Screening for the risk of self-harm in an adult offender population

Michael Colin Horton

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

Academic Department of Rehabilitation Medicine

Leeds Institute of Rheumatic and Musculoskeletal Medicine

December 2017

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Further details of the jointly-authored publications and the contributions of the candidate and the other authors to the work are included on the following page.

Details of Jointly-Authored Publications

Please note that some parts of this thesis have been published in:

Mike Horton, Nat Wright, Wendy Dyer, Alex Wright-Hughes, Amanda Farrin, Zanib Mohammed, Jamie Smith, Tom Heyes, Simon Gilbody, and Alan Tennant. *Assessing the risk of self-harm in an adult offender population: an incidence cohort study.* Health Technology Assessment, 2014. No. 18.64

Work from this publication is contained in the following chapters:

Chapter 2 – Background and Review of Literature; Chapter 3 – Pilot Study; Chapter 4 – Methods; Chapter 5 – Results of Cohort Study; Chapter 6 – Psychometric Analysis of Scales; Chapter 7 - Predictive Analysis; Chapter 8 – Explanatory Model of Self-Harm; Chapter 9 – Discussion and Summary.

Work within the publication which is directly attributable to me:

I undertook some prisoner interviews and led the drafting of the manuscript, the work on the psychometric analysis of scales and the incidence and demographic results, supported by Professor Tennant.

Chapter 1: Background – All written sections

Chapter 2: Design of the study – All written sections except sample size calculations, Confirmatory Factor Analysis, Mokken Scaling and Cox proportional hazards regression modelling

Chapter 3: Results – All written sections except Confirmatory Factor Analysis, Mokken Scale Analysis, Cluster Analysis, Cox proportional hazards regression modelling and Identifying items predictive of self-harm.

Chapter 4: Conclusions – I contributed to this chapter, but this was a jointly produced chapter and it was not solely and directly attributable to me.

NB. All work contained within this thesis that relates to Confirmatory Factor Analysis, Mokken Scaling and Identifying items predictive of self-harm was carried out and written up independently by myself.

Contribution of other authors within the publication:

All the authors were involved in the study management group of the project and collectively took decisions about the direction of the research. All the authors have contributed to the writing and review of this draft final report.

In addition, Jamie Smith and Zanib Mohammed undertook the majority of the interviews with prisoners, Alex Wright-Hughes undertook the Cox regression analysis, supported by Professor Farrin. Nat Wright took the lead on the clinical implications of the project and managed the research at one of the male prisons. Wendy Dyer managed the research at the two other prisons. Professor Tennant oversaw the whole project, and within the HTA report completed the sections relating to Confirmatory Factor Analysis, Mokken Scale Analysis, Cluster Analysis and Identifying items predictive of self-harm.

Clarification of contributions to HTA self-harm project and extensions to PhD work

The work contained within this thesis was carried out as part of a National Institute for Health Research (NIHR) collaborative group project, under the Health Technology Assessment (HTA) funding stream. For further clarification, I provide here a summary of my role within the HTA project on which my PhD was based, my contributions to the final HTA report, and what I did additionally for my PhD thesis.

My contributions to HTA project:

- Involved from second-stage grant application process
- Helped complete the necessary forms for the research ethics committee
- Carried out scoping study for instrument identification
- Underwent full prison staff training to enable data collection
- Collected approx. 5% of cohort study data, plus observed approx. another 3%
- All data management and collation
- Majority of basic descriptive and incidence statistics
- All Rasch analysis
- All AUC predictive analyses
- Significant proportion of final HTA report

What I did not contribute towards in the HTA project:

- Come up with original research question
- Design the study, or write the protocol
- Decide the most appropriate methodologies
- Carry out power calculations for the study
- Collect the majority of the data
- Carry out CFA and Mokken scaling analysis
- Carry out Cluster Analysis (not included in PhD thesis)
- Carry out Logistic Regression
- Carry out Cox Proportional-Hazards modelling (not included in PhD thesis)
- Provide the information relating to the practical implications of the study within a prison setting

Work that I carried out additionally for my PhD, which is included in my thesis:

- Extended background and literature review of HTA report
- Provided a more complete methodology section
- Independently re-did all CFA and Mokken scaling analysis
- Re-did all Logistic Regression, with a slightly different methodology
- Extended the analysis to explore the explanatory process of self-harm, using a Path Analysis approach within a Structural Equation Modelling (SEM) framework
- Provided a more in-depth discussion section

Acknowledgements

With thanks to:

My wonderful supervision team – Firstly, Alan Tennant and the dearly departed Bipin Bhakta, who got me into this in the first place. They have both provided me with unwavering support, offered me many travel and networking opportunities, and taught me so much through hands-on experience. It was not only research knowledge that they bestowed, but also humility, working style and collaborative team dynamic. Secondly, Rory O'Connor and Anne Maree Keenan, who stepped into the shoes of Alan and Bipin and pretty much picked up where they left off, even though they were big boots to fill. Your help, support, guidance and realism has been much appreciated, and thanks for seeing me over the line.

The Study Research Team, for all of their hard work and input – Nat Wright, Wendy Dyer, Jamie Smith, Zanib Mohammed, Diane Corr and team, Tom Heyes, Simon Gilbody, Alex Wright-Hughes, Vikki Wilkinson & Alan Tennant.

My statistical support gurus – Liz Hensor and Karl Bang Christensen, who have both provided way more help and support than merely a bit of statistical advice every now and then.

Tanya Meade for providing an insight into the complex and confusing world of self-harm, and to Shane Latimer, whose help was invaluable in piecing things together and trying to make a bit of sense of it all.

Nick Preston – thanks for the template, the support, the encouragement, the motivational speeches and the kicks up the backside, along with the much appreciated provision of many, many cups of tea and the occasional iced-bun. We're going for a beer when this is handed in!

Penny Wright, for her support and belief in me, and for protecting some of my time demands so that I could work on my thesis.

All those people who are, or who have been, extremely supportive and continually ask me if I'm done yet. – Including all of the ERTG group, the Mech Eng guys, Andrew

Bateman, Amanda Perry, Jan Boehnke, Anne Chamberlain, Mark Wilberforce and many unspecified others!

Mum and Dad, for enquiring with interest how things were going, but not so often as to annoy me!

Gary & Hilary, for all of the help with the kids at the times when I haven't been around.

Eloise & Ed, who brighten up my every day and are (hopefully) completely oblivious as to what a PhD is. Sorry for missing story time on Thursdays.

And lastly to my lovely Lucy, for basically being brilliant. All of the time. With everything. I couldn't have done it without you.

Many thanks to all of you, and yes, I know I should've finished ages ago!

With thanks also to my examiners; Brian Henson and Paul Tiffin, who provided useful and informative feedback. I believe that the amendments based on their suggestions have undoubtedly improved the final thesis.

Abstract

The reported incidence of self-harm is much higher among prisoners than in the general population. Along with the general increased risk among prisoners, there are also specific times of heightened vulnerability, such as on first reception into the prison environment. Currently in the UK, if an act of self-harm is carried out by a prisoner, if intent to do so is expressed, or if a prisoner is deemed vulnerable by a member of staff or a fellow prisoner, then this prisoner will be subject to a self-harm and suicide monitoring process known as an ACCT (Assessment, Care in Custody and Teamwork). As part of the ACCT process, a member of prison staff will carry out a prisoner assessment to ascertain the risk level of the prisoner, and whether the ACCT needs to remain open. However, there is currently no standardised risk assessment or clinical decision aid used as part of the ACCT process, which is perhaps largely owing to the paucity of validated risk assessment tools or clinical decision aids that are available to identify risk of self-harm in prisoner populations.

The primary aim of the study was therefore to determine whether any pre-existing, standardised instrument would be suitable to use for the purpose of assessing the risk of self-harm among the specific ACCT population.

This thesis describes various different elements of the study. Firstly, a scoping exercise was carried out in order to select the most appropriate pre-existing instruments that had the potential to predict self-harm events. A pilot study was then carried out to test the study process and to refine the instrument selection. Following this, a prospective cohort study was undertaken to assess the predictive validity of the selected instruments, where a six month follow-up determined self-harm occurrence since baseline, and area-under-the-curve (AUC) analysis examined the ability of the instruments to predict future self-harm. Alongside this, a thorough psychometric analysis of each of the instruments was carried out, in order to validate them among the specified ACCT population. Utilising the study dataset, an exploratory logistic regression analysis was carried out in order to identify individual background and instrument items that may prove effective in predicting future self-harm. Finally, structural equation modelling (SEM) was used to explore the relationships between some of the factors that

influence self-harm, in order to contribute towards an understanding of the complexity of the issue of self-harm in prisons.

The key findings from this programme of work can be summarised as follows:

i. Self-harm is a common occurrence among the prison ACCT population, with 29.1% of the study participants deliberately self-harming during the follow up period, although this varied considerably across gender and participating prisons.

ii. Four of the five selected instruments did display a certain level of psychometric validity among the study population; therefore validating the cut-points for the predictive analysis. However, all instruments required some refinement to meet the strict measurement criteria of the Rasch model.

iii. Of the five pre-existing instruments that were selected for the study, none of these displayed a meaningful predictive validity.

iv. Logistic regression analysis did reveal gender-specific item sets, producing predictive algorithms which were statistically significant in predicting future self-harm; however, the operational functionality of these item sets may be limited.

v. Structural equation modelling revealed an insightful explanatory model of the process that may be involved in the culmination of self-harm in prison. Path analysis models supported the view that self-harm capacity and self-harm propensity are integral elements to the self-harm pathway, although the complete explanatory model is likely to be more complex.

Contribution to knowledge

The work contained within this thesis provides the following contributions to the knowledge within the field:

- To date, this is the largest prospective study of its kind, carried out in order to investigate the properties of a number of pre-existing measures with a view to prospectively predicting self-harm within a prison setting.

- This thesis provides evidence that none of the (five) measures that were selected for the study (BSL-23-F, CORE-OM, PHQ-9, PriSnQuest, SHI) are usefully predictive of prospective self-harm within the specifically targeted ACCT population (those already identified as being at an increased risk of self-harm).

- An in-depth psychometric assessment was carried out for each of the instruments within the specific prison-ACCT population. This provided a population-specific validity assessment of the instruments, which has not previously been carried out.

- This thesis also investigated the self-harm predictive utility of each of the individual items within the dataset that was collected. This provided new and further evidence with regard to potential self-harm risk factors and protective factors among the prison-ACCT population.

- This thesis provides an investigation of a simplified interpretation of Ireland & York's explanatory Integrated Model of Self-Injurious Activity. This was investigated using interval-level Rasch transformations of instrument scores, which were incorporated into a structural equation modelling framework. This approach offers a new perspective, with evidence providing new support for the conceptual model.

