is thought that prior notification might prejudice inspection.

Inspections are to be carried out according to specific guidelines, as set out in the Bureau's regulations, with specific, standardised, documentation requirements for each of stage of the inspection process: from documents notifying targets of forthcoming inspection and forms for requisitioning providers' records, to forms detailing possible decisions by the inspection team. These include positive judgements and consequent closing of a case file, requests that a provider rectify practices within a specified period and forms mandating change and notifying providers of impending re-inspection to check compliance. All forms are to be signed in duplicate, by a representative of the Bureau or Sub-Bureau and a representative of the provider. Sub-Bureaus are to compile 'inspection reports' on each case according to detailed specifications, and to pass these up to the Bureau.

Infractions are to be dealt with initially through a request for change from the Bureau or Sub-Bureau, though providers can be fined, if necessary, on the authority of the BOH. In extreme circumstances, providers' right to receive reimbursement from the NCMS can be suspended and/or removed, disbarring them from the NCMS system. Where the law has been broken, it is stipulated that providers be referred to the legal system. Infractions by providers are listed in order of severity, and stretch from the comparatively less serious (refusing patients, incorrect referral, treatment or drug use, non-observance of NCMS pricing policy) to definitely serious (creating fake treatment records, fake drug prescriptions, excessive drug prescription (*da chufang*), conspiring with patients to cheat the NCMS and similar).

In the case of inspection of NCMS implementation agencies (*jingban jigou*), infractions are grouped into minor and serious ones. Minor ones require a warning and criticism. These include: not responding to queries about the NCMS, prejudicing public opinion; delaying NCMS payments; unilaterally changing documents and data; incorrect

disclosure of information; keeping unclear accounts, etc. More serious infractions, liable to a fine (up to 50,000 RMB), where these result in financial loss to the NCMS, or referral to legal departments, include: insufficiently rigorous checking of reimbursement documents; ignoring or disguising problems discovered during oversight of providers; not properly investigating injury cases; incorrect management of village-level reimbursement; deliberately cheating the NCMS.

Possible infractions by NCMS users specified in local documents cover such things as use of another person's NCMS card to seek treatment, deliberate misrepresentation of illness, submission of false receipts for reimbursement and selling on of drugs (illegally making a profit from drugs: *feifa mouli*) prescribed through the NCMS.

While the Bureau has oversight of three separate components of the health system, oversight of providers is clearly considered the most important and common. Bureau and BOH staff stress that most cases are resolved amicably, with providers agreeing to comply with Bureau requirements. The rationale for this is that while, in the analysis of Wang, deficiencies in providers' 'internal' management create a need for increased 'external' management, the aim of external management should be to encourage providers to improve their own, internal, management practices.

6.3.5 Functioning of the Bureau: Cases

Bureau records of two inspections carried out by 'D' Sub-Bureau in early 2010 give an idea of actual management of inspection work. The first of these relates to an inspection of N Village clinic (S Township) in April, following which the Sub-Bureau issued an 'Announcement on reform' (Taoshan County D Sub-Bureau 2010b), reporting problems found and changes required. Problems found were:

- incomplete reimbursement procedures: documentation was examined relating to 160 reimbursements; in two instances, documents lacked a signature from the payee; in 74 cases, a fingerprint had been given in place of a signature, but had not been witnessed by a third party;

- excessive drug prescriptions in 101 out of 160 cases;
- discrepancy between prescription information and reimbursement documents, including number of drugs prescribed, total cost of treatment, etc.;
- reimbursement of patient fees outside the mandated time limit.

In this case, the verdict of the Sub-Bureau is educational, rather than punitive: procedures are clarified (e.g. use of fingerprints in place of signatures must be witnessed); the decision reminds the clinic that prescription and reimbursement information must match, and instructs the clinic to make sure they fully understand NCMS reimbursement procedures. A parallel document relating to the same case sent to S Township Health Centre (Taoshan County D Sub-Bureau 2010c), the THC responsible for N Village, reports details of the inspection and requests that the THC increase oversight and training of this and other village clinics under its jurisdiction. S THC is requested to report in writing on measures taken, and this is to be checked at a later date by the Sub-Bureau.

The second case, also April 2010, is more serious: D Sub-Bureau carried out a check on X Village clinic (N Township), focusing on NCMS reimbursements in March. The initial assessment was presented in a 'decision' in May (Taoshan County D Sub-Bureau 2010a). The decision found:

- incomplete reimbursement procedures: of 323 reimbursements during the period, nine were not signed by the payee;
- excessive drug prescription (*da chufang*): 78% of prescriptions exceeded the recommended level of 30 RMB;
- falsifying prescription receipts: the clinic was found to have modified prescription information, substituting (one presumes, though this is not specified exactly) more expensive drugs for those originally prescribed in order to increase reimbursement from the NCMS to the clinic;

- during door-to-door investigation, problems were found in fourteen cases, including discrepancies between drugs prescribed to patients and those recorded by the clinic, discrepancies between drugs prescribed and the patient's condition, etc.;
- retention by the clinic of users' NCMS cards: door-to-door investigation revealed that the clinic had misused users' NCMS membership cards to reimburse the clinic's outstanding accounts/bills (arising, it is implied, from other patients); the clinic had also frequently retained users' NCMS reimbursement cards, it is presumed for the purpose of illegitimate reimbursement of expenses;
- the inspection also found that previous Bureau recommendations had not been implemented, or had been imperfectly implemented.

The decision stops short of recommending that the provider be excluded from the NCMS, but specifies a fine of around 1,400 RMB, to be deducted from the monies reimbursed by the NCMS to the clinic, and specifies that the manager be criticised and reported to higher level management to be dealt with, and suggests that he be replaced. The decision also recommends that the THC under which this clinic falls use this case locally for propaganda and education purposes. The document is signed by a representative of the clinic and carries chops from both the clinic and D Sub-Bureau. A copy of the decision is to be retained by the Sub-Bureau; other copies are to be sent to the NCMS Office, the clinic in question and to the county-level Oversight Bureau. During fieldwork and visits to Sub-Bureaus, these cases were presented as typical of inspection work carried out by the Bureau.

6.3.6 Functioning of the Bureau: Problem solving

A secondary function of the Oversight Bureau, alongside oversight, is to provide a supplementary channel for feedback on local operation and implementation of policy. Conceptually, this can be seen in two elements of the functioning of the Bureau: the establishment of a 'letters and visits' (*xinfang*) function within the Bureau and Sub-Bureaus, and use of the reporting function of bureaus to provide input to policy modification at the county level.

Local documents describe the letters and visits function of the Bureau as a way to improve relations between the Oversight Bureau and NCMS users, allowing a channel for users to report on incorrect behaviour of doctors and give suggestions to Bureau personnel. It is hard to know the extent to which this functions, though Bureau staff cited several cases in which they had received anonymous tip offs allowing them to prevent misuse of NCMS funds, though such incidents are apparently not common. The Bureau's contact information is listed on public documentation and posters and was included in a 'letter' to NCMS users given out at the time the Bureau was set up (Taoshan County D Sub-Bureau 2009). In addition, records of NCMS reimbursements, including names, conditions, etc., are on public display on information boards, though it is unclear what use is made of this information. According to interviewees, feedback from the Bureau and Sub-Bureaus can be used to help refine policy. As Director Wang states it, feedback allows the BOH to discover problems, and to determine whether these are due to poor implementation of policy (*zhengce bu dao wei*) or deficiencies (*loudong*) in policy.

6.3.7 Management

According to Director Wen, the main cost of operating the Bureau is transportation costs, as staff spend a lot of time in the field. A member of staff of one Sub-Bureau estimated that she spent around half of her time at the grassroots, doing oversight and door-to-door work. Transportation costs, according to Wen, are around 10,000 RMB per year. Excluding office space, provided by the BOH, the main remaining cost is personnel. As above, the Bureau has 25 staff; apart from core people, staff (Sub-Bureau heads and inspectors, etc.) are requisitioned from THCs. They remain formally employed by their original work unit (Taoshan County Bureau of Health 2008), and their salary continues to be paid by their work unit, but these funds are routed through the Bureau, which pays salaries and controls management of Sub-Bureau staff. In this way, it is claimed, the Bureau manages to achieve independence from providers and NCMS management while keeping a lid on (or outsourcing!) operating costs.

6.4. Significance of Taoshan's 'Independent Oversight Bureau'

6.4.1 Effectiveness

How should we judge Taoshan's Oversight Bureau? In the absence of a thorough audit of Taoshan's NCMS scheme, we are forced to rely on county data to evaluate overall effectiveness of the Bureau. A May 2010 report from the BOH, as one would expect, presents evidence showing that the Bureau is having an effect. The report claims, for example, that in a year and a half of operation, the Bureau had carried out 1,800 inspections, and discovered 480 problems of various types (Taoshan County Bureau of Health 2010). An earlier report (Wang 2009a) states that in its first two months, two regular inspections were carried out, compared with 150 'irregular' (*bu ding qi*) inspections. All the above is said to have helped correct behavioural problems (*wei gui xingwei*) of providers and helped perfect NCMS regulation through feedback. The report also credits the Bureau's work with helping reduce costs, increasing the proportion of reimbursable drugs prescribed, and reducing spending on drugs overall (average spending is said to have decreased from around 57 RMB per person per visit to just over 40 RMB per person per visit at the township level, and from around 25 RMB to around 20 RMB at the village level). The Bureau is also credited with improving performance of NCMS implementing agencies, and the percentage of reimbursements conforming to guidelines is said to have increased from 95% to 99% over the period. Other benefits credited to the Bureau are improving the distribution of NCMS spending (reducing spending at the county level, increasing spending at the township level, etc.) and increasing the *de facto* reimbursement rate. Other materials produced by the county tell basically the same story (Bo 2009; Wang 2009a; Zheng 2009), and local media are equally enthusiastic. The data presented here are open to question. What appears to be clear, however, is that a concerted effort to improve scheme management is underway and that this is having some effect, at least, on overall costs.

6.4.2 How independent is Taoshan's Independent Oversight Bureau?

Official documents describe Taoshan's Oversight Bureau as 'independent'. As above, use of this adjective to describe the Bureau refers specifically to its independence from providers and the NCMS Office; the Bureau and its staff are not, and are not claimed to be, independent of the BOH or other government (local or otherwise). Given this, how should we understand what Taoshan has done? First, while the Bureau and its Sub-Bureaus are demonstrably engaged in oversight of providers, to what degree, given the close relations within the Chinese health system between health agencies and providers, can we expect this supervision to be effective? Second, what about the Bureau's other important function: oversight of NCMS implementation agencies? That is to say, can the BOH supervise itself?

Local policy, including the Charter of the Bureau and operating guidelines, present an idealised view of the Bureau, and actual implementation is likely to diverge, at least to some extent, from this representation. The question is the extent of this divergence. Individual cases of oversight examined above are drawn from specific case reports internal to the Oversight Bureau, and as far as they go, these likely reflect progress of actual cases, though it is impossible to be sure, for example, from examination of a small number of examples, the degree to which these are representative of local practice; equally, it is hard to know the degree to which decisions by the Bureau are taken seriously within the local BOH and government, and the extent to which they are ultimately enforced, though interviewees stressed that they are. A full examination of cases pursued by the Bureau would, were such a thing possible, likely show a range of behaviours, from squeaky clean to murky. The *degree* of murkiness is unknowable.

6.4.3 Shared histories; shared norms

Above, I reported concerns that providers and managers have close ties for clear systemic reasons. In addition to this, there are clear reasons to think that ties between managers and providers are close in Taoshan: Taoshan is not a large county, and it is doubtful the extent to which members of different agencies and providers could *not*

know each other – at least past a certain level of seniority: the pool of available talent cannot be large, when one considers that Sub-Bureaus must be staffed with competent staff with a range of specific qualifications and experience. During fieldwork, the close links between members of the local health administration were very clear to see, particularly during informal interaction. These links operate across the nominal dividing line between managers (BOH, NCMS Office and Oversight Bureau) and providers. Individuals in charge of day-to-day work of the Bureau, for example, have long, shared, histories with managers of the THC's they are charged with overseeing.

Bureau staff and heads, at least, of THC's seem to go back a long way: Director Wen is seemingly on good terms with most THC heads and has worked directly with the head of D THC at least. One lunchtime, after a morning of interviewing, Wen organised lunch at a restaurant overlooking a reservoir outside D town. The party consisted of around 15 people, drawn from the surrounding area: staff drawn from local THC's, plus Wen and staff from D Sub-Bureau. We were a large party arriving at a small restaurant. Waiting for lunch to be served, most of those present hung around outside the restaurant, taking it in turns to play cards, smoking, chatting and joking, behaving very much like a group of peers.

When lunch came, we were split between two rooms of the restaurant, a small local place, made up of a single row of whitewashed concrete single dining rooms opening directly onto the gravel roadway overlooking the reservoir. Our room held eight people, including Wen, Sun (head of D THC), the heads/vice-heads of two other THC's, also surnamed Sun, Liu, the head of D Sub-Bureau, a driver, a couple of others, and me. This was a big, boozy, lunch, with four bottles of spirits and four bottles of cherry wine between seven. Wen drank only cherry wine, while compelling others to drink spirits and repeatedly calling for fresh bottles. As he commented, you can tell what someone is like, including his 'work ability' (*gongzuo nengli*), from how much he can drink. Leading the drinking were Wen and Sun, head of D THC. Both drank little, while inflicting this on others.

The lines of power and influence here are clear: the proximity between Wen and Sun,

Wen's authority over all present, and the relationship of dominance of Sun (head at D) over the other two Sun (to whom he is '*lao da*'). Liu, who gives the impression of being straightforward, open and conscientious, left red-faced and uncomfortable. Most headed back to their respective hospitals and offices, hoping to get some sleep before the afternoon's work, but with very little chance of being sober by then. I left with Wen and Sun (D THC), along with the driver, in a car back to the county town, with a much clearer idea of who knows whom, and who calls the shots.

As above, Bureau staff are recruited from within the health system. Sub-Bureau heads were mostly recruited from among THC heads or vice-heads. The head of G Sub-Bureau, for example, was a THC vice-head before being recruited. The head of D Sub-Bureau had worked in the BOH since the mid-1990s before joining the Bureau. Largely speaking, the impression is one of a group of peers: these people have mostly known each other for a considerable time; they know the details of each others' careers well – they know who worked where and with whom, where they moved on to, what they did, and so on. For the most part, they even live in relative proximity: heads of both D and G Sub-Bureaus, for example, live in the county town and commute out to the townships to work (around forty minutes by bus to D); the same applies for the head of D THC (though he has his own car).

How large a block does this constitute to the possibility of effective oversight by the Bureau, especially of effective oversight by the Bureau of the THC level and NCMS implementation agencies? In theory at least, there are systemic attempts in place to introduce controls over Bureau staff: the avoidance rule, above, is one example, as is the stipulation that Sub-Bureau staff should be rotated every two years. Equally, monthly Bureau meetings in theory rotate through the Sub-Bureaus and this opportunity is used, it is said, to conduct periodic audits of Sub-Bureaus: the inspectors inspecting the inspectors. These are concerns which would merit further investigation, were this possible.

The aim of the preceding three sections is to argue that a serious attempt at local reform is underway. Each section presents data which is limited in one way or another.

This is a largely unavoidable consequence of research in a challenging context where it may be impossible to triangulate officially supplied data. The aim of the preceding three sections, however, is not to show effectiveness of Taoshan's Oversight Bureau *per se*, but to make a case for taking this seriously as a local reform. This should not, in other words, be considered just a 'face project' (*mianzi gongcheng*) of significance only in currying favour with superior levels.

6.4.4 Risk: Standing on the edge of the volcano

On the basis of the above, what should we conclude about Taoshan's Oversight Bureau? Despite the above observational evidence, it would be a mistake to write off this initiative, but we should be careful how we understand its function and significance. The important question is not the exact degree of independence achieved by the Bureau, or the degree to which staff manage to operate according to codified rules (rather than within the framework of unwritten rules we see protruding at the lunch described above), or the exact degree to which the Bureau has managed to reduce drug prescription costs. More revealing is to understand the perceptions and motivations of the key actors concerned. Of these, the most important is Director Wang.

All Taoshan policy documents, reports, and similar, present the Bureau simply as an oversight agency, and this is correct – to an extent. The key question, though, is to ask *whose* oversight agency this is. This re-framing brings to prominence an alternative understanding of the Bureau as *an alternative information gathering system established within the BOH*: an evaluation and reporting system operating in parallel to that of the NCMS system as normally constituted, and reporting directly to the leadership (the NCMS Committee – *not* the NCMS Office) of the BOH, and hence in a near-direct line to Director Wang. This is the real significance of what Taoshan has done. In interview, Wang candidly described the thinking behind the setting up of the Bureau. His main motivation, he said, was a feeling of *risk*. Wang described himself as feeling as though he was standing “on the edge of a volcano” before setting up the

Bureau: there were large amounts of money flowing around, he said, and differences in interests between different parties, but he had no way of knowing what was really going on or of maintaining control. If something untoward happened, he said, it would be his responsibility; hence the need for the Oversight Bureau.

There are confused and contradictory echoes of this in interviews with BOH and Bureau staff: while, for example, Wen denied that establishment of the Bureau was related to any specific local event, another interviewee, Xi, was open about there being significant risk in implementing the NCMS: as he put it, there is a lot of government attention to this, and a number of infractions have been discovered, including cases of fraud, in some cases involving large sums of money. People have been punished, including heads of Bureaus of Health and NCMS Offices. Xi talked of a case in Henan in which a hospital, the BOH and the city leadership were all punished. This case, and others, have been publicised, at least internally. The problem is that there are too few people and too much work; every leader (*lingdao*) is preoccupied with this, according to Xi. Director Wang, when asked whether Taoshan's initiative had the potential to be rolled out (see below), said he thought so: he sees a lot of problems, in the media and in internal documents – and these problems are not discovered from within: they're all discovered by external audit committees and the like.

A piece of independent evidence is available to lend credence to this interpretation of risk as the central motivating factor behind the establishment of the Oversight Bureau. Provincial records of staff occupying leadership positions in the health system show that, in the years prior to Taoshan's establishment of the IOB, the whole leadership of Taoshan's health management was changed, including the Party Secretary, heads of hospitals, and others of less note. It is inconceivable that this could have taken place for normal staffing reasons; the only conceivable explanation is of some local problem within the health system prompting intervention by higher levels. By 2008, provincial materials carry citations for good performance of individuals within the BOH leadership, and things appear to be back on track.

6.4.5 Risk and NCMS structure: Making good deficient central policy

Unfortunately, it was not possible to discuss these events locally. It seems clear that Taoshan in some sense 'got burnt' in operating the NCMS, that there was a wholesale change in staff and that the Oversight Bureau was set up soon afterwards by the incoming head of the BOH. To this extent, Taoshan's initiative corresponds to a line of analysis in the literature on local government innovation in China: that local governments tend to innovate in response to crisis (Fewsmith 2008b; Gao 2008b). The story of Taoshan's initiative is more complex than this, though, stemming not just from a local crisis, but more broadly responding to deficiencies in the NCMS. Here, I show how the Bureau in fact responds to a structural deficiency in local NCMS management and oversight. In the next section, I show a pervasive framing of NCMS implementation in terms of risk. Neither structurally-induced (capacity) deficits nor this risk framing should be considered unique to Taoshan.

From a systemic point of view, the IOB responds to failings in the institutional design of the NCMS, which, for very specific and contingent historical and political reasons, concentrates oversight at the county level while giving insufficient attention and resources to this: as discussed above, oversight is hobbled by capacity, personnel and funding constraints and insufficient separation of management and supervision organs. The core of Taoshan's initiative is extremely simple: it strengthens local oversight of scheme functioning by requisitioning resources from THCs and placing these under the managerial control of the BOH and in this way makes good, at least to a degree, local capacity and management funding deficits. Irrespective of the crisis presumed to have occurred in Taoshan, these are systemic deficits: they result from poor scheme design and are in no way unique to Taoshan.

In adopting this solution, what Taoshan has done is little more than formalise practices existing in many places and introduce a degree of managerial rigour into their functioning, if we consider that the co-opting of THC employees into running and overseeing the NCMS is, in fact, common in many places: Taoshan has simply formalised a *de facto* common administrative reallocation of resources in a way

consistent with Chinese government personnel allocation (*bianzhi*) rules.⁵⁶ The link between deficient national policy and Taoshan's innovation is very clear: Director Liu of D Sub-Bureau, for example, argued that an oversight entity like the Bureau was a necessity in running the NCMS – if counties don't strengthen oversight, they will lose money from the NCMS as funds will be mis-spent. It would make sense, Liu argued, to allow a small amount of money from NCMS funds to be used for this purpose as this would save the scheme money. This, however, would require national policy change and is unlikely to happen. Taoshan's initiative is a way of making good deficient national policy.

6.4.6 Risk and NCMS structure: Ubiquity of risk framings

At the risk of downplaying the crisis that Taoshan seems to have experienced, the perception of risk cannot be directly equated to actual occurrence of local crisis or mis-implementation – this perception is mirrored elsewhere and should not be considered unique to Taoshan. Fieldwork in Meijiang City (also X Province) revealed a very similar framing of risk. The head of the NCMS Office described the main risks in implementing the scheme as coming, firstly, from collusion between patients and providers to cheat the scheme and, secondly, in mis-setting of NCMS plans and the corresponding risk of over-spending. NCMS fund setting and fund risk are dealt with in detail in the next chapter.

Meijiang's approach to oversight and intra-systemic transparency is extremely interesting in its own right: from the start of operation of the scheme, Meijiang contracted day-to-day management of the NCMS to a commercial insurance company. This kind of 'model' of NCMS functioning has been described elsewhere (Mao 2005) and one important rationale stressed in most analyses is that this allows Bureaus of Health to buy in actuarial capacity and improve fund management. In the case of Meijiang, though, this arrangement also allows a much greater degree of direct

⁵⁶ A possible criticism of what Taoshan has done is that it in fact formalises the collateral financing burden discussed above, at least to an extent.

oversight of NCMS reimbursement than would be possible by relying on the NCMS Office alone: insurance company staff are posted in major providers to oversee reimbursements, check patients' identities and verify NCMS membership. At this stage, all reimbursements are checked individually. The company's management centre carries out a second, computerised, check of all reimbursements. In the third stage, the NCMS Office checks a randomly chosen proportion of submitted reimbursements as quality control and as a check on the work of the insurance company. (This is in addition to routine evaluation (*kaohu*) of hospitals' work.) Meijiang also carries out door-to-door checks of *all* NCMS reimbursements involving injury claims as NCMS managers believe there to be a high risk of the scheme paying out unnecessarily in cases where there is third party liability. This degree of oversight is clearly beyond the means of most counties operating the NCMS, though Meijiang's framing of risk – and its response to risk – are in fact very similar to Taoshan's. The policy response in both cases corresponds to a similar *repertoire* of official action: a simple quantitative increase in oversight of providers and NCMS reimbursement through an increase in resources devoted to this, rather than through substantially changing the *mode* of oversight.

Where the two localities differ is in their approach to transparency. In the case of Taoshan, Wang's logic is simple: it is better to discover problems yourself than to have someone else discover them for you. The much-vaunted 'independence' of Taoshan's Independent Oversight Bureau is therefore only relative – and, locally anyway, of only relative importance: the Bureau is not intended to provide independent monitoring; rather it should be thought of as *dependent* monitoring, and the creation of a form of limited, intra-systemic, transparency. Meijiang's approach is somewhat different: according to Director Gao of the NCMS Office, the inclusion of a third party in management and oversight of the NCMS in and of itself serves to reduce risk and increase transparency: the insurance company provides a monthly 'fund risk report', analysing changes in NCMS operation, cost increases and reasons for this and suggesting countermeasures. (One report from 2009, for example, analyses spending

and notes that this is very close to targets set by the NCMS leadership group and recommends that the NCMS Office more strenuously implement provider management regulations (X Company Meijiang NCMS Management Centre 2009.) Meijiang NCMS Office has the capacity to do this analysis themselves, but reports from the insurance company are sent both to Director Gao and to city government, creating a degree of local transparency: where Taoshan's IOB creates an extra information channel reporting to the head of the BOH, and greater transparency within the BOH, Meijiang's use of a commercial insurance company creates a three-way relationship (insurance company – NCMS Office – city leadership) and a degree of transparency at a larger scale – that of the city.

Director Gao is clear about the benefits of this: while this limits his personal authority and freedom of action, obliging him to act on the company's recommendations or to explain why he has chosen not to do so, it also reduces his risk in implementing the NCMS by directly involving the city leadership in oversight. It also, he says, provides him with third party legitimacy when discussing scheme management and targets with hospitals in the city (this is important, he says, as he was a doctor before moving to work in the NCMS Office, and consequently knows the heads of all the local hospitals). In this sense, it helps formalise the multiple unofficial relationships existing between Gao, the NCMS Office, and providers. Other localities have visited Meijiang to study their management of the scheme, but most places have failed to implement this well. As Gao explains it, under normal operation of the scheme, authority is too concentrated in his person. Meijiang's management solution provides a way to increase transparency and for Gao to decrease the risk in implementing the NCMS, but he must cede a degree of authority in order to achieve this. This need to cede authority, he says, is the main barrier to other localities' studying Meijiang's model.⁵⁷

⁵⁷ The theme of risk carries through other interviews conducted in Taoshan: discussions of current work and plans by BOH to implement reforms to provider payment mechanisms within the NCMS (*zhifu fangshi gaige*) clearly present this from within a discourse of risk. The next chapter is a detailed analysis of one county's experiments with cost control reforms in response to fund risk.

6.4.7 Risk: *Perceptions and power*

Above, I argued that underlying development of Taoshan's Oversight Bureau is a profound perception of risk and that, while in Taoshan's case we can see clear proximate reasons for development of the IOB, this risk framing is by no means unique and is mirrored elsewhere in local understandings of implementation of the NCMS. Here, I make two supplementary points. First, it is *perceptions* of risk that drive people to action. Second, risk is intimately bound up with exercise of power.

Risk, as perception, deals with "projected dangers of the future": "as conjectures, as threats to the future, as prognoses, [risks] have and develop a practical relevance to preventive actions" (Beck 1992, p. 34). Risk is virtual:

"Risk does not mean catastrophe. Risk means the anticipation of catastrophe. Risks exist in a permanent state of virtuality, and become 'topical' only to the extent that they are anticipated [...] At the moment at which risks become real [...] they cease to be risks and become catastrophes. Risks have already moved elsewhere [...] Risks are always events that are threatening" (Beck 2006, p. 332).

Risk should be seen as a site of conflict and the exercise of power, as "the inequalities of definition enable powerful actors to maximize risks for 'others' and minimize risks for 'themselves'. Risk definition, essentially, is a power game". This as risk "is a socially constructed phenomenon, in which some people have a greater capacity to define risks than others. Not all actors really benefit from the reflexivity of risk only those with real scope to define their own risks" (Beck 2006, p. 333).

Two facets of this risk argument are relevant in examining the case presented here. First, it is *perception* of risk, and not putative objective assessment of probabilities, that underlies operation of Taoshan's Oversight Bureau, Meijiang's approach to oversight and, in subsequent chapters, approaches to fund management and provider payment reforms. The overall apparatus of government whose function is control and oversight of local cadres is huge and defies mapping. In Chapter Five, I discussed NCMS

operation in terms of the pressurised system, supplemented by an analysis of the NCMS funding system and the way this enforces a cellular structure of scheme operation at the county level. This is complemented by any number of systemic oversight mechanisms – innumerable audit committees and bureaus from the central level downwards, as well as media and public scrutiny (to a degree, at least) – whose aim is the enforcement of an ‘audit culture’ (Kipnis 2008) whose ultimate *purpose*, though not amenable to quantitative assessment, is individualisation of responsibility and risk. While we may not be able to assess the extent of the apparatus underpinning this, its effects are visible.

Second, risk is rarely neutrally accepted by those exposed to it. Risk can be channelled through exercise of power. In Taoshan, we see exercise of local authority with the express purpose of reducing risk in one part of the system (the BOH), by enforcing (attempting to enforce) certain standards of behaviour in, and attempting to transfer risk to, other parts of the system (providers, NCMS implementation agencies). In this way, the audit culture whose effects are clearly felt at the county level is replicated and re-deployed further down the administrative hierarchy. Oversight and supervision of providers and NCMS implementation agencies is little more than the extension by the BOH of an already-existing audit culture into an underexploited area. The form this takes – a monitoring agency charged with oversight, and managed by work targets linked to pay and bonuses – is firmly within a certain *repertoire* (Behn 1987; Behn 1988) of official action.⁵⁸ While Taoshan’s initiative may be locally-generated, it is of very little actual *uniqueness* (see below), and is clearly much more consistent with a common official audit paradigm than it is with, for example, the transferring risk through reform of provider payment mechanisms and new forms of contracting (though I argue in Chapter Seven that the same logic holds here also).

⁵⁸ Behn summarises this as “managerial repertoire – patterns of managerial situations and successful actions that can be recalled and applied to new problems” (Behn 1988). I develop this further in subsequent chapters.

6.5 Taoshan's Oversight Bureau, the NCMS and policy change

6.5.1 Taoshan's initiative in the context of calls for structural change in local oversight

Above, I discussed perceived shortcomings in local oversight arrangements presented in Chinese academic and policy discourse. The need for increased NCMS oversight has been presented in very many central documents and speeches over time, and at least a number of authors have made a case for increasing third party oversight of the NCMS. Gao and Zhu, working on a national project on NCMS fund security and management, with support from the MOH Rural Health Department (*Nong Wei Si*), believes, in line with the above, that under current conditions, local oversight is formulaic and non-specialist. They suggest that independent accountants could provide effective third party oversight (Gao and Zhu 2009).

A possibly more complete proposal comes from Wu (2009), who argues for systemic reform, requiring reorganisation: separation of management and oversight functions, the breaking of the relationship of dependence between the health administration and providers (*fu-zi guanxi*, lit. 'father-son relationship') and the need for change in responsibility for hospital management. The oversight system needs to be fundamentally reformed; there is need for a pluralised oversight system, including specialist professional oversight entities, real transparency over NCMS fund use, and broad *de facto* 'social participation' (*shehui duo fang canyu*) in oversight, alongside greater self regulation of providers. Wu offers no concrete solutions to these problems, but does advance certain principles. Most importantly, oversight organs must be independent:

"Oversight organs must be independent of the targets of their oversight; they must be separate from the entities charged with management of providers. What this structural separation (*jiegou fen li*) means is that [oversight organs] should be able to implement oversight policy without being influenced by [particular] interests ..." (Wu 2009, p. 41).

Wu calls for the formation of professionalised groups at the provincial level, made up of people with relevant expertise from large hospitals, health research institutes, audit departments and so forth. These people are to be selected randomly and to make up a temporary oversight group. Wu also calls for a solid transparency framework, setting out rights of access to information by People's Congresses, scheme users, etc., as well as greater media oversight. He also discusses two possible mechanisms for increasing oversight. The first involves NCMS Offices having a larger presence in THCs, using specially selected staff to carry out oversight, and rotating of staff to avoid problems of complicity. Oversight and management (*kaohu*) of the overseers should be carried out by personnel departments of Bureaus of Health, alongside NCMS Offices, with the aim of increasing control. Another, less detailed, proposal calls for greater county-level oversight of township-level overseers (Wu 2009). Chen (2010) calls for similar reforms to Wu, but in some places is possibly more interesting, calling for much greater user engagement in oversight and management, which could be carried out "according to the model rural self-governance", a reference to development of village committees.

Wu and Chen provide interesting thinking on what a more developed and independent oversight regime might look like. One should notice, however, that these ideas remain fairly firmly within a government-centred framework: though there are also numerous calls for increased social supervision (through user participation, as above) and through increased media and other oversight, such statements seem generally to remain at the level of exhortations, rather than detailed plans.

How does Taoshan's initiative compare to these calls for structural reform? My analysis above of the place of Taoshan's Bureau within the local oversight structure in fact shows significant similarities with the kinds of initiative presented by Wu and Chen, or at least that Taoshan's oversight structure is not radically different. However, I have also argued that increased local *intra-systemic* transparency should be considered the goal of this initiative. On this basis, while Taoshan's initiative is likely useful, and possibly a helpful stepping stone to greater oversight, it is doubtful whether this really

fulfils the functions one would hope for from significant structural change to local oversight.

6.5.2 Will Taoshan's Bureau be sustained?

As above, Taoshan's Bureau and similar initiatives elsewhere are rooted in a strong perception of risk on the part of local NCMS managers. The extent to which this is likely to be an enduring phenomenon is not clear. The NCMS has been the subject of great government attention since its inception, and this is likely linked to its being very much in a developmental phase. As in Chapter Five, various authors caution that the NCMS lacks a solid legal foundation, and that this could leave it vulnerable to a shift in official priorities over time (Gao and Han 2007, p. 85; Zhang 2011b). Above, I presented data from various analyses to the effect that local NCMS management is underfunded. These data all refer to early stages of operation of the scheme, and for the most part deal with counties in poorer areas.⁵⁹ Such solutions, though, are beyond the reach of the majority of counties running the NCMS. For local financing of NCMS management to be sustainable and not subject to the vagaries of county-level priorities and budgetary fluctuations, this must be formally included as a budget item in county budgets and not based on *ad hoc* appropriations (Sun and Chai 2009, p. 112; Yan, Raven et al. 2010). The same applies, by extension, to Taoshan's Oversight Bureau. It is hard to know the extent to which this has taken place to date, but the fact that the first provincial-level legislation (*tiaoli*) on the NCMS, released in Jiangsu in 2011, mandates very clearly certain levels of spending on the NCMS (3% of average incomes of local residents), but fails to mandate a similarly concrete level of spending on management, is not encouraging (Jiangsu Sheng Renda Changweihui Gonggao 2011, No. 79). Two of the main counties in which I carried out fieldwork, Taoshan and Feitian, are not well off and there is little reason to think that this kind of situation is

⁵⁹ Some counties clearly have few constraints in funding NCMS management: a number of counties and cities in more developed areas, such as Meijiang, have contracted day-to-day management of the NCMS to commercial insurance companies and pay a management fee (Mao 2005), a large part of the motivation for which is to increase management efficacy.

uncommon. The above are concerns for the NCMS in general and, more specifically, for initiatives such as Taoshan's: with less attention to the NCMS, with greater local budgetary pressure, with a period of crisis-free operation of the NCMS, would the perception of risk driving Taoshan's initiative remain so keen? Would the Bureau survive and continue to have a role? These are concerns, but are unanswerable at present.

6.5.3 *Supra-local influence of Taoshan's Oversight Bureau*

What, to date, has been the impact of Taoshan's Oversight Bureau outside the county? As above, the county was invited to present its experience at a provincial meeting attended by other Bureaus of Health and NCMS managers and it was also included in a national report. The BOH was also invited to submit materials on their initiative to the MOH in preparation for a national meeting at which they were to discuss their work; however, after a delay, the county was informed that the meeting would not be convened. Reasons are unclear, but locally it is speculated that Taoshan's timing was unfortunate: that this innovation, focused on strengthening NCMS oversight and management, had been overtaken by a shift in central attention to provider payment reform (*zhifu fangshi gaige*) in 2009-2010. It is impossible to be sure of the truth of this, but it could be true: national attention to payment reform has increased substantially in recent years. Whatever the truth of this, Taoshan's initiative has received a *degree* of supra-local attention, and higher levels have been instrumental in propagating the initiative (convening a provincial meeting; inclusion of Taoshan's initiative in national reports, conferring legitimacy), but it has neither achieved significant national fame nor been formally included in higher-level policy.

Vertical propagation of Taoshan's initiative has been limited. What, though, of *horizontal* propagation?⁶⁰ In principle, given the national-level publicity derived from

⁶⁰ Horizontal diffusion of innovations and local government practices receives very little attention in the literature on government innovation in China. My fieldwork tends to show that this is likely more common than vertical uptake of local practices. I return to this theme in the next two empirical chapters.

inclusion in a national report, one might expect that Taoshan's initiative should be widely known about, but this appears not to be the case. Following the provincial meeting at which the Bureau was discussed, Taoshan has received visits from a number of other counties, and a number of counties have adopted similar practices.

D County is cited locally as an example of a Taoshan spin-off, having set up an oversight agency following the provincial meeting, though this lacks the staff, and 'independence', of Taoshan's Bureau. Similarly, H County set up an 'NCMS Investigation Team', upgrading its NCMS Office to an NCMS 'Management Centre', including an inspection department which appears similar to Taoshan's initiative (X Province Bureau of Health New Cooperative Medical Scheme Office 2009). Both examples here relate to in-province learning, but Taoshan *may* have had impact further afield: Wen tells a story of receiving a phone call from a county in Sichuan, asking for materials on Taoshan's initiative, though details on the county and its reform are unavailable.

There is no way of ascertaining clearly whether these initiatives are directly linked to Taoshan's reform. They *may* be linked – as Wen states it, everyone is studying everyone else – but determining the degree of correspondence or influence would require fieldwork that I was unable to carry out. Understanding the genesis of such initiatives is hard: Taoshan's Oversight Bureau is not unique, and other places in China have very clearly independently established similar agencies for similar purposes. Qianjiang (Chongqing) District's Accounting Centre, conforms to a similar repertoire of policy enforcement and performs similar (though more extensive) functions to those of Taoshan's Bureau (Zhang, Liu et al. 2007, pp. 48-49). Seen from Taoshan, though, the examples cited here are motivated by the same spirit as the Oversight Bureau, but are believed to lack the effectiveness and force of the 'original'. Paradoxically, according to Wen it is what makes the Bureau effective – local government support and willingness to allocate extra personnel – that may actually be a barrier to effective 'diffusion' of this mechanism, and may lead other places to adopt watered-down solutions.

6.5.4 Linking Taoshan's Oversight Bureau and policy change

Various media and official evaluations attribute novelty to Taoshan's reform. Of more interest is the fact that Taoshan's Oversight Bureau was locally generated – '*zi sheng de*', in the words of Director Wen – and not specifically mandated by higher levels. One article published by Director Wang (Wang 2009a) states that the BOH studied and visited places elsewhere in China, though this was contradicted in interview – there does not appear to have been a specific inspiration or model for Taoshan's reform, though it may have been influenced by models and discourses in general circulation. I return to this below in a discussion of China's 'imaginative geography' (see Section 9.3.3). In discussion, another member of the BOH stated the relevance of the Bureau to supra-local policy in this way: the BOH needs to use funds in a reasonable way, and increase oversight, but there is considerable discretion as to the form this takes locally: "So, 'strengthening oversight and management of the NCMS', that's just a phrase ... In practice, how to do oversight and management ... That's down to each place" (Interview no. 21).

Two things stand out: first, that this is what one could call an 'orphan policy', lacking a clear mirror at the superior level, though it seems, from the perspective of an outside observer, to correspond to an intent frequently expressed in higher level policy. A central document from 2006 on strengthening NCMS management, for example, calls for increase in 'external oversight' among (many) other ways of strengthening management (Weinongweifa 2006, No. 40); such exhortations are common. Second, and more interestingly, orphan though this is, the policy is understood locally as corresponding to the overall objectives of development of the NCMS and as being a legitimate development of the scheme. Director Wang states the issue this way in a Xinhua article:

"The NCMS is a good thing. But good things are hard to do well. County NCMS funds are several tens of millions of *Renminbi*, if not more than a hundred million. How can these be managed well? There is still not a national model for the NCMS, each place has to work with their local conditions and

experiment to develop the scheme” (Ma 2008, p. 23).

Of equal interest is the evaluation of Taoshan’s initiative by the province. The county was invited to present its experience at a provincial meeting attended by other counties. This was a deliberate attempt to showcase Taoshan’s Bureau and was described by the province as ‘extending’ or ‘popularising’ (*tuiguang*) this model, though other counties were not obliged to adopt this. The provincial assessment of both Taoshan and Meijiang’s practices is worth noting: both Taoshan and Meijiang were described as increasing ‘oversight’ (*jianguan*) of providers, and what both places were doing was described as legitimate, though actual methods differed. The important thing was described as effectiveness (*xiaoguo*), not means (*shouduan*) – there is more than one possible management *form* (*xingshi*) for achieving the same result (Interview no. 48).

This judgement is interesting: first, it clearly shows a degree of provincial tolerance of the *form* that implementation takes, within the overall aim of development of the scheme.⁶¹ Second, in attempting to popularise Taoshan’s Bureau, the hope is not necessarily that other counties will replicate this exactly, but rather that other places will be encouraged to strengthen management; the form this takes in other localities may be different. Third, in understanding Taoshan and Meijiang’s practices as equivalent, the province’s understanding is more developed than that of many of my interviewees in Taoshan, who see Meijiang from afar as a rich city, with far more funds at their disposal, and therefore beyond their ability to emulate (whereas the largest difference, in my analysis, lies in the degree of local transparency). (By the time of my second visit to Taoshan, perceptions had changed somewhat; I return to this in the next chapter.) Fourth, the province recognised that, while useful, neither Taoshan nor Meijiang’s model is an adequate solution to the problem of oversight of providers. Both models are highly labour intensive and, in the long run, flawed: medical services are too complex for this kind of direct oversight to be effective – oversight requires too much information, there are too many unknowable quantities and doctors must be

⁶¹ A point of peripheral interest: this tolerance is even more remarkable if one considers that Taoshan’s initiative seemingly has its roots in a local crisis.

allowed a degree of discretion in their behaviour. In my terms, above, this *mode* of oversight is limited: in both cases, these counties are attempting to adapt existing modes of management to problems to which they are *fundamentally* unsuited. Accordingly, the province was planning to start experiments with provider payment reform. I discuss provider payment experiments and what they mean in terms of modes of oversight and repertoires of official action in the next chapter.

Where now for Taoshan's policy? Locally, there is a clear belief that the Bureau performs a useful function. Members of the Bureau and BOH argue for extending its work to include other components of local health system oversight, transforming this from limited NCMS oversight into a much more powerful entity with an expanded role. They argue that this would make sense: the need for oversight is there; the Bureau exists and could quite easily be extended; much more sense to group oversight functions than to separate them. This may happen, but there are no plans for this as it stands. In contrast, as seen from outside the county, Taoshan's initiative may have missed the boat: for a brief moment, the Oversight Bureau had a modicum of provincial and central limelight, but this appears to have passed.

The failure of Taoshan's model, and other equivalent practices in other places, to get serious national attention and to be proactively propagated points to a theme I return to elsewhere: that local practices and models frequently *remain* local despite potentially being useful more generally. Ostensibly useful local practices seemingly frequently fail to receive adequate backing and propagation. In the case of Taoshan's Bureau, this is clear. A very recent MOH document on strengthening NCMS fund management calls for the establishment of just this kind of agency:

“There is a need to build effective mechanisms to restrain management and operation: each province (district, city) must, according to their local conditions, in a defined period of time, ensure the separation of NCMS management and operation at the township level and THC services. [Counties should] implement a system under which county [NCMS] operating agencies second operational and audit staff to the township level; using a principle of avoidance

[of native location], and revolving postings, etc., ensure the independence of NCMS operating personnel. Currently, it is hard to separate NCMS management and operation from service provision at the township level, [so] county-level NCMS operating agencies should strengthen work on auditing of reimbursement” (Weinongweifa 2011, No. 52).

Promotion of separation of management and oversight – around which Taoshan’s model revolves – is specified here in some detail in national policy. While the need for a model such as this remains evident – if nothing else as a stepping stone, a transitional institution on the way to the kind of solution favoured by the province – at this stage it is impossible to say whether Taoshan will be picked up again and boosted nationally, and it is doubtful whether this initiative could have significant supra-local impact without some sort of higher-level intervention and propagation.

There is a degree of ambiguity in such a clear central exhortation, late in its the development of the scheme, for development of management structures basically set out in Taoshan (and elsewhere) several years previously. Not only does this seem to be an admission that this remains a deficit in management of the scheme, but – more generally – it points to the lack of systemic recycling of useful local practices, both Taoshan’s management model and equivalent or similar practices developed elsewhere. Saying this, two points must be borne in mind: first, Taoshan-*type* management practices have achieved at least a degree of sub-national coverage, despite a seemingly low level of specific central attention to this, and part of the reason for this is surely sub-national propagation efforts such as those described above. Second, detailed central direction on separation of management and oversight and the promotion of Taoshan-type agencies, *late though it is*, is likely more important than actual propagation or replication of Taoshan’s model *per se*. In other words, while Taoshan and other sub-national models are not specifically mentioned here, the underlying idea is clearly articulated as a principle of scheme development.⁶²

⁶² My thanks to Professor He Zengke of the CCTB for pointing me towards this realisation, though in respect to local government innovation in China more

6.6 Interim conclusions and looking forwards

This chapter has shown one county's development of a local oversight mechanism within the NCMS and how this came to have a degree of supra-local impact or fame. Scheme oversight has been consistently criticised as weak, and I argue that this largely stems from specific national debates on the 'peasant burden' pre-dating the NCMS, but which were carried through into scheme policy and design. Taoshan's Oversight Bureau results clearly and directly from this conjuncture and, while the county has experienced some kind of crisis in its recent implementation of the NCMS, the overall issue addressed in Taoshan is structural, not purely local and contingent, and this is clearly visible in the way counties frame implementation of the scheme in terms of risk and the existence of similar measures in other places, both inspired by, and completely separate from, Taoshan's oversight agency. The county's reform is presented here as a local innovation to the extent that it was not mandated from above – it is an 'orphan policy', though one which the county and other systemic actors all understand as a legitimate (though insufficient) development of the scheme. While it falls short of being a fully-fledged solution to problems of scheme oversight as discussed by Chinese academics and analysts, it is a step in the right direction and very recent MOH policy calls for just this kind of interim measure. A number of themes arising from this chapter carry through in subsequent chapters: the ubiquity of risk framings, the importance of fund security and the highly local development of multiple policy solutions which, while simple and often falling short of best practice, are nevertheless regarded as legitimate by systemic actors.

generally.

Chapter Seven

Implementing and developing the NCMS: Institutional logic of Feitian's cost control reform

7.1 Introduction and background

This chapter examines an NCMS cost control reform, 'Controlling Cost Growth' (CCG),⁶³ carried out by Feitian County, a relatively poor county in an eastern province. Feitian's reform stems from a need to control costs in order to better implement the scheme, meet implementation targets and control fund risk. The logic of this reform shows very clearly a county-level view of the operation of the pressurised system and the paradox inherent in this: rigorous targets and pressure to implement at the county level, combined with a policy of principles (*yuanze*) in which the form implementation is to take remains largely unspecified, and in which counties must actively *interpret* policy. Feitian's reform is an innovation to the extent that it was locally developed, for local reasons, and not mandated from above, though the reform is understood as falling within the scope of NCMS policy and being a legitimate development of the scheme. The county's reform has received attention at provincial and national levels and in the process been reinterpreted as a 'provider payment' reform, corresponding to a worry expressed since the start of the scheme that exclusive use of 'fee for service' payments under the NCMS will encourage cost growth, undermining scheme effectiveness and sustainability.

Since the beginning of the NCMS, analysts have expressed concern that the scheme does little to contain costs. This has for the most part been framed as a question of provider payment mechanisms and the reliance of the scheme on *post hoc* 'fee for service' payments. While provider payment reform has received a degree of attention

⁶³ Names of places and people have been changed to preserve anonymity. This applies also to published materials where necessary. In some cases, figures and dates have been marginally altered to prevent identification of sites or people. In all cases, anonymising people and places and any other marginal changes have no effect on meaning.

in national policies, and there have been a number of sub-national initiatives on this, serious national attention has been relatively late: national policy opinions on this were only released in 2012 (Weinongweifa 2012, No. 28). Feitian started cost control reforms in 2007-2008, and started to receive provincial and MOH attention shortly thereafter. The county's reforms were developed locally, and were not part of a national pilot project.

This chapter is structured as follows. Section two provides background on growth in costs and medical spending in China and the predominance of fee for service payments under the NCMS. Sections three and four provide a county-level view of the pressurised system and implementation of the NCMS, through a detailed look at local scheme budgeting. Section five describes Feitian's reform and shows this as a direct extension of the logic of policy implementation and section six frames Feitian's reform in the context of provincial, national and other sub-national policy development. Section seven concludes.

7.2 Background to Feitian's innovation

7.2.1 Cost growth and misaligned incentives

Since the beginning of the NCMS, analysts have expressed the concern that the scheme primarily reimburses providers through fee for service (FFS) payments and that this does little to control cost growth. The nub of this is simple: under FFS, healthcare providers treat patients, prescribe tests, drugs, etc., and then present the bill for this to the NCMS Office. Within the scope of procedures, drugs, etc., allowed within the NCMS, the NCMS has, in principle at least, to reimburse costs presented by hospitals.⁶⁴ As in Chapter Six, overseeing providers is hard. For the most part and in most places, the NCMS has no way to oversee all treatments, and relies on overseeing payments that exceed a certain threshold, carrying out checks of a certain proportion

⁶⁴ According to a World Bank policy note, prices of approximately 4,170 items are set in this way. Items are set by the NDRC; prices are then set by provinces. At the time of the World Bank note, the list was being revised (Langenbrunner 2011).

of providers' medical files to see if there are errors, questionable prescriptions, and the like. Under FFS, providers have an incentive to increase the volume of treatments, tests and drugs provided as this benefits their incomes and there is little insurers can do to curb this kind of behaviour, save to increase the quantity of oversight. This, as in Chapter Six, is likely to run up against limits in personnel, funding and technical expertise, as well as be difficult under the current institutional framework.

Globally, it is widely recognised that exclusive use of FFS is inflationary (Meng, Yip et al. 2010). In China, certain specific concerns coincide with this to make analysts further question FFS. There is a degree of consensus over the root of this problem, which lies in the failure of previous reforms in hospital management and funding. Yip and Hsiao's analysis is broadly representative:

“China experienced a drastic reduction in its national revenue during its economic transition, which reduced the government's capacity to fund health care. Government subsidies for public health facilities fell to a mere 10% of the facilities' total revenues by the early 1990s. To keep health care affordable, the government maintained strict price control by setting prices for basic care below cost. At the same time, the government wanted facilities to survive financially, so it set prices for new and high-tech diagnostic services above cost and allowed a 15% profit margin on drugs. This created perverse incentives for providers who had to generate 90% of their budget from revenue-generating activities, turning hospitals, township health centers and village doctors alike into profit seeking entities. Subsequently, providers over-prescribe drugs and tests and hospitals race to introduce high-tech services and expensive imported drugs that give them higher profit margins” (Yip and Hsiao 2009, p. 201).

There is basic consensus over the core of this analysis among health systems researchers: insufficient funding, revenue generation through provision of treatment, and perverse incentives introduced by a distorted pricing mechanism, compounded by

erosion of professional norms in the reform era and before.⁶⁵ In addition, controlled prices create a multiplier in providers' selling of drugs, pushing up volumes, as the "price-setting approach created a leveraging effect, whereby a provider had to dispense seven dollars' worth of drugs to earn one dollar of profit" (Yip and Hsiao 2008). As I will show below, this point is fundamental to Feitian's reform.

There is a degree of consensus regarding results of the poor incentives introduced by this structure: a technological arms race, in which providers invest in new equipment in order to compete for patients, increased hospital stays and difficulty in referral between providers, which seek to hold on to patients (e.g. Eggleston, Li et al. 2008). Kickbacks on drugs and use of out of date and counterfeit drugs, especially at the village level, are other symptoms of this malaise (Yip, Hsiao et al. 2010, p. 462).⁶⁶

Healthcare costs have risen consistently and faster than growth in both GDP and incomes. One review finds annual growth of total health expenditure of 11.5% between 1978 and 2003, compared with 9.6% annual growth in GDP over the same period, and that "a hospital stay in rural areas was 1.8 times as costly in 2005 as in 1995, but average disposable income rose only 1.1 times over the same period" (Hu, Tang et al. 2008, p. 1847); other studies come to equivalent conclusions (e.g. Zhang, Cheng et al. 2010b). Multiple studies have found over-prescription of drugs, high levels of unnecessary or 'irrational' treatment and excessive investment in capacity (Wagstaff, Lindelow et al. 2009b, pp. 24-26; Langenbrunner 2011). All of these analyses are consonant with the picture sketched above.⁶⁷

⁶⁵ Yip and co. point out that, post-1949, physicians "became employees of hospitals, with hospital administrative control replacing professional self-regulation when the Communist Party came to power"; they argue that this is detrimental to maintenance of professional norms (Yip, Hsiao et al. 2010).

⁶⁶ For further reading, see Eggleston, Li et al. (2008), Hu, Tang et al. (2008), Langenbrunner (2011), Wagstaff, Lindelow et al. (2009b), Yip, Hsiao et al. (2010).

⁶⁷ There is ample evidence on this theme. Wagstaff, Yip et al. (2009) and Eggleston, Li et al. (2008), provide recent reviews of available literature. Wagstaff et al. report very high levels of unnecessary drug use, high-tech tests and antibiotic use (more than twice the international average for common colds), swathes of unnecessary procedures, and unnecessarily extended hospitalisation. Eggleston et al. cite a now-infamous 1990s study which found less than 2% of drug prescriptions to be

7.2.2 NCMS and cost growth

The NCMS was introduced at a time of rapid cost growth arising from the overall environment in which it was expected to function and costs have continued to increase since the introduction of the scheme. Based on MOH data, a recent World Bank report shows huge generalised increases in costs at the township/county level in the early years of scheme operation: between 2003 and 2004, average costs per case jumped by 15-20% in THCs and county hospitals, even after adjusting for changes in case mix. Between 2004 and 2005, the increase was around 25% in county hospitals and around one third in THCs (Wagstaff, Lindelow et al. 2009b, p. 24). The 2006 report commissioned by the State Council argued that the NCMS may have spurred cost growth in its own right, by weakening the incentive for users to constrain providers, and that excessive tests and treatments, inflation of diagnoses and similar problems are common (NCMS Pilot Evaluation Group 2006, pp. 98-99). This and other studies find that providers find countermeasures (*duice*) to, and 'make full use of' (*yong zhu*), policy (NCMS Pilot Evaluation Group 2006, p. 107; Gao and Han 2007).⁶⁸ County-specific studies also frequently show rapid growth in both use and costs (Gao and Zhu 2009; Sun and Chai 2009, p. 104; Zhang, Cheng et al. 2010a). Counties in which I carried out fieldwork described the NCMS as having 'liberated' (*jiefang*) previously-constrained health seeking.

7.2.3 NCMS and cost growth: Framings and legitimate scope of action

I wish to underline two things here: first, to a large extent, cost growth has been framed as an issue of provider payment reform. This comes through very clearly in academic studies and in work by international organisations. Criticisms that the NCMS has not, for most of its existence, explicitly included provider payment reform are justified and

'rational' in one township- and village-level study in a Chongqing county. Astonishingly, "[i]n the case of village clinics, only 0.06% of drug prescriptions were deemed reasonable"!

⁶⁸ Some studies have even found that insurance may increase users' financial risk by increasing use of more expensive care, use of more expensive facilities, longer hospital stays and so forth (Wagstaff and Lindelow 2008a; Pan, Dib et al. 2009).

the absence of this is clearly a factor in continued cost growth, and introduction of provider payment reforms under urban insurance schemes, for example, indicates that these have potential to be effective (Eggleston, Li et al. 2008; Yip, Hsiao et al. 2010).⁶⁹ This framing is important: as one interviewee argued, provider payment reform has dominated debates around cost growth in China, and other issues, such as adequate local gatekeeping and referral by primary level doctors, as in the UK, have received less attention (Interview no. 62). Second, despite the absence of payment reforms, there has been rhetorical attachment to cost control from early on in scheme development and other mechanisms for controlling costs and/or controlling providers are common, though insufficient.⁷⁰

The NCMS is in many ways very constrained in its ability to influence cost growth. Various Chinese analyses point out that the NCMS, as a 'public service unit' (*shiye danwei*) attached to the county BOH, lacks teeth of its own. In dealing with providers, for example, the NCMS Office typically has no direct recourse to administrative sanction of providers for overspending, breaking of NCMS rules and the like: methods of Party discipline that can be brought to bear on hospital heads, for example, must be exercised through the BOH and, as in Chapter Six, at the county level, this relationship is complex. One NCMS manager in Meishan stated that they frequently ask for BOH and city intervention with hospitals, often when rapid cost growth in a particular hospital

⁶⁹ Note that while provider payment reform has received comparatively little national policy attention, various local reforms have taken place. These have tended to be limited and there has been little evaluation (Eggleston, Li et al. 2008; Yip, Hsiao et al. 2010).

⁷⁰ Demand side constraints as described in Chapter Four are one example, though less prominent examples include signing of service contracts between NCMS schemes and providers, use of drug lists and development of management mechanisms (cf. Chapter Five). The 2006 interim report lists various measures, including signing of contracts with providers, suspending payments to providers, dismissing hospital staff, etc. (NCMS Pilot Evaluation Group 2006, pp. 104-106). This point is also made in the recent World Bank review of China's health reforms (Wagstaff, Lindelow et al. 2009b, p. 126). While the absence of provider payment reforms is certainly a failing, and these measures should not be thought of as a substitute for payment reform, one can suppose that without them scheme effectiveness and cost growth would almost certainly have been worse.

threatens fund security. This, he said, works well, but it clearly relies on third party intervention.⁷¹ Other means of control over providers are economic (through fines) and legal (through use of contracts) (NCMS Pilot Evaluation Group 2006, pp. 104-106).

The exact scope of NCMS Offices' legitimate activity is unclear: interviewees in Taoshan, for example, bemoaned the lack of a clear policy basis on which to fine providers, claiming they are able to reclaim misspent monies, but that they have no clear policy basis on which to impose fines in excess of this – something they thought could have a deterrent effect – and that they had been forced to adapt an existing, but separate, national policy for this purpose, though the legitimacy of this was dubious (Interview no. 23). One thing that stands out is the importance of contracts, which allow NCMS Offices to specify terms with providers and agree on related sanctions on their own authority. Even this, though, is an imperfect instrument: a hospital may choose to simply accept a fine for breaking the terms of a contract, for example, and the NCMS Office may have little recourse: how to exclude a hospital from the NCMS system, cutting it off from NCMS reimbursements, if people rely on that hospital?⁷² Meijiang was considering removing the NCMS status of a particular hospital due to persistent misbehaviour when I last visited, but this is clearly not something to be done lightly (Interview no. 45).

⁷¹ The 2006 report to the State Council describes this administrative complexity and underlines just how aware local NCMS managers are of the limitations of their legitimate authority. In a perverse twist, the report points out that if the NCMS Office were split from the health system (*weisheng xitong*) (as it is in counties where trials of NCMS management under other government bureaus are taking place, ironically in order to create a degree of separation in the local BOH – insurer – provider matrix, as discussed in the last chapter), then it would have no right to even check hospitals' medical files (NCMS Pilot Evaluation Group 2006, p. 105)!

⁷² It is likely that, in this sense, the NCMS is likely to have less clout in richer areas and where NCMS funds make up an smaller part of overall local health spending and provider revenues. A related point is made in analyses which point out that the leverage that can be exerted by the NCMS through payment reforms differs by place: where the NCMS makes up a large part of provider revenues, its leverage is likely to be large, but where other insurers (urban insurance schemes, etc) are important, its leverage is likely to be less (e.g. Langenbrunner 2010).

7.2.4 Feitian's cost control reform

Cost growth creates a problem for NCMS implementation: not only do increasing costs reduce the amount of treatment that can be reimbursed by the scheme, they also have the potential to create volatility in scheme management (and hence fund risk) and make achievement of scheme implementation targets hard. The core of Feitian's reform is very simple. From the mid-2000s, Feitian's NCMS started experiencing rapid cost growth, making it hard for the county to increase the reimbursement rate to users, to introduce outpatient pooling (which they were under pressure to do), and prejudicing fund safety. The county is one of the poorest in the province: according to county documents, 2008 per capita GDP was around 15,000 RMB and average (annual) peasant incomes were a little over 6,000 RMB. The agricultural population makes up almost 90% of the county's total population (Feitian County Bureau of Health 2007). According to one study, in 2002 (i.e. before introduction of the NCMS) 65% of households in poverty were poor because of illness, while 40% of households that had 'fallen back into poverty' had done so because of illness (Huang 2005).

The logic underlying the county's reform is simple. The BOH recognises that government funding for healthcare providers is insufficient, making providers commercially minded. One result of this, for the NCMS as an insurer, is that rising costs are driven at least in part (and possibly to a large extent) by healthcare providers' need to make money, rather than by demographic factors, release of pent up patient demand and increases in expensive and/or technical treatments. The BOH believes opportunities for hospitals to stimulate demand for services and oversell tests, treatments and drugs, are widespread. From 2007/2008, the county started signing contracts with hospitals, setting fixed budgets for NCMS spending, past which the scheme would not pay, and compensated hospitals for shortfalls, at least to a degree, outside the NCMS proper. In health insurance terms, this equates to the use of global budgets (in Chinese *zong'e fu fei* or *zong'e yu fu* systems) for both inpatient and outpatient spending; in Chinese government discourse, such methods have a clear precedent in 'contracting' (*baogan*) systems. In 2007-2008, Feitian produced a series

of policy documents codifying its reform. It reported its reform to the province and the county was subsequently invited to present its experiences at both provincial and national meetings. As discussed below, the county's reform has been quite successful in restraining cost growth, but simple design means that other aspects of its functioning may be less than ideal.

The next sections show a county-level view of operation of the pressurised system. This is the background to NCMS fund setting and the county's cost control reform presented in the subsequent two sections.

7.3 NCMS as pressurised system: The view from Feitian

In Chapter Five, I argued that the centre sets general policy principles and that sub-national governments, principally counties, are responsible for actual implementation of this policy. NCMS targets are principally set within the health system (*weisheng xitong*), and combine with the enforcement of a cellular funding structure through the finance system. I argued that this implementation structure corresponds broadly to the description in the literature of the 'pressurised system', suitably modified to take account of the importance of the functioning of the fund transfer, aggregation and management system. The result of this, I argued, is the localisation of 'fund risk' (*jijin fengxian*) with implementing counties. This vocabulary and framing of the functioning of the NCMS system is near-ubiquitous in Chinese policy discourse (policy, speeches, academic analyses, official documents, etc.). Importantly, for my analysis, this framing and vocabulary are also central to local understandings of NCMS implementation. Feitian's reform clearly shows the functioning of this system and the ways in which systemic localisation of risk is articulated at the local, management, level. In this section, I examine in some detail fund setting in Feitian as a backdrop to understanding the county's development of cost control / provider payment measures. Fund setting corresponds most closely to what one could consider county *implementation* of the NCMS – counties are required to design and modify their

reimbursement plans and to carry out budgeting within the broad parameters of the scheme.

7.3.1 Implementing the NCMS: National targets and local policy

In Chapter Five, I showed the evolution of NCMS targets over time, from a very low level of funding and a large degree of freedom in plan setting, towards increasing funding and exacting targets for scheme enrolment, reimbursement levels and increasing reimbursement ceilings. I set out central targets for scheme functioning. These targets are basically replicated in provincial documents, and are passed down to counties to be achieved. County scheme operation revolves around local NCMS reimbursement plans (*buchang fang'an*), which counties design and modify based on implementation targets. Main plan variables are funding levels, reimbursement rates, thresholds for reimbursement, payment ceilings, etc. Plans change frequently, reflecting local changes in scheme operation, funding and targets.

Actual NCMS plans – the details of scheme budgeting and deliberations over plan setting – are in most cases not public, though counties issue local policies relating to their schemes. For Feitian, these are available dating back to 2003, when the county issued Opinions on establishment of the scheme (Feitian County People's Government 2003b), interim operating measures (Feitian County People's Government 2003a) and trial regulations (Feitian County S Township People's Government 2003). Generally speaking, these documents show overall scheme parameters: the trial regulations, for example, set out who is eligible to participate in the scheme, sign-up procedures, funding levels and methods, the scope of NCMS reimbursement (things included and excluded, levels of reimbursement, payment ceilings, etc.), procedures for reimbursement as well as providers in which treatment can be sought, etc. In all, such documents give a basic outline of the scheme in the county.

When Feitian started implementing the NCMS, overall funding was 40 RMB per user. As elsewhere, users were obliged to sign up (or not) on a household basis. Initially,

Feitian's plan covered mainly inpatient treatment, though certain expensive treatments (e.g. cancers) were also included on an outpatient basis. According to the policy, reimbursement levels were graded: users faced a 1,000 RMB threshold before the NCMS started to reimburse expenses, and above this threshold, bands applied, with spending of 1,001 RMB to 2,000 RMB eligible for reimbursement at 10%, spending of 2,001 to 3,000 reimbursed at 20% and so on. The maximum reimbursement rate was 60%, for spending above 10,000 RMB and the reimbursement ceiling was set at 20,000 RMB, though policy contains a provision that those with financial difficulties may apply for extra relief.

In 2005, the county issued a revised policy. Funding levels remained the same, but the threshold for reimbursement was lowered, reimbursement bands were modified, and a number of new elements were included, including preferential reimbursement rates for treatment at THC level and below. There were greater revisions to policy in 2006: reimbursement thresholds and bands were revised, provisions were included for those who sought treatment while working outside the county, and changes were made to other management procedures (including issuing of NCMS cards and creating a database of those signed up).⁷³ In 2007, the county increased NCMS funding to 67 RMB, modified reimbursement bands and doubled the payment ceiling to 30,000 RMB. Other significant revisions were the inclusion of students in the NCMS and the introduction of limited outpatient reimbursement (e.g. Feitian County Bureau of Health 2006; Feitian County C Township 2007).

7.3.2 Implementing the NCMS: Targets, funds and plans

Setting NCMS funds principally revolves around levels of funding (*chouzi*) and reimbursement (*buchang*) or, in other words, incomings and outgoings. By 2007, we have data from the BOH on scheme functioning (Feitian County Bureau of Health

⁷³ Electronic swipe ('IC') cards allowing on-the-spot deduction of the component of spending paid by the NCMS at the point of treatment. These have progressively replaced paper booklets and *post hoc* reimbursement in the NCMS Office (and similar venues) over the development of the scheme.

2007). Scheme funding in that year was around 35 million RMB, of which approximately 86% was spent on reimbursement of all kinds; the average reimbursement rate recorded by the BOH was just under 27% and the scheme had managed to accumulate a little over 3 million RMB in surplus (*jieyu*). By 2009, the sign-up rate was almost 95% and overall contributions were 140 RMB per person. As of September of that year, around 25 million RMB had been spent on reimbursement, of which 21 million RMB was spent on inpatient reimbursement and the rest on outpatient reimbursement. The BOH estimated that surplus for the year would be 6 million RMB, allowing them to reach an accumulated surplus of 14 million RMB. By 2010, per person funding had increased again, to 170 RMB. Almost 53m RMB was raised that year, bringing total funds available to spend to 67 million RMB, including accumulated surplus funds. To September that year, almost 28 million RMB had been spent on inpatient reimbursement, and almost 6 million on outpatient reimbursement.

This process of fund setting and modification is an ongoing one. A Feitian BOH fund modification report from 2009 (Feitian County Bureau of Health 2009a), for example, starts by setting out basic operational indicators (sign-up rate, funding levels, etc.) and showing growth in scheme payouts and surplus over time. Provincial targets require that the county increase user contributions. This, along with increased government contributions, would give a markedly enlarged fund, and the reimbursement package must be redesigned to take this into account. The logic of plan setting is very clear: on the basis of the increased funding level, the BOH calculates the overall fund at more than 40 million RMB for the following year. From this, and in line with provincial policy, they earmark 10% as a contribution to the risk fund (*fengxian jijin*). Overall spending for the year is expected to slightly exceed collected funds, based on previous years' figures, meaning the BOH will have to use a little of the accumulated surplus, but that the plan is 'safe' ("*cai yong gai fang'an, jijin yunxing shi anquan de*"). The report states, though, that county government should be ready to modify the plan at short notice if need be:

"Inflation and the unpredictability of medical cost growth, the impact of the new

health reform plan, etc., could all pose a certain risk to this plan. In the case of problems due to factors we cannot control, the county government should rapidly modify the plan or provide a guarantee of funding [for NCMS reimbursement]" (Feitian County Bureau of Health 2009a).

The report ends with an interesting statement: proposed modifications to the NCMS plan will decrease the threshold for reimbursement, increase the reimbursement ceiling, maintain fund 'stability', etc., but will fall slightly short of the provincial reimbursement rate target. The BOH argues that this is due to funding constraints and already-high levels of inpatient spending.

By the following year, implementation goalposts had shifted once more: the 2010 NCMS work report shows that provincial targets required them to increase funding to 185 RMB/person (Feitian County Bureau of Health 2010c). Contracts signed with the city contained a hard target (*ying xing yaoqiu*) specifying that the county's NCMS reimbursement level reach 60% the following year (Y Province Bureau of Health 2010). According to a county assessment, this would require an increase in the funding level to 300 RMB per person, of which 100 RMB would have to come from user contributions. According to this report, this was the most immediate work priority of the BOH (along with modifying the reimbursement plan and increasing cost control work).

Targets are increasing all the time, as interviews make clear. As staff of the NCMS Office describe it, in 2003, when Feitian started running the scheme, inpatient reimbursement rates were around 20% and these were set by the county. By 2010, the target was 40%. In 2011, it was 60%, and in mid-2011 a target of 70% had just been announced. According to the NCMS Office, these are serious targets and the pressure to meet them is great. Both government contracts (contracts signed between county and city governments) and BOH contracts (signed between county and city BOHs) contain health targets. Since the start of the New Health Reform (*Xin Yigai*), though, Feitian has seen an increase in targets included in government contracts, in addition to the targets contained in BOH contracts, and this is understood to be a way to strengthen these targets: if a target is included in the government contract, it is sure to

be included in the BOH contract, though not all BOH targets will be included in the government contract. Reimbursement targets have generally been included in BOH contracts, but in 2011, for example, the 60% target was included in the government contract, and this target was then passed on to the BOH by the county government. This was seen by the BOH as non-negotiable. Starting from 2010, specific cost control targets started to be included in Feitian NCMS contracts, and in 2011, a *zero percent* NCMS cost growth target was included in the BOH contract. The county tried to negotiate with the city BOH prior to signing the contract, arguing that the target was unachievable, but they were unsuccessful: the city argued back that this was a provincial target over which they had no control: it had to be included (interview no. 56).⁷⁴

In mid-2011 the 60% reimbursement target was being discussed locally. At the time, NCMS funding had been increased to almost 300 RMB, with 15 RMB deducted for health checks (*tijian*). According to the NCMS Office, this would be enough to reach the 60% reimbursement target – if they could control costs well. If they couldn't, the county would have to use some of its accumulated fund surplus (around 20 million RMB) to make up the shortfall. Hypothetically, if they really could achieve zero cost growth, it would be possible to hit the 60% target and still accumulate some surplus. Realistically, if the BOH managed to maintain the level of cost growth of the previous year, projected 2011 spending would require they use some of the accumulated surplus to make up the shortfall, though the NCMS Office was sure they would be able to hit the 60% target. According to their calculations, if they wanted to hit the 70% target, the BOH

⁷⁴ Locally, there was discussion as to what this target really meant. According to the NCMS Office, the year before, overall cost growth had been just over 6%, compared with around 10% GDP growth for the county. If the BOH were really to achieve zero cost growth, growth in provider revenues would fall dramatically behind GDP growth. At the time of the discussion, the NCMS Office believed that the county would be allowed to interpret the zero growth target to mean that cost growth should be in line with GDP (i.e. GDP growth plus zero), though this was likely wishful thinking: Director Hu, of the city BOH, stated clearly that the zero percent growth target was just that – zero growth on the previous year's costs, though he wasn't sure how this would be achieved. Achievement of this target will be extremely difficult, and probably impossible.

would need to draw down around 1 million RMB of accumulated surplus per month to achieve this. If they decided to aim to achieve this for the last quarter of the year, for example, the 'cost', in terms of NCMS surplus, would be 3 million RMB. All in all, increasing the reimbursement level and dipping into the surplus would likely mean that Feitian would have to increase the funding level yet again the following year. As the NCMS Office understood it, the 70% target was mandatory. The timing of the target, however, might allow some room for flexibility: even if Feitian reached it in December, this would still count. Ultimately, though, the decision about hitting this target, and timing of this, would be down to the county leadership, not the BOH.

7.3.3 Interim summary

Several points should be highlighted from the preceding sections. First, NCMS policy has clearly changed a lot since the beginning of the scheme. This includes elements of scheme management, such as inclusion of specific illnesses, and so on. Most importantly, though, county-level NCMS management revolves around scheme budgeting, driven to a large extent (though obviously not entirely) by national implementation targets handed down from the centre to the province and then to the county. Targets are specified in contracts signed by both the BOH and government with higher levels. Over time, targets have increased hugely: in mid-2004, Feitian achieved a sign-up level of just over 80%; in 2010, the target was 95%. Scheme funding in Feitian has increased from 40 RMB per person in 2003 to almost 300 RMB in mid-2011. At the outset, average reimbursement levels in Feitian were around 20%; at the time of my last visit to the county, the BOH was discussing how to achieve 60-70% reimbursement rates and whether the province's zero cost growth target was achievable.

Targets are taken seriously: in 2009-2010, it seemed that Feitian would be unable to meet the provincial 35% reimbursement target (though I never found out whether the county managed to hit their target). What one should notice, though, is not that Feitian (probably) failed to meet its target, but that – on the contrary – the county took this

target seriously and tried to meet it, despite limited funds. Equally, the reimbursement targets of 60-70% and the 2011 zero growth target were seen locally as obligatory, though it will be very difficult (or impossible) to reach at least some of these targets. As above, according to the NCMS Office, Feitian had a degree of leeway in the timing of meeting the 70% target. On one hand, this could be taken as evidence of shirking, or of making full use of possible leeway. Equally, though, it is evidence that the county was taking this target seriously and working towards it, despite obvious constraints.