Publications

Mike Horton, Nat Wright, Wendy Dyer, Alex Wright-Hughes, Amanda Farrin, Zanib Mohammed, Jamie Smith, Tom Heyes, Simon Gilbody, and Alan Tennant. *Assessing the risk of self-harm in an adult offender population: an incidence cohort study.* Health Technology Assessment, 2014. No. 18.64

Table of Contents

List of Abbreviations

ACCT	Assessment, Care in Custody and Teamwork
AUC	Area Under the Curve
BDI	Beck Depression Inventory
BHS	Beck Hopelessness Scale
BPD	Borderline Personality Disorder
BSL	Borderline Symptom List
BSL-23	Borderline Symptom List-23
BSL-23-F	Revised Borderline Symptom List-23 (Frequency-based responses)
BSL-23-S	Borderline Symptom List-23-Supplement
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CI	Confidence Interval
CInt	Class Intervals
CORE-10	Clinical Outcomes in Routine Evaluation System – 10 item short-form
CORE-OM	Clinical Outcomes in Routine Evaluation System – Outcome Measure
CTT	Classical Test Theory
DASS-21	Depression, Anxiety and Stress Scales-21
df	Degrees of Freedom
DIF	Differential Item Functioning
DSHI	Deliberate Self-Harm Inventory
DSM	Diagnostic and Statistical Manual of Mental Disorders
FASM	Functional Assessment of Self-Mutilation
HADS	Hospital Anxiety and Depression Scale
IPTSB	Interpersonal-Psychological Theory of Suicidal Behaviour

IQR	Interquartile Range
IRT	Item Response Theory
MTT	Modern Test Theory
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NOMIS	National Offender Management Information System
NPV	Negative Predictive Value
NSSI	Non-Suicidal Self-Injury
PCT	Primary Care Trust
PHQ-9	Patient Health Questionnaire
PORSCH	Prison and Offender Research in Social Care and Health
PPV	Positive Predictive Value
PriSnQuest	Prison Screening Questionnaire
R&D	Research and Development
RDS	Referral Decision Scale
RMSEA	Root Mean Square Error of Approximation
RMT	Rasch Measurement Theory
ROC	Receiver-Operating-Characteristic
SCOPE	Suicide Concerns for Offenders in Prison Environment
SD	Standard Deviation
SEM	Structural Equation Modelling
SH	Self-Harm
SHI	Self-Harm Inventory
TLI	Tucker-Lewis Index
VIF	Variance Inflation Factor
WRMR	Weighted Root Mean Square Residual

List of Figures

List of Tables

1 Introduction

The reported incidence of self-harm (SH) is much higher among prisoners than in the general population. (1, 2) Along with the general increased risk among prisoners, there are also specific times of heightened vulnerability, such as on first reception into the prison environment. (3) Currently in the UK, if an act of self-harm is carried out by a prisoner, if intent to do so is expressed, or if a prisoner is deemed vulnerable by a member of staff or a fellow prisoner, then this prisoner will be subject to a self-harm and suicide monitoring process known as an ACCT (Assessment, Care in Custody and Teamwork). (4) As part of the ACCT process, a member of prison staff will carry out a prisoner assessment to ascertain the risk level of the prisoner, and whether the ACCT needs to remain open. However, within this process, there is currently no standardised risk assessment or clinical decision aid used to help inform the process. This is largely owing to the paucity of validated risk assessment tools or clinical decision aids that are available to identify the risk of self-harm in prisoner populations. (5)

The primary aim of this study was therefore to determine whether any pre-existing, standardised instrument would be suitable to use for the purpose of assessing the risk of self-harm among the specific ACCT population.

The working hypothesis of this study is as follows:

Self-harm within the prison ACCT population can be predicted using a pre-existing screening instrument.

In order to test this hypothesis, a multi-stage project was carried out which consisted of the following elements:

1. A scoping study to identify potentially useful screening instruments
2. A pilot study to test the logistic process and to further refine the instrument selection
3. A prospective cohort study where prisoners on an ACCT will be administered the selected instruments, and then followed-up six months later to determine whether any incidence of self-harm had occurred
4. A psychometric analysis of each of the study instruments, to determine their psychometric properties and the validity of the total instrument score
5. An assessment of the predictive validity of each of the study instruments
6. A path analysis study to assess the relationships between a number of contributing factors and how they impact on the final outcome of self-harm.

The thesis has been structured as follows:

Chapter Two is a review of the literature and provides the background information to the program of work that was undertaken. Although this work is regarding the issue of self-harm within a prison setting, the chapter starts with an introduction to the application of screening principles, as this provides the methodological foundation on which the research project is based. In this project, these screening principles were applied to a specific population of people who carry out self-harming behaviours within a prison environment; therefore this chapter goes on to introduce the topic of self-harm, including some of the differing viewpoints, definitions and issues regarding the subject. Following this, some of the research into self-harm risk factors and prediction is presented, before the self-harm research area is framed in the context of the prison environment.

Chapter Three presents the methodological and logistic basis of the study, leading to the implementation of the pilot study. As there are a lot of individual risk factors for self-harm, there are also a large amount of standardised psychometric instruments that may be potentially predictive of self-harm. This chapter presents the process involved with identifying and refining the selection of instruments that would eventually be used in the cohort study. The results of the scoping exercise and pilot study are then presented, along with the implications of these results for the cohort study.

Chapter Four provides a detailed account of each of the methodologies used within the thesis. Firstly, the study is described in terms of the setting and participants, and the process that was followed in order to collect the relevant data. An introduction to measurement and psychometrics is then provided, including information regarding the confirmatory factor analysis (CFA) and Mokken scaling methodologies, and a more detailed description of Rasch analysis, which was the predominant methodology utilised for the purpose of instrument validation. The methodologies involved in ascribing the predictive validity of each instrument are then described, along with the process used to identify individually predictive items, and how these may be combined into a predictive item set. Finally, an introduction to structural equation modelling (SEM) is provided, which is a methodology that can be used to explore the complex interaction between traits (variables), and how the relationships between a number of traits may influence a final outcome, which in this case is whether or not self-harm occurred.

Chapters Five, Six, Seven and *Eight* present the results of the different separate elements of the study:

Chapter Five presents the descriptive results of the cohort study, including details of study recruitment, participant characteristics and the self-harm incidence that was observed within the study cohort.

Chapter Six presents the results of the psychometric analysis for each of the separate five instruments that were utilised within the cohort study. This includes information regarding the basic psychometric details for each of the instruments, followed by the CFA and Mokken scaling results. A detailed breakdown of the Rasch analysis is presented for each of the instruments, including information relating to each of the individual items (questions) within each instrument. Additionally, a summary of two separate Rasch-based resolutions are presented, where these were appropriate.

Chapter Seven reports the results of the predictive validity element of the study. This section presents the results of the area-under-the-curve (AUC) analysis for each of the instruments that were used within the cohort study, also presenting the receiver-operating-characteristic (ROC) curves where appropriate. Additionally, this chapter reports the results from the exploratory analysis of individual predictors of self-harm, and how these individual items may be combined in order to create a screening algorithm instrument that could be used as part of the (initial) ACCT process.

Chapter Eight explores possible explanatory mechanisms that may contribute to the final outcome of self-harm. A structural equation modelling (SEM) approach offers the opportunity to assess the conceptual relationships between some of the separate constructs that were captured by the study instruments. This chapter presents the findings of the SEM analysis, where a pre-existing conceptual model was used to derive and test a number of different path models. This analysis offers the potential to contribute to the understanding of the process that leads to self-harm, which may ultimately help to inform care pathways and targeted interventions

Finally, *Chapter Nine* presents a discussion of the results in relevance to the existing body of literature, along with the practical implications of the results within the prison setting. The limitations of the different elements of the thesis studies are also presented in this chapter, along with recommendations for further study. This chapter concludes with a concise summary of the research that is presented within this thesis.

2 Background and Review of Literature

Although this thesis generally concerns the issue of self-harm within a prison setting, this chapter starts with an introduction to the application of screening principles, as this provides the methodological foundation on which the research project is based. In this project, these screening principles were applied to the specific population of interest, which is those who carry out self-harming behaviours within a prison environment. Following the initial introduction to screening, this chapter goes on to introduce the topic area of self-harm, including some of the differing viewpoints, definitions and issues regarding the subject. Following this, some of the research into self-harm risk factors and prediction is presented, before the self-harm research area is framed in the context of the prison environment.

2.1 Screening

The term 'screening' refers to a process of sorting and separation. In a medical context the purpose is to separate a population into groups (i.e. those affected and those not affected) through the application of a particular test or instrument. The intention of screening is to identify something potentially harmful earlier than it would normally be identified. The identification may be made before anything harmful has actually occurred, or it may be that a harmful condition could be picked up earlier than usual within its progression. In some instances, the early identification can lead to interventions being put in place that may lead to the prevention of anything harmful occurring. However, screening tests often cannot stop a disease or condition from developing, although the early identification may allow for early treatment, meaning that the impact of the condition may potentially be reduced.

The widely accepted definition of the term ‘screening’ is the one that was first proposed in 1957 by the United States Commission of Chronic Illness:

“the presumptive identification of unrecognised disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment”. (6)

This definition was also used in a landmark 1968 report on screening by Wilson and Jungner, which was commissioned by the World Health Organisation. (7) The report went on to suggest 10 important criteria that should be satisfied when a screening programme is implemented:

1. The condition sought should be an important health problem.
2. There should be an accepted treatment for patients with recognized disease.
3. Facilities for diagnosis and treatment should be available.
4. There should be a recognizable latent or early symptomatic stage.
5. There should be a suitable test or examination.
6. The test should be acceptable to the population.
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood.
8. There should be an agreed policy on whom to treat as patients.
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
10. Case-finding should be a continuing process and not a “once and for all” project.

While these criteria are still viewed as the gold standard, (8, 9) Andermann et al. (2008) (10) proposed a number of additional screening criteria that should be considered:

- The screening programme should respond to a recognized need.
- The objectives of screening should be defined at the outset.
- There should be a defined target population.
- There should be scientific evidence of screening programme effectiveness.
- The programme should integrate education, testing, clinical services and programme management.
- There should be quality assurance, with mechanisms to minimize potential risks of screening.
- The programme should ensure informed choice, confidentiality and respect for autonomy.
- The programme should promote equity and access to screening for the entire target population.
- Programme evaluation should be planned from the outset.
- The overall benefits of screening should outweigh the harm.

The original and additional criteria all provide useful guidance, but some of the individual criteria may be more or less relevant depending on the screening process under consideration.

It is important to note that a screening test is not intended to be used diagnostically, (7) although they are sometimes used in this way. Usually the intention of a screening instrument is simply to group people into those who require a further assessment and those where no further assessment is necessary. As is implied within the definition, the screening process is a gross filtration rather than a perfect separation. (11)

The specifics of how to ascertain the quality of a screening test will be covered later (in Section 4.6), but ideally a screening test would correctly identify and classify all true cases, and similarly the test would identify and classify all individuals who are not at risk (i.e. often termed ‘healthy’ or ‘disease free’ individuals).

The possible outcomes of an individual screening test are as follows:

- True positive: the individual IS at risk (or has the condition) and the test is positive.
- False positive: the individual IS NOT at risk (or does not have the condition) but the test is positive
- True negative: the individual IS NOT at risk (or does not have the condition) and the test is negative
- False negative: the individual IS at risk (or has the condition) but the test is negative (12)

While the ideal situation is for a test to obtain all true positives and true negatives, in reality no screening test is perfect; therefore the intention is to maximise the true positives and true negatives, whilst minimising the number of false positives and false negatives.

False negatives are problematic as the screening instrument has failed to identify a true case, meaning that the health implications for the individual with the false negative are undesirable. False positives are problematic in a different way, as they may lead to unnecessary treatment, which is undesirable in terms of resource allocation as well as the potentially harmful treatment or psychosocial consequences for the individual with the false negative result (11).

Because no screening test is perfect, and because both the benefits and drawbacks can be health-related or economical, it is usual that a trade-off is necessary between these two elements, with the terms varying depending on the specific conditions of each individual situation. In some circumstances missing a single case could be crucial, so the associated false positives are seen as a side-effect that is necessary in order to identify all of the true cases. In other circumstances, a large amount of unnecessary treatment may overwhelm a healthcare system, so the intention of the screening programme would be to reduce the number of false positives. The associated increase in false negatives would be viewed as a side-effect that is necessary in order to reduce the burden of the healthcare provision system.

The principles of screening are generic, and widely applicable to a range of situations and specific populations. In the case of this project, the intention is to apply the principles of screening to a specific population of people who carry out self-harming behaviours within a prison environment. It is therefore not only appropriate, but also necessary, to outline some of the background, along with differing viewpoints and definitions, regarding self-harm.

2.2 Self-Harm

Self-harm is a human behaviour that has been around for a long time, perhaps even as long as humans have existed, in different cultures and in various geographical locations. (13) At present, a basic understanding of self-harm, including its classification, diagnosis, and treatment, is still lacking, (14) and it has been recognised that *"there is so much information on self-harm in medical, nursing, women's studies and psychology literature as to be almost overwhelming"*. (15)

The complexities regarding self-harm seem to have been around ever since the behaviour has been documented, and even within these early documentations, the ambiguity with regard to the nature of self-harm is present. (16) With regard to this ambiguity concerning what 'self-harm' actually is, the term 'self-mutilation' was used in a study by Emerson in 1913 (17), and in 1938, Menninger (18) attempted to sub-categorise self-mutilative behaviour separately from other suicide-related behaviours. Although much research has been done since then, the ambiguity and inconsistency remains with regard to how self-harm should be viewed. It has been suggested (16) that the field of self-harm research is divided into those who consider self-harm to be a broad continuum of self-injurious behaviours, irrespective of the underlying intent, and those who are in favour of firm categorisation of self-harm into that with and without suicidal intent. The viewpoint that self-harm belongs to a broad continuum of behaviours, regardless of the intent, is based on Kreitman's (19, 20) seminal work on parasuicide. Kreitman's terminology of 'parasuicide' was intended to classify all non-accidental hospital-treated self-poisoning and self-injury that did not result in death, regardless of the intention of the act. (20) The opposing viewpoint that self-harm is distinctly separate from suicide largely stems from the early work of Beck and colleagues. (21, 22)

This epistemological separation broadly remains within the literature, but the area of definition and terminology used in self-harm research continues to be somewhat confusing and inconsistent. This inconsistency has given rise to a large number of different terms for self-harm, and there are often variable definitions which relate to each of these terms. Some of the different terms used to describe the phenomenon of self-harm can be seen in Figure 1.

Figure 1. Alternative terms for self-harm

All of these definitions relate to the specific intention of the act to harm oneself, although the motivating factor underlying the self-harm may differ. However, there are also certain terms which are more closely associated with suicide, where the final outcome of death is seen as the motivating factor. These behaviours are also often classified as self-harm behaviours, although the terminology relates more specifically to the suicide intention:

- Suicide attempt
- Suicidal behaviour
- Suicidal gesture
- Suicide ideation

This proliferation of different terms often leads to confusion with regard to what is actually being studied. Even when the self-harm definition and classification information is explicitly stated, it often remains difficult to compare information across studies as the different definitions mean that the study populations are not consistent. It has been recognised (23) that consistency among terms used to describe self-harm is an area that could benefit from integration, as the variety of different names and terminologies creates confusion regarding which specific construct is under investigation. (24, 25) This view is supported by others (26), and, while using some of these alternative terms, part of the difficulty in understanding self-injurious behaviour is due to the multiple terms used to describe the behaviour and the confusion surrounding whether or not self-injurious behaviour represents a suicide attempt. (27)

This lack of consistency in self-harm definition is nicely summarised by Silverman (28):

"There is no uniform set of terms, definitions, and classifications for the range of thoughts, communications, and behaviours that are related to self-injurious behaviours, with or without the intent to die. Nor is there an agreed taxonomy that encompasses the full spectrum of what is clinically defined as suicide-related behaviours. As a result, researchers cannot easily compare their study populations or results, and clinicians have difficulty in translating research findings into practical applications when working with patients at risk for suicidal behaviours."

2.2.1 Defining Self-Harm

As has been acknowledged, there are a large number of different terms and definitions used to describe self-harming behaviours. Some of the definitions applicable to self-harm are as follows:

"The term self-harm covers a spectrum of behaviour. The most serious forms relate closely to suicide, while behaviours at the milder end of the spectrum merge with other reactions to emotional pain."

Skegg, 2005. (29)

"…the deliberate destruction or alteration of body tissue without conscious suicidal intent."

Favazza, 1989. (30)

"Self-injury is a behaviour that involves deliberately injuring one's own body, without suicidal intent and with or without pain."

Duffy, 2006. (31)

"…an intentional act resulting in bodily injury to oneself in a direct and socially unacceptable manner."

Favazza & Rosenthal, 1990. (32)

"…self-poisoning or self-injury, irrespective of the apparent purpose of the act."

NICE, 2004. (33)

These differing statements highlight the apparent lack of consistency in how self-harm is defined, and even within each definition, it may still be difficult to classify various self-harm related behaviours.

Smith & Kaminski (34) qualified that Favazza's (30) definition includes moderate self-harm acts such as cutting, scratching, or burning the skin; hitting oneself; pulling one's hair; reopening one's wounds or breaking one's bones. Also included are more severe self-harm acts such as eye enucleation, face mutilation and amputation of limbs, breasts and genitals. Latimer et al. (35) further identified other behaviours that are classified as self-harm; these include a lack of self-care (36), swallowing dangerous objects (37) and using acid to burn skin. (38) Furthermore, some deliberately reckless behaviours that are intended to cause harm are also classified as self-harming behaviours. (29) Examples of these behaviours include sexual risk-taking, (39) intentional overuse of drugs (40) and reckless driving. (41) These behaviours introduce a further complication in that not all of them induce direct self-harm. There is also ambiguity as to where the line is drawn with indirect self-harm, as it has been stated that that indirect forms of self-injury such as chronic alcoholism or smoking are not classified as self-harm because the harm caused is temporally remote. (42)

It should be noted that Favazza (30) excluded from his definition the common expressive forms of body modification such as tattooing and piercing, and this is reflected within another term and definition which includes a social acceptability aspect, along with including the qualifier of the behaviour being repetitive:

"Non-suicidal self-injury (NSSI) represents the direct, repetitive, intentional injury of one's own body tissue, without suicidal intent, that is not socially accepted."

Lloyd-Richardson et al., 2007. (43)

Within the range of different working definitions, it has been identified (16) that there are broadly four main elements that characterise self-harm, and that different definitions will relate in different ways to these four elements of intent, method, lethality, and outcome.

Intent

Intent refers to the motivation behind any self-harming action, and more specifically whether the self-harmer had suicidal intentions when the self-harm was carried out. This element is key within a number of self-harm definitions, (21, 22) but it has been argued that intent is hard to measure reliably as it can be assessed incorrectly as well as being subject to recall bias. (16) Also, many people are unclear about their intent when self-harming, and suicidal and non-suicidal intent and behaviour can often coexist within the same individual, (16) along with multiple motives being present in both suicide attempts and self-harm without suicidal intent. (28) Many patients who self-harm deny that they have an intent to die, (28) and it has been shown that a vast majority of so-called suicide attempts are in fact episodes of self-injury without any suicidal intent. (44)

Lethality

Typically, lethality refers to the medical or biological danger to life, (45) and although it is a key component of the nomenclature, it remains more closely linked with suicide than self-harm.

When it is applied to the assessment of risk of death by suicide, it reflects the potential for death associated with the means used to attempt suicide. (46) Although an association between objective lethality and risk of dying by suicide has been recognised, (47) it has also been found that a third of cases reporting with near-fatal self-harm had no suicidal intent, (48) and that the association between the degree of suicide intent and the medical lethality of suicide attempt was minimal. (49)

Despite lethality being more closely associated with suicide, it is acknowledged within the broader definitions of self-harm and falls within Skegg's (29) defined spectrum of behaviours.

Method

Method refers to the process used by an individual in order to induce self-harm. Each method could be categorised as a different self-harm behaviour, and people who self-harm often employ a number of different methods and behaviours. The method may be related to the lethality of the act, although the level of lethality may often be under or over estimated by the individual self-harming. (50)

The method of self-harm has been used as a way to define a particular research population (51-53) as it is a way of providing consistency, although defining purely on method would be likely to restrict the study population in question.

The number (count) of different methods or behaviours employed to carry out self-harm is believed to be an important factor when assessing the risk of future self-harm. It has been found that when compared to frequency or recency of any single method, the count of different methods has the strongest association with psychopathology, (54) and that the number of different methods or behaviours is the best single predictor of future self-harm. (55)

Outcome

Outcome is the consequence of the self-harm event, and this can result in one of three possible states: death, survival with injuries or survival without injuries. (16) Although these outcomes may broadly cover the range, further information may also be required with regard to the extent of the injuries and any recovery timescale that might be involved. Although outcome is a necessary component of a self-harm nomenclature, it should not be regarded as a sufficient classifier when considered by itself. (56)

These four elements of self-harm are recurring throughout the literature, and despite the problems identified within each of these elements, the main area of difference and point of contention between the various definitions relates to whether self-harm comprises behaviours relating to suicide or not. There is literature supporting each of these viewpoints, which leads to an array of research that is often inconsistent and confusing.

2.2.1.1 Self-Harm and Suicide as separate entities

The viewpoint that self-harm is distinctly separate from suicide largely stems from the early work of Beck and colleagues. (21, 22) This notion is widely supported, (23, 30, 34, 57, 58) and many authors share the conceptual viewpoint that self-harm should be viewed as a distinct entity, where any form of self-harm behaviour with suicidal motivation is excluded (or treated separately) from the (non-suicidal) self-harm research population. The key difference of intent also relates to some of the different terminology that is used. For example, the term 'Non-Suicidal Self Injury' (NSSI) is applied to this specific group to differentiate from a different self-harm group that may include suicidal motivations.

This viewpoint is predominant among the majority of the research that has been conducted within North America, and it has been stated (34, 59) that attempted and completed suicides should be treated as aetiologically distinct from self-harm. In fact, it has been pointed out (23) that a lack of distinction between those who are attempting suicide and those who are self-harming with no intent to die is particularly concerning. It is suggested that differentiating between these two groups is of utmost importance when examining functions or explanations of the behaviour, as research that does not differentiate between these groups assumes that there is no difference between them, which will therefore confound results and obscure relevant findings. (23, 25)

Within the context of this distinct separation, it is observed that four characteristics stand out in terms of the self-harming behaviour that is being described: deliberateness, tissue damage without intent to die, social unacceptability, and typical repetitiveness. (23)

Despite research suggesting that self-harm and suicide attempts are two separate entities, (26, 60) confusion remains regarding the terminology and the apparent overlap between the two phenomena.