As an outside observer, it is impossible to understand the process of fund setting *overall*: target setting and implementation is an ongoing process in which there is no endpoint: a given target in a given year is just one in a succession of targets – increasingly demanding targets, if Feitian’s experience is anything to go by. For the county, this is a long process of management evolution, in which pressure is a constant and in which significant and difficult decisions are made all the time. From outside, it is clearly impossible to understand this process in its totality; of more interest is to understand the logic of the process – here, the logic of fund setting. In the next section, I examine Feitian’s fund setting in 2007, as a prelude to showing the development of the county’s cost control work.

7.4 Implementing the NCMS: The logic of fund setting

In this section, I examine Feitian’s fund modification process in 2007. Funding, and fund pressure, combined with rapidly increasing costs underlie Feitian’s starting to carry out cost control work. BOH reports and interviews carried out locally show how seriously fund setting is taken: targets from outside (and, to a degree, local pressures) are clearly felt by, and acted upon by, the BOH. The importance of fund modification in 2007 (as opposed to any other year) is that this is the background to development of cost control and provider payment mechanisms by the county, starting that year.

7.4.1 Feitian: The 2007 report

In mid-2007, the BOH submitted a report to the county government making the case for modification of the NCMS fund (Feitian County Bureau of Health 2007). The argument for a change to the reimbursement plan is based on two main things. First, despite continual growth in use of the NCMS, average reimbursement rates remain stagnant at twenty-plus percent. The report argues that user acceptance and support are key to sustained scheme development, and says that for several years the local People's Congress and People's Consultative Committee have called for the introduction of outpatient pooling (*menzhen tongchou*) in the NCMS reimbursement plan, but that this has not been possible to date due to limited funds and limited management capacity (mainly lack of a computerised oversight/management system allowing direct oversight of THCs), meaning that only a very limited number of outpatient items have been included in NCMS reimbursement. Second, with an increase in NCMS funds in 2007, the county may have enough funds to start outpatient pooling on a limited basis and this coincides with provincial targets for outpatient spending (Y Province Government Office 2007). It is also argued that including outpatient reimbursement should increase attractiveness of the scheme to users and promote treatment seeking at the THC level for minor ailments.

The BOH report proposes three possible outpatient reimbursement scenarios. In all three, outpatient treatment in THCs is to be covered, with no user co-payment and debate centres on reimbursement levels and ceilings for reimbursement. Based on a baseline survey and estimated cost growth rates, the report settles on a plan setting a 20% reimbursement level and a 500 RMB ceiling. For the county, this is the most expensive plan; it is thought that other plans, while cheaper, could run up too great a surplus, and it would be hard to set a lower reimbursement rate, the report argues, as neighbouring counties have reimbursement rates of 20%. The proposed plan, while expensive, does not create a large fund risk ("*jijin zhichu suiran [...] gao [...], dan fengxian bu da*") and would have the advantage of being in line with neighbouring

counties (euphemistically phrased as “reducing work pressure”, “*gongzuo yali qing*”).⁷⁵ The report recommends township-level pilots (*shidian*) prior to full-county adoption. The catch, however, is that introducing outpatient pooling, combined with huge growth in use over the previous two years (almost 40% year-on-year person/time growth), will be expensive for the NCMS overall and will require modification of the inpatient reimbursement plan if the scheme is not to go into the red within the current year (“*cunzai dang nian chizi fengxian*”). Once again, the BOH gives alternative proposals, using alternative reimbursement thresholds, ceilings, etc.

A number of points emerge from the report. First, in calculations underlying different scenarios, clear prominence is given to numbers of users who would benefit from the modification (*shou yi ren shu*). Second, adjustment of the scheme (reimbursement bands, thresholds, etc.) is carried out with an awareness of the likely reaction of scheme users and township cadres: it is believed that one plan, with a high reimbursement level in one reimbursement band, would be popular with many (and could help with publicising the scheme). Third, while increasing reimbursement levels will be popular with users, a degree of continuity is seen as being easier for users and township cadres to understand, and it is felt that reducing reimbursement levels within certain bands could be unpopular with certain users, and could cause misunderstandings. Fourth, and relatedly, maintaining the status quo would be popular with users and lower-level cadres, but would bankrupt the scheme by 2010. Fifth, fund

⁷⁵ Fieldwork in Meijiang revealed similar local comparisons in setting of NCMS indicators: one NCMS manager there argued that they must take neighbouring counties into account when designing their scheme – if their reimbursement levels, for example, were significantly either lower *or higher* than other roughly comparable counties, this would cause questions to be asked. Similarly, Meijiang’s scheme funding has consistently exceeded average provincial levels. In principle at least, this should cause difficulties in target setting: against what can the county’s performance be evaluated? Director Gao, the head of the NCMS in Meijiang, stated that he benchmarks performance of the NCMS against both performance in previous years and against performance of the city’s urban insurance scheme (Interview no. 38). Both of these examples seem to reveal a degree of comparative assessment falling largely outside the framework of target-driven implementation presented in the majority of the literature on implementation in China (cf. Edin 2003; O’Brien and Li 1999).

security is a major feature of all analyses: two plans, for example, could have an influence on fund security if current levels of growth in particular reimbursement bands continue.

The conclusion, from the alternatives presented, is that the county must raise the level of NCMS funding or encounter fund risk under any feasible plan by 2008. Simply persevering with the current reimbursement plan would require the county BOF to increase funding to the NCMS by 8% annually in order to maintain fund security (*quebao jijin yunxing anquan*). This need to increase funding coincides with a provincial requirement that counties increase funding levels.⁷⁶ A funding increase would also allow the county to build up the NCMS surplus and buffer against overspending. Aside from NCMS budgeting, the report requests funding for a networked reimbursement system in THCs. This is not only a necessary precondition for carrying out outpatient pooling, but could help reduce fund risk arising from outpatient pooling/reimbursement. 'Informatisation' (*xinxihua*) is also a provincial priority.

7.4.2 Feitian: Cost growth and the logic of cost control

In Feitian, the drive to control NCMS costs was directly rooted in a perception of risk in implementing the NCMS, combined with increasing implementation targets. Inpatient spending is the most important component of NCMS spending, in Feitian making up 75% of the total. In the years running up to the development of CCG, overall NCMS inpatient spending increased a lot, by almost 15% in 2005 and by more than 40% in 2006 and total NCMS inpatient reimbursement necessarily also increased dramatically, by around 25% in 2005 and almost 40% in 2006. Average (per patient/time) inpatient spending also increased substantially, by 9.5% in 2005 and by 12.5% in 2006 (Feitian County People's Government 2010). Qiu, the head of Feitian BOH, in a report on cost

⁷⁶ While the province required that counties increase funding, the source of that funding was left to counties' discretion. The BOH opted to pass on the entire funding cost increase to users. The rationale was that, while this would mean users paying a greater proportion than before, an assessment of users who had dropped out of the scheme had found that the main reasons for this were not economic.

control work, attributes these rapid increases to a number of factors: an increase in demand due to rising living standards and changing behaviour; demographic factors (including ageing and increase in chronic disease); increasing use of new equipment; the NCMS stimulating demand among users and – something that is repeatedly stressed in BOH analyses – under conditions of insufficient government investment, that providers resort to selling unnecessary tests, using expensive equipment, etc., in order to make money (Feitian County Bureau of Health 2010). This dramatic increase in costs, both total and average, combined with only a limited increase in NCMS funding, resulted in the overall NCMS reimbursement rate perpetually hovering at around 25%. 2006 was the year of greatest cost growth, apparently due to an increase in treatment seeking as users became used to the scheme and started to get treated for problems they would previously not have sought treatment for (Interview no. 56), though neither Qiu nor the NCMS Office could give a definitive reason for this growth.

These explanations are likely true in part: as above, rapid growth can be seen elsewhere. In Meijiang, for example, the switch from use of family accounts to outpatient pooling (see Chapter Five) provoked very rapid outpatient cost growth, risking overspending.⁷⁷ Other factors can also lead to rapid and unpredictable cost growth, risking destabilising the NCMS fund, not least misbehaviour by providers: in one county where I carried out fieldwork, one THC director, Jin, described to me how one village doctor in his township had used up the entirety of funds allocated to the township in 2007. When representatives of the county went to investigate, they found that the village doctor had been falsifying receipts in collaboration with village cadres. Jin described this as ‘terrifying’ (*hen kongbu*) (Interview no. 70).

⁷⁷ The county initially set a conservative outpatient reimbursement rate, but found that allocated funds weren’t being used, resulting in a large surplus. The following year, the county increased the reimbursement rate and the ceiling twice to try to stimulate use, leading to a rapid and unsustainable increase. The NCMS Office cracked down through increased oversight of outpatient spending, followed by a reduction in reimbursement rates. According to scheme managers, the problem was lack of an evidence base and insufficient understanding of users’ ‘consumption habits’ (*xiaofei xiguan*), making it hard to predict how the county’s policy modification would function.

There is a dual logic in carrying out cost control: increasing the benefit to NCMS users, and reducing fund risk. This is clearly visible in discussions with Feitian BOH and in BOH reports and materials. As a report by Qiu states, “The final aim of CCG is to protect the interests of the masses and to reduce NCMS fund risk”. A presentation used by the BOH at meetings showcasing the county’s initiative makes this point even more clearly: the basic motivations (*jiben dongyin*) of cost control work are that increasing costs “increase patients’ burden and increase NCMS funding pressure”; if cost growth is not reined in, it will “have a direct impact on the benefit users derive from the scheme and pose quite a large threat to NCMS fund safety”. This argument is common to materials produced by Feitian describing their reform.

Interviews confirm this analysis: according to Qiu, CCG has two aims – it is good for NCMS users, but as much as this, it is about controlling cost growth, which risks eroding the NCMS fund. As Qiu phrases it, medical spending has the potential to be a bottomless pit for any country (!) or government, including Feitian. The NCMS Office give a much more concrete view of the importance of cost growth for Feitian: there are two reasons for carrying out CCG: increasing benefit to users and increasing fund security (*jijin anquan*). In fact, increasing user benefit is only possible if you can control costs. The core of the equation, according to the NCMS Office, is that in places like Feitian, with limited amounts of funding, it is hard to increase the reimbursement rate due to lack of funds. Increasing the rate requires either increasing the amounts of funding in the scheme (principally through increasing local government contributions, given the structure of NCMS funding) or controlling costs. CCG reduces risk by reducing costs, allowing the BOH to increase the surplus, which functions as a buffer against overspending.⁷⁸ According to the NCMS Office, risk in implementing the

⁷⁸ Accumulation of a surplus (*jiayu*) is one of the main ways for the scheme to hedge against fund risk. As in Chapter Four, both surpluses and risk funds are specified in national NCMS policy. From the county point of view – in Feitian, at least – there is a difference in functioning, however: the surplus is usable on the authority of the BOH, whereas the risk fund requires that the BOH write a report to the BOF, and use of this must be reported to the next level up. Feitian reports show the county has consistently built up its surplus over time (to approximately 20 million RMB by 2010), and this is consistent with the picture of risk-averse counties presented here. An

scheme is in overspending or running out of money. If the scheme has money, then there is no risk. If the alternative to overspending is running a surplus, this is preferable as the NCMS can always carry out Second Round Reimbursement (see Chapter Eight). At the time of interview, Feitian's NCMS fund was around 80 million RMB, but this had been much lower in 2005-2006, increasing the risk to the overall fund of growth and/or fluctuations in spending and making it hard to increase reimbursement levels. As required reimbursement levels increase, however, the pressure on funds, and fund risk, increases despite the overall increase in funds (Interview no. 57).

7.4.3 Other views of fund risk

Above, I have argued that Feitian's main framing of NCMS management is in terms of fund risk, and that this risk framing strongly informs ways the county thinks about fund setting and cost growth. This framing should not be thought of as confined to Feitian. Interviews with the head of the NCMS in Meijiang revealed a similar framing: as Director Gao explained it, from 2005 (when scheme funding started to increase) he started to think of the goals of the NCMS in two ways. The basic goal was fund security; the higher level goal was to increase the effectiveness of NCMS spending (*jijin xiaolv*) and increase benefit to users. For Gao, the goals are separate: the aim is to increase the benefit of the scheme, but the over-riding consideration is to maintain fund security. (Fascinatingly, as Gao explained it, this thinking underlying operation of the scheme is his own rationalisation of scheme management, though it is nowhere set out in policy.) I asked Gao what would happen if the scheme overspent or went bust. His reply was that the government would not allow the scheme to go bust, but that in a place like Meijiang the risk was in overspending. If the scheme overspent, the government would want a good explanation as to why this had occurred and would, at a minimum, punish (*chuli*) him and the heads of large public hospitals. Interviews in Taoshan also described the risk inherent in fund setting. One NCMS manager described the NCMS plan as "quite hard to calculate": there are a lot of factors to take into consideration and

interesting parallel in existing literature is Li's discussion of 'autonomy-preserving behaviour' of local governments (Li 2005).

a small mistake can lead to problems. As he described it, you can't "blindly increase the reimbursement rate" (Interview no. 32).

7.5 Feitian: Cost control and the functioning of 'Controlling Cost Growth'

In 2007, Feitian started cost control reforms, the main component of which was setting budgets for treatment in major hospitals and THCs. As above, interlinked concerns of increasing costs, fund security and stagnating scheme reimbursement levels, combined with the need for large changes to scheme operation to accommodate outpatient pooling, all played a role in increasing the county's understanding of the need to control costs. This is a direct expression of the pressurised system: increasing implementation targets combine with the localisation of fund risk at the county level to create this conjuncture. The county's approach to cost control, later re-interpreted as provider payment reform, is a direct extension of the logic of the pressurised system. In this section, I examine the system Feitian put in place.

7.5.1 Understanding Feitian's cost control reform

Feitian's cost control reform is first discussed in the 2007 report cited above, in which it receives a brief mention, and the main policy document setting out CCG appeared in 2008, though the county had in fact started cost control work in 2007. The 2008 document shows CCG as part and parcel of county efforts to improve management and spending of NCMS funds: overall, the policy includes a number of elements alongside CCG, including more rigorous checking of patients' identities and NCMS credentials, more rigorous oversight of NCMS funds, and increased efforts to steer patients toward treatment at the THC level. CCG work is explicitly subsumed under a need to strengthen fund management (*jiaqiang jijin guanli*) (Feitian County Bureau of Health and Bureau of Finance 2008).

The backbone of Feitian's reform is the signing of NCMS management contracts with providers and the attached audit (*kaohé*) framework. Main elements of the 2010 contract cover providers' basic NCMS management/implementation, cost control and

provisions on BOH reimbursement of providers. Basic management revolves around putting in place management and hospital information systems, publicising the NCMS, and correctly carrying out reimbursement, reporting on cases and spending, preventing fraud and similar. Adherence to CCG carries most weight in audit of hospitals' implementation of the NCMS.

Main components of CCG are the setting of budgets for hospitals' NCMS-related spending. These include a number of elements, though the most important (due to its weight in overall spending) are targets for inpatient costs. Targets for both inpatient and outpatient spending, both overall spending budgets and growth in average (per case) costs, are written in to hospitals' management contracts. Overall budgets are calculated on the basis of previous years' budgets and allow a defined percentage growth in average costs per case. Main measures are as follows:

- Outpatient costs: initially, hospitals were given targets for average outpatient costs based on costs of outpatient treatment of non-NCMS members: NCMS outpatient costs were not to exceed non-NCMS outpatient costs by more than a fixed percentage.⁷⁹ Subsequently, a target for the rate of growth of average outpatient costs was also set, at around eight percent (set to be roughly in line with GDP growth). In 2011, contracts signed with hospitals required zero percent growth in average costs.
- Inpatient costs: initial inpatient cost growth targets were set based on data from the preceding three years, allowing a maximum average cost growth of around eight percent and setting maximum budgets for each hospital on this basis. Subsequently, year-on-year targets were set based on figures from the previous year and were reduced to six percent year-on-year growth in average costs and then, in 2011 and in line with the zero cost growth target, at zero.

⁷⁹ In some ways, this is a very clever mechanism. It is recognised that the NCMS, in weakening the incentive for patients to control costs, has weakened cost control overall. Feitian's solution was to benchmark NCMS outpatient costs against spending in which patients *do* have an incentive to control costs (though there is no way of knowing whether this has had an impact on non-NCMS outpatient costs).

- Total budgets are also calculated for each hospital, based on historical spending and an allowed growth rate (around eight percent in early stages). This is, in health insurance terms, a 'global budget': if providers' spending is less than budgeted, they may retain the difference, but they must also bear any overspend.
- Various other measures were included in CCG: fixed payments for certain disease-specific payments (mainly hospital-based deliveries, though the number of conditions covered has been expanded subsequently) and capped use of frequently-prescribed antibiotics. The NCMS Office also sets maximum 'self pay' rates on a hospital-by-hospital basis, capping the proportion of user spending that can fall outside the scope of NCMS-reimbursable spending.

The aim of these measures, overall, is to control growth in *average* costs. The use of a global budget is intended as an incentive to increase efficiency. In an attempt to prevent refusal of patients and shifting of spending from the NCMS to users' private spending, the NCMS Office collects figures on numbers of cases treated and sets targets for users' self-pay rates. Hospital contracts specify the right of the BOH to oversee and audit (*kaohu*) hospitals' NCMS expenditure and to retain a proportion of hospitals' NCMS reimbursement if incorrect practices are discovered. This is possible as NCMS monies are reimbursed to the hospitals every quarter for the previous three months and, as above, greater oversight of provider-level data was achieved with the introduction of a new information system around the same time as CCG. All expenses (cases) exceeding 10,000 RMB are to be checked specifically by the BOH. Hospitals are also to be held responsible for illegitimate expenditure due to patients abusing the scheme (using a false ID to claim reimbursement, for example) and hospitals' own mismanagement and/or incorrect behaviour. For example, if hospitals are discovered attempting to have expenses reimbursed which do not qualify for NCMS reimbursement, they can be fined multiples of the amounts concerned. The BOH is to randomly check 10% of all inpatient and outpatient cases every month as a control (Feitian County Bureau of Health 2010a).

Contracts are backed up by audit (*kaohu*). Weighting of components in this shows the priority attached to different elements of hospitals' NCMS management. By far the most important (weighted more than 50%) is CCG. Within this, control of inpatient costs receives the greatest weighting, and marks are deducted for exceeding the average spending target and for exceeding the overall budget. Equivalent targets apply to outpatient reimbursement. Other content covers more general NCMS management.

CCG has changed over time. As well as changes in the rates of cost growth allowed, management has been strengthened. Hospital contracts, for example, were revised in 2011 (Feitian County Bureau of Health 2011). Main differences are changes in targets (zero cost growth) and in a strengthening of penalties. This contract, for example, specifies the right of the BOH to remove a hospital's designated provider status (i.e. exclude providers from NCMS reimbursement) for up to six months as a penalty for submitting false claims, where the 2010 contract specified only a fine. The 2011 contract also includes specific provisions in an attempt to prevent hospitals discharging and re-admitting patients in order to reduce, on paper, costs per case.

7.5.2 The logic of Feitian's cost control reform

The core of Feitian's reform is in the setting of maximum NCMS spending targets: If hospitals exceed their targets, they must – in principle, at least – pay the excess spending themselves; if they can come in below budget, according to the terms of CCG, they may retain the difference between actual and budgeted spending. The logic of this is simple. As above, hospitals receive only a small proportion of their operating costs from government and must generate revenue through selling services. For the NCMS, though, this poses a problem: hospitals' need to generate revenues translates into spiraling reimbursement costs. This is the problem CCG attempts to solve. Qiu explains the logic of CCG as follows: under FFS, due to low profit margins in hospitals and cost sharing between the patient and the insurer, for every increment in hospitals' revenues, the NCMS, as insurer, together with the patient, must collectively pay a multiple of this increment. In other words, even small increases in hospital revenues

must be paid for in high multiples by the NCMS and users. This is a structural problem, due to hospital margins and the structure of NCMS reimbursement and corresponds clearly to the leveraging effect described above (Yip and Hsiao 2008). What is more, over time, absolute figures involved increase (for the NCMS), as schemes increase their reimbursement rates. Qiu believes that the NCMS has caused an increase in illegitimate spending: before there was insurance, there was little room for doctors to overcharge as the people they were treating had little money; this has now changed with the introduction of a third party insurer.

Feitian's solution to this problem is to set budgets for NCMS reimbursement and to simultaneously top up hospitals' 'legitimate' revenues outside the NCMS reimbursement channel through use of a separate adjustment fund, rather than forcing all hospital revenue generation through the multiplier of the NCMS. This is the logic, and deal, underlying CCG. County analyses present this as a good outcome for all three parties concerned (users, providers and the NCMS). Feitian's reform contains two quite radical adjustments to NCMS functioning. First, hospitals which manage to make savings within the target budget set by the BOH are allowed to keep these savings. Second, the BOH retains an adjustment fund of 5% of NCMS funds which it can use to compensate hospitals for policy change, changes in disease incidence and similar. Probably most importantly, the adjustment fund allows the NCMS Office to reward hospitals for 'complying with' (*peithe*) policy. The clearest statement of this is given in a 2010 report on CCG, where Qiu writes that, while superficially the county's cost control reform appears to disadvantage hospitals, their interests are not in fact damaged as the county allows them 10% yearly growth in spending and the adjustment fund can be used to increase hospitals' overall NCMS reimbursement budget if patient numbers and spending have increased significantly as a result of addition of new treatments (*xiangmu*), or an increase in the level of services (*fuwu nengli*), as long as average costs are controlled and the hospital's services are in order (Feitian County Bureau of Health 2010b).⁸⁰ In doing this, the clear aim of the NCMS

⁸⁰ A similar provision is the use of a self-pay rate for hospitals. Hospitals' are set a

Office is to push hospitals to control average costs and to compete on volume of services and treatment delivered for a share of the residual – the adjustment fund.⁸¹

Most fundamentally, what one should notice is that CCG makes good a deficiency in the overall design of the NCMS: it allows a degree of fund retention by providers and employs an adjustment fund to route funds to providers outside the NCMS payment-for-service multiplier. *Prima facie*, the legitimacy of these measures seems questionable. In terms of national NCMS policy, both measures seem to contradict the stipulation that NCMS funds be earmarked exclusively for spending on treatments and check ups. Qianjiang, which has also carried out provider payment reforms, sought permission from its parent municipality (Chongqing) in order to operate a similar global budget system at the village level and allow clinics to retain funds if they could come in under budget (interviews no. 74). In Feitian, surprisingly, the decision to allocate a proportion of NCMS monies to an adjustment fund was made at the county level by the Bureau of Health, the Bureau of Finance and the local Audit Office (*Shenji Ju*). According to the NCMS Office, hospitals find it hard to achieve the targets, let alone make savings, and this appears to be borne out to a degree by available hospital-level data (see below). One should also notice, though, that the county is not shy about this: the report prepared for the MOH specifies clearly how this functions and the rationale

percentage target for spending outside the scope of NCMS reimbursement. In BOH reports, this is phrased as “preventing the legitimate interests of users from being infringed”, the subtext of which is that this allows patients to get treatments, drugs, etc, that they could not get through the NCMS. From a hospital’s point of view, however, a higher target is clearly beneficial, allowing use of higher tech services and more expensive drugs. In both cases, one should notice that these are variables under control of the NCMS Office / BOH and function as levers to exact policy compliance (Feitian County People’s Government ND).

⁸¹ The terms of this fund retention have changed over time. Where the 2010 contract specified that any surplus can be retained as operating income (*yewu shouru*) by providers and that overspending must, ‘in principle’, be borne by providers (Feitian County Bureau of Health 2010a), the 2011 contract provides a clearer formula for dealing with hospitals’ overspending very clearly linked to results of audit. This appears both more equitable (penalties for overspending are reduced somewhat) and to more clearly link retention and overall performance: audit scores of less than 90% make hospitals ineligible to retain any fund and make the hospital liable for a large share of any overspend (Feitian County Bureau of Health 2011).

behind this.

7.5.3 Functioning of CCG

The main component of NCMS spending in Feitian is inpatient reimbursement and measures to control inpatient spending have the greatest weight in overall cost control and audit (*kaohé*) of providers. The NCMS Office sets budgets for individual hospitals based on analysis of previous spending and the self-pay percentage allowed. Initially, this was based on data from the previous three years; subsequently, increases have been calculated on a year-on-year basis.⁸² In 2008, the county's calculations allowed for around 10% growth in overall spending on the average of the previous three years; from 2009, annual cost growth targets were reduced to around 8%. The clearest picture of functioning of CCG is given in a 2009 report, which gives 2008 cost control targets and results for a sample of seven hospitals and THCs.⁸³

In the case of Feitian People's Hospital, available data appear to show CCG functioning relatively effectively. In 2008, a 10% target for growth in total hospital spending was set, alongside a 12% target for growth in NCMS inpatient reimbursement (Feitian County Bureau of Health 2009b). Actual NCMS spending exceeded targeted spending by around 100,000 RMB, and the hospital applied for funds from the adjustment fund to make up this shortfall. The BOH reimbursed 94% of the sum applied for, and a note states that the hospital had added a new MRI scanner. Based on 2008 figures, a year-on-year growth rate of 8.5% was set for 2009. From

⁸² Given that the reform was started at a time of rapid cost growth – presumably some legitimate and some illegitimate – this approach would seem to lock in this cost growth. Surprisingly, this kind of approach (and the use of a three year baseline period) seems to have been used in many places for similar purposes (e.g. Yip, Hsiao et al. 2010).

⁸³ As stated elsewhere, this research set out to examine local innovation processes. A fully-fledged impact analysis would have had to be carried out very differently and sufficient access would have been impossible. From reports and data provided locally, however, and from interviews with the BOH and with hospitals, it is possible to draw tentative conclusions on the impact of CCG. For the most part, data presented here are derived from BOH materials from 2009, so relate to the period 2007-2009.

available data, this case appears to be broadly representative.

Overall, available figures show a number of points. Hospitals' spending was *broadly* in line with BOH targets and results of BOH evaluations were positive (all hospitals evaluated in the report for the MOH got scores of 90-94%). The adjustment fund seems to have been used selectively: of the seven hospitals for which figures are included in the report, three were compensated for policy changes and/or additions of new equipment. In all cases, compensation was less than that requested (and in one case, no compensation was given, despite being requested). Some hospitals overspent and were not compensated for this; others under-spent and some were permitted to retain this surplus. One hospital, which saw a dramatic reduction in spending due to a policy change had its funding reduced commensurately. *De facto* reimbursement rates (*shiji buchang bi*) range from 25 to 29% (2008 figures), excluding one poor performer languishing at 19%.

According to BOH materials, in its first year of operation, CCG reduced the rate of inpatient cost growth by around 30% (though this would still leave a high rate of growth overall). In 2008, average inpatient spending decreased by 4.5% allowing an increase in the overall reimbursement rate to 30%. Different hospitals fared differently: average (per person/time) inpatient spending in Feitian People's Hospital decreased by around 5.5%, for example, whereas in another large county hospital it increased by around 6%. At the THC level, average costs appear to have fallen by between around 12% and 21% year-on-year. The average inpatient reimbursement rate for January to November 2009 was 32% and the county was hoping to reach 40% in 2010 (data not available). As Qiu stresses, all parties (users, the NCMS *and* providers) must benefit from any reform. BOH reports claim that between 2007 and 2008, hospital revenues increased almost 10% on average and revenue derived from drugs decreased by 5% over the same period.

As above, it is impossible to carry out a full analysis of the impact of cost control due to limited data. What figures presented here indicate, however, is that a real effort was underway to control costs, with what appears to be a reasonable degree of success,

and that this has, according to BOH data and analyses, translated into both increased reimbursement levels and hospital revenues.

7.5.4 Functioning of CCG: Feitian Chinese Medical Hospital

Interviews with Feitian Chinese Medical Hospital show an ambiguous picture. The hospital produced two policy documents in 2010 governing NCMS implementation and cost control. (These are interesting in their own right: they are internal to the hospital and its management, but embody and extend the management logic of CCG policy.) Hospital policy sets maximum average costs for NCMS patients. Targets were in fact set almost 20% lower than the targets received from the BOH. In interviews, this was described as an attempt to prevent over-spending, which is seen as a large risk in terms of the hospital's overall budget and its ability to pay staff salaries. As one hospital Vice-Director, Director Tan, explained this, an over-spend of, say, 200 RMB per patient for a thousand patients would translate into a 200,000 RMB bill that the NCMS would not reimburse. According to him, the hospital simply doesn't have that much profit and this figure is equivalent to several people's salaries.

Hospital-level targets are passed on to departments within the hospital, and department heads are ultimately responsible for cost control. According to Tan, the hospital fines departments 1,000 RMB per 1% they overspend, and with average salaries around 6,000 RMB for a department head, this is a significant amount of money. Department heads must then decide how to apportion fines among staff in their departments. As Director Tan phrased it, the aim of this is to increase pressure (*yali*) on staff, and all department heads have been fined. Department heads I interviewed said they were broadly supportive of CCG: although this has made their jobs harder, as they must take more factors into consideration in their work and are less free to prescribe as they would like, patient numbers have increased a lot, and they report an increase in the willingness of THCs to refer patients to the hospital, which they attribute to lower costs. One department head said he thought the policy would limit doctors' freedom to carry out tests and prescribe freely, but that this should push doctors to increase their

level of expertise. His rationale is interesting: carrying out multiple tests (or over-prescribing), as a doctor, is a conservative strategy and a way to limit personal risk – if you have done all possible tests (prescribed all possible drugs), you can't be blamed for not having been thorough. Under the new policy, though, doctors have to be more selective in testing and prescribing and exclude non-essential items and this requires more expertise and judgement. This department head thought that overall this was a good thing: it would be good for good doctors, but bad for poorer ones.

There is a negative side to this story, however. As above, in 2010, Feitian Chinese Medical Hospital issued two internal documents on implementing cost control within a period of several months (Feitian County Chinese Medical Hospital 2010a; Feitian County Chinese Medical Hospital 2010b). The second policy was an attempt to strengthen implementation: in the first half of the year, growth in average costs had been more than 30%. In mid-2010, the BOH intervened and clamped down on cost growth, with the result that the hospital brought overall cost growth for the year down to around 6-7%. Two things stand out from this: first, the existence of hospital policy in and of itself may be insufficient to prompt actual change without strict enforcement; second, while interviewees denied that CCG had resulted in patients being refused treatment/reimbursement, it is hard to see how, at least in this case, such a dramatic reduction in costs could have been achieved while ensuring users' access, though according to the NCMS Office, the main reason for cost growth had been the opening of a new ward in the hospital, meaning that the spending forecast based on 2009 data was no longer reasonable, and the adjustment fund had been used to compensate the hospital. It is also likely that the Chinese Medical Hospital is not unique: this is one of the larger hospitals in the county and, one presumes, one of the better managed. According to BOH data, in previous years, cost control had been quite successful in this hospital, at least compared to the township level.

7.5.5 Interim summary: CCG

As above, in health insurance terms, Feitian's reform revolves around use of a 'global budget' (though I argue below that it closely resembles the pressurised system from which it developed). The aim of this is to encourage providers to reduce costs of treatment and, it is hoped, increase efficiency within an overall budget. Without careful management, this can have negative consequences: providers may refuse patients or refuse care in order to save (or make) money. Feitian attempts to control this through the use of maximum 'self-pay' levels for each hospital, as above, in an attempt to stop hospitals pushing patients to use more treatments outside the scope of NCMS reimbursement than previously, and punishments for violating this are included in audit. In addition, the BOH collects data on patient/time numbers in hospitals in an attempt to make sure hospitals do not refuse patients. It is unclear exactly how effective these measures have been, but available evidence is broadly positive: a 2010 report shows large increases in inpatient volumes (person/time) from 2007 to 2008, though the NCMS Office deny having come across cases of patients being refused treatment and management and department heads of the Chinese Medical Hospital also denied that this could take place and stated that treatment volumes (and workload) have risen, not fallen.

More far-reaching impacts of Feitian's reform, such as how it has affected actual treatments delivered and the relative effect of the reform on different hospitals, for example, are unclear, given limitations of data and fieldwork. Indeed, the simplicity of this reform means that it does not systematically take into account variations between hospitals, the quality of services they offer, disease incidence, the constitution of their catchment areas, transport conditions and the like. Feitian's reform also fails to specify what services should be provided: this is the space providers are expected to exploit to control costs and actual strategies used by providers are not specified by the NCMS Office. The consequences of this are far beyond the reach of this study.

Set against best practice solutions discussed by the World Bank and others (Wagstaff, Lindelow et al. 2009b; Meng, Yip et al. 2010), CCG is clearly a simple, pragmatic, and

less than optimal solution to a complex problem. Even compared to the range of practices in developing and managing global budgets, CCG appears simple (Dredge 2009). More concretely, though, Feitian's reform was conceived of as a cost control reform and only later specifically framed as a provider payment reform, and while elements of it are likely not optimal, seen through the lens of provider payment reform, in terms of its original purpose, it is a low tech response to a very specific conjuncture: a specific design flaw in the NCMS (exclusive use of funds for reimbursement of user costs) and rising costs closely linked to under-investment in providers and skewed pricing policy.

7.6 Feitian's reform and policy change: Risk, cost control and payment reform

Feitian's reform was locally initiated, but is linked in multiple ways with policy at different levels and in different places in the Chinese policy system.

7.6.1 Risk and the logic of provider payment reform

Feitian's drive to control costs is ambiguous: it was both an attempt to reduce the county's fund risk *and* to increase reimbursement levels and benefit to users. Both motivations are apparent in local documents and in interviews with the BOH. Both are, to a large extent, a product of the pressurised system of ever-increasing targets and localised financial/implementation risk. The approach adopted by Feitian is a direct extension of this system. This is particularly visible in the congruence of various measures in managing fund risk all dedicated to the same end (cost control, retention of surpluses, buffering, etc.).

As above, use of global budgets is the backbone of Feitian's cost control programme. Global budgets (in Chinese *zong'e fu fei* or *zong'e yu fu* systems) are conceptually simple and have a clear precedent in contracting (*baogan*) systems in the Chinese management repertoire. (The simplicity of this approach is a major reason for this choice in Feitian; I discuss this below.) This is a direct extension of the logic of the

pressurised system across the NCMS-provider interface, and functions by transferring risk in the operation of the scheme from the NCMS Office and BOH to providers. It does this by enforcing a (fairly) hard budget constraint on providers' NCMS spending, while leaving the means of attaining cost reductions to the providers themselves.⁸⁴ The logic of this was expressed very clearly to me by one interviewee, who described this as being based on 'results orientation' (*jieguo daoxiang*): the aim should be to set hospitals' budgets in such a way that providers maximise their own interests through controlling costs. Under this logic, the county NCMS Office is not required to directly oversee *how* hospitals control costs, so long as costs are controlled and key indicators are met. This is expressed extremely clearly in contracts signed between the NCMS Office and providers and in audit (*kaoh*) forms used by the NCMS Office: in the section on audit of outpatient costs, for example, the principle (*yuanze*) of cost control is described as being set by the NCMS, but specific practices (*juti cuoshi*) are the responsibility of providers (Feitian County Bureau of Health 2008). This logic of results orientation mirrors the logic of counties' achievement of targets handed down in policy. This framing of provider payment reform is not unique to Feitian. Use of global budgets entails, from a county point of view, a very different distribution of risk to use of FFS reimbursement. As above, under FFS, the NCMS has an obligation to reimburse legitimate treatment costs presented by providers. This was expressed very clearly in Taoshan in discussions of provider payment reform: at the time of fieldwork (late 2010), with most payment being through post-treatment FFS, Taoshan scheme managers described risk in implementation of the scheme as lying disproportionately with the NCMS Office. According to Director Ma of the BOH, under FFS, as long as costs

⁸⁴ BOH materials show selective use of the adjustment fund to compensate hospitals for various reasons and without more complete data, it is impossible hard to know how hard a constraint this really is. Probably most important is the fact that the adjustment fund is under the direct administration of the NCMS Office / BOH, which has a degree of discretion in its use. A related concern is that one rationale of the adjustment fund is to compensate providers for changes in treatment volumes, costs, etc, brought about by improvement in facilities. This seemingly creates an incentive for providers to continue investing in high tech equipment – as above, one of the frequently acknowledged causes of growth in medical spending in China.

passed by providers to the NCMS Office for reimbursement fall within the scope of permissible procedures, drugs, etc., the NCMS Office has an obligation to reimburse these. The problem comes, according to Ma, from the fact that providers can stimulate demand and that scheme users have a certain ability to pay for treatment, and that this could lead to increasing payouts for the scheme (a direct mirror of Qiu's judgement, above). Under FFS, fund risk rests with the NCMS Office, and not with providers. Payment reforms, such as contracting and prepayment of outpatient services (*menzhen zong'e yu fu*), which Taoshan were considering at the time I last visited the county, are accordingly seen as not just a method of controlling costs, but also of shifting responsibility and risk to providers, or at least rebalancing risk between the two parties.

Seen from the other side, the case of Feitian Chinese Medical Hospital cited above very clearly shows the extension of a government discourse of risk into internal hospital management and, seemingly, a considerable degree of embedding of this rationality, not just at the management level, but also at the level of department heads and below. This framing in terms of responsibility is mirrored in terms used to describe providers' behaviour: Feitian's hope, as expressed in reports on their reform and in interviews, is that use of global budgets will change provider behaviour from 'passive' (*beidong*) to 'active' (*zhudong*) in controlling costs (Feitian County People's Government ND). This framing was also evident in other counties where I discussed this with NCMS managers: Director Bo of the NCMS Office in a provider payment reform pilot county in a western province described that county's replacement of FFS with a prepayment system as a shift from detailed management in which the NCMS Office had to oversee all receipts presented for reimbursement to more 'macro' management (*hongguan de guanli*). Bo described that county's use of a global budget as a *baogan* (as above: contracting) system, under which the NCMS Office calculated budgets and providers were responsible for "actively controlling fund [use] themselves" (Interview no. 49).

7.6.2 From cost control to provider payment reform

Feitian's reform was a local initiative. The county is not wealthy and this clearly contributed to pressure in implementing the NCMS. In this sense, the county is similar to many counties implementing the scheme.⁸⁵ Equally, rapid cost growth is by no means unique to Feitian. Feitian's is a very *local* reform, however: while cost control and payment reform 'fit' with policy at the provincial and national levels, concrete measures used by Feitian were decided on by the BOH. When Feitian started CCG, they did so as a management 'exploration' (*guanli tansuo*); this was a routine, or near routine, part of the work of the BOH. The county government required that the BOH introduce measures to control costs, and CCG was not deemed to require specific approval from outside the county (though support was required locally for the introduction of the adjustment fund). The county did not visit (*kaocha*) other counties in preparation for carrying out CCG, and nor did it have external technical assistance (with policy design, etc.), though the BOH did study materials and policies from other places and Feitian did receive a visit from an external academic expert who carried out an assessment which may well have helped highlight the importance of cost control. According to the NCMS Office, the main reason Feitian did not seek outside assistance in developing CCG was that, as far as they were concerned, they were simply changing scheme management methods, not attempting to develop a national model for cost control or payment reform. This was considered routine work within the scope of the county's legitimate authority.

Fascinatingly, this is reflected in the way Feitian's understanding of its own reform has changed over time. When the county was planning CCG, they were largely unaware of 'payment reform' as a concept (they were not aware of reforms in Lufeng, for example; see below). According to the NCMS Office, at that time, there was little attention to provider payment reform in national debates, whereas this is now commonly discussed

⁸⁵ This was recognised by the province: pressure for this kind of reform is greatest in places like Feitian, where resources are most constrained. This also corresponds broadly to a characterisation mentioned in the last chapter that local government innovation in China is frequently a response to crisis.

as a 'global problem' (*shijie nanti*). At this time, as far as Feitian was concerned, this was not the case: the county was simply changing management methods. When Feitian started CCG, this was firmly conceived of as 'cost control' (*kongzhi feiyong*) work, and the language of early Feitian documents is firmly rooted in 'cost control'. Only later was this reinterpreted as 'payment reform' (*zhifu fangshi gaige*): by 2009, when CCG was starting to receive a degree of recognition, the language changes to one of 'payment reform', not 'cost control'. When I first visited Feitian, I attended a meeting between the BOH and a visiting delegation from another county. A member of the NCMS Office chaired the meeting. One of the items discussed was Feitian's work on payment reform, what the county had previously called – he added, for the benefit of the visitors – 'cost control'.

The most interesting indicator of this evolution in terminology and thinking is the 2008 policy in which CCG is first significantly set out. The title of the document, Opinions on strengthening NCMS fund management (Feitian County Bureau of Health and Bureau of Finance 2008), is revealing. In this, CCG is one section of five, all relating to various fund management measures. The 2008 policy draws on provincial policy for legitimacy, citing a phrase from a 2008 document which encourages counties to "increase oversight and management of all kinds of medical providers, enthusiastically promote 'control of overall costs, control of average inpatient costs, limiting of average outpatient costs, limiting of average prescription fees and volumes, and other cost control measures" (Y Province Bureau of Health, Bureau of Finance and Civil Affairs Bureau 2008). According to the 2008 Feitian policy, CCG is *in the spirit of* this provincial document. The phrase quoted in Feitian's 2008 policy frames the county's reform very specifically in the language of cost control, not payment reform, and it is this idea that is taken up and developed by the BOH and for which the provincial policy is cited to provide legitimacy. In Feitian's policy, though, this phrase appears out of context. In its original context, the phrase is followed directly by an exhortation that counties "experiment with payment reform, gradually implement 'prepayment systems,

capitation systems, single-disease payments' and similar payment reforms" ⁸⁶. Faced with exhortations, side by side, but nevertheless separate, to carry out both cost control and payment reform, the county embarked on a programme *explicitly* conceived of as cost control. Around the same time, the BOH attended a meeting outside the county on cost control, but at this point, discussions of this remained, as far as Feitian was concerned, largely at the level of concepts. From 2009, there started to be more practical attention to cost control and payment reform nationally, and that is when Feitian started to think of what they were doing in terms of payment reform.

7.6.3 Implementing and interpreting policy: The logic of 'results orientation'

Feitian's reform has a clear relation to provincial policy, but equally cannot simply be considered direct 'implementation' of that policy. The county specifically phrased its reform in terms of provincial policy which, while not obliging the county to carry out cost control / provider payment reform, encouraged this and provided legitimacy for what the county did: specific cost control targets were not included in Feitian's responsibility contract until 2010, but this was included as an item in provincial policy from 2007, as was provider payment reform. In 2008, however, when the county issued formal policy on CCG, they pegged their reform to the provincial document of the same year.

The rationalisation of this by the NCMS Office is interesting: the provincial policy was described as a 'guidance document' (*zhidao wenjian*), whose existence meant that the county did not need special authorisation to carry out CCG. The legitimacy of CCG derived from the fact that higher level policy contained this provision ("*kongzhi yiliao feiyong zengzhang – [you] yi ju hua*"). On this basis, the county believed that CCG was legitimate and not in contravention of policy at higher levels ("*[bu] weifan shang mian*

⁸⁶ I use 'single disease payment' (SDP) as a simplified translation of two closely-related Chinese terms, *dan bing zhong ding'e fu fei* and *dan bing zhong xian'e fu fei*. These are similar to diagnosis-related groups as used elsewhere, but simpler both in means of calculation and in scope – DRGs are exhaustive, covering all diseases, whereas SDPs cover a limited number of conditions, and many diseases remain subject to FFS payment (Meng, Yip et al. 2010).

de zhengce)” and that carrying out this reform was not risky as there was a clear direction from the province, though the *means* to achieve this end were not specified: “the province wanted to control costs, [but] how to do that – that that was up to you” (*“shang yi ji yao kongzhi feiyong [...] daodi zenme zuo – [shi ni] ziji de shiqing”*).

City and provincial interpretations of CCG are interesting for two reasons. First, Director Hu of the city BOH gave the same interpretation of CCG as the county: limited funds mean that the county must control costs if it is to increase reimbursement levels. Second, Hu described Feitian’s practices – controlling costs and making better use of funds – as ‘correct’ (*you daoli*). Most interestingly, though, Hu presented Feitian’s practices as largely unremarkable. His assessment was that there are a number of counties under that city, and other places are also working on cost control: one county is doing well in introducing single-disease payments, another is developing a stratified reimbursement model (*fen duan baoxiao*),⁸⁷ and so on. As he stated it, “all [these] counties have some specific practices”. Feitian’s practices, while correct, are similar to those in other places. The motivations of all these places are correct, though the specific methods used are different (*“bu guan zenme shuo, mudi shi dui de”*).

More broadly, the county has received a fairly significant degree of provincial and national recognition. Feitian started to get recognition for this reform in around 2009 when the county reported CCG to the province as a component of its NCMS work, and the province then reported this to the MOH. This reporting seems, from discussions with both the county and the province, to have been entirely routine and to not need any particular explanation. In 2009, the county discussed CCG at a provincial meeting and subsequently attended a national meeting on payment reform and received a provincial delegation, wanting to understand Feitian’s practices with a possible view to propagating (*tuiguang*) these.

Provincial analyses mirror Hu’s analysis: cost control was a response to pressure, and pressure and fund risk are greatest in economically less developed places. According

⁸⁷ As in the description of Feitian’s plan setting above, this refers to setting reimbursement bands and setting varying reimbursement levels across bands.

to the province, counties have a lot of autonomy in the way they manage the NCMS, and this is presented as both necessary and desirable: county leaders are most attuned to specific local circumstances and are able to judge what will work in that county far better than the province. On this basis, they should be making concrete decisions as to how to manage the NCMS. Many things that are immediately apparent to a county would not be so to the province, overseeing scores of counties. This approach to management is rationalised as one of 'results orientation' (*jieguo daoxiang*): the role of the province should be to set goals within which counties have autonomy as to means and to assist in areas in which counties are unable to solve problems themselves. Successful practices at the county level, once discovered, can be refined and rolled out by the province. The province itself, as well as lacking the local knowledge needed to micromanage county NCMS schemes, also lacks the resources (personnel and time) to do this.

7.6.4 CCG, sub-national models and changing repertoires

As well as being clearly linked to provincial policy, Feitian's reform exists in relation to reforms elsewhere in China. As above, Feitian's reform is a simple attempt at a solution to a complex problem. I have argued above that use of a contracting (*baogan*) system both conforms to an existing technique in the Chinese government repertoire (see Chapter Six) and that this is simpler than alternative approaches to provider payment reform. This simplicity is presented as a virtue by BOH staff, and CCG is said to 'suit' Feitian largely because it is low tech and feasible for a county BOH to put in place independently. According to Qiu, this simplicity means that CCG also has relevance elsewhere. When Feitian presented its experience outside the county, it was well received, which Qiu attributes to its being relatively easy to implement ("*bijiao keyi caozuo*"). As one interviewee rationalised this, this type of solution is not best practice (*zui you*), it is second best (*ci you*), but it remains a useful development of the scheme. This is clearly linked to Feitian's development of cost control measures independently and largely without external assistance and oversight. It is, at best, a *provisional* solution to a complex problem. I return to this theme in more detail in Chapter Eight.

While Feitian's reform was locally developed, without outside expertise, it exists, at least to an extent, in 'conversation' (Gudeman and Rivera 1990) with reforms elsewhere.⁸⁸ Though the BOH looked at materials and policies from other counties in preparation for CCG, it does not seem that any one place was a specific inspiration for Feitian's reform; rather that, having examined materials from other places, a global budget seemed more feasible than other alternatives. *Per diem* (*an chuang ri fu fei*) and capitation (*an ren tou fu fei*)⁸⁹ payment systems were judged to involve complicated calculations and to be too technically complex.

This is in contrast to Lufeng. In discussions of payment reform in Feitian, and in other fieldwork counties, Lufeng inevitably comes up in conversation. Lufeng County, in Yunnan, is the site of a provider payment reform pilot overseen by Wang Lusheng of the China National Health Development Research Centre⁹⁰ and is nationally famous. From 2006-2007, Lufeng started to reform provider payment mechanisms based on a system of global budgets for outpatient spending and a *per diem* payment system for inpatient treatment, alongside introduction of single-disease payments for 32 specific conditions. Lufeng's system of calculations is more complex than that used in Feitian. Outpatient budgets, for example, are set on a township/town basis and budgets for individual THC's are based on the population covered, the 'service level' (*fuwu nengli*) of providers, sub-county economic growth levels and the pattern of use of health services, and growth is allowed based on change in the consumer price index. Inpatient budgets are calculated based on an assessment of data on providers at the

⁸⁸ Gudeman and Rivera are anthropologists who study 'folk voices' and marginal economic practices in Colombia. They show how the marginal economic practices, framings and discourses of Colombian smallholders and peasants are linked in multiple ways with mainstream economic theory and practices – how centre and periphery, 'folk voice' and 'centric text', are bound up in the same conversations: they speak to each other; their framings, categories and theories are mutually informing (though written histories (of economics) inevitably elide the folk voice in favour of the centric text).

⁸⁹ *Per diem*: an approach to pricing that sets per day and per bed costs and uses these as a basis for paying providers. Capitation: sets payments based on size of population and observable characteristics.

⁹⁰ Formerly the Chinese Health Economics Institute, a research institute under the Ministry of Health.

county and township levels, including inpatient spending and average length of stay, in order to arrive at daily rates for both township and county providers, which can then be used to calculate budgets for individual providers based on volumes of treatment. Frequent audit (*kaohu*) appears to have been heavily stressed in the development of the Lufeng model (Ji, Wang et al. 2011) especially as regards treatment volumes and providers refusing patients (Lufeng Xian Weisheng Ju 2009a; Lufeng Xian Weisheng Ju 2009b).

Of counties in which I have conducted fieldwork, a majority, including Feitian, have visited Lufeng on investigation (*kaocha*) trips. It is instructive to see Lufeng in the mirror of county Bureaus of Health: the overwhelming perception in Feitian and elsewhere is one of complexity. The NCMS Office contrast Lufeng's reforms with CCG: the Lufeng model is described as a 'project' (*keti*) developed by an external expert and beyond the ability of most counties to develop independently. The conclusion of the NCMS manager who attended Wang Lusheng's training session was that the Lufeng model was too technically complex to be easily implementable. In contrast, CCG is simple and feasible.

Meijiang and Taoshan Bureaus of Health both visited Lufeng around the time I carried out fieldwork. Taoshan's reaction to Lufeng's experience is particularly interesting. Two members of the BOH visited Lufeng in 2010 as part of a group organised by the county's parent city. Zhou, an NCMS manager from Taoshan, was one of the people who visited. At the time of the visit, he was already familiar with Lufeng's policy, having attended a training session run by Wang Lusheng. He described the visit as allowing him to understand the 'operation' (*caozuo*) of Lufeng's model. The group stayed for one day and Zhou thought this was enough to understand operational issues. One of the Vice Directors of Taoshan BOH also visited Lufeng at the same time, and his reaction was very similar to that of Feitian's NCMS Office: that Lufeng's model was interesting, but too technically complex to be developed by a county without expert support. The county were seemingly under no obligation to adopt Lufeng's practices.

Taoshan's visit to Lufeng coincided with my first visit to the county and I only met the

members of staff who visited Lufeng on my second visit there. My two visits show a change in understanding in Taoshan of both Lufeng's model and of payment reform more generally. At the time of my first visit, Director Wang and others were sceptical of the feasibility of implementing Lufeng's reforms in Taoshan for two main reasons: first, that these were too complex; second, they also doubted whether what Lufeng was doing was significantly different to, or better than, what they were doing in Taoshan. Zhou, for example, described the county's oversight agency and payment reform, as promoted by the MOH, as variant ways of doing the same thing, of 'increasing regulation' (*guifan yunxing*) of the NCMS. By the time of my second visit to Taoshan, discussion seemed to have changed and there was a growing understanding of payment reform as a more *efficient* solution to their problems than a simple quantitative increase in monitoring. Representatives of the county BOH had attended a training meeting on provider payment reform and were talking about introducing a number of single-disease payments. At the meeting, they had had a presentation from a Shaanxi official who had discussed the province's experience and promised to introduce Taoshan to a county they could study.⁹¹ As Zhou explained it, before the training session, he had understood that provider payment reform was about controlling costs, but hadn't thought through the implications of this. In going to study another county's experiences with single-disease payments what he was most interested in was not policy, but questions of detailed management (*caozuo*), including how to set up computer systems to operate this. From discussions, it seemed that the BOH was short on exact plans, but that payment reform had become much more central to their thinking. Taoshan is not an official pilot for payment reform so, according to Zhou, they are free to choose how they do this themselves within the 'direction' (*fangxiang*) of overall policy, rather than having to follow a specific approach. Taoshan's changing understanding remains within a discourse of risk and risk mitigation, however: as Zhou described it, the essence of payment reform is to transfer oversight (*jianguan*) from the BOH to providers. Providers should internalise oversight and, once this is achieved,

⁹¹ Shaanxi started development of single-disease payments very early, in 2006 (Zhang 2011b).

“there is basically no fund risk” (“*jiben shang bu cunzai jijin fengxian*”); oversight by the BOH is replaced by providers’ own internal management (*neibu guanli*).

What stands out in the case of both Feitian and Taoshan is changes in the way the two counties framed what they were doing. Feitian’s approach corresponds to a different *repertoire* of behaviour to that discussed in Chapter Six: there, I argued that both Taoshan and Meijiang’s management reforms corresponded to a simple *quantitative* increase in monitoring of scheme reimbursements and that this, while likely a useful interim measure, is ultimately a flawed approach to NCMS management. Feitian’s reform is fundamentally different: rather than a quantitative increase in scheme monitoring, as in Taoshan, Feitian’s reform requires a change in the *mode* of monitoring, and this is clearly visible in changing understandings of NCMS managers and providers. Initially Feitian understood what they were doing as ‘cost control’. Over time, the county has started to reconceptualise what it is doing as ‘payment reform’ and to understand this differently. A similar, and equally marked, process of upgrading of management repertoires is visible in Taoshan: at the time of my first visit, the county predominantly understood their Oversight Bureau and the provider payment experiment being carried out in Lufeng as equivalent, whereas by the time of my second visit, thinking had changed and they had started to understand these two things as conceptually distinct approaches to management of risk.

7.6.5 Interim summary

The preceding paragraphs show several things: first, implementation implies development of policy: Feitian’s reform is clearly linked to provincial policy, but is nonetheless equally clearly a county project, in which the specific form and mechanisms used to achieve a broad policy aim were locally developed. Second, Feitian BOH *understands* its systemic role as consisting in not just implementing, but also developing, policy. Development of policy is necessarily a creative process, rather than simple decoding of the intention of provincial policy. Third, this understanding appears to hold at the city/provincial level also: at least in this case, there is a

significant degree of common understanding of the systemic role counties are expected to play. Fourth, there appears to be a significant degree of tolerance of divergent local solutions to similar or equivalent problems, as shown by Hu's assessment of Feitian's practices as essentially equivalent to practices in many other local counties: multiple local reforms are understood to be legitimate interpretations of provincial and systemic aims. Fifth, these judgements directly mirror county and provincial understandings shown in Chapter Six.

7.7 Interim conclusions and looking forwards

The argument presented here is unavoidably long and complex. The story starts with an anomaly: a peripheral, rural county seemingly embarking on an innovative programme of provider payment reform on its own initiative and receiving fairly substantial recognition for this – at a time when provider payment remained more a topic of debate among academics and analysts than a core component of national NCMS policy. The story shows a more complex series of events than this, however: that, while Feitian's reform was indeed a county initiative, it sprang directly and unavoidably from the county's implementation of the NCMS, which I have described here as an extension of the pressurised system discussed in previous chapters. Target-driven implementation, combined with localisation of fund risk, created a conjuncture in which the county was forced to control costs if it was to implement the scheme effectively, and this underlies Feitian's reform. Only subsequently was this re-interpreted as 'provider payment' reform, in line with evolving national and sub-national reform debates. Feitian's reform is fairly simple, and is understood by the county as simple – something that a peripheral county such as Feitian could do on its own, without high-powered external assistance (as in Lufeng), and through consulting policies and documents from other places. The systemic role played by Feitian is not just that of simple implementer of pre-cast policy, but also that of sub-national, 'frontline', developer of policy solutions. This is understood as normal and natural by all actors concerned, and this understanding of roles shows direct continuity with the discussion of national and sub-national roles by Wu Yi discussed in Chapter Five, and

with evidence presented in Chapter Six. Evidence from the city indicates that Feitian is far from unique. Feitian's reform is understood as a second best (*ci you*) solution. Could the county have developed a best practice (*zui you*) reform? This is clearly unlikely, given limited capacity and the fact that Feitian's reform was carried out very much on the periphery of national reform currents. Provincial attention, though, shows this as still relevant and valid, and as a part of, overall, an evolving approach to management of the NCMS, not just in Feitian, but both nationally and sub-nationally, in which national policy and sub-national practices and models form part of an ongoing conversation of reform.

In the next chapter, I discuss a county-level innovation which shares many features with those shown here and in Chapter Six, and which has been successful in being taken up in central policy. In Chapter Nine, I draw together empirical threads presented in empirical chapters into an integrated analysis.

Chapter Eight

Transitional institutions and fill-in development: The Second Round Reimbursement policy

“Planned Brasilia is, in a thousand ways, underwritten by unplanned Brasilia”

(Scott 1998, p. 348).

8.1 Introduction

This chapter examines a ‘mechanism’ (*jizhi*) within the New Cooperative Medical Scheme, the ‘Second Round Reimbursement’ (SRR) policy, developed by several counties during early stages of implementation of the NCMS. The policy allows counties to reduce fund risk (*jijin fengxian*), specifically risk of overspending, during scheme implementation, and to maintain relatively high levels of fund use (and reimbursement) despite generally low levels of capacity and unreliable funding. From its origins as a locally developed mechanism, SRR was adopted as first provincial, then national, policy.

I argue that local capacity deficits and uncertainty over funding, combined with pressure to implement the NCMS, pushed certain counties to develop SRR, a mechanism allowing them to carry out reimbursement in two stages: an initial, conservative (‘first round’) reimbursement, followed by a second reimbursement later in the year when scheme spending commitments become clearer. Under this, NCMS users receive a second reimbursement at the end of the year if the scheme is running a surplus. This functions as a way to increase reimbursement to NCMS users while reducing the risk of overspending by the county BOH. The SRR is a direct extension of the logic of the pressurised system, as set out in previous chapters. The argument in this chapter follows on directly from cases examined in preceding empirical chapters and should be seen as an extension, and elucidation at the systemic level, of dynamics visible at the local level in previous chapters.

As in Chapter Five, I use 'mechanism' to render '*jizhi*' as used in Chinese policy discourse. 'Mechanism' exists in relation to key terms including 'policy' (*zhengce*) and 'system' (*zhidu*), in which it is the junior partner, and should be understood as referring to a component and dependent element of a larger, and separately instituted, policy. In this case, 'mechanism' is used to make clear that SRR is not a free-standing entity, but operates within the umbrella of the NCMS: it is one particular method of managing reimbursement. Saying this, it is hard to draw a clear theoretical dividing line between the notion of 'mechanism' and the larger policy of which it forms a part. At what point in the official acceptance and propagation of a discrete local mechanism for, for example, carrying out reimbursement, should one start to conceive of it as a 'policy' or a 'system'? Given terminological difficulties, for the most part, in the text I refer to the SRR as a 'policy', not a 'mechanism', but it should be remembered that SRR is a dependent component of the NCMS more generally.

This chapter contributes to existing literature in the following ways. First, it provides a case study of an unexamined mechanism within the NCMS and shows how a simple local innovation achieved widespread coverage and national uptake and contributed to development of the NCMS under conditions of limited management capacity. In doing this, it employs what LI has described as a "'back to time line' narrative [of] the processes leading to the translation of reform ideas into [...] practice" (Li 2006b, p. 78). Second, the chapter argues that we should see the SRR as a significant 'transitional institution': while not the most efficient reimbursement method, it was a useful second best, a low-tech solution to a complex problem and a stepping stone in scheme development. Finally, I argue that this kind of transitional institution is not unique; rather, the approach to reform adopted in development of the NCMS can create space for the emergence of 'appropriate' innovations (Conway, Waage et al. 2010) and transitional institutions through decentralised experimentation at a distance from centres of political power and technical knowledge. Studies of public sector innovation have largely ignored considerations of 'appropriateness' in favour of leading edge, best practice, solutions. I argue that this derives from a decontextualised approach to

innovation and that appropriate, transitional, innovations merit more attention, especially in peripheral settings.

The chapter is mainly based on analysis of Chinese policy documents and academic literature, supplemented by interviews carried out during fieldwork.

8.2 NCMS reimbursement, local capacity and rationale of the SRR

8.2.1 NCMS reimbursement schemes: Local design; low payouts

From the outset, a key stipulation of NCMS policy has been the balancing of scheme revenues (government and individual contributions) and expenditure/payouts. This is described as 'setting payments according to receipts' (*yi shou ding zhi*), the aim being maintaining scheme sustainability over time (Guobanfa 2003, No. 3). That is to say that the scheme must remain within its means. Fund pooling is at the county level, and it is at this level that this balance must be maintained, though regulations introduced in 2004 specify that counties should maintain a 'risk fund' (*fengxian jijin*) equivalent to approximately 10% of total NCMS monies for any given year. Schemes may also run a rolling surplus, but the total amount permitted is quite limited (detailed below). Use of the risk fund is permitted only in the event of overspending resulting from significant changes in disease incidence and the like: it is an emergency fund and is not to be used for routine stabilisation of local spending (Caishe 2004, No. 96). The NCMS is a pay-as-you go, repeating annual programme. Users sign up (or don't) and pay their contributions each year, for that year. Monies collected in a given year must be used in that year (with the exception of the risk fund and monies in 'family accounts', *jiating zhanghu*).

As in Chapter Seven, NCMS plans are set at by counties. The World Bank questions whether local design of reimbursement schemes is a good thing; worries include local scheme 'capture' (routing of funds to the better off), and concerns over local capacity:

“... benefit package design is highly technical, requiring intensive analysis of epidemiological and financial data. Governments often rely on national and

international experts for help with this task when it is particularly complex. Although provincial-level institutions will be capable of providing technical leadership and ensuring that benefit packages have broad-based benefits in most instances, central level institutions may have a residual role to play in backup support" (Wagstaff, Lindelow et al. 2009b, p. 100).

Counties face considerable difficulties in scheme design: they must juggle not just disease burden, but also a variety of co-payment measures (thresholds for payment, payment ceilings, user co-payments for treatment, etc.), as well as appropriate levels of inpatient and outpatient coverage, appropriate levels of reimbursement for treatment at different levels of facilities (village, township, county, outside-county), etc., in order to reimburse users' medical costs without bankrupting their schemes which, for the most part, operated, in the early stages of the scheme, on budgets of only 20-30% of total per capita rural health spending (Brown and Theoharides 2009; Wagstaff, Lindelow et al. 2009a). In many cases, the difficulty of this is compounded by lack of adequate data on such things as disease burden, etc. The design of local reimbursement packages shows great variation (e.g. Brown, de Brauw et al. 2009; Wagstaff, Lindelow et al. 2009a); Brown et al. find that "[r]eimbursement regimes for health expenditures vary tremendously, *even in neighbouring counties*" (Brown, de Brauw et al. 2009, p. 317, emphasis mine).

Over the history of the NCMS, concerns have consistently been expressed about low reimbursement levels. According to a survey by Yi et al, in 2004 actual levels of reimbursement were 14 RMB per person, as compared to a total investment from all sources of 35 RMB per person (Yi, Zhang et al. 2009b, p. S122). Also analysing 2004 data, Yan et al. find that "97 percent of health expenditures of individuals that were participating in NCMS were covered by their own income, savings or borrowing" (Yan, Zhang et al. 2006, p. 18). This is to say that, in addition to low reimbursement levels arising directly from low levels of scheme funding, setting of reimbursement packages was, at least initially, also very conservative even compared to available resources. Mao, in a 2005 review, found counties to be running 'huge' fund surpluses, with

“27.28% of counties in eastern provinces, 32.51% in central provinces, and 55.98% in western provinces [having] large surpluses” (Lei and Lin 2009, p. S41).

There is a difference of opinion as regards the root of this problem; Yi et al. ask,

“So what is the source of the problem? There are at least three possible explanations. First, policy makers do not know how to design the benefit package. Second, policy makers did not earmark enough funding to cover the specified benefit package. Costs have risen so much that although originally funding was sufficient, by the time reimbursement was being made, there were not enough funds. And, finally, the promised funds actually did not get down to NCMS operation level for reimbursement” (Yi, Zhang et al. 2009a, p. S126).⁹²

In contrast, Lei and Lin see this conservatism as arising from a desire to minimise risk to the scheme resulting from localised (and unpredictable) disease outbreaks and the like, combined with overly-narrow pooling:

“[surpluses arise] partly because the risk pooling level for the NCMS at the county level is too narrow. There is a chance that the population in an entire county might be subjected to a widespread disease outbreak. Thus, the NCMS officials are worried that if this type of situation were to happen, the NCMS might go bankrupt. Therefore, the NCMS officials set the deductibles too high and the copayment rates too low in order to guard against bankruptcy of the NCMS system” (Lei and Lin 2009, p. S41).

Such low levels of reimbursement – lower than strictly necessary, given scheme funding – are of obvious concern for a scheme whose expressed aim is to reduce the burden of health seeking for its users (Yan, Zhang et al. 2006; Brown, de Brauw et al. 2009).

The above analyses rely on early data, and by around 2007 one can see changes in reimbursement levels: not only had scheme funding increased, but total

⁹² These authors dismiss the final possibility, calculating that over 90% of funds earmarked for payments were used for this purpose in their data.

reimbursement as a percentage of scheme funding (payments / receipts) had also increased, from 40% in 2004 to 94%, according to one analysis (Yi, Zhang et al. 2009b), while later data show fund use rates of nearly 100% (Barber and Yao 2011).

These analyses likely contain a lot of truth; in the next section, however, I give a supplementary interpretation, close in spirit to that of Lei and Lin, in which capacity deficits and funding difficulties conspire to create local conservatism in setting reimbursement plans. One result is the enormous surpluses identified by Mao and others. Another result is an innovative local coping mechanism: the Second Round Reimbursement policy.

8.2.2 Capacity deficits, surpluses and conservatism

As noted above, reimbursement design is highly technical and requires specific capacity. This capacity has been, and remains, a question at the county level. Given China's size and variation in levels of development, capacity levels clearly vary by place. Brown and de Brauw give an indication of this variation:

“experience and training of county-level administrators varies widely, suggesting that some programmes are likely to be better designed and more sustainable than others. For example, 19% of the top health care administrators in counties covered by our survey had at least 25 years of experience in health care management, whereas 15% had three or fewer years of experience. Similarly, almost a quarter of these administrators had not attended university. *Indeed, in one county visited during survey pre-testing, responsibility for determining the reimbursement schedule was subcontracted to a junior high school maths teacher*” (Brown, de Brauw et al. 2009, p. 310, footnote 29, emphasis mine).

Capacity levels and deficits are obviously not uniform: more developed counties and cities are likely to have better trained personnel, and, as above, one rationale for contracting management of the NCMS out to commercial insurers is to buy in experience which the local health administration lacks, most county-level Bureaus of

Health having had no experience of managing an insurance programme before the NCMS (Interview no. 36).

Chinese language analyses of the SRR describe low capacity levels, lack of experience, lack of adequate data (on disease incidence, use of facilities, etc.), and delays in counties' receiving central and provincial NCMS transfers as the main reasons for local conservatism in setting of reimbursement plans (Song 2008; Xiang, Song et al. 2008a; Xiang, Song et al. 2008b; Xiang, Su et al. 2008; Xiang, Liu et al. 2010).

This conservatism comes through in low levels of fund use. Analysis of the SRR carried out under the auspices of the MOH shows enormous variation in levels of *de facto* fund use in 2006: while Zhejiang was managing an overall (averaged provincial) level of fund use of 92.9%, Fujian was only managing 28.3%. Other poor performers were Ningxia (43.9%), Liaoning (55.4%) and, surprisingly, Tianjin (62.4%), while the average provincial use rate was 71.5% and 25 out of 30 provincial-level units were not achieving the 85% use level soon to be recommended by the MOH (Weinongweifa 2007, No. 253). 2006 data show differences between eastern, central and western regions, but these are less significant than differences between counties in different waves of the NCMS process. For example, in 2006 only 17.6% of counties which started implementing the NCMS in 2003 used less than or equal to 70% of available funds, with remaining counties exceeding this level. This figure increases to 19.6% for counties which started in 2004, 23.2% for those which started in 2005, and a huge 56.1% for those which started in 2006 (Xiang, Song et al. 2008b). This clearly shows lack of experience, and experience sharing, to be a factor underlying conservative reimbursement setting.⁹³

While counties in the first wave of the NCMS experiment were to carry out baseline assessments, by later stages of development of the scheme (and particularly during

⁹³ Despite the above, overspending has not been uncommon: an interim report on development of the NCMS shows that this can, in fact, be quite significant (NCMS Pilot Evaluation Group 2006, p. 31).

the accelerated country-wide roll out from 2006) it is believed that much less attention was paid to scheme design, with some counties even cutting corners and ‘borrowing’ assessments from their neighbours (Xiang, Song et al. 2008b; Sun and Chai 2009), all of which undoubtedly led to poorer design than would otherwise have been the case.

Jumping ahead in time, the above picture is complicated by national and provincial changes in regulation of reimbursement models. In 2007, the MOH required provinces to offer 1-2 standardised reimbursement models for use by counties under their purview, with the aim of increasing reimbursement levels and narrowing the gap between counties (Weinongweifa 2007, No. 253) (though some provinces jumped the gun; Jiangxi, for example, released guidelines in 2005 requiring counties to adopt one of three reimbursement models). One result of this attempt at standardisation of reimbursement models was, in fact, divergence in the effective levels of fund use in counties, depending on factors such as available levels of services, costs of services, and so on, and leading to an increase in surpluses in counties effectively ‘constrained’ by the provincial plans (Xiang, Song et al. 2008b). As Xiang et al. state, there are only two ways to make sure counties get to the 85% central requirement for fund use: allow them to set reimbursement plans according to local conditions (*yin di zhi yi*) and accept that this will make coordination hard, or impose standardisation but allow use of mechanisms for adjustment of fund use (*jijin shiyong tiaojie jizhi*), including SRR.

Overall, a convergence of factors, from local capacity deficit and lack of experience, delays and irregularities in NCMS fund transfers, and muddled central and provincial coordination attempts, created conservatism in local Bureaus of Health when setting reimbursement plans: faced with uncertainty, they would rather play safe and set low reimbursement levels than expose themselves to fund risk and risk over-spending.

My fieldwork substantiates this. One NCMS manager in Taoshan reported having used SRR in the early years of scheme development:

“At the start, we were all conservative, all over the country. [At the beginning] we were worried the money wouldn’t be enough. If there was excess, you

could do SRR, but if you over-spent, who could take responsibility for that?"
(Interview no. 25).

Taoshan used SRR a couple of times, but no longer uses it. Fund risk remains, but the county is now more comfortable setting reimbursement plans (and, at the time of last interview, in late 2010, overall NCMS funding was stable, obviating the need to keep recalibrating the plan). Fieldwork in Meijiang, which has contracted day-to-day NCMS management to a commercial company, reveals that they have never employed SRR as they have the actuarial capacity (through the insurance company) to accurately project fund use (Interview no. 36).

8.2.3 Rationale, functioning and effectiveness of SRR

The Second Round Reimbursement policy is described by one author as an "effective way of getting by" (*'you xiao de bu jiu banfa'*) (Song 2008, p. 29). The core of this policy is use of a second round of distribution of NCMS reimbursement monies to scheme users who have claimed against the scheme in a given year so as to use up surplus (*jiyeu*) monies and attempt to comply with central targets for overall levels of fund use. In other words, SRR allows Bureaus of Health to set conservative reimbursement plans ('First Round Reimbursement' (FRR) plans, *yi ci buchang*), thereby minimising fund risk (*jijin fengxian*), but to subsequently distribute excess monies (the difference between monies actually distributed according to local, conservative, plans, and – at least in theory – the 85% use target) later in the year when they have a better idea of disease incidence, total funding commitments, etc. Under SRR, NCMS users get two reimbursements during the year: a first (conservative) reimbursement at the time of the initial claim and a second reimbursement later in the year to bring overall reimbursement up to a level more or less consistent with overall fund use targets.

This, at least, is the theory. Predictably, SRR varies considerably by place. This variation is charted by Song (2008), whose Master's thesis is an investigation of the SRR, and in a number of papers by Xiang (Xiang, Song et al. 2008a; Xiang, Song et al. 2008b; Xiang, Su et al. 2008), who led an MOH-backed evaluation of this. Variation

exists along a number of axes: variation in the ratio of SRR to FRR, variation in illnesses eligible for SRR (all illnesses, versus certain, specified, conditions), variation in those eligible for SRR (all NCMS users, all those whose spending exceeds a certain threshold, specific poor and/or vulnerable groups, etc.), variation in the management (for example, the time at which SRR takes place: at the end of the year, at the beginning of the second year, etc.). The object of Xiang's investigation was to determine how best the SRR should be run, from a welter of local variants. Below, I return briefly to the question of local variation within the SRR in the context of Xiang's research and the MOH drive to standardise the policy.