2.2.1.2 A Continuum of Self-Harm

Contrary to the perspective distinctly separating self-harm with and without suicidal intentions, it has been reported (59) that some authors see both of these phenomena as on a continuum of lethality, and that these authors consider any differentiation to be irrelevant, confusing, and possibly even dangerous. (61, 62)

'Self-injurious behaviour' has been described as any behaviour in which a person directly and deliberately inflicts injury upon the self, including both suicidal and non-suicidal self-injury, (63) and this broad classification of self-harming behaviours has been stated to include actions ranging from stereotypic skin-rubbing to completed suicide. (64-67) It has been identified (68) that according to the proponents of the continuum model, that regardless of the intent of the behaviour, suicidal and non-suicidal self-harm behaviours share common experiential qualities, and that both are intentional acts causing direct bodily harm to oneself. (66)

The issues with determining intent have already been discussed (in Section 2.2.1), and it has been suggested (29) that to identify and describe a self-harming behaviour prior to clarifying the intent is a more realistic approach than trying to label the intent of behaviours from the outset. This supports the finding that most people admitted to hospital after an overdose neither want nor expect to die. (69)

It has been identified that that people engaging in self-harm are a heterogeneous group, and that self-harm behaviours and motivations will differ; therefore caution is needed when generalising about self-harm. (29)

2.2.1.3 Linkage between Self-Harm and Suicide

It is recognised (23) that the matters of distinction between suicidal and non-suicidal self-harm are further complicated as self-harmers are more at risk of attempting suicide, having more suicidal thoughts and a history of suicide attempts. (26, 70) This supports the previous finding that approximately 55%-85% of 'self-mutilators' have a history of at least one suicide attempt. (66) A strong statistical connection between self-harm and subsequent suicide has also been reported, with an estimation that around a quarter of suicides are preceded by self-harm in the previous year, (71, 72) and that people who deliberately harm themselves have a 30-fold increased risk for completing suicide compared to those who do not self-harm. (73)

Although it has been recognised that prior self-harm is one of the strongest predictors of suicidal attempts, both cross-sectionally and longitudinally, (68) it has been argued that the studies which indicate self-harm as a precursor to suicide do not differentiate between the suicidal intent behind the self-harm behaviours. (16, 68) It remains unclear whether non-suicidal self-harm specifically increases the risk of suicidal behaviour, as an increased risk of completed suicide has not been established on a purely non-suicidal self-harm sample (16) and previous non-suicidal self-harm could easily be incorrectly classified as a suicide attempt. (68)

Despite these doubts regarding the classification of suicidal versus non-suicidal self-harm, it has been shown that both groups are actually at a higher risk of a subsequent suicide attempt. (54, 74)

Although it might appear to be relatively easy to differentiate between suicidal and non-suicidal self-harm, it has proven, both practically and empirically, to be difficult to separate these phenomena. (58) Given the high co-occurrence of these behaviours, it has been concluded that although suicidal and non-suicidal self-harm may differ in important ways, these behaviours are also related. (75)

In an attempt to bridge the gap between suicidal and non-suicidal self-harm, some explanatory theories have been suggested. These include the 'Gateway Theory', where non-suicidal self-harm is seen as a 'gateway' form of self-harm that leads to more extreme forms (that may be considered as suicide attempts); and the 'Third Variable Theory', which suggests that the association between non-suicidal and suicidal self-harm behaviour is spurious, and that a third (changeable) variable accounts for any co-occurrence of non-suicidal and suicidal self-harm behaviours. (68)

Alternatively, a more widely supported (16, 68) theoretical framework is Joiner's Interpersonal-Psychological Theory of Suicidal Behaviour (IPTSB). (76) This theory describes three main domains which contribute towards an increase in the risk of suicide: the feeling of being a burden or liability to loved ones (perceived burdensomeness); the sense of isolation and a lack of connection to others (thwarted belongingness); and the learned ability to hurt oneself (acquired capability), which reflects the degree to which an individual is able to enact a lethal suicide attempt. (16, 68, 76)

Joiner's theory of acquired capability (76) is supported by the finding that the number of different self-harm behaviours is significantly associated with subsequent suicide attempts, but not the number of episodes. (54) This suggests that a continuum of self-harm behavioural severity may be present, and that some people find a method of self-harm that works for them and that they continue using this single method. These self-harm events become habitual, but offer immediate gratification to the person self-harming. This form represents typical non-suicidal self-harm, and it is not progressive. The individuals engaging in these non-progressive behaviours are also unlikely to be experiencing feelings of perceived burdensomeness or thwarted belongingness. (68) Some people, however, seek increasingly diverse and severe methods of self-harming, and the increased use of behaviours would mark a progression along the continuum, with the user becoming accustomed to more severe forms of self-harm behaviour. This progression along the spectrum results in an individual acquiring more capability to carry out a suicide attempt. As an individual progresses along the continuum, a shorter transition is necessary for a self-harm behaviour to turn into a suicidal behaviour, regardless of the intent behind it. (77)

In addition to offering a linkage between self-harm and suicide, Joiner's IPTSB framework has also been applied and extended within a prison setting, where Ireland & York (78) used it as the basis for their Integrated Model of Self-Injurious Activity.

2.2.1.4 Defining Self-Harm within this study

It is recognised that the study of self-harm and suicidal behaviour is difficult due to the conceptual and empirical confounding of the variables when different definitions are applied. (79) This view is echoed by Silverman, where it is stated that the literature remains replete with confusing terms, definitions, and classifications that make it difficult, if not impossible, to compare and contrast research, epidemiological or clinical studies. (28)

It is clear from the differences in definition throughout the literature that there is a lack of consistency in what exactly is meant by 'self-harm'. Various different definitions may be better suited to some of the different theoretical frameworks of self-harm, and perhaps the context of the research study has a bearing on the definition of self-harm that is used, as in some cases it would be highly impractical to separate out suicidal versus non-suicidal self-harm events. Although the definition of self-harm might vary among researchers, it seems important to state the active definition of self-harm that is used for a study, so at least some attempt can be made to compare and contrast studies which use an equivalent basis for the classification of self-harm.

Although suicidal and non-suicidal self-harm behaviours may be separated by the motivational intent, this may be irrelevant to primary care teams and authorities who are charged with dealing with any sort of self-harming behaviour, regardless of the prior motivating factor. This view is supported by Lanes, (80) who stated that it is important to note that self-harmers generally distinguish between self-harm and genuine suicidal intent, but this does not qualify as a basis for judging the potential outcome of threatened or enacted self-harm. Despite the motivational and aetiological differences between suicidal and non-suicidal self-harm, as the final outcome is likely to be similar in terms of treatment cost and impact, it may make sense, from a public healthcare commissioning perspective, to group all self-harm behaviours together, regardless of the intent.

Considering the public health implications that are present in the prison setting of this study, the definition of self-harm as provided by the National Institute for Health and Care Excellence (NICE) (33) may potentially be the most appropriate; here, it is described as:

"…self-poisoning or self-injury, irrespective of the apparent purpose of the act".

This also corresponds to the definition of self-harm used within prison custody, where it is defined as, *"any act where a prisoner deliberately harms themselves irrespective of the method, intent or severity of any injury"*. (81)

This definition is all inclusive, and thus relates more closely to epidemiological outcome events; it covers a broad range of behaviours, and makes no assumptions with regard to the underlying intent. At a practical level, within a prison setting an episode of self-harm would be recorded and treated (medically) in the same way, regardless of the underlying intent. In commissioning terms, to separate out self-harm events based on intent would be illogical; therefore it is felt that this broader definition is the most appropriate for this research setting.

Although much of the self-harm literature relates to the distinct separation of suicidal and non-suicidal self-harm, prison-based outcome event statistics may not distinguish between the two without a degree of more in-depth information being available. Despite this restriction, to make the distinction between suicidal and non-suicidal self-harm in the prison setting would probably not be helpful given that prison authorities are ultimately concerned with preventing death as well as self-harm. (82)

The NICE definition of self-harm (33) is to be taken as the active definition within this study, but reference may also be made to other self-harm terminology that is used within the literature. The NICE definition may be different to the self-harm definitions that were used within other studies, and therefore results of such studies will not be directly comparable. When studies are cited, they should be taken in the context of their own specific definition of self-harm and the corresponding associated terminology.

2.2.2 Theories of Self-Harm

With regard to the reasons for and motivations behind self-harm, there are many theories that exist in the literature, some of which are complementary, others not. Again, there is some crossover with self-harm and suicide theory, but various studies suggest that self-harm differs from suicide attempts in clinically important ways. (28) There are various overlapping reasons for carrying out self-harm that have been reported, and these include: to die; to escape from unbearable circumstances; to influence others; to feel better; to feel relief from anxiety or tension; to temporarily reduce anger, sadness, depression or shame; as a form of self-punishment; as relief from self-blame or self-loathing; as a form of anger expression; to serve as a form of emotion regulation; and as a form of distraction. (16, 28, 49, 83-85) Suicide attempts on the other hand, are thought to relate more closely to the perceived notion of making someone else better off. (49) Again, the key factor to differentiate the behaviours of self-harm and suicide attempts relate to the underlying intent to die, but it is important to note that multiple motives often underpin both sets of behaviours. (28)

There are many triggers for an act of self-harm, including mental disorder, psychological distress arising from reactivated memories of past trauma, current distress, financial difficulties, and social difficulties. (15, 86) Theoretical models have been proposed that explain self-harm as a combination of biological sensitivity arising from trauma and activated by social factors. (86) Some of the more historically significant explanatory models of self-harm were reviewed by Messer & Fremouw, (23) and these include the following:

The sexual/sadomasochistic model

This model suggests that non-suicidal self-harm relates to sexual development and concerns of sexuality, and it has also been associated with body image and sexual confusion. (23, 87) It has been suggested that self-cutting can be viewed as a means of gratifying oneself sexually whilst also punishing any feelings of sexual desire and self-gratification. (88, 89)

The depersonalisation model

This model focuses on the psychological state of dissociation, or depersonalisation, which is assumed to stem from feelings of abandonment or isolation. It has been suggested that self-harm may be carried out in order to re-establish a sense of identity or to regain a sense of self, and also that scars from self-harming may serve as reminder of existence and identity. (23, 90)

The interpersonal/systemic model

The interpersonal model and the systemic model are actually separate models, but they are similar in that they both involve the role of others as a contributor to self-harming behaviour. These models are symptomatic of family or environmental dysfunction, where the ‘system’ could be viewed as the family or a residential home, hospital, or prison environment. (89)

The suicide model

The suicide model considers that acts of self-harm (self-mutilation) should be viewed as attempts to avoid suicide. On a continuum presented by Firestone and Seiden, (91) it is suggested that self-harm and thoughts to injure oneself without an intent to die should be viewed as "microsuicides".