How effective is SRR? Should this be considered a useful policy tool, or rather an unfortunate artefact of the pressurised system and the cellular funding structure? Overall, analyses by Song and Xiang acknowledge imperfections in functioning of the SRR, but conclude that it can be useful if practised properly. For these authors, the usefulness of SRR lies in its increasing reimbursement rates to users, helping with the provincial integration process, increasing fund use and helping prevent fund risk (from overspending, etc.). Xiang et al. provide a before-and-after comparison of ten counties which implemented SRR in 2006. Of these, two increased overall fund use by less than 10 percentage points, two increased by 10-15 percentage points, four increased by 15-20 percentage points, and two increased by more than 20 percentage points (23.5% and a stunning 40%). Of this sample, most counties ended up with reimbursement rates of between 70% and 87%, though two overshot and ended up overspending (Xiang, Song et al. 2008b, Table 4). Similar analyses show increases in actual reimbursement rates for scheme users, as opposed to overall fund use levels (Song 2008, Table 8), and a large-scale national review of the NCMS carried out in 2006 (see below) found the problem of surplus funds to be becoming progressively less serious, especially in 'old' NCMS counties (NCMS Pilot Evaluation Group 2006, p. 49).

SRR, then, can increase fund use levels, and actual levels of reimbursements received by users. This is clearly a good thing. But it can also have downsides. Xiang and Song

(2008b) discuss various problems in the way SRR has been used: it can be too targeted, giving reimbursement only to the poor (felt to not be in line with the goals of the NCMS); if SRR publicity (*xuanchuan*) is not handled well, misunderstandings and distrust can result; in some places, management is too complex, wasting time and money; in some places, timing is poor; in some places, bizarrely, funds for SRR have been earmarked at the *beginning* of the NCMS year, defeating the point of the policy and artificially reducing already-limited funds available for reimbursement of health expenses.

In contrast to these problems of implementation, the real criticism lies elsewhere: the policy delays reimbursements to users. In implementing SRR, localities in fact carry out a *partial* first round reimbursement, and users must wait until later in the year to receive the remainder of the reimbursement that they would have, in theory and if calculations could have been done better in the first place, received earlier in the year. In the meantime, they are effectively owed money by the NCMS. Timely reimbursement is important: sums of money may well be significant, and retention could well influence further health seeking by users whose money is 'tied up' within the scheme.

This is to say that SRR is not an *efficient* way of reimbursing NCMS users. Much better would be to design an effective first round reimbursement plan. But, as above, the origin of SRR lies, for the most part, in local capacity deficits which make (or made) effective plan design hard: SRR is precisely a means of overcoming, or at least partially mitigating, these constraints. It may not be an ideal solution to NCMS reimbursement, but it has performed a useful function in overcoming capacity shortfalls in development of the scheme, and is an *appropriate* mechanism for the improvement of reimbursement procedures. I return to the notion of appropriateness below.

8.3 Local roots and uptake: Provinces, the centre and transition to urban insurance

8.3.1 County-level innovation and sub-national spread

According to one observer linked to the MOH review process, SRR started more or less simultaneously in 2004 in a number of counties/cities, including Ningguo City (Anhui), Danjiangkou and Laohekou (both Hubei). Other counties were likely doing the same or similar things, but these counties came to central attention as they reported their local practices upwards. The motivation of these counties was to reduce surplus and thereby increase both reimbursement levels and users' enthusiasm (personal communication; January 2011). I have been unable to locate any description or analysis of the Hubei examples, but there is an early record showing operation of SRR in Ningguo City in 2004; this derives from an investigation by the Party School of Xuancheng (Ningguo's parent city) into local NCMS development (Project Group of Xuancheng City Party School 2004). This report gives little detail on SRR, but describes Ningguo's development of SRR as "an innovative way of getting by" (*"bu jiu cuoshi, hen you chuangyi"*), and authors of the report suggest improvements to Ningguo's practices. It is likely that this 'model' would now be frowned upon, however, as it specifically benefits disadvantaged groups rather than the broad constituency of NCMS users (Ningzheng 2007, No. 102; Wang 2007).

Counties around China started implementing SRR at different stages. Counties studied by Xiang mostly started in 2006-2007, but many started earlier, and by this stage 80% of provinces already recognised SRR and in these provinces a majority of counties were implementing SRR (Xiang, Su et al. 2008). Jiangxi produced detailed provincial regulations on SRR very early, in 2006 (Ganheyibanzi 2006, No. 20). Stated reasons accord with those discussed above, and SRR is credited with increasing provincial fund use to 80% overall for the whole year of 2006 (Xiao 2007), compared to a eager aggregate 56% fund use for the first eight months of the year (Xiao 2006). In 2006, Anhui produced guidance on reimbursement policy, including mention of SRR

(Weisheng Bu Nongcun Weisheng Guanli Si 2006).

8.3.2 SRR on the central agenda

It is not until late 2007 that SRR can be seen to be part of the central agenda, and central opinions on this were only produced in 2008. Even as late as 2005-2006, SRR was receiving little central attention, if any: the two large-scale State Council mandated interim NCMS evaluations together run to more than five hundred pages. In neither report does SRR receive significant attention. The 2006 report gives marginally more information than the 2007 report, but in neither does the SRR receive more than a couple of lines. This brief mention tells of the existence of SRR and reasons for this (in line with those above). Despite dealing with reimbursement design, overspending and surplus funds in considerable detail (indeed, surveys used and given in the appendix include content on reimbursement, use of health checks and even use of risk funds – that is to say, the main categories into which fund use can be broken), there is no significant explanation of SRR, how it functions or how widespread it is (NCMS Pilot Evaluation Group 2006, pp. 7-9; MOH Center for Statistics Information 2007, p. 120). In other words, the two largest interim evaluations, whose stated purpose is to understand development of the scheme to that date, contain no significant content relating to SRR.

Equally, before 2007, when serious central attention to SRR starts, there are numerous references to reimbursement plans in central policy and speeches, but SRR is conspicuous by its absence. A typical example is the 'Notice of the Office of the State Council's forwarding of the Ministry of Health and other ministries' guiding opinions on doing well the New Cooperative Medical Scheme' (Guobanfa 2004, No. 3). This encourages experimentation with reimbursement methods; localities are warned against both overspending and retaining too large a surplus and provinces are encouraged to reduce the discrepancy between counties implementing the NCMS where possible, but SRR is not mentioned.

As far as can be ascertained, serious central attention to SRR starts in 2007, at least in

as much as this is when SRR becomes a part of central discourse and policy. In January 2007, Wu Yi (Vice Premier) emphasised the importance of increasing reimbursement levels within the NCMS at the 2007 National NCMS Work Meeting; she describes reimbursement as the core (*hexin*) of the NCMS. She mentions that some places have used the surplus from the NCMS (and the Medical Financial Assistance scheme) to carry out SRR, and that counties should not let their fear of overspending reduce benefit to NCMS users. She states that the MOH will soon provide guidance for provinces on regulation of county-level reimbursement plans (Wu 2007).

Guidelines were released in September of that year (Weinongweifa 2007, No. 253). These are 'guiding opinions' (*zhidao yijian*), and they reiterate the importance of the NCMS reimbursement plan (*tongchou buchang fang'an*), once more described as the 'core' of the NCMS; the aim of the Opinions is to orient provinces in their oversight of counties, with the aim of increasing standardisation. Much content is not relevant for our purposes; however, the Opinions appear to contain the first central guidance on use of SRR, which it is suggested can be used in cases where counties have a high level of fund surplus. The Opinions give little *de facto* regulation of practices, but do make clear that SRR is to benefit all NCMS users, not just poor groups, etc.

Other official statements on SRR were made around the same time; at an MOH press briefing in February 2008, a Ministry spokesman responded to a journalist's question on nationwide levels of fund surplus by stating that the MOH required localities to use various methods, including SRR, to reduce this (Weisheng Bu 2008a).⁹⁴ Fully-fledged SRR policy, however, was not released until August 2008 (Weinongweifa 2008, No. 65). Again, this is a guidance document, but it is the first central document to specifically deal with SRR. The document fits within the overall framework of the 2007 Opinions and is supplementary to them.

⁹⁴ As well as SRR, these also include check ups (*tijian*), which in many ways perform the same function of using up surplus funds; check ups and SRR are collocated in numerous documents (Weinongweifa 2007, No. 253), and are even completely conflated rhetorically in some places (Zhao 2009). Central policy on health checks was released in 2008 (Weisheng Bu 2008b).

8.3.3 Standardising and regulating SRR

The MOH SRR policy of 2008 gives the first national guidance on SRR and attempts to regulate and standardise its functioning. It follows the review by Xiang which, as mentioned above, was supported by the MOH (the MOH NCMS Research Centre, *Weisheng Bu Xinxing Nongcun Hezuo Yiliao Yanjiu Zhongxin*). These opinions set out how the SRR should function; main provisions are as follows:

- SRR is a method of using surplus (*jiayu de*) NCMS funds, and maintaining overall fund surplus within certain bounds (15% of NCMS funds in a given year, or 25% of historically accumulated funds); SRR is to be given to NCMS users who have received reimbursement for 'serious illnesses' (*da bing buchang*); SRR is to be regarded as a stopgap, and is not to be a substitute for proper reimbursement planning; SRR monies are not to be included in NCMS planning at the start of the year.
- Depending on the amount of the surplus, this can be distributed to all NCMS serious illness claimants or to those with a proportionally larger burden; the principle, however, is that SRR should not target specific groups.
- Additional provisions cover methods of calculation, transparency, actual SRR payment methods, gearing SRR to different levels of providers, the need for good publicity work (*xuanchuan*), etc.

These opinions build on the review by Xiang, especially as regards the timing of payment, the prohibition on allocating monies for SRR at the beginning of the year, the guidance on recipients of SRR and the emphasis on publicity. One of the most important provisions in these opinions, presumably reflecting central concern that localities use SRR to avoid improving setting of ('first round') reimbursement plans, is that the SRR is to be a temporary measure only and not a permanent fixture. This should be understood as set against great heterogeneity in functioning of SRR, captured by both Xiang and Song. Here, academic review feeds directly into central policy-making: the terms of the debate and the categories of analysis are essentially

the same across Xiang and Song's work and subsequent central policy.

Central support for SRR has been reiterated many times since the release of the opinions. A review of NCMS work in the first half of 2009 and the MOH's 2009 statement on consolidating the NCMS both reiterated support for SRR (Weibannongweifa 2009, No. 108; Weisheng Bu Xinwen Bangongshi 2009). Capacity building work in western China in 2010 was to include SRR, presumably as this is where the largest capacity deficits remain (Weibannongweifa 2009, No. 228), though Zhejiang was still advocating SRR for schemes running large surpluses in 2009 (Zheweibannong 2009, No. 4). In addition to direct central impact, SRR has been taken up in some places for use in *urban* insurance schemes (*Chengzhen Jumin Yiliao Baoxian*), notably in Hainan, Jiangxi and Fujian (Wang 2009c).

8.3.4 Interim summary

Above, I have shown the local emergence of a discrete NCMS reimbursement mechanism at a very early stage, largely as way to cope with local capacity deficits. This mechanism was a local 'innovation' in the sense that it was not a centrally-scripted development, but a local response to local issues. This mechanism, and its variant forms, went on to achieve very substantial national coverage before becoming the subject of an MOH investigation, and subsequent release of official guidance aimed at standardisation and regulation. Additional impact can be seen in the inclusion of SRR in MOH capacity building in western China and in the spread of the mechanism from the rural to the urban insurance system.

8.4 Capacity constraints, appropriate innovation and importance of context

To what extent can the SRR be called an 'innovation', and on what basis can it be described as 'appropriate'?

8.4.1 Innovation

The concept of innovation has been tackled from different angles by a number of different literatures (Greenhalgh, Robert et al. 2004), and application of the term is highly uneven. On the one hand, if the qualifying bar is set too low, a wide range of practices of varying degrees of novelty can be described as innovation (Hartley 2006, p. 27); much business literature falls into this trap, attempting classifications of practices according to their perceived degree of novelty and the context of that novelty: products which are new to the world, products which are new to the market but not new to the firm, and so on; in some cases, “product or service modifications” may even be deemed to qualify (Roper 2010). There are parallel considerations in public sector innovation, with some authors promoting attention to processes of continuous, marginal innovation (Bessant 2005), while, as above, other writers maintain that innovation must be a thoroughly radical phenomenon, implying “original, disruptive, and fundamental transformation of an organisation’s core tasks” (Lynn 1997, cited in Hartley 2006, p. 25).

Clearly, while setting the bar too low risks seeing all change as innovation, setting the bar too high risks being restrictive and excluding emergent practices.

Is SRR an innovation? Similarly to initiatives in Taoshan and Feitian, SRR is an orphan policy, not mandated from above: it is a locally developed mechanism, though this responded to, and was dependent on, a parent *policy*. While the SRR is clearly consonant with NCMS policy, it has its own institutional coherence, marking it out from ‘first round’ reimbursement, the baseline method of reimbursement specified in the NCMS.

There is little specific focus on policy ‘mechanisms’ within the China policy literature. Here, I want to draw a parallel between the SRR and the literature on public sector innovation more globally. This literature places a lot of weight on developing typologies of innovation. While such typologies are, obviously, heuristic devices, one notion is useful here: ‘process innovation’ (frequently opposed to ‘product innovation’, ‘service

innovation', and similar) (Hartley 2006). This is very close to the place occupied by 'mechanism' in Chinese policy discourse. Hartley describes process innovation as "new ways in which organisational processes are designed". What stands out here is the implied focus on *means*, not *ends*. (In Hartley's typology, a focus on ends is clearest in 'strategic innovation', which implies "new goals or purposes of the organisation".) SRR conforms to this characterisation: the mechanism is a *means* allowing implementation of the NCMS and achievement of NCMS *ends* (providing reimbursement for medical treatment, etc.). SRR does not, at least explicitly, set out to change aims of the NCMS.

While it is possible to describe the SRR principally as a process innovation, or a largely functional mechanism within the NCMS, it is not completely devoid of any normative component. The recommendations arising from Xiang's review of the SRR (see Section 2.3, above) have two strands. The first comprises relatively straightforward recommendations on technical functioning of the SRR: timing must be correct, publicity must be handled well, etc. This relates to fine-tuning of the mechanism. The second strand, however, is potentially more interesting. This concerns beneficiaries of SRR, specifying that SRR should not target specific groups at the expense of the broad constituency of NCMS users. I noted above that this recommendation was largely a reaction to specific local implementations of the SRR singling out vulnerable groups as the recipients of SRR reimbursement.⁹⁵ Xiang argued that this was illegitimate, and this interpretation prevailed in subsequent policy. What can be seen here is not *just* fine tuning, but also a *normative realignment* of the SRR, based on an understanding of the legitimate function of the NCMS. In this, both state and non-state actors played a role. Warping a formulation from Tang Tsou, it seems – at least in this case – that localities have the freedom to *propose* (in this case a new mechanism), but that the centre has the freedom to *dispose* (i.e. to adjudicate as to correctness and legitimacy) (Tsou 1986, Chapter 7).

⁹⁵ One provincial NCMS manager I interviewed argued that this was a missed opportunity: SRR could have been used as a way to carry out top-up insurance for NCMS users who had made large payments.

To summarise: first, the SRR can be mainly considered a 'process innovation', though it potentially has a distributional and normative component. Second, 'process innovation' conforms closely to development of 'mechanisms' in Chinese policy discourse. Third, this analysis should be seen as confirming the assertion in Chapter Five that the 'mechanism' level is a fruitful place to examine local innovation in China, especially in defined policy spaces such as the NCMS.

8.4.2 Innovation on the periphery: Appropriateness, context, metis, reified knowledge

In this section I make three inter-related points about the SRR: first, as a local innovation, the SRR originated on the periphery of power/knowledge, and this helps explain its low-tech and 'appropriate' character. Second, the SRR can be thought of as a 'transitional institution' – ephemeral but important. Third, one must ask whether the SRR has more widespread applicability or replicability in other environments or health systems.

The SRR, as an innovation, can be considered an 'appropriate innovation', developed on the periphery of the Chinese state policy making apparatus, in a context of capacity constraints arising from imperfect decentralisation of responsibilities for implementation of health system management. The SRR developed, largely organically, far from the centre (or centres) of political power and technical knowledge.

Innovation on the periphery has received little specific attention. Underlying much literature on public sector innovation is a desire to decode the secrets of good public performance with the aim of replicating this elsewhere (this trend underlies very many studies; for examples, see Greenhalgh, Robert et al. 2004; Albury 2005). This, combined with this literature's importing of thinking and methods from business studies and management, has to a large extent given a focus on *purposive action* (horizon scanning, trialling, replication, etc.) as the basic mode of innovation. Underlying this is an elitist understanding of innovation that privileges a technical worldview at the expense of a recognition of the importance of context in institutional development.

In contrast with the above, Greenhalgh, et al. caution against focusing exclusively on purposive innovations at the expense of 'good ideas' arising 'locally' and on the periphery, whose development may be less programmatic and controlled. In a discussion of diffusion of innovations, they state:

"Most diffusion research has addressed proactively developed innovations (e.g. technologies or products developed in formal research programs) whose main mechanism of spread is centrally driven and controlled [...]. But many innovations in service delivery and organization occur as 'good ideas' in local services, which spread informally and in a largely uncontrolled way [...]" (Greenhalgh, Robert et al. 2004, p. 603).

The SRR fits this description, though in this case capacity, as well as location (the periphery, the 'local'), has been a major factor influencing development. While more generally, capacity levels of local units may be a concern during decentralisation (Green 2005) and it is easy to understand that local capacity deficits have the potential to undermine possible increases in responsiveness that could be derived from decentralisation of government functions, especially in developing countries, my argument points to another possible outcome: appropriate innovation.

As above, the SRR has been instrumental in increasing levels of reimbursement and fund use for NCMS users, though it has not been *efficient* when considered against the benchmark of a properly functioning 'first round' reimbursement scheme. It is a 'second best' (*ci you*), rather than optimal (*zui you*), policy solution, similar to those examined in previous chapters. However, such a level of efficiency was likely unattainable in the majority of places implementing the NCMS, at least in early stages and if one excludes more developed counties/cities that were able to buy in management capacity or draw on external support. Low levels of provision (and funding) for management, as discussed elsewhere, were likely also unhelpful. This is, in fact, mirrored in Schumpeter, according to whom, innovation and creative destruction produce a range of possible, and non-optimal, solutions: a lot of things that are tried will not work and this is an inevitable consequence of working in the dark, in a dynamic system:

“the ‘best method’ of producing in the theoretical sense is to be conceived as ‘the most advantageous among the methods which have been empirically tested and become familiar’. But it is not the ‘best’ of the methods ‘possible’ at the time. If one does not make this distinction, the concept becomes meaningless [...]” (Schumpeter 1934, p. 83).

Effectiveness, in other words, must be situational, and the SRR is an *appropriate* response to the constraints faced by localities, given their obligation to implement the NCMS. Mirroring Feitian’s reform, it is low-tech, requires little capacity and is simple to run, and this likely accounts for its popularity and widespread acceptance long before receiving explicit national policy backing. In the absence of better techniques, it allowed localities to increase reimbursement levels.

Use of the term ‘appropriate’ is based on Conway et al., who discuss scientific innovation and its application in the developing world. They argue that technological innovations originating in developed/industrialised countries have often failed to produce hoped-for results when applied in developing countries (and highlight, among other things, capacity deficits as hindering successful implementation). The lesson to be learnt from this, they argue, is that technologies, and innovation, must be *appropriate* to the environment in which they are to be employed:

“Whatever the source and wherever the application, the important feature of technologies is that they are locally appropriate. For a technology to be appropriate in a developing country, as anywhere, it has to: [be] readily accessible and affordable; [be] easy-to-use and maintain; [serve] a real need; [be] effective” (Conway, Waage et al. 2010, p. 27).

Conway et al. stress the importance of appropriateness in technological innovation, extending a concern, deriving from the appropriate technology movement, for the appropriateness of technology *per se*. These authors deal with scientific innovation and its technical applications. The point, however, applies more broadly: insurance schemes such as the NCMS are organisational technologies, and are *extremely*

context-dependent, making appropriateness even more important, if anything, than for the kinds of technologies discussed by Conway et al.

A similar position is taken by Bloom and Standing (2008) who, discussing 'future health systems', argue that China and India are likely to become important sources of institutional innovations in healthcare. They argue that basic structures of existing health systems were largely developed in Europe and North America in the late nineteenth and early twentieth centuries and that structures underlying these systems have a preponderant weight in current debates on health system change in poorer and transitional countries, often "with little questioning of their appropriateness and adaptability". These authors argue that health systems are socio-political artefacts whose organisation is not immutable, but the result of specific historical processes of development.

This concern with appropriateness is closely linked to the idea of *metis* as used by Scott, who contrasts what he terms the 'synoptic project' of modernist state planning with *metis*, which he refers to variously as 'practical knowledge', 'situated knowledge', 'the art of the local', and similar. *Metis* is an ability to act correctly under *specific* circumstances: the context of the action is indissociable from the action itself (Scott 1998).

The SRR bears the imprint of *metis*, as local and contextual knowledge, and this gives the SRR its appropriate and peripheral character compared to more technocratic understandings of innovation. Reflecting this, a paradigmatic view is given by Roper, citing von Stamm, in course materials relating to innovation from the University of Warwick's highly-regarded MBA programme:

"Creativity is not something where someone who has never worked in that field before suddenly gets this marvelous idea. Creativity is relating a concept to a particular body of knowledge. The existing body of knowledge is as important as the novel idea and creative people spend years and years acquiring and refining their knowledge base – be it music, mathematics, arts,

sculpture or design” (von Stamm, 2003, cited in Roper 2010, p. 1).

Roper continues,

“This quote illustrates well some of the key facets of creativity which will recur [in examining innovation]. First, creativity is a conscious process; it requires determination, commitment and organisation. Second, knowledge – and lots of it – provides the basis for creativity. Third, creativity implies the new – some element of inspiration which adds something novel” (Roper 2010, p. 1).

What emerges clearly from this characterisation is the assumed intimate link between innovation, creativity and knowledge. The point to notice here, though, is that knowledge appears as decontextualised and technocratic, leaving no room for considerations of *metis* or appropriateness, or for the emergence of innovation from the periphery.

8.4.3 Transitional institutions: Two conceptions of efficiency

Criticisms of the SRR above rely on an evaluation of the limited *efficiency* of the SRR in performing its function. At first blush, this seems entirely reasonable: SRR is a second best, less efficient, way of carrying out NCMS reimbursement than FRR would have been, had it been possible to use this consistently. This is the logic underlying understandings of innovation and creativity expressed by Roper, von Stamm and others. The idea of efficiency underlying this is *allocative efficiency*. However, one should set against a relative failure in allocative efficiency the usefulness of the SRR in terms of development of the NCMS overall. Douglass North provides a helpful distinction for dealing with this, contrasting *allocative* and *adaptive* efficiency, the latter to be understood as the ability of systems (economic, political, etc.) to adapt to change and to deal effectively with novel situations over time. For North, this is key to long-term survival (North 2005).

This follows a highly Schumpeterian logic as regards efficiency. In a description of the process of creative destruction, Schumpeter argues that capitalism should be analysed

not for how it “administers existing structures”, but rather how it “creates and destroys them”. Efficiency over time is more important than efficiency in the here and now:

“A system – any system, economic or other – that at every given point in time fully utilizes its possibilities to the best advantage may yet in the long run be inferior to a system that does so at no given point in time, because the latter’s failure to do so may be a condition for the level or speed of long-run performance” (Schumpeter 1976, p. 83).

The importance of the SRR in this view, is as a *transitional, or provisional, institution* – ephemeral but important in ensuring adaptation in a dialectical policy development process. The SRR as a transitional institution is just one example, however, of a broader category of halfway house institutions and stepping stones which conform to the logic set out here. Reforms in Taoshan and Feitian also conform closely to this logic and pattern. One example is given by Bloom, who discusses the use of ‘household accounts’ (*jiating zhanghu*) for managing outpatient reimbursement within the NCMS:

“A number of counties have introduced household accounts into which individuals pay their contributions and then draw them down to fund minor [medical] purchases. This is not an efficient design for risk-sharing, but if the main aim has been to build trust in a new institution it is understandable that people should begin with this kind of contribution” (Bloom 2011, p. 1305).

Both the SRR and household accounts are examples of appropriate and provisional structures in the development of the NCMS. Most importantly, perhaps, is that the NCMS has not got ‘stuck’ with these interim solutions, but has managed to transcend them and move on to more effective ways of administering the scheme⁹⁶ – in many ways, a real-life example of Schumpeterian creative destruction in operation in Chinese policy making and system reform. Household accounts have been largely replaced by outpatient pooling (*menzhen tongchou*), while SRR has largely ceded to

⁹⁶ My thanks go to Gerald Bloom for this observation.

better-calibrated 'first round' reimbursement plans, though with changes in funding levels and management still continuing, SRR will likely continue to play a role for some time. As Scott says in his discussion of the limits of synoptic planning and the importance of local, unplanned, *metis*-heavy development as an adjunct to this, "Planned Brasilia is, in a thousand ways, underwritten by unplanned Brasilia" (Scott 1998, p. 348).

Are there lessons here useful for other countries in the process of health system reform? Does the SRR have widespread applicability or replicability in other environments or health systems? As argued above, there is little sense in aiming for second best, so the answer must, realistically, be 'no'. For reasons given above, however, I see this question as less important than the transitional significance of the SRR. More broadly, the type of institution represented here by the SRR raises questions about the generalisability of China's reform experience and replicability of discrete institutions developed during this process. I return to this in Chapter Nine.

To summarise, I argue that the SRR is not an innovation if judged by the criteria put forward by scholars such as Lynn, von Stamm and Roper, above. It is not best practice and if it can be considered an innovation, it is clearly an ugly duckling – ungainly and technically unsophisticated. In terms of allocative efficiency, it is a step backwards from alternatives already in existence. My use of *appropriateness* is an attempt to come to terms with this. I relate this to the development of the SRR on the periphery and to the importance of local, grounded knowledge – *metis* – in this process. Furthermore, I argue that a criticism based solely on allocative efficiency misses a key dimension of analysis: the SRR functioned as a provisional institution allowing implementation of the NCMS despite obstacles, and its importance for adaptive efficiency should be set against its poor allocative efficiency. That this can be considered a transitional structure is shown by the fact that the SRR is now becoming less common and less useful as counties become more adept in implementing the NCMS. I argue that the SRR is not an isolated example; rather, it is an example of a fairly common phenomenon in the development of the NCMS.

8.5 Concluding comments

I have shown in this chapter that the emergence and usefulness of the SRR is inseparable from the institutional context of early development and roll out of the NCMS. This policy was not a 'scripted' policy tool developed by the centre and then applied nationwide; rather, it was a bottom-up, fill-in, response to immediate problems, later taken up by the centre and incorporated into national policy. Its importance is not in its being an efficient method of conducting reimbursement, but in making possible reimbursement levels that would otherwise have been hard to achieve, given capacity (and other) problems and as a transitional institution in the development of the NCMS.

In Chapter Five, I argued that the scope of the reforms taking place in China, and the piecemeal design of many, militate against overall study – that the NCMS, as a national programme is ungraspable in its entirety, and that the majority of sub-national activity within an established policy domain such as this takes place at the level of mechanisms. These have their own genealogies, understanding which requires detailed examination of policy trajectories, debates and national and sub-national levels and detailed process tracing. My study of the SRR in this chapter shows both development and spread of a specific policy mechanism from local to national level, complementing the local-level studies in previous chapters and justifies the assertion that understanding of policy trajectories has to be focused at this level, rather than at the level of large-scale synthetic constructs such as 'the NCMS'.

I have also argued for consideration of *appropriateness* in thinking about policy innovation. Considerations of appropriateness require attention to context and contextualised understandings of organisational technologies. Such technologies may, in fact, be a step backwards in terms of efficiency, at least when considered in isolation from context. In the real world, however, efficiency can only be measured locally, and not against any rarefied, absolute, standard. Our conception of innovation must be de-rarefied and contextualised to deal with this and to understand innovation in peripheral settings: paradoxically, perhaps, it is not necessary to be at the leading edge in order to innovate. Adaptation and maintaining performance over time may well be

more important than achieving best practice in the here and now.

Chapter Nine

Counties: Implementers and innovators in national policy schemes, part two

9.1 Introduction

This chapter develops the analysis pursued throughout this thesis. I draw together themes and analysis from empirical cases discussed in previous chapters and integrate these with the theoretical framework started in Chapters Two and Four. The analysis falls into several main sections. In the first section, I build on analysis in Chapters Four and Five and on empirical cases examined in Chapters Six and Seven to advance the argument that counties must be seen not just as (imperfect) implementers of policy, but also as frontline policy developers and that both roles are systemically recognised and legitimate. This has implications for the way we see county behaviour, the risk associated with implementation/innovation and the locus of systemic judgements of legitimate practice.

The following section returns to the analysis of the cadre responsibility system and policy discourse in Chapter Four and develops this further, to argue that counties are, of necessity, engaged in interpretation of policy, not simply in implementing 'pre-cast' central policy. This formalises analysis in previous chapters regarding the pressurised system, and argues that in enforcing implementation but leaving the form of that implementation largely unspecified and subject to loosely-structured county interpretation, this acts as a motor for production of differential local practices. Rather than being just imperfect reflections of central policy, local practices exist as part of an ongoing 'conversation' of reform, in which county practices, central policy and a range of sub-national models all play a role. In this interpretation, the aim of much national policy and centrally-promoted sub-national models is to reorient counties' thinking, and to upgrade the 'repertoires' they employ in decentred management and development of the NCMS. The next chapter concludes, addressing functioning of the system

described here and usefulness of sub-national practices discussed in this thesis.

9.2 County roles: Risk, tolerance and legitimate action

9.2.1 Policy development and risk

In Chapter Five, I showed rhetorical commitment to decentralised initiative in development of the NCMS. I argued that this stems from three notions: first, that centre and localities should have distinct roles – the centre in setting the principles (*yuanze*) of reform, and the localities in developing working methods (*caozuo*, *cuoshi* and similar) and specific policy ‘mechanisms’ (*jizhi*); second, that this derives from a belief both that China’s size requires that localities have a degree of freedom to differentially implement policy and that designing reform in advance is impossible, making adaptation and problem solving an inevitable part of reform; third, that tolerance of multiple means of achieving policy goals is necessary. This elaborates on analyses of differential policy implementation and tolerance discussed in Chapter Two, and presents these as the espoused principles underlying development of the NCMS.

Previous empirical chapters clearly show counties *developing* policy, or mechanisms, under the NCMS umbrella, and I have argued in all cases that this cannot be simply considered implementation, if what we mean is something akin to ‘compliance’ (LI 1998a). Of more significance, though, are understandings of actors involved in this reform process. In both local cases examined, understandings of county roles show a high degree of consistency across relevant levels (county, city, province) and with nationally-espoused principles of policy development and reform. In saying this, I am arguing that, to a reasonable approximation, these cases show a common understanding of county roles in reform from the top to the bottom of the government system, though this is not strictly codified, but rather “informal but highly institutionalised” practice (Zhou 2010, p. 50).

A second empirical thread running through the preceding chapters is *risk*. Taoshan and Feitian both show perception of risk to be a major factor shaping implementation

decisions, and I have argued that this is not confined to these cases (and cross-referenced other examples) and shown the same structural risk as underlying development of a systemically-adopted mechanism, the SRR. In Chapter Four argued that framings of local government innovation in China are excessively restrictive, seeing this as risky, radical and system-changing, and focusing on outliers, such as the household responsibility system and democratisation initiatives, perceived to be of great significance at the expense of a larger volume of 'marginal' innovation (Bessant 2005). I also addressed Heilmann's treatment of risk in innovation, arguing that his positing of specific mechanisms allowing innovation in China derives from a similar framing of innovation as potentially deviant and inherently risky. I argued that the default analytical standpoint is that deviation, flexible implementation and innovation are risky.

The significance of examining implementation and innovation side by side and in looking at counties' parsing of risk involved in both activities (to the extent that these are separable), is that it allows a partial revision of this risk-in-innovating thesis. In both empirical cases examined, risk is an unavoidable consequence of implementation, and I have argued in Chapter Five that this is a systemic feature: risk is systemically *located* at the county level through operation of the pressurised system and the NCMS financing system. In contrast, risk *in innovating* is, at least in these cases, not great: as above, the policy development role of counties is well understood. In this, my analysis comes close to Chung's insistence on real central commitment to decentralised initiative and the related requirement: systemic tolerance (Chung 2000).

This points to a need to reframe local government innovation in China. Innovation need not be systemically challenging. In 'volume' terms at least, most is clearly *not* systemically challenging. Chung argues that there is a greater degree of tolerance for local innovations in the economic sphere than in the political sphere, and much evidence would seem to bear this out. Chinese vocabulary of reform, though, points to a supplementary understanding: much reform is phrased in terms of *tizhi* (literally, 'system'). The NCMS can, for example, be described as a 'system', as can the medical

system (*yiliao tizhi*), political system, and so on. Reform is frequently phrased as reform of a system: *yiliao tizhi gaige* is reform of the medical/health system, for example. This vocabulary also transfers into understanding of components of reform: a given initiative can be 'within the system' (*tizhi nei*) or 'outside the system' (*tizhi wai*). 'The system' is rarely definable, at least fully, but what this points to is the existence of (admittedly loose) criteria for the making of judgements as to the legitimacy of reforms and practices. Within-system reforms have, by definition (even if we can't fully delineate the system they fall within), a degree of legitimacy. Outside-the-system reforms, *a priori*, much less. In Bessant's vocabulary, within-system reforms are likely to be marginal and continuous, in that they fall within the scope of reform thinking and discourse. Outside-the-system reforms are much more likely to be challenging, radical and *discontinuous*.

This is a more useful categorization of reforms than a loose division into 'economic' and 'political'. Jiangyin City (Jiangsu), for example, received the 2010 Central Compilation and Translation Bureau Local Government Innovation Prize for a reform to the city's government target-setting system, which now includes a range of indicators relating to residents' 'satisfaction levels' (with job, accommodation, etc.) based on simple quantitative surveying, under a 'Happy Jiangyin' (*Xingfu Jiangyin*) umbrella (Anon 2011).⁹⁷ The relevance of this is that while this is a political reform (admittedly minor), it is in no way systemically challenging; rather, it fits clearly within an overall Hu-Wen discourse of 'taking people as the base' (*yi ren wei ben*) and similar (Anon 2008). Similarly, Chen and Yang, analysing local innovations submitted for and/or awarded the CCTB prize, show that since the early 2000s, very many have been in the broad area of social services, in line with expressed reform priorities of the Hu-Wen leadership (Chen and Yang 2009). There is discursive continuity between these initiatives and central reform direction. This is not to say that all developments of policy are legitimate. As in Chapter Five, many borderline legitimate practices likely exist, and

⁹⁷ I was fortunate enough to attend the prizegiving ceremony, held in Jiangyin in 2010, at the invitation of the CCTB.

specifically illegitimate practices clearly exist also (cf. Guangnan example). The extent of this is unknowable.

9.2.2 Legitimate roles and judgements of legitimacy

The core of the preceding paragraphs is that there is a degree of consensus over legitimate county roles in the policy process and corollary systemic tolerance of plural local practices deemed to further systemic ends. This is clear from previous chapters. Once we discard a compliance view of implementation and acknowledge the developmental role of counties, we risk analytical confusion: the range of *de facto* existing county practices is vast, and the range of *possible* practices is vaster still. How, under such circumstances, to understand congruence between local developments of policy and expressed policy intentions or principles at higher levels? In other words, once we accept that direct homology is only one possibility among many (and an unlikely, if not impossible, one) in policy implementation, how are we to know what constitutes implementation and what constitutes mis-implementation or deviance? I have pointed out the importance of 'discursive continuity' between local initiatives and central reform principles. How, though, to understand this? Landry, addressing what 'faithful' implementation might mean, says that the relevant criterion is not whether a given decision is specifically cleared in Beijing, but "whether local decisions conform to the broad requirements set by China's central leadership" (Landry 2008, p. 17). This is likely correct, but it is analytically challenging – intentions and principles are hard to pin down.

The broader point is that if there is no such thing as syllogistic implementation (see below), local practices and iterations of policy *must* be subject to judgements as to their usefulness, legitimacy, appropriateness and novelty. The significance of this observation is that normative judgements as to the legitimacy of local practices are constantly being made at multiple scales and levels within the policy system: this is what we see in counties' developmental role and in provinces' and cities' approval (and presumably disapproval in other cases) of county practices – ongoing processes of

adjudication of legitimacy in policy development. If we are not dealing with syllogistic implementation, or compliance, this is an inevitable conclusion. This gives two interconnected levels of analysis: counties (and other sub-national governments) operate in a distributed system, in which they are expected to develop policy. In doing so, they must necessarily make judgements as to the legitimacy of variant policy solutions. The next section attempts to account for this theoretically.