The physiological/biological model

Most explanatory models of self-harm focus on the psychological element, but this model suggests that physiological factors also play a role in self-harm. It has been suggested that people engaging in self-harm may have a biological vulnerability to self-harm due to a dysfunctional neurotransmitter system or an abnormal psychophysiological response to self-harm which involves tension reduction. (23, 66, 92)

The affect regulation model

This model is also known as the 'expression model', the 'emotional regulation model', or the 'mood regulation model'. This model suggests that self-harming behaviour is used as a way of controlling or regulating various emotional states such as tension, anxiety, anger, hostility, and depersonalisation. Increased levels of these emotional states have all been found to precede self-harming acts, which is then used as a regulator to control the emotional state. (23, 66, 89, 93)

The behavioural/environmental model

This model suggests that certain environmental factors may initiate and maintain self-harm behaviour. Acts of self-harm may be reinforced through external gain from the environment, increased attention from others, inclusion in a group, or, tying in with the affect regulation model, through an internal relief or escape. (23, 89, 94, 95)

Understanding the motivations and catalysts for the initiation of self-harming behaviour is a complex task, and it is likely that aspects of several of the explanatory models are likely to contribute to understanding the phenomenon. (23, 96)

A biopsychosocial vulnerability-stress model

Although all of the above models have been separately proposed, they are not necessarily mutually exclusive, and different models may be more, or less, appropriate for each individual case. These competing models serve to remind us that self-harm is a complex issue where a single explanatory theory may not exist, as each individual will be varyingly effected by different aspects from each model. A biopsychosocial vulnerability-stress model has recently been proposed specifically for an incarcerated population (97), and this combines several elements of different models to create a complex explanatory framework of individual vulnerabilities and stressors that may contribute to self-harm.

2.2.3 Predictors of Self-Harm

Self-harm behaviours have always previously been listed as a diagnostic symptom of borderline personality disorder (BPD) in the DSM-IV. (98) However, 'non-suicidal self-injury' (NSSI) was added to the DSM-V as its own disorder, which can occur independently of BPD, such as in patients with depression or even in those with no other diagnosable psychopathology. (99) Individuals that self-harm are diagnostically heterogeneous, and may have a range of other psychological disorders other than borderline personality disorder. (100, 101) These co-occurring diagnoses include major depression, anxiety disorders, alcohol abuse, substance abuse, eating disorders, post-traumatic stress disorder, bipolar affective disorder, factitious disorder (Munchausen Syndrome), schizophrenia, body dysmorphic disorders, and several personality disorders. (14, 15, 100-104) Although diagnostic disorders are common among individuals that self-harm, self-harm behaviours are also occur in nonclinical samples. (14, 100, 101)

Given the overlap between non-suicidal and suicidal self-harm, it is unsurprising that several studies have identified shared risk factors for non-suicidal and suicidal self-harm. These include depression; borderline personality disorder; physical or sexual abuse; externalizing behaviours; impulsivity; and family problems. (68) There are also a range of stand-alone associated risk factors for the singular outcomes of completed suicide and (non-fatal) self-harm. Additionally, when self-harm has previously occurred, the two main outcomes that are of particular importance are repetition of self-harm and suicide, (105) and there are also associated risk factors for these two outcomes. The associated risk factors for these four specific outcomes (suicide, non-fatal self-harm, repetition of self-harm, and suicide following self-harm) are summarised in Table 1.

Table 1. Associated risk factors of suicide, non-fatal self-harm, repetition of self-harm, and suicide following self-harm

Associated Risk Factor	Suicide	Non-fatal self-harm	Repetition of self-harm	Completed suicide after self-harm
adverse childhood experiences		■		
adverse social circumstances		■		
alcohol or drug-related problems	■	■	■	■
antisocial personality			■	
anxiety disorder		■		
at present address for less than 1 year		■		
bereaved				■
certain occupations (e.g. farmers, doctors, dentists, lawyers)	■			
criminal record			■	
current involvement with police	■			
debt		■		
depression	■	■		■
externalising behaviours	■	■		
family problems	■	■		
female sex		■		
higher number of adverse life events		■		
homosexual or bisexual orientation		■		
hopelessness			■	■
impending or recent job loss	■			
impulsivity	■	■		
intoxication		■		
lack of co-operation with treatment			■	
living in unsuitable, overcrowded accommodation		■		
male sex	■			■
older age	■			■
perceived lack of social support		■		
personality disorder	■	■		■
physical or sexual abuse	■	■		
previous history of self-harm	■		■	■
prison	■			
probable psychosis		■		
psychiatric history	■		■	
recent discharge from psychiatric care	■			
schizophrenia				■
separated, widowed, or divorced relationship status	■	■		
serious physical illness	■			■
social isolation	■			■
sociodemographic disadvantage		■	■	
suicidal intent			■	■
under 25 years old		■		
unemployment	■	■	■	■

*Information obtained from (29, 68, 86, 105-109)

Despite the wide recognition of these risk factors for repetition of self-harm and completed suicide following self-harm, it is acknowledged that these risk factors only have a limited everyday usefulness due to their poor predictive value. (106) However, it has also been suggested that people who harm themselves by self-cutting are at a greater risk of repetition and eventual suicide than those who self-harm using other methods. (73, 110)

It has been stated that self-harm is one of the strongest predictors of completed suicide, (86) and that underlying depression is the most important contributor in the progression from self-harm to suicide. (86, 111) However, again the issue with regard to 'self-harm' definition comes to the fore, as 'suicide attempt' and 'self-harm' may be categorised as the same or different behaviours depending on the definition. With this in mind, it has been recognised that a prior suicide attempt is statistically the best predictor of future suicide attempts and death by suicide. (28, 112) This is specifically making reference to 'suicide attempt' rather than 'self-harm', but again this could lead to some crossover or confusion within the areas of research. A prior 'suicide attempt' could well be a drug overdose, which could be recorded as a 'self-harm' event. This would appear to make self-harm events predictive of future suicide, but this would be due to the overdose being categorised as a self-harm event rather than a suicide attempt. Again, this terminology of what constitutes a suicide attempt is inconsistent, and it has been identified that it is imperative for clinicians, researchers, and epidemiologists to have a clear and consistent definition of what is a suicide attempt, as such a standardised definition does not presently exist. (28) In recognition of this potential ambiguity in definition, it has been stated that non-suicidal self-injury is seen as a risk factor for suicide behaviours, but suicide attempts and behaviours are not predictive of non-suicidal self-injury, (68) and although the link between self-harm and suicide is well established, the link between non-suicidal self-harm and suicide remains to be shown. (112)

2.2.3.1 Evidence of screening for self-harm

A number of studies have looked into developing tools or rules to help to predict or classify the risk of further self-harm or suicide following an initial occurrence of self-harm. There is evidence to suggest that a number of the common risk factors could combine to create a tool that is useful for application in a clinical setting or an emergency department. The Manchester Self-Harm Rule (113) comprises of four correlates: any history of self-harm, previous psychiatric treatment, benzodiazepine use in current presentation, and any current psychiatric treatment. The confirmation of any one of these factors would classify a patient as being at 'high' risk of repeating. This was shown to predict 94-97% of self-harm repeaters during a 6-month period (sensitivity), (113) and the predictive performance of this rule was also shown to be favourable to the global assessments of clinicians. (114)

Based on the same principles, and with the same objective in mind, the Manchester Self-Harm Rule was developed further into the ReACT Self-Harm Rule, (115) which also comprises of four elements: recent history of self-harm (in the past year), living alone or homelessness, cutting used as a method of self-harm, and treatment for a current psychiatric disorder. Again, the confirmation of any one of these factors would classify a patient as being at 'high' risk of repeating. This was shown to predict 90-95% of self-harm repeaters during a 6-month period (sensitivity), and the predictive performance of this rule was deemed to be favourable to the Manchester Self-Harm rule due to favourable specificity values. (115)

Both of these studies aimed to derive a useful rule from a range of individual risk-factors that were extracted from existing information following a self-harm presentation in a hospital emergency department, and the aim was to find a rule that was useful in a clinical setting. An alternative approach to determine the risk of (subsequent) self-harm is to examine the predictive utility of a pre-existing instrument, and an example of this is provided by McMillan et al., (116) where a meta-analysis was carried out to determine the predictive utility of the Beck Hopelessness Scale (BHS). (117) Data were pooled from a range of populations, although some history of self-harm was usually present. Again, there was a degree of support for identifying cases of subsequent self-harm (sensitivity), and it was acknowledged that an instrument or tool with a high sensitivity could be used as a preliminary guide to help those in need of a more in-depth assessment. (116) However, the issue with the BHS, along with the Manchester and ReACT Self-Harm Rules, is that a high sensitivity value comes at the expense of a low specificity value, thus resulting in the identification of a high degree of false positives and meaning that the guidance of where to concentrate resources is extremely limited. (116)

2.2.4 Self-Harm in the General Population

Prevalence and incidence estimates are likely to be affected by the different definitions, classifications and terminology used when quantifying self-harm, along with what is judged to be a meaningful history of self-harm. This will vary depending on whether the active definition of self-harm includes just suicide attempts, all suicidal and non-suicidal self-harm behaviours, or just non-suicidal self-harm. (112) Depending on these classifications, self-harm behaviours may range from chewing your lip or lightly biting the inside of your mouth, right through to a genuine suicide attempt. This is difficult to quantify, and to directly compare estimates would also require the definitions of self-harm to be explicitly stated and to remain consistent between studies. Other factors that will influence the prevalence and incidence estimates include: the method of data collection (anonymised, self-report, interviewer administered); the population studied; the age range of the population studied; the survey type (cross-sectional, retrospective or prospective); and the time frame covered (e.g. lifetime, within the last year, within the last two weeks). (112) The type of questions used to gain the information will also have an impact as open-ended questions will tend to under-report the incidence of self-harm, whereas specific, directed (leading) questions will tend to over-report the incidence of self-harm. (112)

Also, until fairly recently, researchers have almost totally ignored studying non-hospital-treated self-harm. As a result of this, very little is known about the incidence and prevalence of self-harm that occurs in the community, (28) although the true incidence of self-harm is probably impossible to determine because it is so often a private and unreported activity. (31)

With all of these factors in mind, the prevalence of reported self-harm is highly variable. Jacobson & Gould (118) reviewed eight studies; two involved adults and six adolescents (broadly defined as “mainly high school students”), and they reported varying 12-month prevalence rates of 2.5% to 12.5% and lifetime prevalence rates of 13.0% to 23.2%. Muehlenkamp & Gutierrez (26) report that estimates of self-injurious behaviour among adolescents range from 5.1% to over 40%, and Skegg (29) states that 5-9% of adolescents in western countries report having self-harmed within the previous year, with lifetime prevalence ranging from 13-30%. It has also been reported (14) that self-harm occurs in 4% of the general population (119) and 14% of college students. (120) Furthermore, Gratz (25) reported that 35% of college students have carried out at least one self-harm behaviour in their lifetime.