9.3 Policy: Implementation, interpretation, and conversations of reform

In Chapter Four, I argued for seeing the CRS and policy discourse as having both constraining and enabling functions. Implementation-as-compliance analyses (Li 1998a), which largely dominate thinking on implementation, focus on the constraining effects of these systems. My focus here is on the potential of policy to *enable* action. This follows directly from the empirical cases discussed in previous chapters, in which I show the operation of the pressurised system and how this forces policy implementation, but leaves the *form* implementation is to take underspecified in a 'policy of principles' (*yuanze*), and in which local practices are invariably situated in relation to policy or practices at different scales of the Chinese policy system. Here, I give a framework for understanding this. I argue three main things: first, that the enabling function of policy discourse revolves around the creation of indeterminacy in local government action while simultaneously forcing action. This is a mirror of my analysis of the pressurised system in Chapter Five. In my case, Bureaus of Health and local governments *have* to act, but lack a detailed template for action. It is in this space that multiple local practices are produced. This extends, but does not negate, related analyses by Göbel, Heberer and Senz and others. Second, I argue that compliance-type analyses rely on an underdeveloped notion of textual specificity which ignores the multiple judgements, readings and interpretations that *must* take place in the process of implementation. Understanding this involves understanding the role of text, authorship and interpretation in implementing policy and my analysis draws on critical and literary theory to advance an argument about policy implementation. Third, I argue that understandings and accepted repertoires of action are in constant

evolution in something approaching a disjointed 'conversation' (Gudeman and Rivera 1990) of reform.

9.3.1 Textual specificity: Policy discourse and the role of interpretation

There is a legitimate question as to the degree to which authors control the reception and interpretation of their texts, and this is as true of policy writers as it is of novelists or poets: all are engaged in a process of communication susceptible to disruption and misunderstanding (cf. Wedeman 2001; see Chapter Four). There is a linked question, however, as to the degree to which (some) authors (even) *attempt* to control the reception and interpretation of 'their' texts. Much has been written about the 'role of the reader' in interpretation of text.⁹⁸ I do not discuss in any absolute sense how meaning is generated in the process of interpretation: this can be debated at length, but is largely peripheral to my analysis.⁹⁹ Here, I question not what texts *mean*, but what they *do* (Fish 1980). I argue that NCMS policy texts are constructed as 'open texts' which *force* interpretation. 'Syllogistic readings' are not possible and counties are necessarily engaged in a process of interpretation of policy. For me, this understanding of the *function* of text is a way to get past a language of 'implementation' and 'compliance', which tends to imply transmission of an inert message between parties to the policy formulation/implementation process by attributing to policy texts a definite and knowable intention (Bowe, Ball et al. 1992, p. 10).¹⁰⁰ Problematising the text and

⁹⁸ Note that 'text', 'work' and similar are frequently understood to include phenomena and artifacts susceptible of interpretation, though we need not go as far as Rosenau, who describes 'the text' as "all phenomenon [sic], all events. Post-modernists consider everything a text" (Rosenau 1992, p. xiv).

⁹⁹ See Rosenau (1992) and Fish (Fish 1980, p. 12) for introductions to this question.

¹⁰⁰ These authors critique a vocabulary of 'implementation' in analysing UK education policy: "[t]he language of 'implementation' strongly implies that there is within policy, an unequivocal governmental position that will filter down through quasi-state bodies and into the schools....[We argue] that it is not simply a matter of implementers following a fixed policy text and 'putting the Act into practice". Elsewhere, they state that, "[p]olicy authors cannot control the meaning of their texts even if they do try" (Bowe, Ball et al. 1992, p. 10). This has a parallel in Long's discussion of 'demythologising planned intervention' and the existence of a 'cargo' metaphor underlying planning and executing of many development interventions (Long 2001).

reader is a way to clarify the role of the implementer in interpreting policy. Hawkes, using Barthes, offers a useful starting point:

“literature may be divided into that which gives the reader a role, a function, a contribution to make, and that which renders the reader idle or redundant, ‘left with no more than the poor freedom either to accept or reject the text’ and which thereby reduces him to [...] an inert consumer to the author’s role as producer” (Hawkes 1992, p. 113).

Umberto Eco describes texts of the first kind as ‘open texts’, whose meanings are not definitively defined in advance of reading, but deliberately left ‘open’ to multiple interpretations; these

“reject the definitive, concluded message [...]. They appeal to the initiative of the individual performer, and hence they offer themselves not as finite works which prescribe specific repetition [...] but as ‘open’ works, which are brought to their conclusion by the performer” (Eco 1989, p. 3).

Open texts do not prescribe a single, definitive, meaning; they open the way for multiple meanings (Eco 1989, p. ix); they open up a “field of possibilities” within which readers generate meaning. This implies, for Eco, “a revised vision of the classic relationship posited between cause and effect as a rigid, one-directional system” and “the discarding of a static, syllogistic view of order, and a corresponding devolution of intellectual authority to personal decision, choice, and social context” (Eco 1989, p. 14). Open texts introduce *discontinuity*: meaning is not, and cannot be, directly, *sylogistically*, deduced or derived from the text: the reader *cannot but* participate in bringing the text to its conclusion and participate in the creation of meaning.¹⁰¹

This mirrors the language of principles (*yuanze*) and methods (*caozuo*, *cuoshi*, etc.), *and the gap between the two*, in Chinese policy discourse. A large amount of Chinese

¹⁰¹ One could, equally, draw on Barthes, who presents a similar analysis, describing texts as either ‘writerly’ (*scriptable*) or ‘readerly’ (*lisible*) (Barthes 1977; Rosenau 1992, p. 27).

policy discourse is constructed as open text with the expressed intention (cf. speeches by Wu Yi, quoted in Chapter Five) of forcing interpretation and creativity and introducing a *creative discontinuity* into the implementation process. In my interpretation (diverging from Schurmann; see Chapter Four), implementation as interpretation is not a process of exegesis which unlocks a privileged ('right') meaning within policy; rather, it creates one possible meaning among many from a field of possibilities. 'Controlling costs', for example, could involve many things: refusing users treatment, capping doctors' salaries, reducing drug costs by reforming purchasing arrangements, instituting some kind of payment reform, and so on (some of which are legitimate readings and some of which are not). The discontinuity of the open text, though, ensures that reading/implementation is a creative process, the outcome of which is not determinable (at least fully) in advance: "The kind of expectation aroused by a message with an open structure is less a prediction of the expected than an *expectation of the unpredictable*" (Eco 1989, p. 80, italics mine).

In specifying principles, but failing to specify means, Chinese policy discourse separates *form* and *function* of policy: function (improving management, increasing oversight, reducing costs, carrying out payment reform, etc.) is expressed, while the concrete institutional form, or mechanism, used to achieve this rarely is, or is under-specified. Counties must develop means (*caozuo*, *cuoshi*, etc.) for the achievement of ends, or principles (*yuanze*), largely set elsewhere. This is a major motor of differential production of practices in policy implementation in China. This extends Göbel's analysis of both 'high expectations' and 'vague instructions' (Chapter Three) and my explanation largely sidesteps the question of voluntarism Göbel raises: here, interpretation is a structurally embedded process.

9.3.2 Legitimate readings: Some readings are better than others

If texts are not final arbiters of meaning, how to understand what constitutes a legitimate reading (implementation, interpretation, etc.)? At first glance, discarding the primacy of supposed authorial intention risks interpretive relativism: if it is no longer

possible to posit a fixed meaning of a given text or policy discourse by reference to authorial intention, then what is to constrain interpretation?¹⁰² Clearly, while many interpretations are legitimate, not all are. We should see open texts as shaping a field of possibilities of legitimate interpretations. The open work is not “an amorphous invitation to indiscriminate participation”. Rather,

“the author offers the interpreter, the performer, the addressee, a work *to be completed*. He does not know the exact fashion in which his work will be concluded, but he is aware that once completed the work in question will still be his own. It will not be a different work, and, at the end of the interpretative dialogue, a form which is *his* form will have been organized, even though it may have been assembled by an outside party in a particular way that he could not have foreseen” (Eco 1989, p. 20).

While interpretation involves autonomy and authority, and is a necessarily creative process, it is not arbitrary.¹⁰³ Above, I posit ‘discursive continuity’ between local implementations and central principles. This is what is at stake here.

Hall has famously analysed policy in terms of paradigms and orders of change (Hall 1993). The risk in such an approach is that policy is objectified, and that the analyst plays the role of arbiter, looking for a turning point in policy development and attributing significance to specific developments (Zittoun 2009). Such an approach has limitations for understanding the legitimacy of local practices within the grey zone of NCMS implementation/innovation. More useful is to understand *systemic* judgements of local practices – to understand how actors involved in policy processes see usefulness and

¹⁰² As Rosenau states it, ‘post modern’ approaches “recognize [that] an *infinite number* of interpretations (meanings) of any text are possible because [...] one can never say what one intends with language, ultimately all textual meaning, all interpretation, is undecidable” (Rosenau 1992, p. 119, italics mine), and all readings must be individual: “for any text a plurality of readers must necessarily produce a plurality of readings” (Codd 1988).

¹⁰³ Bondanella quotes Eco as saying, “I accept the statement that a text can have many senses. I refuse the statement that a text can have every sense” (Bondanella 1997, p. 130).

legitimacy of peripheral practices. In other words, legitimacy must be an empirical, rather than a theoretical, question. This is what we see in counties' appropriation and use of policy discourses and sub-national models in the preceding empirical chapters, where county practices and models are profoundly local, but are understood by all concerned to remain legitimate scheme developments.

For the most part, we should see the legitimacy of specific interpretations as governed by a loosely-defined 'interpretive community' of policy makers/implementers rather than by strict notions of textual fidelity, impossible in the case of the open text. For Fish, interpretation is conventional: "there is no single way of reading that is correct or natural, only 'ways of reading' that are extensions of community perspectives": community norms allow certain readings, while closing off others, and this helps account for a degree of 'stability of interpretation' (Fish 1980, pp. 15-16). Understanding, though, what constitutes a legitimate or illegitimate interpretation is impossible in any absolute sense: judgements are situational and temporal. Counties are clearly not 'naïve readers' (Bowe, Ball et al. 1992, p. 22) of policy, but neither are they Eco's 'model reader', infinitely attuned to the full range of possible meanings or interpretations of the text (Eco 1990); rather, they are somewhere in between. My discussion of changing management repertoires in previous chapters is an attempt to deal with this.

9.3.3 *Repertoires, models and conversations*

'Repertoire', as a general term, is employed by various authors, but is rarely clearly specified. Behn provides a working understanding of managerial repertoires as "patterns of managerial situations and successful actions that can be recalled and applied to new problems" (Behn 1988, p. 659). Elsewhere, he uses an analogy to explain this term: "Chess is taught and learned through a combination of analysis and practice [...] By studying and experimenting with various previously established approaches, each player builds a repertoire of moves" which must be adapted "to the specific problem at hand" (Behn 1987, p. 204). Behn is concerned with management

practices. We can extend this, however, to a more directly developmental application: Douglass North's use of 'artifactual structures' (see Section 4.2.4) draws on a notion of repertoires as used in Heiner (Heiner 1983). The ways in which actors construct regularities to deal with their environments are mediated by the artifactual structure – the repertoires – within which they function.

Previous empirical chapters show the circulation of discourses of reform and of discrete managerial solutions. In the cases examined, specific county-level practices exist in relation both to policy at superior levels *and* to sub-national practices and experience elsewhere. As in Chapter Seven, sub-national cost control and provider payment reforms across China exist in relation, to some degree at least, with Lufeng's provider payment model. This is a particularly prominent example, but the scope of horizontal linkages is very significant. Lufeng, for example, has borrowed various practices from Qianjiang, including use of 'coupons' (*juan*) as a way to reimburse providers for public health services, while Qianjiang has also visited and studied Lufeng (interview nos. 74, 75).¹⁰⁴ In previous chapters, I have discussed provincial propagation of practices I have examined. As one NCMS manager rationalised this process, everyone is studying everyone else. One BOH manager in Taoshan, talking about that county's reforms, said that the county knew their reforms were legitimate because they had received significant attention (visits, etc.) from elsewhere.

This linking is not necessarily physical; it also has a representational component. China's 'imaginative geography' (Said 2003, p. 71) is as important, and probably more important, than its physical geography in this respect. NCMS managers in Meijiang, for example, when questioned about their links with other counties, said that they tend to look to Shanghai for good reform practices: while Shanghai budgets are clearly greater

¹⁰⁴ Residents get coupons for certain public health services, which they may redeem with providers, who are reimbursed according to the number of coupons they collect, introducing a market-type mechanism into delivery of a publicly-funded service. Qianjiang has been very central in China's health reforms since the 1990s, when it was included in the MOH/World Bank/DFID 'Health VIII' project. It has been extensively studied and the MOH developed a range of training materials based on Qianjiang's experience (see Zhang, Liu et al. 2007).

then theirs, Shanghai benefits from having good universities and technical input, which Meijiang lacks to the same extent. Meijiang has also visited various other countries to study their experiences and, while one can be cynical about such visits, there appears to be at least a degree of analysis of those experiences in terms of China's reforms. One manager in Meijiang concluded that one developed country they had visited on a study trip was not entirely suitable as a model as patients' 'quality' (*suzhi*) (see Murphy 2004) was higher than in China, meaning that health seeking behaviour would inevitably be different, though certain mechanisms might be transferable. In some cases, though, flows of learning are entirely unpredictable. One NCMS Office head, when asked how he finds relevant experience and models to study, said he used government reports and briefs, but also described how that county's studying of Jimo County's model of medical financial assistance (on Jimo, see e.g. Song, Zhang et al. 2006) came about as a result of a quick Baidu search!¹⁰⁵ The county went on to use Jimo's 'experience', though with a degree of adaptation (Interview no. 48).

Counties must continually make decisions as to appropriateness of specific reforms. As in Chapter Six, Taoshan largely perceives Meijiang's model of implementation as not worth studying – economically, the counties are too far apart for Meijiang's experience to be useful. Pieke has described cadre perceptions of their role in a village in which he conducts fieldwork in Yunnan. Drawing on Bourdieu, he describes local cadres as having a "specific *habitus* that privileges certain responses, reactions, and decisions over others". In doing this, he sees cadres' thinking as oriented in relation to a "nationally standardized discourse of 'models of [economic] development'", though he says such models are treated by his interviewees as '*doxa*' and uncritically applied (Pieke 2004, p. 530).¹⁰⁶ In this case, models *limit* thinking, by emphasising a 'development is elsewhere' mentality. The importance of *habitus* and *doxa* in Pieke's

¹⁰⁵ Baidu is a popular Chinese search engine.

¹⁰⁶ Pieke cites specifically "the Wenzhou model (predominantly private enterprise), the Jiangsu model (predominantly collective enterprise), and the Zhujiang model (predominantly foreign investment)" – models which, "[a]t the national level these models are often mere propaganda tools or easy but messy analytical concepts" (Pieke 2004).

use is that they are restrictive: cadres' behaviour is relatively rigidly patterned, and supra-local reform discourses (here, models) are "taken for granted and unreflexively applied".

My fieldwork shows a very different picture, which I describe as a 'conversation' of reform (Gudeman and Rivera 1990). A 'conversation' in which multiple discourses, practices and models circulate, are applied, changed, and re-applied, is a much more *open* reading of reform and sources of reform initiative than Pieke's use of *doxa*. As elsewhere in this thesis, I see the *quantity* of sub-national policy development and innovation as very large. *Conversation* is an attempt to deal with a very messy picture: peripheral practices discussed here are clearly part of an overall process of national reform, despite their shortcomings and despite the fact that most do not explicitly become part of national policy or an avowed component of central reform discourse. I have argued that we should not see most local solutions as 'best practice'; rather, we should see them as extremely situational 'second best' solutions, if seen in terms of allocative efficiency. Nevertheless, they form part of a larger conversation of reform: it is impossible to draw a boundary between the discourses of reform in a county implementing the NCMS and related central discourses. As Gudeman and Rivera phrase it, the notion of 'bounded communities' does not hold where there is a "link between inscriptions of the past and practices of the present [...] and between theorists of the core and practitioners of the periphery". In a moment of beautiful analysis, these authors discuss the language used by Colombian smallholders in talking about their economic practices: "Their carefully used words startlingly called forth the texts of Adam Smith and John Stuart Mill, of Aquinas and Aristotle, and of many others, so that these forced themselves upon the conversation" (Gudeman and Rivera 1990, p. 14). In various fieldwork encounters, I have witnessed county NCMS managers tongue-tied in talking about local reforms as they grapple with a not-yet-familiar vocabulary of health systems and economics, alternating this with a more routine vocabulary of Chinese administrative practice. The rationalisation of global budgets as *baogan* (contracting) systems is an example of this.

In contrasting 'conversations' and Pieke's use of *doxa*, I want to stress the geographical plurality of this process. In Pieke's analysis, China's reform experiences are consolidated into a limited number of tropes which, among his interviewees, limit thought. In my cases, I see changing 'imaginative geographies' of reform, in which multiple stylised representations, both national and sub-national, are mingled and in which localities retain significant scope for interpretation and re-interpretation. Said describes "a set of representative figures, or tropes" underlying (Orientalist) discourse, though which,

"[s]omething patently foreign and distant acquires a status more rather than less familiar. One tends to stop judging things either as completely new or as completely well known; a new median category emerges, a category that allows one to see new things, things seen for the first time, as versions of a previously known thing" (Said 2003, p. 58).

This is the process visible in the changing framings of oversight, cost control and provider payment reforms described in my empirical chapters, in which newly-acquired 'representative figures' are used to re-rationalise administrative goals and practices, to change or upgrade managerial and policy repertoires. With the caveat that the broad range of forms of implementation of the NCMS is unknowable, it seems striking the degree to which multiple places converge on a relatively small number of policy solutions, visible, for example, in the multiple places that have developed oversight solutions similar to Taoshan's: there is, in Fish's term, a degree of 'stability of interpretation' (Fish 1980). Seen in this light, provider payment reforms such as that developed in Lufeng are attempts to reconfigure China's imaginative geography, at least as regards development of the NCMS. How these attempts function is clearly variable: Lufeng's model has had a lot of coverage and has changed local understandings but, at least in my fieldwork counties, its direct impact has been limited as NCMS managers opt for simpler solutions.

9.3.4 Synthesis

This chapter develops the analysis started in Chapters Two and Four, and for which empirical evidence is provided in subsequent empirical chapters: counties develop, not just implement, policy. This is systemically expected, and the rationale underlying this is both Hayekian and Schumpeterian: counties are expected to adapt policy locally and it is hoped they will develop policy solutions of supra-local significance. There is a large degree of consensus over the roles of counties in the policy implementation/development process, within which counties have a large degree of freedom, or discretion, as to the means of reform, if not the ends. The developmental role of counties is not simply a question of voluntarism, as in Göbel or Chung; nor is the unique contribution of localities to the policy development process that of a brake on reform, as in Göbel's reading of O'Brien and Li. Rather, as in Chapter Five, counties are forced to implement the NCMS through the setting of top-down targets and the operation of the NCMS fund system. This corresponds to the 'control' function of the CRS and policy as discussed in Chapter Four. During implementation, however, indeterminate targets and policy create a disjuncture in the policy process which I describe as a motor for local production of variant practices and reforms. This corresponds to the 'communication' function of the CRS and policy. Both of these functions are essential to an understanding of the NCMS, and this understanding must be phrased in terms of both implementation *and* innovation.

This chapter has formalised an analysis of policy interpretation given in previous chapters, to show the textual/discursive basis of production of differential practices and the range of ideas in circulation on which counties draw, and to which they contribute, in an ongoing conversation of reform. Counties are not rational agents developing best practice policy solutions; rather, they exist in a sea of ideas and discourses about ends and means of reform. This stress on ideas builds on my analysis of control in previous chapters. As Kingdon states, ideas are important: "[p]olitical scientists are accustomed to such concepts as power, influence, pressure and strategy", but this is insufficient to understand public policy: "[t]he contents of the ideas themselves [...] are integral parts of decision making in and around government" (Kingdon 1984, p. 125). The mediating

term in my analysis, the 'content' of the ideas and the expression of this in policy and locally-developed mechanisms, is *form*. Underlying this is a belief in institutional pluralism: "effective institutional outcomes do not map into unique institutional designs [...] there is no unique mapping from *function* to *form*" (Rodrik 2007, p. 190, emphasis mine). Ideas, or institutional form, are a core concern of innovation, from an understanding of 'combinations' in Schumpeter (1934) to 'new ideas that work' (Mulgan and Albury 2003). In the case of the NCMS, counties implement policy, but the *form* of that implementation exists largely outside policy, rigidly defined, and more in a loosely articulated sphere of representation.

Chapter Ten

Conclusions: New ideas that work?

Chapter Nine concludes an analysis started in Chapters Two and Four and developed through subsequent empirical chapters. The analysis in these chapters has shown, in the case of the NCMS, a distinct policy implementation process, in which counties are expected to play an important role. Local policy development is not (exclusively) voluntaristic; rather, it is systemically required and embedded and is a characteristic of China's 'state system' (Huang, 2010a). This conclusion draws together evidence from the previous chapters with arguments most fully developed in Chapter Eight to address the question of the usefulness of sub-national practices I have examined. If innovation is about creating 'new ideas that work' (Mulgan and Albury 2003), one must ask whether these ideas do, in fact, 'work'. I examine these practices in relation to the dual Hayekian/Schumpeterian framework set out in Chapter Four, questioning their usefulness both in terms of allowing decentred or adaptive management of the NCMS, and in terms of allowing development of novel policy solutions capable of contributing to the dynamic policy development process espoused by the centre. I also question the degree to which practices examined here have relevance for other countries and health systems, and their likely significance in changing overall policy paradigms, whether Chinese or global.

10.1 Adaptive development or making good central policy?

I have argued at length that counties develop policy and innovate under national policy frameworks, and this is expected and, to a large extent, structurally inevitable. This claim does not mean that county practices are necessarily 'good': we should expect a range of practices to emerge from implementation structured in this way. We should expect deviance, shirking, feigned compliance, thoughtless implementation and all the other problems the literature would make us expect. We should also, however, expect the kinds of practices seen here. In previous chapters, I have argued that examples I

consider are not 'best practice' solutions to problems of China's healthcare reform: they are limited, highly contextually dependent, second best solutions to complex problems. This observation holds across both chapters which examine local cases and is addressed most thoroughly in the case of the SRR. In that chapter I also advanced an analytical framework to deal with this, arguing that SRR is an 'appropriate' development of policy, given capacity constraints, uncertainty over funding and the structure within which counties are required to implement policy. I argued, in vocabulary from North, that while SRR was not 'allocatively efficient', it had played a role in development of the NCMS overall and that one should set this mechanism's importance in underpinning adaptive efficiency against its second best allocative efficiency. I argued that this can be considered Schumpeterian, in that performance over time and adaptation are prioritised over achieving optimal efficiency in the here and now.

Analysing local development of policy solutions in terms of both allocative and adaptive efficiency is useful, but leads to a theoretical impasse: where should the balance lie? I clarify the bases of the institutions and processes visible in county-level implementation/development of the NCMS, but it is impossible to draw up a balance sheet and meaningfully weigh one factor against the other. I argued in relation to the SRR that this was a stepping stone, and the argument for adaptive efficiency is robust in this case. At some level, the same judgement can be made regarding both Taoshan's and Feitian's reforms: while these are less obviously stepping stones, locally at least, they were important mechanisms in ensuring scheme functioning and this is enough to allow a broadly positive evaluation on the grounds of adaptive efficiency. The question that arises, though, is whether this could have been different.

Three separate points stand out. First, there is a legitimate question as to the degree of local design and the degree of central specification necessary in a reform process such as this. I have argued that a central stress on county initiative relies on two related notions: development in line with local conditions (*yin di zhi yi*) and the potential of counties to contribute to overall reforms. One of these is a Hayekian argument about

decentralisation; the other is a Schumpeterian argument about adaptation. None of the cases I have examined, however, clearly falls into either of these categories. None is so location-specific that it required input that could only be found locally, and neither do counties where I carried out fieldwork see themselves as pioneering local reformers. Most are seemingly pleased that their reforms have gained some recognition (not least because this confirms the legitimacy of these), but none set out, it seems, to develop ways of working that would be nationally applicable – all set out to develop responses to local implementation problems. Moreover, all are responding to problems that should be considered generic within the NCMS: in early periods of implementation, the NCMS operated on reimbursement rates that were lower than necessary, even given low levels of scheme funding, as capacity-constrained counties hoarded funds rather than risk overspending within a risk-based pressurised system (cf. SRR). Similarly, costs have risen very significantly, partly at least, due to the lack of serious scheme control of providers' prescribing/treatment behaviour (cf. Feitian), and oversight of providers has been less effective than it could have been (cf. Taoshan). These are deficiencies in scheme design. It is reasonable that some aspects of scheme design should require local piloting and development. Provider payment reforms are likely a case in point: how to create functioning systems to ensure that patients are treated well and at low cost is a complex empirical question. In contrast, Taoshan's oversight system and Feitian's cost control reforms revolve around, in one way or another, 'making good' central policy. They are appropriate local innovations, or policy developments, but it is less clear that such systems needed to be developed locally.

Second, it is reasonable to ask whether counties *should* have been given this large a role in such a technically complex area. As elsewhere in this thesis, doubts have been expressed about the suitability of requiring that something this technically complex be carried out at the county level. In 2002-2003, China started development of the NCMS more or less from scratch. Pre-reform, 'old', CMS is very much a part of the discourse of China's health reform, but most county Bureaus of Health now responsible for implementing the NCMS had little capacity in running health insurance schemes by the

early 2000s. Despite this, they were expected to play a very considerable role in both managing and developing this policy. Aside from questions of fund pooling, in which there is consensus favouring increasing the degree of centralisation (Wagstaff, Lindelow et al. 2009b), the degree of centralisation/decentralisation (and capacity) necessary in any given policy area is something for empirical determination (Tang and Bloom 2000). All my cases show the hallmarks of limited capacity, of peripherality, and of counties attempting to overcome this with the means at their disposal. In showing counties trying hard to implement central policy despite constraints, these cases provide a corrective to the control-centred, good centre vs. deviant localities, 'above there is policy, below there is counter-policy' default analytical position on central and local roles, though my research does not give an adequate basis on which to judge overall whether this is, in terms of allocative efficiency, a success.

Third, county mechanisms I have discussed are coloured, to a degree at least, by their origins in the pressurised system: they are 'ambiguous' – this is particularly visible in Feitian's cost control, which was both about increasing benefit of the scheme *and* reducing implementation risk. The risk framing of local reforms, carries over, in other words, from the overall structuring of implementation in terms of pressure and the localising of risk, and has consequences for the types of reforms that are produced, though in all cases, I should stress, counties I examine appear to be acting in good faith.

10.2 Differentiated responsibilities?

Addressing these issues is an empirical matter. I have addressed these for the specific cases I have examined, but an examination at the level of the NCMS overall is impossible, for reasons discussed in Chapter Five. The ability of counties to develop workable policy solutions appears, from central discourse (and provincial interviews), to be an article of faith, and this sits uneasily with the arguments above and with the empirical evidence presented in previous chapters. Both Li and Cao pose the question of responsibility of different levels of government in China's reforms. Li argues that the

centre must both enable and regulate:

“Central actors have a responsibility in (re)defining the major parameters of behaviour, and specifying the new rules of the game [...] It is insufficient for the central actors to announce the reform objectives, only to leave the specific mechanisms as to how to get there entirely to the provincial and local actors, on the pretext that all details are ‘routine’ implementation matters” (Li 2006a, p. 173).

Cao argues that localities have a great degree of freedom in the way they govern locally as long as they do not contravene the overall direction of central policy, but that the corollary of this is that policy solutions are negotiated locally.¹⁰⁷ Cao examines local land conversion practices and argues that a range of outcomes are visible, and that these are conditioned by processes of accommodation between government and citizens at the local level, in which the centre rarely intervenes, which he sees as a central risk-reduction strategy (Cao 2011).

These analyses coincide with Huang’s notion of ‘centralised minimalist’ government:

“if we look at the past 30 years of Reform from a history-of-practice perspective, a crucial point that emerges is that the Right’s economic miracle and the Left’s social crisis have both come from one and the same source – namely, the distinctive state system (*guojia tizhi*) of the Reform period” (Huang 2010a, p. 116).

Huang sees this state system as having its roots in an imperial tradition of ‘centralised minimalism’, “a minimalist state under highly centralized imperial power that resorted widely to ‘semiformal governance’ by the use of quasi-officials drawn from the communities and by the state bureaucracy’s interfering only in the event of disputes”, a

¹⁰⁷ Cao’s thesis revolves around a separation of governing of officials (*zhi guan*) and governing the people (*zhi min*). For him, the role of the centre (and proxies) is in governing officials, through the personnel system. Local officials, however, actually manage local affairs. The system functions to control people, rather than specify of tasks. Cao’s analysis is reminiscent of Landry (2008).

pattern of government (governance) that he claims extends through the Republican era and beyond. The point of this critique is the existence of a large realm of only minimally regulated activity: Huang describes Imperial China as

“very high in despotic power, given the concentration of authority in the person of the emperor, but low in infrastructural power: the official bureaucratic apparatus reached only the level of the county magistrate” (Huang 2008, p. 23).

Underlying all these arguments is something that very closely mirrors the under-specified space and the logic of ‘results orientation’ at the centre of policy implementation/innovation, at least in the case of the NCMS. Tied up with this is a question of responsibility – a realisation that the policy space left under-specified can have both positive and negative consequences.

Very prosaically, though, one can ask what the alternative was. First, could China have built sufficient capacity at the county level *prior to* initiation of the NCMS? This seems unlikely, both given the number of counties in which the NCMS was to be rolled out, the complexity of the issues involved, and the speed of the development of the scheme (‘pilots’ to national roll out in five years). It seems doubly unlikely given that the NCMS was a prelude to other, harder, reforms, such as reform of hospital management. Had it been possible to carry out greater *pre hoc* capacity building for the NCMS, would it have also been possible for all the other reforms that followed in its wake? Second, could more money have been made available for management and could implementation have been overall better guided, helping to ensure that locally produced policy solutions were closer to being ‘allocatively efficient’? (Just because we recognise the importance of adaptive efficiency doesn’t mean that we shouldn’t aim for greater allocative efficiency.) Probably. Third, could elements of the reform have been better designed in advance? Almost certainly. These are all clear criticisms and have largely been expressed in the broader literature.

Saying this, in 2003 very few counties in China ran any kind of rural health insurance

scheme or had any experience of this. Ten years on, rural health insurance has been rolled out nationwide, covering around 2,700 counties, and huge parallel health reforms are underway. This is a very great achievement, in which experimentation and piloting as well as top-down implementation have all played a role. Counties have played a very important role in this process, both in implementing and developing the scheme, despite constraints, and because of their position within China's 'distinctive state system'. Seen from the county level, or at least from my fieldwork counties, this has not been a straightforward process: on the contrary, it has been fraught and difficult.

10.3 Learning from China?

Within the local reforms examined here, are there mechanisms which could be considered useful for other countries in the process of health system reform? This is unlikely, though I see this question as less important than the contextual transitional significance of the cases I have examined. More broadly, though, the type of institution I have examined here raises questions about the generalisability and replicability of China's reform experience. Naughton, for example, asks what lessons can be drawn from the broad sweep of China's reform period and what other countries can learn from China's experience. While acknowledging China's dynamism, he is skeptical about the usefulness of China's experience for other countries:

"among all the remarkable institutional improvisations in China, among all the institutional solutions that China devised to tricky transitional problems, there is none that is so successful, and so robust to context, that we would feel comfortable recommending it to other countries" (Naughton 2009, p. 8).

The important idea here is *context*. All the reforms examined here are peripheral, appropriate, and highly contextual. If we accept that context is of unavoidable and foundational importance in shaping such institutions, we must question their 'transferability'. This reasoning militates against a 'search for solutions' or a 'what works' view of Chinese reform processes: what works in one case may not be what

works elsewhere and at other times, often for very mundane and highly contextually-specific reasons. In the foreword to a special issue of *Health Economics* in 2009 reporting on China's health reforms, Hsiao and Maynard state that,

“[b]y innovating and translating theory into practice, China is creating a research laboratory offering rich opportunities for research and the improvement of the evidence base about ‘what works’ efficiently in health care” (Hsiao and Maynard 2009, p. S2).

While the importance of the trialling, research and experimentation going on now in China is unchallengeable, we should be cautious about the degree to which Chinese experiences are generalisable, for all the reasons given here. This is not a pessimistic view, but a positive statement which should direct attention to contextualised *processes* of institutional innovation and learning, rather than a restricted focus on outcomes or observable institutions or policies.

10.4 Where have all the innovations gone?

In Chapter Eight, I argued that it is not necessary to be at the technical leading edge in order to innovate. All the local reforms examined here can be considered second best policy solutions but, as above, this should be balanced against their importance in system adaptation. A parallel question exists, however, regarding the possibility of local reforms changing overall policy paradigms (Hall 1993; also Kuhn 1996). I have argued that we can distinguish between within-system (*tizhi nei*) and outside-the-system (*tizhi wai*) initiatives in Chinese reform discourse, and that the first corresponds to what is often called ‘continuous’ or ‘marginal’ innovation and the second to ‘discontinuous’ or ‘breakthrough’ innovation. By definition, discontinuous, outside-the-system, reforms are more likely to have paradigmatic significance than within-system reforms. In all cases examined here, counties are innovating, but their innovation is profoundly within-system. In Chapter Four, I argue that the failure of much analysis to clearly specify terms and clarify what is meant by ‘innovation’ leads to problems of mis-identification of much local government activity. The more significant point,

however, is the high degree of conformity of local practice with systemic ends and the absence of radical ruptures.

China's pre-reform health system gained great international attention for delivering low cost, broad coverage healthcare and was held up as a model for developing countries, as well as attracting the attention of developed country observers looking for alternatives to technologically-driven approaches to delivery of healthcare (Sidel and Sidel 1983). One of the most significant critics of institutionalised, technological medicine of the time was Ivan Illich (1975), who argued that medicine as an institution has progressively 'medicalised' pain and suffering: that all societies and cultures have means of coming to terms with, and dealing with, suffering, but that "[m]edical civilisation replaces the culturally determined competence in suffering with a growing demand by each individual for the institutional management of his pain". Removed from a larger cultural framing of suffering,

"pain has come to pose only a technical question for industrial man – what do I need to get in order to have my pain managed or killed? If the pain continues, the fault is not with the universe, God, my sins, or the devil, but with the medical system. Suffering is an expression of consumer demand for increased medical outputs. By becoming unnecessary, pain has become unbearable" (Illich 2003, p. 920).

Institutional and technological management of health, in other words, has reduced individuals' autonomous ability to deal with suffering: "The new suffering is not only unmanageable, but it has lost its referential character. It has become meaningless, questionless torture" (Illich 2003, p. 920). While this framing of suffering and the role of medicine is far from mainstream, Illich remains a significant, and by no means unique, critic of institutionalised medicine. Roy Porter, a prominent medical historian, for example, argues that "medicine has become the prisoner of its success. Having conquered many grave diseases and provided relief from suffering, its mandate has become muddled. What are its aims? Where is it to stop?". For Porter, this is an institutional problem:

“[t]hanks to diagnostic creep or leap, ever more disorders are revealed [...] The root of the problem is structural. It is endemic to a system in which an expanding medical establishment, faced with a healthier population, is driven to medicalising normal events like menopause, converting risks into diseases, and treating trivial complaints with fancy procedures” (Porter 1997, pp. 717-718).

There is, surprisingly, common ground between Illich, Porter and other critical institutional analyses and mainstream analyses that question the sustainability of health spending in developed countries due to ever-increasing technological advances. According to a recent McKinsey Quarterly analysis,

“Throughout the world, leaders of government health agencies, heads of health care companies, and even patients [...] behold the growth of health care spending with alarm [...] If current trends persist to 2050, most OECD countries will spend more than a fifth of GDP on health care. By 2080 Switzerland and the United States will devote more than half of GDP to it – and by 2100 most other OECD countries will reach this level of spending” (Drouin, Hediger et al. 2008, p. 1).

The point here is not to seriously question the role that medicine or the medical professions should play in the creation and management of health, but rather to clarify the *type* of solutions that I have examined in this study. In stating in Chapter Nine that what we see being developed are within-system solutions and pointing to the discursive continuity between local policy development and central intentions, my aim was to clarify county and central roles and show the importance of role differentiation for the process of reform. The reverse side of this, however, is that this discursive continuity to a large extent precludes paradigm-busting innovations that might, for example, lead to a genuine re-framing of health and the management of health.

Seen in this way, conclusions must be ambiguous: what I have shown are small, peripheral, solutions which contribute in small, highly contextually dependent, ways to

China's adaptive policy development process. This is a good thing. The appropriate and peripheral policy solutions examined here fall short of the technical leading edge solutions that Roper et al. or Hsiao and Maynard would like to see, for reasons I have dealt with in detail. This is understandable. They also, however, fail to translate into genuinely discursively transgressive innovations capable of reframing overall debates. Within-system local innovation, from evidence presented here, is clearly an important component of China's reform, but its contribution to larger, boundary-spanning, and radical change is likely to be limited.