Additionally, large surveys suggest that 4.6% of the population in the USA and 4.4% in the UK have previously self-harmed. (105) These results are similar to those of Meltzer et al., (108) who reported that a National Interview Survey showed that 14.9% of respondents had contemplated suicide at some point in their life, and that 4.4% of respondents had attempted suicide at some point in their life. Two per cent of all respondents stated that they had deliberately harmed themselves without suicidal intent. This was a large, national (UK) study involving a representative sample (n=8,450), and should, therefore, provide a fair representation of the adult population (age 16-74). It should be noted, however, that these results are based on a singular self-harm question; therefore an element of subjective judgment may be present, along with the recall bias limitations of retrospective studies.

The best current UK estimate of hospital attendance due to self-harm is 400 per 100,000 hospital attendances (0.4% of all hospital attendance). (121) The current incidence of self-harm is also estimated as between 300 and 600 cases per 100,000 per year. (122, 123) Despite difficulties in diagnostic classification, self-harm is one of the commonest reasons for admission to a medical ward, with around 200,000 hospital attendances per year in the UK, with the majority of these cases (80%) involving self-poisoning. (105) However, it is widely recognised that prevalence rates of self-harm behaviour in the general population are difficult to estimate given that the self-harm may go unreported and will not result in a hospital attendance. (71, 86, 105) Among the general population (who do not routinely present at A&E), physical self-harm is more common, with cutting being the most common form. (118)

Although all of these reported estimates are variable, they are also all likely to be limited to some extent by one or more of the factors outlined at the start of the section. This difference in methodological factors in the reporting of self-harm prevalence has been investigated, where it was shown that the methodological factors accounted for 51.6% of the heterogeneity in the prevalence estimates. (124) When looking specifically at non-suicidal self-injury in non-clinical samples, a meta-analysis showed a pooled lifetime prevalence of 17.2% among adolescents, 13.4% among young adults and 5.5% among adults. (124)

When considering the outcomes of repeat self-harm and suicide that are of particular importance following self-harm, (105, 106) the 1-year repetition of self-harm among hospital presentations is 16%, (71) and this repetition rate would be closer to 33% if episodes rather than individuals were taken as the unit of analysis. (110) The repetition of self-harm events tends to occur quickly, with a quarter of repetitions occurring within 3 weeks, with the median time to repetition being 12 weeks. (71) Additionally, follow-up studies show rates of suicide to be 1.8% in the year after a self-harm episode, 3% at 4 years, and 6.7% for periods longer than 9 years. (71)

Self-harm is often believed to be more common among females, but the self-harm prevalence between genders has been shown to be equivalent. (16, 29, 31, 124) While self-harm can be found across the entire population, it is more common among those who are socio-economically disadvantaged and who have limited social support. (108, 112) Self-harm is especially prevalent in psychiatric populations, (54) with those having mental disorders being 20 times more likely to report having harmed themselves in the past. (108) Among respondents who had reported a lifetime prevalence of self-harm, 57% were categorised as having a neurotic disorder, 6% as having a psychotic disorder, 24% as alcohol dependent and 16% as drug dependent. (108)

Additionally, there are also other specific groups among whom self-harm behaviour appears to be more common. These specific high-prevalence populations include people that are homeless, people who associate strongly with the Goth subculture, homosexual and bisexual individuals, and those who are in prison. (59, 80, 112) There are most likely different reasons for the higher occurrence of self-harm within each of these populations, but given the increased prevalence of self-harm in those from socio-economically disadvantaged areas, and in those with mental health problems, it is not surprising that self-harm presents a significant problem within prisons due to the overlap of these factors within the prison population. (125-128)

2.3 Self-Harm in Prisons

As mentioned in Section 2.2.1.4, the definition of self-harm used within prison custody is "any act where a prisoner deliberately harms themselves irrespective of the method, intent or severity of any injury".(81) It is noted that those who do self-harm often do so covertly; therefore such self-harm will often go undetected in a community setting. In prisons, such incidents are more likely to be detected and counted although there will still be some incidents that are overlooked. (129)

Within offender populations, certain groups are recognised to be at greater risk of self-harm, including those who are psychiatrically ill, those with long-sentences and 'poor copers', who are defined as acutely vulnerable prisoners whose major problems are unrelated to psychiatric illness or the nature of their offence. (130) 'Poor copers' tend to be young offenders (under 26) who have committed acquisitive crimes and have a poor coping ability to being in prison. (131) Even controlling for the characteristics of a prison sample, rates of self-harm in prisons seem to be much higher than the general population. (2)

2.3.1 Prevalence of Self-Harm in Prisons

There are differing estimates of self-harm incidence and prevalence within offender populations or corrective institutions. Again, these differing estimates are possibly due to different definitions of a 'self-harm' event. Appelbaum et al. (57) identified that published research has estimated that 30% of prisoners engage in self-harming behaviour. (132) It has also been reported that 41% of female inmates and 28% of male inmates engage in suicide-related behaviour (non-fatal self-harm), (133) and that 50% of female prisoners have a history of self-harm. (134) The proportion of prisoners engaging in self-harm in American prison systems during 2008 varied from 0.03% to 8.93% across prison systems, with an overall rate of 0.71%. (57) This is markedly different from the results among Greek male prisoners, where self-harm behaviour was reported among 49.4%. (135) Potential reasons for this discrepancy may be the differing classifications of self-harm, differences in the samples (cultural, diagnostic, offender demographic etc.), and the mode of data collection. It may be worth noting that the Greek data (135) were derived from face to face prisoner interviews, whereas the American prison system data (57) were derived from recorded events within prison institutions.

Given this discrepancy in reported prevalence rates, it is important to note how self-harm data is gathered. In the UK, the most complete data is likely to come directly from the offender management statistics. (136) These statistics are regularly published and are therefore likely to be the most up-to-date estimates that are available. Although there may be some deviation between individual institutions, these statistics relate to actual recorded self-harm events, so the classification of a 'self-harm event' is likely to be broadly consistent across all institutions. However, it should be noted that unreported and untreated self-harm events will not be accounted for.

The number of incidents of self-harm in UK prisons rose rapidly between 2003 and 2005. By 2005 there were 23,781 incidents of self-harm in UK prisons, rising from 16,393 incidents in 2003. This rise of 45% was over eleven times the rise in the overall UK prison population for the same period, which was just over 4%. Since 2005, the incidence of self-harm in prison seems to have largely stabilised (see Figure 2). This stabilisation could possibly be due to the prison response to the previously observed rise.

According to the Ministry of Justice, (129) in 2014 there were 25,775 incidents of self-harm reported, with roughly three-quarters of these attributed to the male inmate population. These self-harm events were carried out by 7,722 individuals, with 86% of these being males.

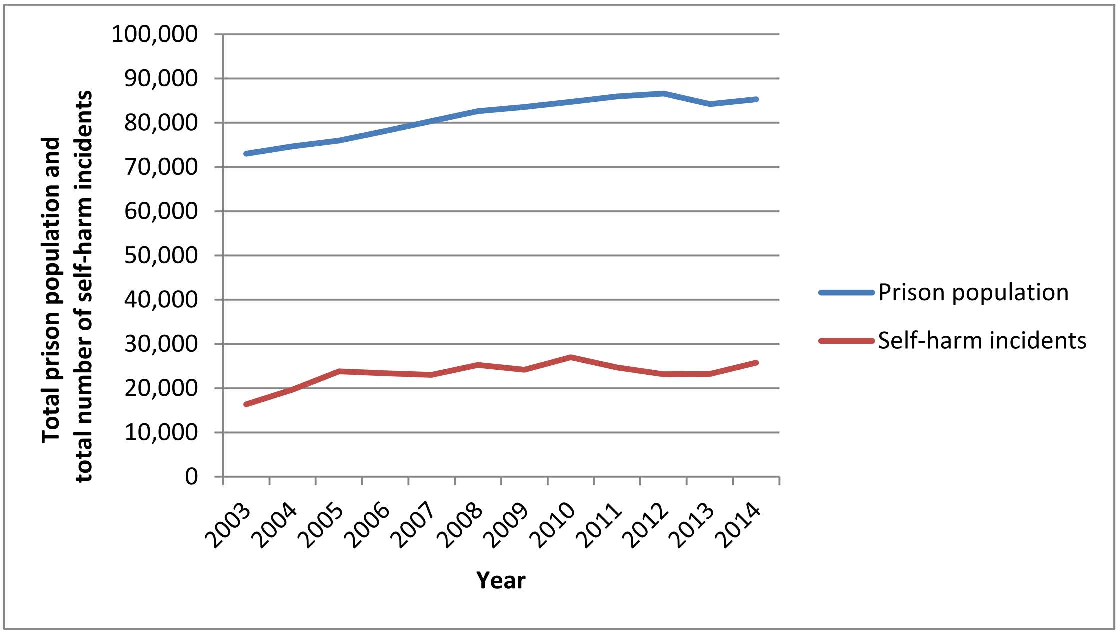

Figure 2. Changes in the overall UK prison population and the number of reported self-harm incidents 2003-2014.

An overall incidence rate cannot be accurately calculated due to the transient nature of prisoners within the system and the lack of available statistics regarding the turnover of prisoners. However, using the number of individuals self-harming, along with the average number of prisoners within the prison system in any given year, an overall approximate yearly incidence of self-harm within prisons can be calculated. This rate has shown a general trend of overall growth from 7.02% in 2004 to 9.05% in 2014, which is a relative increase of 29% over this time period.

For males, this rate has shown the same general trend of overall growth as the total rates, steadily increasing from 5.7% in 2004 to 8.13% in 2014, which is a relative increase of 43% over this time period. As males make up the vast majority of the prison population, this similarity in trend with the total prison system is not surprising.

For females, the rate showed a general growth from 27.9% in 2004, to a peak rate of 37.7% in 2009 (a 35% increase). Since 2009, this rate steadily decreased back to around the 2004 rates, with a low of 26.9% in 2013, and a rate of 28.3% in 2014 (See Figure 3).

Although prison turnover has not been taken into account, these values are approximately twice those reported in the Corston report (137), where it was stated that 16% of women self-harm in prison compared to 3% of men.

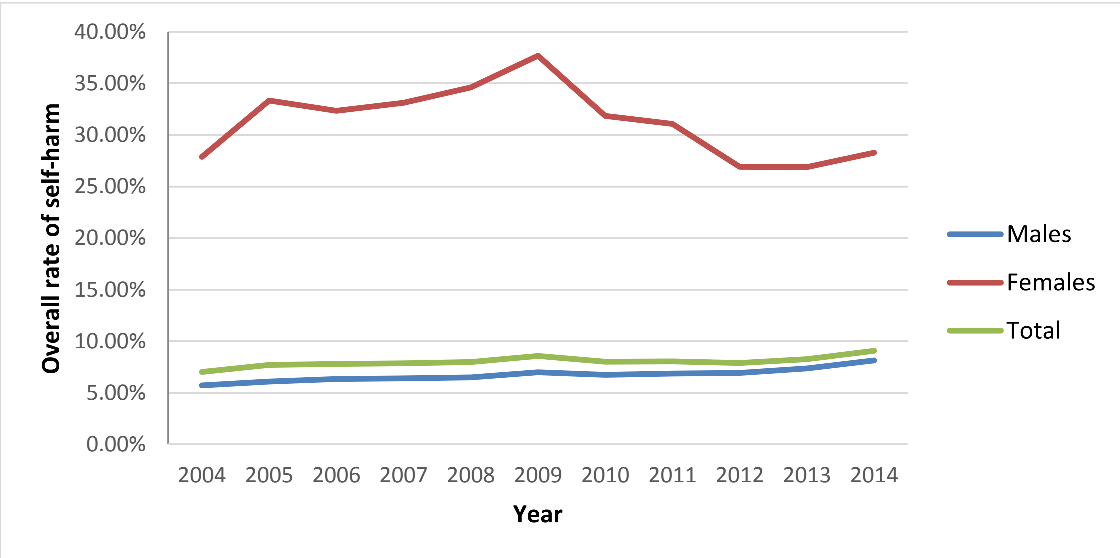

Figure 3. Overall rate of self-harm in prisons from 2004 to 2014.

In terms of self-harm incidents, although the total amount of self-harm incidents has remained reasonably stable around 25,000 per year, it should be noted that there has been a shift in trend for the amount of incidents attributable to males and females.

For males, the rate has risen from 141 self-harm incidents per 1000 prisoners in 2004, to 233 self-harm incidents per 1000 prisoners in 2014, marking a 65% increase over the time period (see Figure 4).

For females, since 2010 the rate has dramatically dropped from 2,982 self-harm incidents per 1000 prisoners, to 1,546 self-harm incidents per 1000 prisoners in 2013, marking a 48% decrease over the time period. In 2014 the rate increased again slightly to 1,736 self-harm incidents per 1000 prisoners (see Figure 5).

Taking the 2014 values, among the individuals that self-harm, males report an average of 2.9 self-harm incidents per individual and females report an average of 6.1 self-harm incidents per individual.

It should be noted that over the entire period of 2004-2014, although the proportion of total self-harm incidents that require hospital attendance has risen slightly from 5.5% in 2004 to 6.8% in 2014 (a 24% increase), this proportion has remained fairly stable within each of the gender groups. For males, around 9% of incidents require hospital attendance, and for females around 2% of incidents require hospital attendance (see Table 2).

In addition to the self-harm incidents that have occurred within prison, there is also around 1 self-inflicted death per 1,000 prisoners each year (aggregated rate across all prisoners, 2004-2014 = 0.88 self-inflicted deaths per 1,000 prisoners). (136)

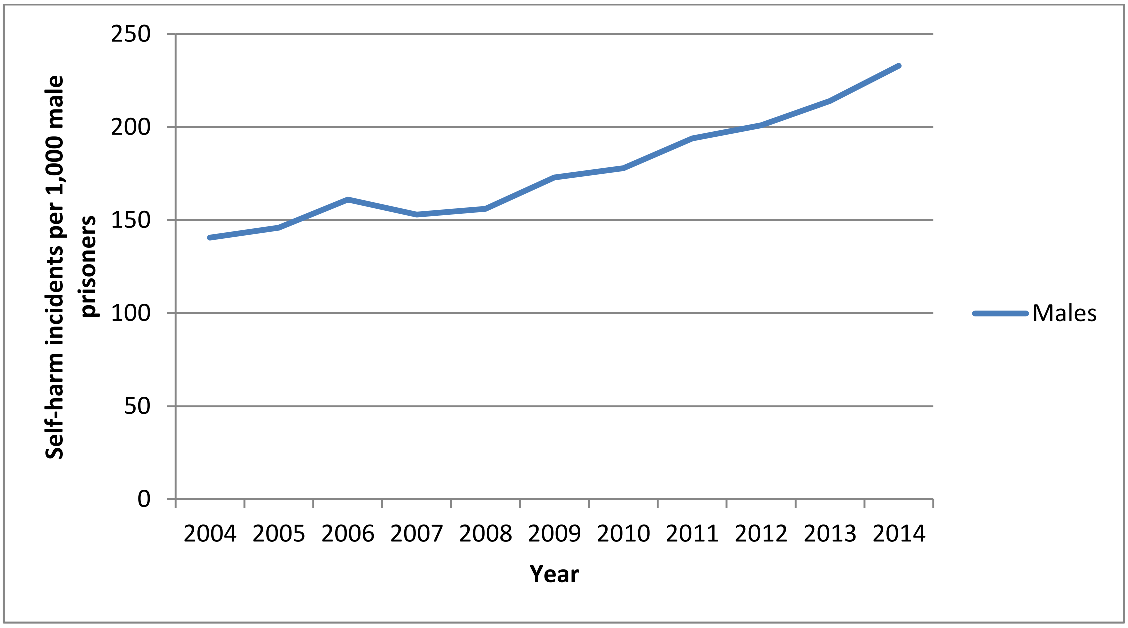

Figure 4. Number of self-harm incidents per 1,000 male prisoners by year, 2004-2014.

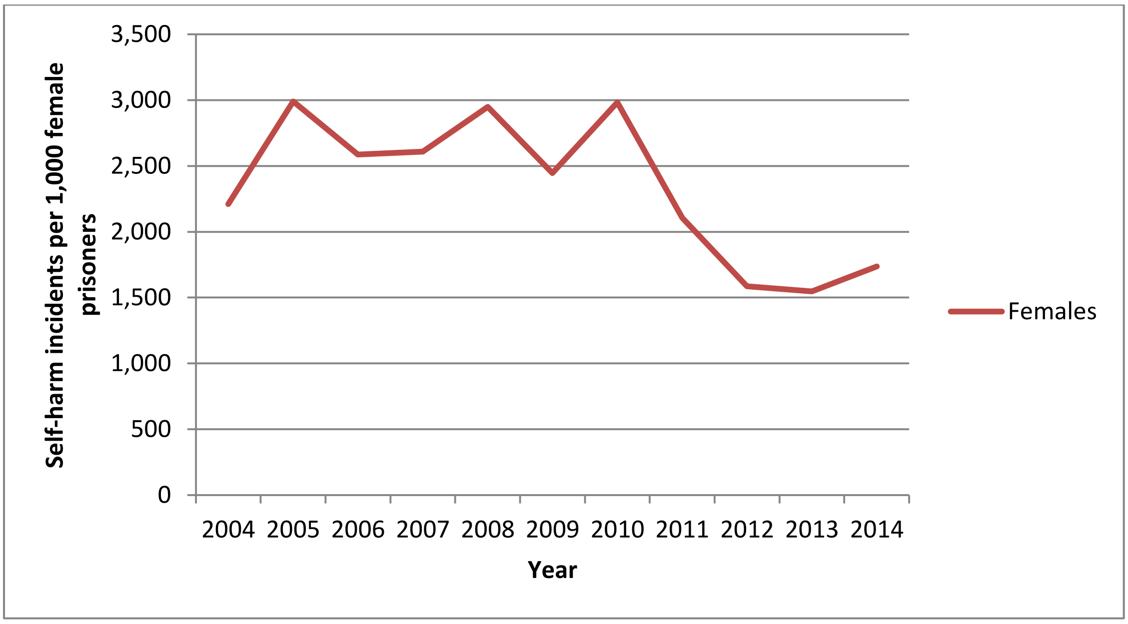

Figure 5. Number of self-harm incidents per 1,000 female prisoners by year, 2004-2014.

Table 2. Self-harm incident information by year and sex for all prisoners in England and Wales, 2004-2014.

	Year										
	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Total Prison Population	**74,658**	**75,980**	**78,151**	**80,380**	**82,636**	**83,559**	**84,725**	**85,951**	**86,634**	**84,249**	**85,307**
Male	70,209	71,513	73,703	76,006	78,222	79,277	80,489	81,763	82,481	80,359	81,402
Female	4,449	4,467	4,448	4,374	4,414	4,283	4,236	4,188	4,153	3,890	3,905
Self-harm incidents	**19,702**	**23,781**	**23,400**	**23,000**	**25,234**	**24,184**	**26,979**	**24,648**	**23,158**	**23,230**	**25,775**
Males	9,874	10,420	11,899	11,592	12,219	13,706	14,347	15,829	16,567	17,216	18,995
Females	9,828	13,361	11,501	11,408	13,015	10,478	12,632	8,819	6,591	6,014	6,780
Individuals self-harming	**5,243**	**5,837**	**6,090**	**6,296**	**6,586**	**7,149**	**6,767**	**6,907**	**6,821**	**6,942**	**7,722**
Males	4,003	4,348	4,652	4,847	5,058	5,535	5,418	5,606	5,703	5,897	6,618
Females	1,240	1,489	1,438	1,449	1,528	1,614	1,349	1,301	1,118	1,045	1,104
Self-harm incidents per 1,000 prisoners	**264**	**313**	**299**	**286**	**305**	**289**	**318**	**287**	**267**	**276**	**302**
Males	141	146	161	153	156	173	178	194	201	214	233
Females	2,209	2,991	2,586	2,608	2,949	2,447	2,982	2,106	1,587	1,546	1,736
Individuals self-harming per 1,000 prisoners	**70**	**77**	**78**	**78**	**80**	**86**	**80**	**80**	**79**	**82**	**91**
Males	57	61	63	64	65	70	67	69	69	73	81
Females	279	333	323	331	346	377	318	311	269	269	283
Self-harm incidents per individual	**3.8**	**4.1**	**3.8**	**3.7**	**3.8**	**3.4**	**4.0**	**3.6**	**3.4**	**3.3**	**3.3**
Males	2.5	2.4	2.6	2.4	2.4	2.5	2.6	2.8	2.9	2.9	2.9
Females	7.9	9.0	8.0	7.9	8.5	6.5	9.4	6.8	5.9	5.8	6.1
Self-harm related hospital attendances	**1,093**	**1,219**	**1,214**	**1,290**	**1,290**	**1,304**	**1,369**	**1,533**	**1,547**	**1,603**	**1,749**
Males	873	985	1,001	1,104	1,083	1,131	1,193	1,375	1,391	1,484	1,617
Females	220	234	213	186	207	173	176	158	156	119	132
Percentage of self-harm incidents requiring hospital attendance	**5.5%**	**5.1%**	**5.2%**	**5.6%**	**5.1%**	**5.4%**	**5.1%**	**6.2%**	**6.7%**	**6.9%**	**6.8%**
Males	8.8%	9.5%	8.4%	9.5%	8.9%	8.3%	8.3%	8.7%	8.4%	8.6%	8.5%
Females	2.2%	1.8%	1.9%	1.6%	1.6%	1.7%	1.4%	1.8%	2.4%	2.0%	1.9%

Adapted from (136).

Cutting or scratching is the most common form of self-harm within prisons, for both males and females, accounting for 66% of the male incidents and 50% of the female incidents. The rates of all other recorded types of self-harm are much lower across both genders, with the exception of self-strangulation which accounts for 28% of the female incidents (see Figure 6).

Figure 6. Self-harm by type, aggregated across years 2004-2014.

The age distribution of the individuals self-harming is fairly evenly spread, with the largest majority (24% for male and female groups) being in the 30-39 age bracket. The distribution is almost exactly the same across both genders (see Figure 7).