10.5 Concluding comments

Starting in Chapter Two, I argued that existing approaches to central-local relations and policy implementation fail to account for the range of local action I see within the NCMS. The following chapters have followed on from this to develop an analytical framework to account for the NCMS as both national and sub-national system and for the local reforms I studied in detail. My analysis is an addition to existing literature, but does not preclude other analytical approaches. My focus is limited to one policy area and, largely speaking, to government functioning at the county level. Other policy areas and analysis of other levels of government will certainly show different features. Saying this, the county is a very important level of government, possibly the most important in many cases, and dynamics of implementation/innovation within the NCMS are unlikely to be a world away from dynamics of other social policy programmes.

In Chapter Five, I argued that 'the NCMS' is not graspable in its entirety, and this makes for very messy analysis. Two implications of this are a focus on mechanisms and on discrete 'moments' in reform. My focus on mechanisms is important for allowing analytical clarity, but masks as much as it reveals: in management terms, and discursively, mechanisms are important. This focus, though, risks masking the fact that while these are important, they are, on their own and in the absence of proper coordination, insufficient to bring about substantial change in operation of any given county's health system – these are small components of much larger and very complex

reform processes, the outcome of which must, finally, be judged on access to affordable and effective healthcare. The reforms I analyse here are all limited in scope: all are reforms that could be managed by a BOH acting more or less independently; none spans boundaries between policy areas or coordinates health management with institutionally separate issues such as environmental management (Holdaway 2010), or was part of a concerted drive to develop a leading-edge county model. Such a 'whole county' model of NCMS functioning would have to be much more coordinated than any of the local reforms presented here. Qianjiang's health reform is a likely candidate, having been extremely purposively developed from a very good start in the 1990s with seed money and technical inputs under the Health VIII Programme (see note, p. 252) and shows a high degree of coordination of reforms across areas of local health system functioning. Without the opportunity for detailed fieldwork, however, it is not possible to assess this. Equally, this study really only grasps a limited number of 'moments' in development of the NCMS. All the examples I analyse are very contextually and historically dependent and their future usefulness is unclear, given changes in the scheme overall, and changes in the way that the scheme interfaces with health reforms more generally.

My analysis would have almost certainly been different if I had conducted fieldwork in different counties, been differently guided in early stages of research, or examined different mechanisms. Other local reforms, many more systemically significant than those I examine, are also taking place, as are systemically-driven experiments. Broadly speaking, though, my county examples show implementers – county BOH and NCMS managers – working within a very demanding national policy framework and trying to make the scheme work despite systemic limitations and scant design. To a large extent, this is related to my choice of fieldwork sites, most of which are not rich, and are 'peripheral' to policy debates and national experiments. They do not universally do a good job, and all local reforms here should be considered second best, but in their attempts to solve complex systemic problems, they show a better face of local government in China than the broad literature on this would lead us to expect.

Appendix

List of interviews

Interview no.	Date	Indicative description
01	05/2009	Academic, UK
02	06/2009	Academic, UK
03	10/2009	Academic, China
04	10/2009	Academic, China
05	11/2009	Academic, China
06	11/2009	Academic, China
07	11/2009	Academic, China
08	11/2009	Academic, UK
09	11/2009	Academic, China
10	11/2009	Academic, China
11	11/2009	Academic, China
12	12/2009	Academic, China
13	12/2009	Academic, China
14	04/2010	Academic, China
15	04/2010	Academic, China
16	05/2010	Academic, China
17	01/2011	Academic, China
18	06/2011	Academic, China
19	06/2011	Academic, China

20	05/2010	Official(s), Taoshan
21	05/2010	Official(s), Taoshan
22	05/2010	Official(s), Taoshan
23	05/2010	Officials/hospital staff/doctor(s), Taoshan
24	05/2010	Official(s), Taoshan
25	05/2010	Official(s), Taoshan
26	05/2010	Officials/hospital staff/doctor(s), Taoshan
27	05/2010	Official(s), Taoshan
28	05/2010	Official(s), Taoshan
29	10/2010	Official(s), Taoshan
30	10/2010	Official(s), Taoshan
31	10/2010	Official(s), Taoshan
32	10/2010	Official(s), Taoshan
33	10/2010	Official(s), Taoshan
34	10/2010	Official(s), Taoshan
35	10/2010	Official(s), Taoshan
36	06/2010	Official(s), Meijiang
37	06/2010	Officials/hospital staff/doctor(s), Meijiang
38	06/2010	Official(s), Meijiang
39	06/2010	Officials/hospital staff/doctor(s), Meijiang
40	07/2011	Official(s), Meijiang
41	07/2011	Official(s), Meijiang

42	07/2011	Official(s), Meijiang
43	07/2011	Official(s), Meijiang
44	07/2011	Officials/hospital staff/doctor(s), Meijiang
45	07/2011	Official(s), Meijiang
46	07/2011	Official(s), Meijiang
47	07/2011	Officials/hospital staff/doctor(s), Meijiang
48	06/2010	NCMS official(s), Province X
49	06/2010	Official(s), County P
50	11/2010	Official(s), Feitian
51	11/2010	Official(s), Feitian
52	11/2010	Official(s), Feitian
53	11/2011	Officials/hospital staff/doctor(s), Feitian
54	11/2010	Official(s), Feitian
55	06/2011	Official(s), Feitian
56	06/2011	Official(s), Feitian
57	06/2011	Official(s), Feitian
58	06/2011	Official(s), Feitian
59	06/2011	Officials/hospital staff/doctor(s), Feitian
60	06/2011	Official(s), Feitian
61	06/2011	Official(s), Feitian
62	06/2011	NCMS official(s), Province Y
63	06/2011	NCMS official(s), Province Y

64	07/2011	Official(s), Qianjiang
65	07/2011	Officials/hospital staff/doctor(s), Qianjiang
66	07/2011	Hospital staff/doctor(s), Qianjiang
67	07/2011	Official(s), Qianjiang
68	07/2011	Hospital staff/doctor(s), Qianjiang
69	07/2011	Hospital staff/doctor(s), Qianjiang
70	07/2011	Hospital staff/doctor(s), Qianjiang
71	07/2011	Hospital staff/doctor(s), Qianjiang
72	07/2011	Officials/hospital staff/doctor(s), Qianjiang
73	07/2011	Officials/hospital staff/doctor(s), Qianjiang
74	07/2011	Official(s), Qianjiang
75	07/2011	Official(s), Qianjiang

List of references

1. Chinese policy documents

Feitian County C Township (2007). Notice on forwarding of Feitian NCMS regulations.

Caijian (2004) No. 91. *Caizheng Bu guanyu yinfa 'Caizheng Jiancha Zhuanyuan Banshichu dui Zhongyang Caizheng Xinxing Nongcun Hezuo Yiliao buzhu zijin shenhe jiandu caozuo guicheng' de tongzhi* (Ministry of Finance Notice on the issuing and circulation of Rules on the operation of the Financial Oversight Office audit of Central Finance support funds for the New Cooperative Medical Scheme), 06/12/2004.

Caishe (2003) No. 112. *Caizheng Bu, Weisheng Bu guanyu Zhongyang Caizheng zizhu Zhong Xi Bu diqu nongmin canjia Xinxing Nongcun Hezuo Yiliao zhidu buzhu zijin bofu you guan wenti de tongzhi* (Notice of the Ministry of Finance and the Ministry of Health regarding the payment of Central Finance funds supporting peasants' participation in the New Cooperative Medical Scheme in Central and Western Areas), 25/08/2003.

Caishe (2004) No. 96. *Caizheng Bu, Weisheng Bu guanyu jianli Xinxing Nongcun Hezuo Yiliao fengxian jijin de yijian* (Opinions of the Ministry of Finance and Ministry of Health on building the New Cooperative Medical Scheme risk fund), 22/10/2004.

Caishe (2008) No. 8. *Caizheng Bu, Weisheng Bu guanyu yinfa Xinxing Nongcun Hezuo Yiliao jijin caiwu zhidu de tongzhi* (Notice of the Ministry of Finance and Ministry of Health on the printing and distribution of the New Cooperative Medical Scheme finance system), 22/01/2008.

Caishefa (2007) No. 5. *Caizheng Bu, Weisheng Bu guanyu tiaozheng Zhongyang Caizheng Xinxing Nongcun Hezuo Yiliao zhidu buzhu zijin bofu banfa you guan wenti de tongzhi* (Ministry of Finance and Ministry of Health Notice on questions relating to adjusting payment of Central Finance contributions to the New Cooperative Medical Scheme), 31/01/2007.

Caizhuganjian (2008) No. 39. *Jiangxi Zhuanyuanban guanyu 2008 niandu Zhongyang Caizheng Xinxing Nongcun Hezuo Yiliao buzhu zijin shenhe gongzuo de zongjie baogao* (Summary report of the Jiangxi *Zhuanyuanban* 2008 audit of Central Finance assistance to the New Cooperative Medical Scheme), 20/05/2008.

Feitian County Bureau of Health (2006). Feitian County NCMS procedures (2006 revised version).

Feitian County Bureau of Health (2007). Report on adjustment of the NCMS plan. Unpublished report.

Feitian County Bureau of Health (2008). Notice on the issuing of Feitian NCMS cost control audit procedures. Unpublished report.

Feitian County Bureau of Health (2009a). Report on the adjustment of the NCMS plan in 2010. Unpublished report.

Feitian County Bureau of Health (2009b). NCMS cost control. Unpublished PowerPoint presentation.

Feitian County Bureau of Health (2010a). Feitian County NCMS cost control provider responsibility contract. Unpublished document.

Feitian County Bureau of Health (2010b). Implementation of NCMS cost control work. Unpublished report.

Feitian County Bureau of Health (2010c). Report on NCMS work in 2010. Unpublished report.

Feitian County Bureau of Health (2011a). Feitian County NCMS cost control provider responsibility contract. Unpublished document.

Feitian County Bureau of Health (2011b). Feitian County New Cooperative Medical Scheme Operating procedures (2011 revision). Unpublished document.

Feitian County Bureau of Health and Bureau of Finance (2008). Opinions on strengthening NCMS fund management. Unpublished document.

Feitian County Chinese Medical Hospital (2010a). Notice on controlling average and per day inpatient costs. Unpublished document.

Feitian County Chinese Medical Hospital (2010b). Notice on supplementary policy on controlling average and per day inpatient costs. Unpublished document.

Feitian County People's Government (2003a). Interim New Cooperative Medical Scheme operating procedures.

Feitian County People's Government (ND). Implementing NCMS cost control reforms. Unpublished report.

Feitian County People's Government (2003b). Opinions on setting up the New Cooperative Medical Scheme.

Feitian County People's Government (2010). NCMS cost control. Unpublished PowerPoint presentation.

Ganheyibanzi (2006) No. 20. *Guanyu kaizhan 2006 nian Xin Nong He di er ci buchang gongzuo de tongzhi* (Notice on 2006 NCMS Second Round Reimbursement work), 28/09/2006.

Guobanfa (2003) No. 3. *Guowuyuan Bangongting zhuanfa Weisheng Bu deng bumen guanyu jianli Xinxing Nongcun Hezuo Yiliao zhidu yijian de tongzhi* (Notice of the Office of the State Council on the forwarding of the Opinions of the Ministry of Health and other ministries regarding the building of the New Cooperative Medical Scheme), 10/01/2003.

Guobanfa (2004) No. 3. *Guowuyuan Bangongting zhuanfa Weisheng Bu deng bumen guanyu jin yi bu zuo hao Xinxing Nongcun Hezuo Yiliao shidian gongzuo zhidao yijian de tongzhi* (Notice of the Office of the State Council on the forwarding of the Guiding Opinions of the Ministry of Health and other ministries on advancing New Cooperative Medical Scheme pilot work), 13/01/2004.

Guobanfa (2011) No. 8. *Guoban yinfa yiyao weisheng tizhi wu xiang zhongdian gaige 2011 niandu anpai* (Office of the State Council issue and circulation of organisation of

main work on the five main components of health system reform in 2011), 13/02/2011.

Guobanhan (2010) No. 67. *Yiyao weisheng tizhi wu xiang zhongdian gaige 2010 niandu zhuyao gongzuo anpai* (Organisation of main work on the five main components of health system reform in 2010), 06/04/2010.

Guofa (1993) No. 70. *Guowuyuan zhuanfa Caizheng Bu deng bumen 'Liangshi fengxian jijin guanli zanxing banfa' de tongzhi* (Notice on the State Council's forwarding of the 'Temporary regulations for the management of grain risk funds' of the Ministry of Finance and other departments).

Guofa (2009) No. 12. *Guowuyuan guanyu yinfa yiyao weisheng tizhi gaige jin qi zhongdian shishi fang'an (2009-2011 nian) de tongzhi* (State Council Notice on the issuing and circulation of the near-term (2009-2011) health system reform implementation plan), 18/03/2009.

Hunan Sheng Hezuo Yiliao Ban (2006). *Xinxing Nongcun Hezuo Yiliao jijin guanli* (New Cooperative Medical Scheme fund management), 09/2006.

Jiangsu Sheng Renda Changweihui Gonggao (2011) No. 79. *Jiangsu Sheng Xinxing Nongcun Hezuo Yiliao tiaoli* (Legislation of Jiangsu Province on the New Cooperative Medical Scheme), 24/03/2011.

Lufeng Xian Weisheng Ju (2009a). *Chuangxin jizhi, shenru shixian, nuli kaichuang Xinxing Nongcun Hezuo Yiliao gongzuo xin jumian* (Innovate mechanisms, thoroughly bring about and diligently pioneer a new phase in the New Cooperative Medical Scheme), 12/05/2009.

Lufeng Xian Weisheng Ju (2009b). *Jiji tansuo da dan shixian: Lufeng Xian nuli tuijin Xin Nong He fu fei jizhi gaige* (Experiment enthusiastically and implement courageously: Lufeng County diligently implements reform of NCMS payment reform mechanism), 03/04/2009.

Ningzheng (2007) No. 102. *Guanyu yinfa Ningguo Shi Xinxing Nongcun Hezuo Yiliao guanli banfa de tongzhi* (Notice on the circulation of Ningguo City regulations on

management of the New Cooperative Medical Scheme), 21/11/2007.

Nongweisi (2008). *Weisheng Bu Bangongting guanyu queding Xinxing Nongcun Hezuo Yiliao you guan shidian zhongdian lianxi diqu de tongzhi* (Notice of the Office of the Ministry of Health confirming main linking areas for New Cooperative Medical Scheme pilots), 27/07/2008.

Feitian County S Township People's Government (2003). Notice on forwarding Feitian County New Cooperative Medical Scheme detailed regulations (trial).

Taoshan County Bureau of Health (2008). Implementation plan on the establishment of Taoshan's Inspection Bureau. Unpublished document.

Taoshan County Bureau of Health (2009a). New Cooperative Medical Scheme work handbook, 12/2009. Unpublished document.

Taoshan County Bureau of Health (2009b). Regulations on Taoshan County's New Cooperative Medical Scheme inspection work, Unpublished document.

Taoshan County Bureau of Health (2010). Perfect mechanisms, manage scientifically, promote the effective and stable operation of the New Cooperative Medical Scheme). Unpublished document.

Taoshan County D Sub-Bureau (2009). A letter from D Sub-Bureau to farmers. Unpublished document.

Taoshan County D Sub-Bureau (2010a). New Cooperative Medical Scheme inspection document: Case decision, No. XXXX. Unpublished document.

Taoshan County D Sub-Bureau (2010b). New Cooperative Medical Scheme inspection document: Announcement on reform, No. XXXX. Unpublished document.

Taoshan County D Sub-Bureau (2010c). New Cooperative Medical Scheme inspection document: Announcement on reform, No. XXXX. Unpublished document.

Tianjin Shi Laodong he Shehui Baozhang Ju (2009). *Cheng-xiang jumin jiben yiliao baoxian* (Urban and rural residents' basic health insurance), 02/06/2009.

Weibanjifufa (2003) No. 147. *Weisheng Bu Bangongting zhuanfa Yunnan Sheng Weishengting guanyu Wenshan Zhou Guangnan Xian A'ke Xiang Xinxing Nongcun Hezuo Yiliao shidian gongzuo zhong bu dang zuofa de qingkuang tongbao de tongzhi* (Notice of the Office of the Ministry of Health on the forwarding of the Yunnan Bureau of Health report on inappropriate practices in New Cooperative Scheme pilot work in A'ke Township of Guangnan County, Wenshan Prefecture), 25/11/2003.

Weibanjifufa (2003) No. 47. *Weisheng Bu Bangongting guanyu zuo hao Xinxing Nongcun Hezuo Yiliao shidian gongzuo de tongzhi* (Notice of the Office of the Ministry of Health on doing New Cooperative Medical Scheme work well), 23/07/2004.

Weibannongweifa (2004) No. 46. *Weisheng Bu Bangongting guanyu chengli Weisheng Bu Xinxing Nongcun Hezuo Yiliao jishu zhidao zu de tongzhi* (Notice of the Office of the Ministry of Health on setting up a New Cooperative Medical Scheme technical guidance group), 01/04/2004.

Weibannongweifa (2009) No. 108. *Weisheng Bu Bangongting guanyu zuo hao 2009 nian xia ban nian Xinxing Nongcun Hezuo Yiliao gongzuo de tongzhi* (Notice of the Office of the Ministry of Health on doing well New Cooperative Medical Scheme work in the second half of 2009), 29/06/2009.

Weibannongweifa (2009) No. 228. *Weisheng Bu Bangongting guanyu yinfa '2009 nian zhong xi bu diqu Xinxing Nongcun Hezuo Yiliao guanli nengli jianshe xiangmu guanli fang'an' de tongzhi* (Notice of the Office of the Ministry of Health on circulation of the '2009 central and western regions New Cooperative Medical Scheme management capacity building project plan), 22/12/2009.

Weibannongweifa (2010) No. 53. *Weisheng Bu Bangongting guanyu guifan Xinxing Nongcun Hezuo Yiliao jijin shiyong guanli de tongzhi* (Notice of the Office of the Ministry of Health on regularising use and management of New Cooperative Medical Scheme funds), 06/04/2010.

Weinongweifa (2005) No. 319. *Guanyu zuo hao Xinxing Nongcun Hezuo Yiliao shidian you guan gongzuo de tongzhi* (Notice on doing New Cooperative Medical Scheme pilot

work well), 11/08/2005.

Weinongweifa (2006) No. 13. *Guanyu jia kuai tuijin Xinxing Nongcun Hezuo Yiliao shidian gongzuo de tongzhi* (Notice on speeding up promotion of New Cooperative Medical Scheme pilot work), 10/01/2006.

Weinongweifa (2006) No. 40. *Weisheng Bu, Caizheng Bu guanyu jiaqiang Xinxing Nongcun Hezuo Yiliao guanli gongzuo de tongzhi* (Notice of the Ministry of Health and the Ministry of Finance on increasing NCMS management work), 25/01/2006.

Weinongweifa (2007) No. 253. *Weisheng Bu, Caizheng Bu, Guojia Zhongyiyao Guanli Ju guanyu wanshan Xinxing Nongcun Hezuo Yiliao tongchou buchang fang'an de zhidao yijian* (Guiding opinions of the Ministry of Health, the Ministry of Finance and the State Administration for Traditional Chinese Medicine on perfecting New Cooperative Medical Scheme pooling and reimbursement plans), 10/09/2007.

Weinongweifa (2007) No. 304. *Weisheng Bu guanyu yinfa 'Quan guo Xinxing Nongcun Hezuo Yiliao tongji diaocha zhidu' de tongzhi* (Notice of the Ministry of Health on the issuing and circulation of a 'National New Cooperative Medical Scheme statistical investigation system'), 18/12/2007.

Weinongweifa (2008) No. 17. *Guanyu zuo hao 2008 nian Xinxing Nongcun Hezuo Yiliao gongzuo de tongzhi* (On doing New Cooperative Medical Scheme work well in 2008), 30/03/2008.

Weinongweifa (2008) No. 65. *Weisheng Bu guanyu guifan Xinxing Nongcun Hezuo Yiliao Er Ci Buchang de zhidao yijian* (Guiding opinions of the Ministry of Health on the standardisation of New Cooperative Medical Scheme Second Round Reimbursement), 04/12/2008.

Weinongweifa (2009) No. 62. *Weisheng Bu guanyu zai sheng ji he shequ shi ji Xinxing Nongcun Hezuo Yiliao dingdian yiliao jigou kaizhan jishi jiebao gongzuo de zhidao yijian* (Guiding opinions of the Ministry of Health on carrying out immediate reimbursement work at province, community, and city-level designated New

Cooperative Medical Scheme providers), 24/06/2009.

Weinongweifa (2009) No. 68. *Guanyu gonggu he fazhan Xinxing Nongcun Hezuo Yiliao zhidu de yijian* (Opinions regarding the consolidation and development of the New Cooperative Medical Scheme), 02/07/2009.

Weinongweifa (2011) No. 52. *Guanyu jin yi bu jia qiang Xinxing Nongcun Hezuo Yiliao jijin guanli de yijian* (Opinions on strengthening management of New Cooperative Medical Scheme funds), 25/05/2011.

Weinongweifa (2012) No. 28. *Guanyu tuijin Xinxing Nongcun Hezuo Yiliao zhifu fangshi gaige gongzuo de zhidao yijian* (Guiding opinions on advancing New Cooperative Medical Scheme provider payment reform work), 15/05/2012.

Weisheng Bu (2008a). *2008 nian 2 yue 18 ri Weisheng Bu lixing xinwen fabu hui shi lu* (Record of the routine press conference of the Ministry of Health of the 18th February 2008), 19/02/2008.

Weisheng Bu (2008b). *Weisheng Bu guanyu guifan Xinxing Nongcun Hezuo Yiliao jiankang tijian gongzuo de yijian* (Opinions of the Ministry of Health on regularising New Cooperative Medical Scheme health check-ups), 24/10/2008.

Weisheng Bu Nongcun Weisheng Guanli Si (2006). *Nongcun weisheng gongzuo jianxun (2006 nian di 13 qi)* (13th Rural Health Work Brief of 2006), 24/11/2006.

Weisheng Bu Nongcun Weisheng Guanli Si (2010). *Nongcun weisheng gongzuo jianxun, di 4 qi (zong di 100 qi)* (Rural health work brief, No. 4 of 2010 (No. 100 overall)), 04/2010.

Weisheng Bu Xinwen Bangongshi (2009). *Weisheng Bu deng wu bumen yaoqiu gonggu he fazhan Xin Nong He zhidu* (Ministry of Health and five other ministries require consolidation and development of the NCMS), 13/07/2009.

X Company Meijiang NCMS Management Centre (2009). *Meijiang City 2010 NCMS reimbursement and management plan*. Unpublished document.

X Province Bureau of Health New Cooperative Medical Scheme Office (2009). X Province New Cooperative Medical Scheme Brief.

Y Province Bureau of Health (2010). Notice on doing NCMS work well in 2011.

Y Province Bureau of Health, Bureau of Finance and Civil Affairs Bureau (2008). Opinions on doing NCMS work well.

Y Province Government Office (2007). Opinions on perfecting the New Cooperative Medical Scheme.

Zhewuibannong (2009) No. 4. *Guanyu zuo hao 2009 nian xia ban nian Xinxing Nongcun Hezuo Yiliao de tongzhi* (Notice on doing well New Cooperative Medical Scheme work in the second half of 2009), 31/07/2009.

Zhongfa (2002) No. 13. *Zhonggong Zhongyang, Guowuyuan guanyu jin yi bu jiaqiang nongcun weisheng gongzuo de jueding* (Decision of the Party Centre and the State Council on continuing to increase rural health work), 19/10/2002.

Zhonggongzhongyang Guowuyuan (2009). *Guanyu shenhua yiyao weisheng tizhi gaige de yijian* (Opinions of the Party Central and the State Council on deepening health system reform), 17/03/2009.

Zhonghua Renmin Gongheguo Shenjishu Bangongting (2011) No. 7. *Zhonghua Renmin Gongheguo Shenjishu shenji jiegou gonggao* (Public announcement of audit results of the National Audit Office of the People's Republic of China), 16/02/2011.

Zhonghua Renmin Gongheguo Weisheng Bu (2012). *Xin Nong He gongzuo 2011 nian jinzhan he 2012 nian zhongdian* (NCMS progress in 2011 and priorities for 2012), 27/02/2012.

2. Other references

Albury, D. (2005). "Fostering innovation in public services." Public Money and Management, January 2005: 51-56.

Anon (2008). "*Yi ren wei ben de kexue neihan* (The scientific meaning of 'taking people as the base')." 01/08/2008. Retrieved March 2012, from *Renmin Wang* (People's Daily Website).
<http://theory.people.com.cn/GB/40557/130316/130317/7600245.html>

Anon (2010). "*Weisheng Bu Xinxing Nongcun Hezuo Yiliao Yanjiu Zhongxin keti yanjiu baogao hui bian* (Compilation of Ministry of Health New Cooperative Medical Scheme Research Centre research reports)." Retrieved March 2012, from *Weisheng Bu Xinxing Nongcun Hezuo Yiliao Yanjiu Zhongxin* (Centre for China Cooperative Medical System).
www.ccms.org.cn/article.aspx?id=240

Anon (2011). "*Di ba jie Zhongguo Zhengfu Chuangxin Jiang luntan ji 'Xingfu Jingyin' Guoji Xueshu Yantao Hui zhaokai* (The Eighth Chinese Government Innovation Prize Forum International Scholarly Discussion Meeting, Happy Jiangyin, opens)." Retrieved March 2011, from *Zhongguo Zhengfu Chuangxin Wang* (China Innovations).
<http://www.chinainnovations.org/Item/24125.aspx>

Babiarz, K. S., G. Miller, et al. (2010). "New evidence on the impact of China's New Rural Cooperative Medical Scheme and its implications for rural primary healthcare: Multivariate difference-in-difference analysis." British Medical Journal, 341: c. 5617.

Bachman, D. (1987). Implementing Chinese tax policy. Policy implementation in post-Mao China. D. M. Lampton, Ed. Berkeley; London, University of California Press.

Barber, S. L. and L. Yao (2010). Health insurance systems in China: A briefing note. Geneva, World Health Organisation.
http://www.who.int/healthsystems/topics/financing/healthreport/37ChinaB_YFINAL.pdf

Barber, S. L. and L. Yao (2011). "Development and status of health insurance systems in China." The International Journal of Health Planning and Management, 26: 339-356.

- Barthes, R. (1977). The death of the author. Image, music, text. S. Heath, Ed. London, Fontana.
- Baum, R. and A. Shevchenko (1999). The 'state of the state'. The paradox of China's post-Mao reforms. M. Goldman and R. MacFarquhar, Ed. Harvard, Massachusetts, Harvard University Press.
- Beck, U. (1992). Risk society: Towards a new modernity. London, Sage.
- Beck, U. (2006). "Living in the world risk society." Economy and Society, 35(3): 329-345.
- Behn, R. D. (1987). "The nature of knowledge about public management: Lessons for research and teaching from our knowledge about chess and warfare." Journal of Policy Analysis and Management, 7(1): 200-212.
- Behn, R. D. (1988). "Management by groping along." Journal of Policy Analysis and Management, 7(4): 643-663.
- Bekedam, H. (2006). Health: Challenges and opportunities in China. Beijing, World Health Organisation.
- http://www.wpro.who.int/NR/rdonlyres/6ED75855-1A7D-42CB-96E5-0F42D8605CD7/0/Henk_Challenges_and_Opportunities_eng.pdf
- Bernstein, T. P. and X. Lü (2000). "Taxation without representation: Peasants, the central and local states in reform China." The China Quarterly, 163: 742-763.
- Bessant, J. (2005). "Enabling continuous and discontinuous innovation: Learning from the private sector." Public Money and Management, January 2005: 35-42.
- Bloom, G. (2011). "Building institutions for an effective health system: Lessons from China's experience with rural health reform." Social Science and Medicine, 72: 1302-1309.
- Bloom, G. and J. Fang (2003). China's rural health system in a changing institutional context. Brighton, Institute of Development Studies. IDS Working Paper.
- <http://www.ids.ac.uk/ids/bookshop/wp/wp194.pdf>
- Bloom, G. and X. Gu (1997). "Health sector reform: Lessons from China." Social Science & Medicine, 45(3): 351-360.

- Bloom, G., L. Han, et al. (2000). How health workers earn a living in China. Brighton, Institute of Development Studies. IDS Working Paper 108.
- Bloom, G. and H. Standing (2008). "Future health systems: Why future? Why now?" Social Science & Medicine, 66: 2067-2075.
- Blumenthal, D. and W. Hsiao (2005). "Privatization and its discontents: The evolving Chinese health care system." The New England Journal of Medicine, 353(11): 1165-1170.
- Bo, J. (2009). A guardian angel for the NCMS, bringing down cost of drugs. Farmers' Daily, Beijing.
- Bondanella, P. (1997). Umberto Eco and the open text: Semiotics, fiction, popular culture. Cambridge, Cambridge University Press.
- Bourdieu, P. (1992). The logic of practice. Cambridge, Polity.
- Bowe, R., S. J. Ball, et al. (1992). Reforming education and changing schools: Case studies in policy sociology. London, Routledge.
- Brehm, J. and S. Gates (1997). Working, shirking, sabotage. Ann Arbor, University of Michigan Press.
- Brown, P. H., A. de Brauw, et al. (2009). "Understanding variation in the design of China's New Co-operative Medical System." The China Quarterly, 198: 304-329.
- Brown, P. H. and C. Theoharides (2009). "Health-seeking behavior and hospital choice in China's New Cooperative Medical Scheme." Health Economics, 18: S47-S64.
- Bryman, A. (2001). Social research methods. Oxford, Oxford University Press.
- Cao, Z. (2011). "Zhongguo shang xia fen zhi de zhili tizhi ji qi wending jizhi (Vertically decentralised authoritarianism and mechanisms of political stability in China)." Shehui Xue Yanjiu (Sociological Studies), 25: 1-40.
- Carrin, G., A. Rona, et al. (1999). "The reform of the rural cooperative medical system in the People's Republic of China: Interim experience in 14 pilot counties." Social Science & Medicine, 48: 961-972.

Chang, X. (2010). "Canhe nongmin zai Xin Nong He jijin jianguan guocheng zhong suo fahui zuoyong de anli yanjiu (Case research on the effectiveness of NCMS user participation in fund oversight and management)." Zhongguo Nongcun Weisheng Shiye Guanli (Chinese Rural Health Service Administration), 30(5): 323-325.

Chen, F. (2010). "Xinxing Nongcun Hezuo Yiliao jiandu zhidu cunzai de wenti ji qi duice tantao (Problems in oversight of the New Cooperative Medical Scheme and investigation of countermeasures)." Zhonggong Fujian Sheng Wei Dangxiao Xuebao (Journal of the Party School of Fujian Province), 2.

Chen, G. and C. Wu (2006). Will the boat sink the water? The life of China's peasants. New York, Public Affairs.

Chen, S. (2009). Taoshan's independent inspection bureau. X City Daily, X City.

Chen, X. and X. Yang (2009). Difang zhengfu gonggong quanli chuangxin: Jingyan yu qushi (Local government public management innovations: Experience and trends). Changchun, Jilin Daxue Chubanshe (Jilin University Press).

Chen, Y. (2005). "Xinxing Nongcun Hezuo Yiliao shidian gongzuo chu tan (Initial exploration of New Cooperative Medical Scheme pilot work)." Caizheng Jiandu (Financial Inspection), 2.

China Financial Publishing House, Ed. (2002). Zhongguo caizheng nianjian (Financial yearbook of China). Beijing, Zhongguo Caizheng Zazhishe (China Financial Publishing House).

China Financial Publishing House, Ed. (2005). Zhongguo caizheng nianjian (Financial yearbook of China). Beijing, Zhongguo Caizheng Zazhishe (China Financial Publishing House).

China Financial Publishing House, Ed. (2007). Zhongguo caizheng nianjian (Financial yearbook of China). Beijing, Zhongguo Caizheng Zazhishe (China Financial Publishing House).

Christiansen, F. and H. X. Zhang (2009). The political economy of rural development in China: Reflections on current rural policy. Duisberg, Universität Duisburg-Essen. Duisburg Working Papers on East Asian Studies.

<http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=22717>

- Chung, J. H. (1998). Study of provincial politics and development in the post-Mao reform era: Issues, approaches and sources. Provincial strategies of economic reform in post-Mao China. P. T. Cheung, J. H. Chung and Z. Lin, Ed. Armonk, New York; London, East Gate, ME Sharpe.
- Chung, J. H. (2000). Central control and local discretion in China. Oxford; New York, Oxford University Press.
- Chung, J. H. (2001). Reappraising central-local relations in Deng's China: Decentralisation, dilemmas of control and diluted effects of reform. Remaking the Chinese state: Strategies, society and security. C.-m. Chao and B. J. Dickson, Ed. London, Routledge.
- Chung, J. H. (2011). Central-local dynamics: Historical continuities and institutional resilience. Mao's invisible hand: The political foundations of adaptive governance in China. S. Heilmann and E. J. Perry, Ed. Cambridge, Massachusetts; London, Harvard University Press.
- Chung, M. H. (1992). "China's future: Regionalism, federation or disintegration." Studies in Comparative Communism, 25(3): 211-227.
- Codd, J. A. (1988). "The construction and deconstruction of educational policy documents." Journal of Education Policy, 3(3): 235-247.
- Conway, G., J. Waage, et al. (2010). Science and innovation for development. London, UK Collaborative on Development Sciences.
- Cox, W. M. and R. Alm (2008). Creative Destruction. The Concise Encyclopedia of Economics. George Mason University, Library of Economics and Liberty.
- Downs, A. (1967). Inside bureaucracy. Boston, Little Brown.
- Dredge, R. (2009). Hospital global budgeting. Designing and implementing health care provider systems. J. Langenbrunner, C. Cashin and S. O'Dougherty, Ed. Washington DC, The World Bank.
- Drouin, J. P., V. Hediger, et al. (2008). "Health care costs: A market-based view." McKinsey Quarterly, September 2008.
- Duckett, J. (2011). The Chinese state's retreat from health: Policy and the politics of

retrenchment. London, Routledge.

Eco, U. (1989). The open work. Cambridge, Massachusetts, Harvard University Press.

Eco, U. (1990). The limits of interpretation. Bloomington; Indianapolis, Indiana University Press.

Edin, M. (2000). Market forces and communist power: Local political institutions and economic development in China. Department of Government. Uppsala, Uppsala University. Dissertation for the degree of Doctor of Philosophy in Political Science.

Edin, M. (2003a). "Local state corporatism and private business." Journal of Peasant Studies, 30(3): 278-295.

Edin, M. (2003b). "State capacity and local agent control in China: CCP cadre management from a township perspective." The China Quarterly, 173: 35-52.

Eggleston, K., L. Li, et al. (2008). "Health service delivery in China: A literature review." Health Economics, 17: 149-165.

Feng, L., Y. Wang, et al. (2009). "*Xinxing Nongcun Hezuo Yiliao de nongmin jiandu: Yi Chongqing Shizhu Xian wei li* (Peasant oversight of the New Cooperative Medical Scheme: A case study of Shizhu County, Chongqing)." Zhongguo Weisheng Zhengce Yanjiu (Chinese Journal of Health Policy), 2(5): 54-56.

Fewsmith, J. (2004). "Promoting the scientific development concept." China Leadership Monitor, 11.

Fewsmith, J. (2006). "Institutional innovation at the grassroots level: Two case studies." China Leadership Monitor, 18.

Fewsmith, J. (2008a). "A new upsurge in political reform? Maybe." China Leadership Monitor, 24.

Fewsmith, J. (2008b). "What happened in Maliu township?" China Leadership Monitor, 25.

Fish, S. E. (1980). Is there a text in this class? The authority of interpretive communities. Cambridge, Massachusetts, Harvard University Press.

Flyvbjerg, B. (2006). "Five misunderstandings about case study research." Qualitative Enquiry, 12(2): 219-245.

Foster, K. (2005). "Chinese public policy innovation and the diffusion of innovations: An initial exploration." Chinese Public Administration Review, 3(1/2): 1-13.

Gao, G. (2008a). *Gonggu he fazhan juyou Zhongguo tese de Xinxing Nongcun Hezuo Yiliao baozhang zhidu* (Consolidate and develop the New Cooperative Medical Scheme with Chinese characteristics). Yiliao weisheng lübishu: Zhongguo yiliao weisheng fazhan baogao 4 (Green book of health care: Annual report on China's healthcare, No. 4). L. Du, Ed. Beijing, *Shehui Kexue Wenxian Chubanshe* (Social Sciences Academic Press). 2008 edition.

Gao, G., Ed. (2010). Xinxing Nongcun Hezuo Yiliao peixun jiaocai (New Cooperative Medical Scheme training materials). Beijing, *Zhongguo Guji Chubanshe*.

Gao, G. and Y. Han (2007). *Wo guo Xinxing Nongcun Hezuo Yiliao zhidu de fazhan* (Development of China's New Cooperative Medical Scheme). Yiliao weisheng lübishu: Zhongguo yiliao weisheng fazhan baogao 3 (Green book of health care: Annual report on China's healthcare, No. 3). L. Du, Ed. Beijing, *Shehui Kexue Wenxian Chubanshe* (Social Sciences Academic Press). 2007 edition.

Gao, G., L. Zhao, et al. (2006). "Xinxing Nongcun Hezuo Yiliao zhidu de yunxing chengben he guanli chengben tantao (Investigation into operation and management costs of the New Cooperative Medical Scheme)." Zhongguo Weisheng Jingji (China Health Economics), 276: 24-26.

Gao, G. and J. Zhu (2009). "Wo Guo Xinxing Nongcun Hezuo Yiliao jijin guanli guocheng zhong cunzai de wenti yu yingdui celüe (Problems in China's New Cooperative Medical Scheme fund management and policy countermeasures)." Zhongguo Weisheng Jingji (China Health Economics), 28(1): 52-55.

Gao, J. (2009). "Governing by goals and numbers: A case study in the use of performance measurement to build state capacity in China." Public Administration and Development, 29: 21-31.

Gao, X. (2008b). "Difang zhengfu chuangxin yuan he nan chixu: Yi Chongqing Shi Kai Xian Maliu Xiang wei li (Why is local government innovation hard to sustain? A case study of Maliu Township, Kai County, Chongqing)." Zhongguo Gaige (China Reform), 5: 29-32.

Göbel, C. (2007). The rural tax and fee reform and central-local relations in China: A policy analysis. Duisburg/Essen, University of Duisburg-Essen. Dissertation for the degree of Doctor of Philosophy.

Göbel, C. (2011). "Uneven policy implementation in rural China." The China Journal, 65(January 2011): 53-76.

Goldman, M. and R. MacFarquhar (1999). Dynamic economy, declining party-state. The paradox of China's post-Mao reforms. M. Goldman and R. MacFarquhar, Ed. Harvard, Massachusetts, Harvard University Press.

Goodman, D. S. (1986). Centre and province in the People's Republic of China. Cambridge, Cambridge University Press.

Goodman, D. S. (1994). The politics of regionalism: Economic development, conflict and negotiation. China deconstructs: Politics, trade and regionalism. D. S. Goodman and G. Segal, Ed. London; New York, Routledge.

Green, A. E. (2005). Managing human resources in a decentralized context. East Asia decentralizes: Making local government work. World Bank, Ed. Washington DC, World Bank.

Greenhalgh, T., G. Robert, et al. (2004). "Diffusion of innovations in service organizations: Systematic review and recommendations." The Milbank Quarterly, 82(4): 581-629.

Gu, X. (2008). Quanmin yiliao baoxian zou shang zhenggui (Universal health insurance gets onto the right track). 2008 nian Zhongguo shehui xingshi fenxi yuce (2008 analysis and forecast of China's social trends). X. Ru, X. Lu and P. Li, Ed. Beijing, Shehui Kexue Wenxian Chubanshe (Social Sciences Academic Press).

Gudeman, S. (1992). "Remodeling the house of economics: Culture and innovation." American Ethnologist, 19(1): 141-154.

Gudeman, S. and A. Rivera (1990). Conversations in Colombia: The domestic economy in life and text. Cambridge, Cambridge University Press.

Hall, P. A. (1993). "Policy paradigms, social learning, and the state: The case of economic policymaking in Britain." Comparative Politics, 25(3): 275-296.

Hansen, M. H. (2006). In the footsteps of the Communist Party. Doing fieldwork in China. M. Heimer and S. Thøgersen, Ed. Copenhagen, NIAS Press.

Hartford, K. (1985). Socialist agriculture is dead; long live socialist agriculture! Organizational transformations in rural China. The political economy of reform in post-Mao China. E. J. Perry and C. Wong, Ed. Cambridge, Massachusetts; London, Harvard University Press.

Hartley, J. (2006). Innovation and its contribution to improvement: A review for policy-makers, policy advisers, managers and researchers. London, Department for Communities and Local Government.

Hawkes, T. (1992). Structuralism and semiotics. London, Routledge.

Hayek, F. A. (1945). "The use of knowledge in society." The American Economic Review, 35(4): 519-530.

He, B. and S. Thøgersen (2010). "Giving the people a voice? Experiments with consultative authoritarian institutions in China." Journal of Contemporary China, 19(66): 675-692.

He, Z. (1997). "*Muqian wo guo xian ji zhengzhi tizhi zhong cunzai de wenti, chengyin ji jiejuе duice* (Current problems in the county-level political system: Causes and policy remedies)." Tianjin Shehui Kexue (Tianjin Social Sciences), 6.

Heberer, T. and G. Schubert (2012). "County and township cadres as a strategic group. A new approach to political agency in China's local state." Journal of Chinese Political Science, 17(3): 221-249.

Heberer, T. and A. Senz (2011). "Streamlining local behaviour through communication, incentives and control: A case study of local environmental policies in China." Journal of Current Chinese Affairs, 40(3): 77-112.

- Heilmann, S. (2008a). Experimentation under hierarchy: Policy experiments in the reorganization of China's state sector, 1978-2008. Cambridge, Massachusetts, Centre for International Development, Harvard University.
- Heilmann, S. (2008b). "From local experiments to national policy: The origins of China's distinctive policy process." The China Journal, 59: 1-30.
- Heilmann, S. (2008c). "Policy experimentation in China's economic rise." Studies in Comparative International Development, 43(1): 1-26.
- Heilmann, S. (2009). "Maximum tinkering under uncertainty: Unorthodox lessons from China." Modern China, 35(4): 450-462.
- Heilmann, S. and E. J. Perry (2011). Embracing uncertainty: Guerrilla policy style and adaptive governance in China. Mao's invisible hand: The political foundations of adaptive governance in China. S. Heilmann and E. J. Perry, Ed. Cambridge, Massachusetts; London, Harvard University Press.
- Heimer, M. (2006). Field sites, research design and type of findings. Doing fieldwork in China. M. Heimer and S. Thøgersen, Ed. Copenhagen, NIAS Press.
- Heiner, R. A. (1983). "The origin of predictable behavior." The American Economic Review, 73(4): 560-595.
- Herd, R., Y.-W. Hu, et al. (2010). Improving China's health care system. Paris, Organization for Economic Cooperation and Development.
- Hill, M. and P. Hupe (2009). Implementing public policy. London, Sage.
- Holdaway, J. (2010). "Environment and health in China: An introduction to an emerging research field." Journal of Contemporary China, 19(63): 1-22.
- Hsiao, W. (2008). "When incentives and professionalism collide." Health Affairs, 27(4): 949-951.
- Hsiao, W. C. and A. Maynard (2009). "Foreword." Health Economics, 18: S1-S2.
- Hsing, Y.-t. (2006). "Brokering power and property in China's townships." Pacific Review, 19(1): 103-124.

- Hu, B. and J. Lei (2004). "Xinxing Nongcun Hezuo Yiliao shidian cunzai de wenti ji gajin jianyi (Problems in the New Cooperative Medical Scheme and suggestions for improvement)." Caizheng Jiandu (Financial Oversight), 12: 18-19.
- Hu, S., S. Tang, et al. (2008). "Reform of how health care is paid for in China: Challenges and opportunities." The Lancet, 372: 1846-53.
- Huang, L. (2005). "Feitian works at developing the NCMS." XXXX, XX(X).
- Huang, P. C. (2008). "Centralised minimalism: Semiformal governance by quasi officials and dispute resolution in China." Modern China, 34(1): 9-35.
- Huang, P. C. (2010a). "Beyond the right-left divide: Searching for reform from the history of practice." Modern China, 36(1): 115-133.
- Huang, P. C. (2010b). "Introduction to 'Constitutionalism, reform, and the nature of the Chinese state: Dialogues among western and Chinese scholars, III'." Modern China, 36(1): 3-11.
- Huang, Y. (1999). Inflation and investment controls in China: The political economy of central-local relations during the reform era. Cambridge, Cambridge University Press.
- Illich, I. (1975). Medical nemesis: The expropriation of health. London, Calder and Boyers.
- Illich, I. (2003). "Medical nemesis." Journal of Epidemiology and Community Health, 57: 919-922.
- Jackson, S., A. C. Sleight, et al. (2005). "Health finance in rural Henan: Low premium insurance compared to the out-of-pocket system." China Quarterly, 181: 137-157.
- Ji, X., L. Wang, et al. (2011). "Yunnan Sheng Lufeng Xian Xinxing Nongcun Hezuo Yiliao menzhen zong'e yu fu zhidu yanjiu (Research into use of outpatient global budgets in the New Cooperative Medical Scheme in Lufeng County, Yunnan)." Zhongguo Weisheng Zhengce Yanjiu (Chinese Journal of Health Policy), 2(4): 27-33.
- Jiang, S. (2010). "Written and unwritten constitutions: A new approach to the study of constitutional government in China." Modern China, 36(12): 12-46.
- Jin, L. (1995). "Caizheng jiancha zhuanyuan banshi jigou de zhineng zhuanbian ji 1995 nian de

gongzuo renwu (Transformation in the work of dedicated financial inspection organs is the work task for 1995)." Caizheng (Finance), 355: 5-8.

Kingdon, J. W. (1984). Agendas, alternatives, and public policies. Boston; Toronto, Little Brown and Company.

Kipnis, A. B. (2008). "Audit cultures: Neoliberal governmentality, socialist legacy, or technologists of governing." American Ethnologist, 35(2): 275-289.

Klotzbücher, S., P. Lässig, et al. (2010). "What is new in the New Rural Cooperative Medical System? An assessment in one Kazak county of the Xinjiang Uyghur Autonomous Region." The China Quarterly, 201: 38-57.

Kohli, A. (1990). Democracy and discontent. Cambridge, Cambridge University Press.

Kuhn, T. S. (1996). The structure of scientific revolutions. Chicago; London, University of Chicago Press.

Lü, X. (1997). "The politics of peasant burden reform in China." The Journal of Peasant Studies, 25(1): 113-138.

Lam, T.-c. (2010). The county system and county governance. China's local administration: Traditions and change in the sub-national hierarchy. J. H. Chung and T.-c. Lam, Ed. London, Routledge.

Lampton, D. M. (1987a). The implementation problem in post-Mao China. Policy implementation in post-Mao China. D. M. Lampton, Ed. Berkeley; London, University of California Press.

Lampton, D. M. (1987b). Policy implementation in post-Mao China. Berkeley; London, University of California Press.

Landry, P. F. (2008). Decentralized authoritarianism in China: The Communist Party's control of local elites in the post-Mao era. New York, Cambridge University Press.

Langenbrunner, J. (2010). Models and next steps for provider payment reforms in rural China. Washington DC, The World Bank.

Langenbrunner, J. (2011). Models and next steps for provider payment reforms in rural China. Washington DC, The World Bank.

Lei, X. and W. Lin (2009). "The New Cooperative Medical System in Rural China: Does more coverage mean more service and better health?" Health Economics, 8: S25-S46.

Li, C. (2004). Political localism versus institutional restraints: Elite recruitment in the Jiang era. Holding China together. B. J. Naughton and D. L. Yang, Ed. Cambridge, Cambridge University Press.

Li, L. C. (1998a). Central-provincial relations: Beyond compliance analysis. China Review 1998. J. Y. Cheng, Ed. Hong Kong, Chinese University of Hong Kong.

Li, L. C. (1998b). Centre and provinces: China 1978-1993. Oxford, Clarendon Press.

Li, L. C. (2005). "Understanding institutional change: Fiscal management in local China." Journal of Contemporary Asia, 35(1): 87-108.

Li, L. C. (2006a). "Differentiated actors: Central-local politics in China's rural tax reforms." Modern Asian Studies, 40(1): 151-174.

Li, L. C. (2006b). "Path creation? Processes and networks: How the Chinese rural tax reform began." Policy and Society, 25(1): 61-84.

Li, L. C. (2008). "State and market in public service provision: Opportunities and traps for institutional change in rural China." The Pacific Review, 21(3): 257-278.

Lieberthal, K. (1978). Central documents and Politburo politics in China. Ann Arbor, Center for Chinese Studies, University of Michigan.

Lieberthal, K. (2004). Governing China: From revolution through reform. New York; London, WW Norton & Co.

Lieberthal, K. G. (1992). The 'fragmented authoritarianism' model and its limitations. Bureaucracy, politics, and decision making in post-Mao China. K. G. Lieberthal and D. M. Lampton, Ed. Berkeley; Oxford, University of California Press.

Lieberthal, K. G. and M. Oksenberg (1988). Policy making in China: Leaders, structures, and

processes. Princeton, Princeton University Press.

Lindblom, C. E. (1959). "The science of 'muddling through'." Public Administration Review, 19(2): 79-88.

Liu, M., J. Wang, et al. (2009). "The political economy of earmarked transfers in a state-designated poor county in Western China: Central policies and local responses." The China Quarterly, 200: 973-994.

Liu, W. (2008). Xinxing Nongcun Hezuo Yiliao jijin caiwu kuaiji zhidu jianqie (Explanation of the New Cooperative Medical Scheme fund finance and accounting systems). Beijing, *Zhongguo Caizheng Jingji Chubanshe* (China Financial and Economic Publishing House).

Liu, Y., K. Rao, et al. (2008). "China's health system performance." The Lancet, 372: 1914-1923.

Long, N. (2001). Development sociology: Actor perspectives. London; New York, Routledge.

Ma, X. (2008). Getting medical expenses reimbursed and getting rid of the agricultural tax. Beijing.

Manion, M. (1991). "Policy implementation in the People's Republic of China: Authoritative decisions versus individual interests." The Journal of Asian Studies, 50(2): 253-279.

Manion, M. (1992). The behavior of middlemen in the cadre retirement policy process. Bureaucracy, politics, and decision making in post-Mao China. K. G. Lieberthal and D. M. Lampton, Ed. Berkeley; Oxford, University of California Press.

Mao, T.-t. (1977). On the ten major relationships. Selected Works of Mao Tse-tung, Ed. Peking, Foreign Languages Press. Volume 5: 284-307.

Mao, Z. (2005). Pilot programme of NCMS in China: System design and progress. Sichuan, China, Background Paper for World Bank China Rural Health Study.

McCormick, B. L. (1987). Leninist implementation: The election campaign. Policy implementation in post-Mao China. D. M. Lampton, Ed. Berkeley; London, University of California Press.

Meessen, B. and G. Bloom (2007). "Economic transition, institutional changes and the health system: Some lessons from rural China." Journal of Economic Policy Reform, 10(3): 209-231.

Meng, Q., W. Yip, et al. (2010). Health provider payment reforms in China: What international experience tells us. Washington DC, The World Bank. China Health Policy Notes, No. 3.

MOH Center for Statistics Information, Ed. (2007). Zhongguo Xinxing Nongcun Hezuo Yiliao jinzhan ji qi xiaoqiao fenxi: 2005 nian Xinxing Nongcun Hezuo Yiliao shidian diaocha baogao (Progress and effectiveness evaluation of the New Cooperative Medical Scheme: 2005 NCMS pilot investigation report). Beijing, *Zhongguo Xiehe Yike Daxue Chubanshe* (Peking Medical University Publishing House).

Mulgan, G. and D. Albury (2003). Innovation in the public sector (version 1.9). London, Strategy Unit, The Cabinet Office.

<http://www.cabinetoffice.gov.uk/media/cabinetoffice/strategy/assets/pubinov2.pdf>

Murphy, R. (2004). "Turning peasants into modern Chinese citizens: 'Population quality' discourse, demographic transition and primary education." The China Quarterly, 177: 1-20.

Naughton, B. (1987). The decline of central control over investment in post-Mao China. Policy implementation in post-Mao China. D. M. Lampton, Ed. Berkeley; London, University of California Press.

Naughton, B. (2009). "Singularity and replicability in China's developmental experience." China Analysis, 68.

Naughton, B. J. and D. L. Yang (2004). Holding China together: Introduction. Holding China together. B. J. Naughton and D. L. Yang, Ed. Cambridge, Cambridge University Press.

NCMS Pilot Evaluation Group (2006). *Fazhan zhong de Zhongguo Xinxing Nongcun Hezuo Yiliao: Xinxing Nongcun Hezuo Yiliao shidian gongzuo pinggu baogao* (The New Cooperative Medical Scheme under development: New Cooperative Medical Scheme pilot work report. Beijing, *Renmin Weisheng Chubanshe* (People's Medical Publishing House).

North, D. (2005). Understanding the process of economic change. Princeton, Princeton University Press.

- O'Brien, K. J. (2010). "How authoritarian rule works." Modern China, 36(1): 79-86.
- O'Brien, K. J. and L. Li (1999). "Selective policy implementation in rural China." Comparative Politics, 31(2): 167-186.
- O'Brien, K. J. and L. Li (2000). "Accommodating 'democracy' in a one-party state: Introducing village elections in China." The China Quarterly, 162: 465-489.
- Oi, J. (1989). The state and peasant in contemporary China. Berkeley, University of California Press.
- Oi, J. C. (1995). "The role of the state in China's transitional economy." The China Quarterly, 144: 1132-1149.
- Oi, J. C. (1999). Rural China takes off: Institutional foundations of economic reform. Berkeley, University of California Press.
- Paine, L. (1992). The educational policy process: A case study of bureaucratic action in China. Bureaucracy, politics, and decision making in post-Mao China. K. G. Lieberthal and D. M. Lampton, Ed. Berkeley; Oxford, University of California Press.
- Pan, X., H. H. Dib, et al. (2009). "Absence of appropriate hospitalization cost control for patients with medical insurance: A comparative analysis study." Health Economics, 18: 1146-1162.
- Pieke, F. (2004). "Contours of an anthropology of the Chinese state: Political structure, agency and economic development in rural China." Journal of the Royal Anthropological Institute, 10(3): 517-538.
- Pieke, F. (2007). Market Leninism: Party Schools and cadre training in contemporary China. British Inter-University China Centre. BICC Working Paper Series.
- Porter, R. (1985). "The patient's view: Doing medical history from below." Theory and Society, 14(2): 175.
- Porter, R. (1997). The greatest benefit to mankind: A medical history of mankind from antiquity to the present. London, HarperCollins.
- Project Group of Xuancheng City Party School (2004). "Xinxing Nongcun Hezuo Yiliao shidian

- de diaocha (Investigation into the NCMS pilot)." Lilun Jianshe (Theory Research), 90: 61-64.
- Rodrik, D. (2007). One economics, many recipes: Globalization, institutions and economic growth. Princeton and Oxford, Princeton University Press.
- Rong, J. (1998). Cong Yalixing Tizhi xiang minzhu hezuo tizhi de zhuanbian: Xian xiang liang ji zhengzhi tizhi gaige (From pressurised system to democratic cooperative system: Reform of the county and township political system). Beijing, *Zhongyang Bianyi Chubanshe* (Central Compilation and Translation Press).
- Rong, J. (2009). "Bian 'ling he boyi' wei 'shuang ying jizhi': Ruhe gaibian Yalixing Tizhi (From a zero-sum game to a win-win mechanism: How to change the pressurised system)." Renmin Luntan (People's Forum), 02.
- Roper, S. (2010). Innovation and creativity in organisations. Warwick, University of Warwick. The Warwick MBA for distance learning, Lesson 1.
- Rosenau, P. M. (1992). Post-modernism and the social sciences: Insights, inroads, and intrusions. Princeton, Princeton University Press.
- Saich, T. (2011). Governance and politics of China. Basingstoke, Palgrave Macmillan.
- Saich, T. and X. Yang (2003). "Innovation in China's local governance: 'Open recommendation and selection'." Pacific Affairs, 76(2).
- Said, E. (2003). Orientalism. London, Penguin.
- Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. Cambridge, Massachusetts, Harvard University Press.
- Schumpeter, J. A. (1976). Capitalism, socialism and democracy. London; New York, Routledge.
- Schurmann, H. F. (1966). Ideology and organization in Communist China. Berkeley, University of California Press.
- Scott, J. C. (1998). Seeing like a state: How certain schemes to improve the human condition have failed. New Haven; London, Yale University Press.

- Shue, V. (2008). "Rule as repertory and the compound essence of authority." Modern China, 34(1): 141-151.
- Sidel, R. and V. Sidel (1983). The health of China. London, Zed.
- Solinger, D. J. (1995). "Despite decentralisation: Disadvantages, dependence and ongoing central power in the inland - the case of Wuhan." The China Quarterly, 145: 1-34.
- Solinger, D. J. (2004). Policy consistency in the midst of the Asian crisis: Managing the furloughed and the farmers in three cities. Holding China together. B. J. Naughton and D. L. Yang, Ed. Cambridge, Cambridge University Press.
- Song, P. (2008). *Xinxing Nongcun Hezuo Yiliao Er Ci Buchang shishi xianzhuang pingjia ji duice yanjiu* (Evaluation of the state of the NCMS Second Round Reimbursement policy and policy suggestions). Shehui Yixue yu Weisheng Shiye Guanli Zhongxin (Department of Social Medicine and Health Management). Jinan, Shandong University. Dissertation for the degree of Master of Arts.
- Song, X., Z. Zhang, et al. (2006). Yong ai dianliang xiwang (Use love to kindle hope). Qingdao Ribao (Qingdao Daily), Qingdao, 14/01/2006, p. 2.
- Sun, L. (2009). "Ningxia 'yi yuan qian kan bing' moshi shou dao daibiao weiyuan guanzhu (Ningxia's '1 RMB treatment' model gets attention from People's Congress representatives and members)." Retrieved 03/2010, from *Ningxia Xinhua Wang (Ningxia Xinhua Web)*.
http://www.nx.xinhuanet.com/special/no1/2009-03/13/content_15946490.htm
- Sun, S. and Z. Chai, Eds. (2009). Xinxing Nongcun Hezuo Yiliao zhidu de quifanhua yu lifa yanjiu (Research on New Cooperative Medical Scheme regularisation and legislation). Beijing, *Falü Chubanshe (Law Press China)*.
- Taleb, N. N. (2007). The black swan: The impact of the highly improbable. London, Allen Lane.
- Tang, S. and G. Bloom (2000). "Decentralizing rural health services: A case study in China." International Journal of Health Planning and Management, 15: 189-200.
- Tang, S., Q. Meng, et al. (2008). "Tackling the challenges to health equity in China." The Lancet,

372: 1493-1501.

Ter-Minassian, T. and A. Fedelino (2008). Fiscal policy and reforms: Toward realizing a harmonious society. Public finance in China: Reform and growth for a harmonious society. J. Lou and S. Wang, Ed. Washington DC, The World Bank.

Tsou, T. (1986). The Cultural Revolution and post-Mao reforms: A historical perspective. Chicago, University of Chicago Press.

Unger, J. (1985). "The decollectivization of the Chinese countryside: A survey of twenty-eight villages." Pacific Affairs, 58(4): 585-606.

van Doorslaer, E., O. O'Donnell, et al. (2007). "Catastrophic payments for healthcare in Asia." Health Economics, 16: 1159-1184.

van Doorslaer, E., O. O'Donnell, et al. (2006). "Effect of payments for health care on poverty estimates in 11 countries in Asia: An analysis of household survey data." The Lancet, 368: 1357-1364.

Wagstaff, A. and M. Lindelow (2008a). "Can insurance increase financial risk? The curious case of health insurance in China." Journal of Health Economics, 27: 990-1005.

Wagstaff, A. and M. Lindelow (2008b). Health reform in rural China: Challenges and options. Public finance in China: Reform and growth for a harmonious society. J. Lou and S. Wang, Ed. Washington DC, The World Bank.

Wagstaff, A., M. Lindelow, et al. (2009a). "Extending health insurance to the rural population: An impact evaluation of China's New Cooperative Medical Scheme." Journal of Health Economics, 28(1-19).

Wagstaff, A., M. Lindelow, et al. (2009b). Reforming China's rural health system. Washington DC, World Bank.

Wagstaff, A., W. Yip, et al. (2009). "China's health system and its reform: A review of recent studies." Health Economics, 18: S7-S23.

Wang, H. (2009a). "Taoshan strengthens NCMS management and oversight: An effectiveness

analysis." XXXX, XX(X).

Wang, H., D. Gu, et al. (2008). "Factors associated with enrollment, satisfaction, and sustainability of the New Cooperative Medical Scheme program in six study areas in rural Beijing." Health Policy, 85(32-44).

Wang, S. (1995). The rise of the regions: Fiscal reform and the decline of central state capacity in China. The waning of the communist state: Economic origins of political decline in China and Hungary. Walder, Ed. Berkeley, University of California Press.

Wang, S. (2009b). "Adapting by learning: The evolution of China's rural health care financing." Modern China, 35(4): 370-399.

Wang, S. (2011). Learning through practice and experimentation: The financing of rural health care. Mao's invisible hand: The political foundations of adaptive governance in China. S. Heilmann and E. J. Perry, Ed. Cambridge, Massachusetts; London, Harvard University Press.

Wang, X. (2007). "Ningguo Shi Xin Nong He zhidu sheji he shidian jiben jingyan ji pingjia (Basic experience and evaluation of Ningguo City's NCMS)." Zhongguo Nongcun Weisheng Shiye Guanli (Chinese Rural Health Service Administration), 27(2): 96-98.

Wang, Y. (2008). The policy process and context of the Rural New Cooperative Medical Scheme and Medical Financial Assistance in China. Health and Social Protection: Experiences from Cambodia, China and Lao. B. Meessen, X. Pei, B. Criel and G. Bloom, Ed. Brighton, Institute of Development Studies.

Wang, Y. (2009c). 'Er Ci Buchang' jian cheng yibao zhidu 'er ci buchong' (Second Round Reimbursement becomes a supplement to the medical insurance system). Beijing Shangbao (Beijing Business Today), Beijing, 14/07/2009, p. 2.

Watson, A. (1987). "The family farm, land use and accumulation in agriculture." The Australian Journal of Chinese Affairs, 17(1): 1-27.

Wedeman, A. (2001). "Incompetence, noise, and fear in central-local relations in China." Studies in Comparative International Development, 35(4): 59-83.

White, L. T. I. (1997). Unstately power, Volume 2: Local causes of China's intellectual, legal and political reforms. Armonk, New York, ME Sharpe.

Whiting, S. (2001). Power and wealth in rural China: The political economy of institutional change. Cambridge, Cambridge University Press.

Whiting, S. H. (2004). The Cadre Evaluation System at the grass roots: The paradox of Party rule. Holding China together. B. J. Naughton and D. L. Yang, Ed. Cambridge, Cambridge University Press.

WHO (2004). Implementing the New Cooperative Medical Schemes in rapidly changing China: Issues and options. Beijing, Office of the World Health Organization Representative in China.

Wong, C. (1987). "Between plan and market: The role of the local sector in post-Mao China." The Journal of Comparative Economics, 11: 385-398.

Wong, C. (2009). "Rebuilding government for the 21st century: Can China incrementally reform the public sector?" The China Quarterly, 200: 929-952.

World Bank (2010). A generic drug policy as cornerstone to essential medicines in China. Washington DC, The World Bank. China Health Policy Notes, No. 4.

Wu, Y. (2004). *Wu Yi Fuzongli zai quan guo Xinxing Nongcun Hezuo Yiliao shidian gongzuo huiyi shang de jianghua, 2003 nian 12 yue 4 ri* (Speech of Vice Premier Wu Yi at the national New Cooperative Medical Scheme pilot meeting, 4th December 2003). 2004 nian Zhongguo weisheng nianjian (2004 China health yearbook). X. Liu, Ed. Beijing, *Renmin Weisheng Chubanshe* (People's Health Publishing): 91-96.

Wu, Y. (2005). *Tongyi sixiang, jiji tansuo, xunxu jianjin, wenbu tuijin Xinxing Nongcun Hezuo Yiliao shidian gongzuo: Guowuyuan Fuzongli Wu Yi zai 2004 nian quan guo Xinxing Nongcun Hezuo Yiliao shidian gongzuo huiyi shang de jianghua* (Unify thinking, enthusiastically experiment, proceed step-by-step and progressively develop New Cooperative Medical Scheme pilot work: Speech of View Premier Wu Yi at the 2004 national New Cooperative Medical Scheme pilot work meeting). 2005 nian Zhongguo weisheng nianjian (2005 China health yearbook). S. Tao, Ed. Beijing, *Renmin Weisheng Chubanshe* (People's Health

Publishing): 60-65.

Wu, Y. (2007). "Guowuyuan Fuzongli Wu Yi: Quanmian tuijin Xinxing Nongcun Hezuo Yiliao fazhan (Vice President of the State Council, Wu Yi: Comprehensively promote the development of the NCMS)." 16/03/2007. Retrieved February 2011, from *Renmin Wang* (People's Daily Website).

<http://theory.people.com.cn/GB/49169/49171/5479218.html>

Wu, Z. (2009). "Lun Xinxing Nongcun Hezuo Yiliao jianzhi chuanguan (Discussion of innovating mechanisms for oversight of the New Cooperative Medical Scheme)." *Zhongguo Fujian Sheng Wei Dangxiao Xuebao* (Journal of the Party School of Fujian Province), 12: 37-43.

Xiang, L., H. Liu, et al. (2010). "Xin Nong He Er Ci Buchang moshi gonglüe (NCMS Second Round Reimbursement Policy models strategy)." *Zhongguo Yiyuan Yuanzhang* (China Hospital CEO), 11: 61-63.

Xiang, L., P. Song, et al. (2008a). "Xinxing Nongcun Hezuo Yiliao Er Ci Buchang moshi yanjiu (NCMS Second Round Reimbursement Policy models)." *Zhongguo Weisheng Jingji* (China Health Economics), 27(7): 8-11.

Xiang, L., P. Song, et al. (2008b). "Xinxing Nongcun Hezuo Yiliao Er Ci Buchang biyaoxing fenxi (Analysis of the necessity of the NCMS Second Round Reimbursement Policy)." *Zhongguo Weisheng Jingji* (China Health Economics), 27(4): 29-31.

Xiang, L., J. Su, et al. (2008). "Xinxing Nongcun Hezuo Yiliao Er Ci Buchang shishi qingkuang yanjiu (Implementation of the NCMS Second Round Reimbursement Policy)." *Zhongguo Weisheng Jingji* (China Health Economics), 27(8): 33-35.

Xiao, Y. (2006). *Jiangxi Sheng shixing canhe nongmin Er Ci Buchang* (Jiangxi carries out Second Round Reimbursement for NCMS members). *Jiankang Bao* (Health News), Beijing, 24/10/2006, p. 7.

Xiao, Y. (2007). "Jiangxi Sheng Xinxing Nongcun Hezuo Yiliao shishi Er Ci Buchang de shijian tansuo (Jiangxi's experience in implementing NCMS Second Round Reimbursement)."

Weisheng Jingji Yanjiu (Health Economics Research), August: 41-42.

Yan, F., J. Raven, et al. (2010). "Management capacity and health insurance: The case of the New Cooperative Medical Scheme in six counties in rural China." International Journal of Health Planning and Management, 26: 357-378.

Yan, Y., L. Zhang, et al. (2006). Insuring rural China's health? An empirical analysis of China's New Collective Medical System. Freeman-Spogli Institute for International Studies working paper. Palo Alto, California, Stanford University.

Yang, D. (1994). Reform and the restructuring of central-local relations. China deconstructs: Politics, trade and regionalism. D. S. Goodman and G. Segal, Ed. London; New York, Routledge.

Yang, D. L. (2004a). Economic transformation and state rebuilding in China. Holding China together. B. J. Naughton and D. L. Yang, Ed. Cambridge, Cambridge University Press.

Yang, T. (2004b). *Zhongguo Shehui zhengce xingjin zhong de ji ge jiben wenti: Yi Xinxing Hezuo Yiliao zhengce wei li* (Basic problems in China's social policy: The example of the New Cooperative Medical Scheme). Beijing, Social Policy Research Centre, Chinese Academy of Social Sciences.

http://www.chinasocialpolicy.org/Paper_Show.asp?Paper_ID=166

Yi, H., L. Zhang, et al. (2009a). Good news, bad news: Results from a national representative panel survey on China's NCMS. Social Protection Asia.

<http://www.socialprotectionasia.org/pdf/CCAP-SPA-WP01.pdf>

Yi, H., L. Zhang, et al. (2009b). "Health insurance and catastrophic illness: A report on the New Cooperative Medical System in rural China." Health Economics, 18: S119-S127.

Ying, X. (2009). China: Essential health-care provision. Promoting sustainable strategies to improve access to health care in the Asian and Pacific Region. S. D. Division, Ed. Bangkok, UNESCAP.

Ying, Y. (2006). "Xinxing Nongcun Hezuo Yiliao jijin fengbi yunxing moshi jixi (Analysis of closed management of New Cooperative Medical Scheme funds)." Weisheng Jingji Yanjiu

(Health Economics Research), 9: 10.

Yip, W. and W. Hsiao (2008). "Market watch: The Chinese health system at a crossroads." Health Affairs, 27(2): 460-469.

Yip, W. and W. Hsiao (2009). "Non-evidence based policy: How effective is China's new cooperative medical scheme in reducing medical impoverishment?" Social Science & Medicine, 68: 201-209.

Yip, W. C.-M., W. Hsiao, et al. (2010). "Realignment of incentives for health-care providers in China." The Lancet, 375: 1120-1130.

Yu, K., Ed. (2009). Zhongguo difang zhengfu chuangxin: Anli yanjiu baogao, 2007-2009 (Chinese local government innovation: Case study report, 2007-2009) Beijing, Beijing Daxue Chubanshe (Peking University Press).

Zeng, X. (1999). "Baoquan Zhongyang Caizheng shou-zhi bixu jiaqiang Zhuanyuanban gongzuo (Maintaining the integrity of Central Finance receipts and payments requires strengthening the work of the Zhuanyuanban)." Shiye Caikuai (Career Financial Accountant), 6: 50-51.

Zhang, L., X. Cheng, et al. (2010a). "Balancing the funds in the New Cooperative Medical Scheme in rural China: Determinants and influencing factors in two provinces." International Journal of Health Planning and Management, 25: 96-118.

Zhang, L., X. Cheng, et al. (2010b). "How effectively can the New Cooperative Medical Scheme reduce catastrophic health expenditure for the poor and non-poor in rural China?" Tropical Medicine and International Health, 15(4): 468-475.

Zhang, L., Y. Liu, et al., Eds. (2007). Chongqing Shi Qianjiang Qu nongcun weisheng zhengce li guanli gaige: Anli yanjiu (A case study of rural health policy and management reform in Qianjiang District, Chongqing). Beijing, Zhongguo Caizheng Jingji Chubanshe (China Financial and Economic Publishing House).

Zhang, L., H. Yi, et al. (2011). *Caring for one billion: Assessing the new social protection programmes in rural China*. Brighton, Centre for Social Protection, Institute of Development Studies.

<http://www.socialprotectionasia.org/Conf-program-pdf/7-SPA-Final-Paper-No-07.pdf>

Zhang, X. (2011a). Rural healthcare system: Challenges and opportunities. Unpublished presentation given at the 8th East Asian Social Policy conference, Hong Kong. August 2011.

Zhang, Z., Ed. (2011b). Zhongguo nongcun weisheng fuwu tixi jianshe yu fazhan (China rural health service construction and development). Beijing, *Renmin Weisheng Chubanshe* (People's Health Publishing).

Zhao, T. (2009). *Xinxing Nongcun Hezuo Yiliao luoshi mianfei tijian (Er Ci Buchang) shishi fang'an (New Cooperative Medical Scheme implements free check-ups (second round reimbursement) implementation plan)*. Dongchuan Nianjian (Dongcheng yearbook), Ed. Dehong, *Dehong Minzu Chubanshe* (Dehong People's Publishing).

Zheng, F. (2009). "Strengthen oversight and management, promoting NCMS stability and effectiveness." X Province Journal of Health Protection, XXXX(X).

Zheng, Y. (2007). De facto federalism in China: Reforms and dynamics of central-local relations. Singapore, World Scientific.

Zhong, J. a. (2008). "Zuzhi ganbu yao zuo gaige chuangxin de biaoshuai (Organisation cadres must be an example of reform and innovation)." 29/07/2008. Retrieved January 2009, from *Zhongguo Gongchandang Xinwen Wang* (Chinese Communist Party News Website).

<http://cpc.people.com.cn/GB/117092/117104/7579930.html>

Zhong, Y. (2003). Local government and politics in China. Armonk, New York; London, ME Sharpe.

Zhou, X. (2010). "The institutional logic of collusion among local governments in China." Modern China, 36(1): 47-78.

Zhu, L., Y. Ye, et al. (2010). "Nongmin canyu Xinxing Nongcun Hezuo Yiliao jianbu fangshi

*diao*cha (Investigation into peasants' participation in oversight of the New Cooperative Medical Scheme)." Zhongguo Gonggong Weisheng (Chinese Journal of Public Health), 26(8): 1031-1032.

Zittoun, P. (2009). "Understanding policy change as a discursive problem." Journal of Comparative Policy Analysis: Research and Practice, 11(1): 65-82.