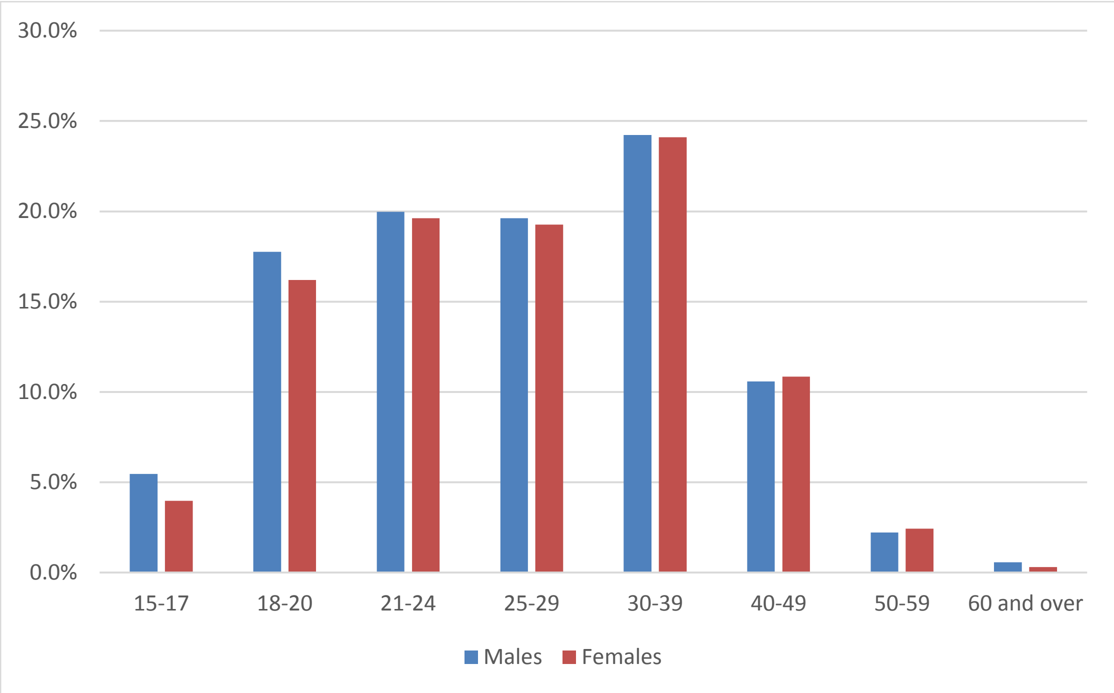

Figure 7. Age distribution by sex of individuals self-harming, aggregated across years 2004-2014.

The majority of self-harm incidents take place among prisoners who have already been sentenced, for both males (77% of incidents) and females (66% of incidents). This slight difference across genders means that females (21%) have a higher proportion of incidents occurring in remand than males (13%) (see Figure 8).

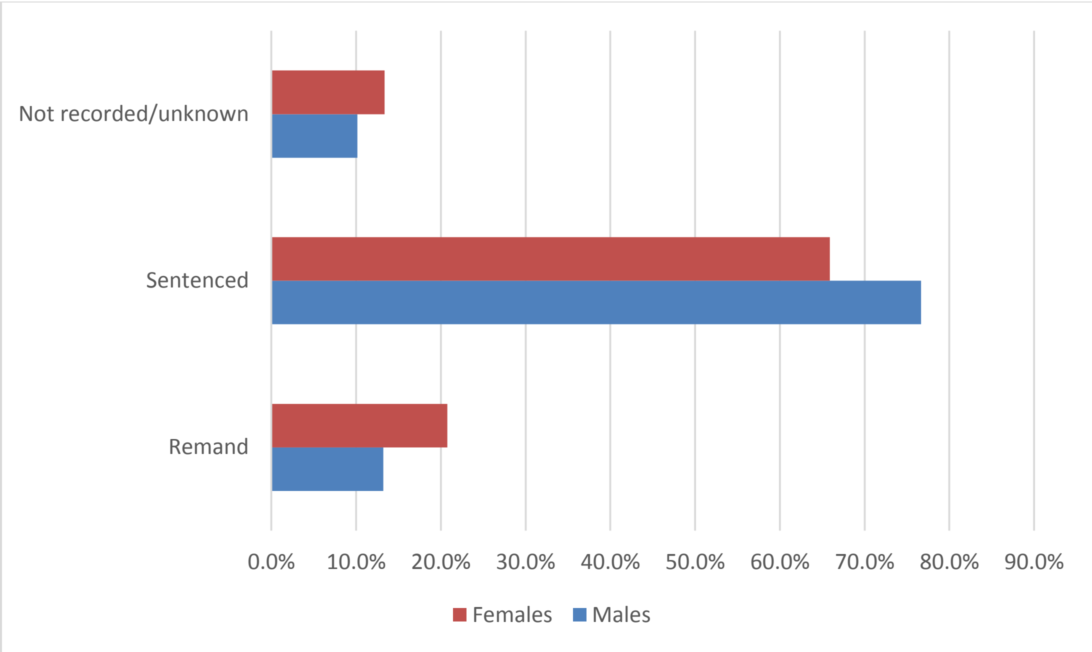

Figure 8. Prisoner sentencing status of individuals self-harming, aggregated across years 2004-2014.

It should be noted that the overall proportion of incidents occurring by prisoners on remand was fairly stable at around 30% until 2010. Since then, this proportion has halved, meaning that 15% of incidents were carried out by prisoners on remand in 2014. However, this drop in the proportion of remand incidents results in an increase of the same magnitude for the proportion of incidents occurring among sentenced prisoners (see Figure 9).

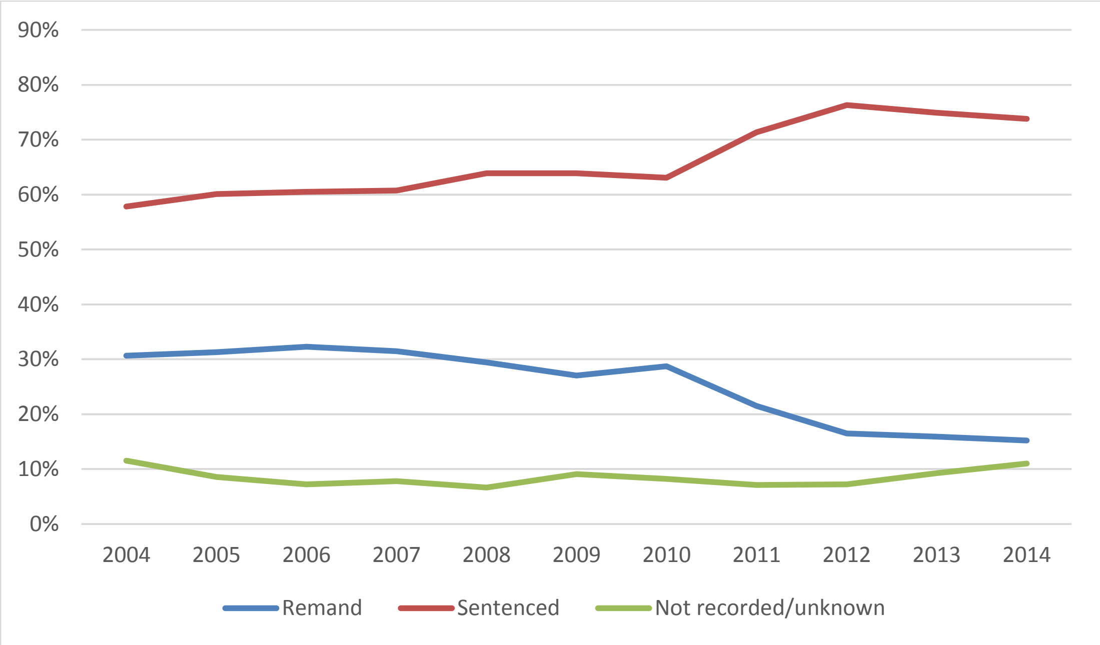

Figure 9. Proportion of self-harm incidents by prisoner sentencing status, by year, 2004-2014.

In terms of when self-harm incidents occur, 2% take place on the day of arrival into prison, with a further 3% occurring within the first or second full day in prison. This cumulatively contributes to 11% of incidents occurring within an individual's first week in prison. 28% of incidents occur within an individual's first month of being in prison, with a further 24% of incidents occurring between one and three months into a prison stay (see Figure 10).

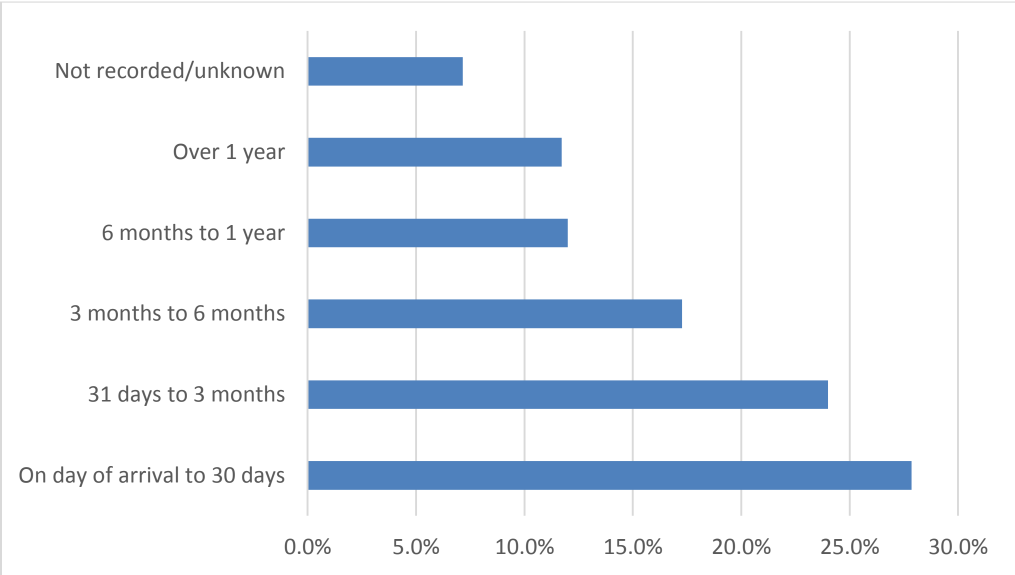

Figure 10. Proportion of self-harm incidents by time period within prison stay, aggregated across years 2004-2014.

The average number of self-harm incidents per self-harming individual can be seen in Table 2. Overall, this value has ranged from 3.3 incidents (in 2013 and 2014) to 4.1 incidents (in 2005) per self-harming individual. The female average is more than double that of males, with the male average between 2004 and 2014 ranging from 2.4 incidents to 2.9 incidents, and the female average ranging from 5.8 incidents to 9.0 incidents per self-harming individual. Despite this discrepancy, it can be seen that the majority of self-harming individuals carry out less than 3 incidents, with 62% of males and 45% of females only carrying out a single self-harming incident (see Figure 11). However, 1% of male and 7.5% of female self-harming individuals are responsible for more than 20 repeat incidents of self-harm, which will skew the overall average values.

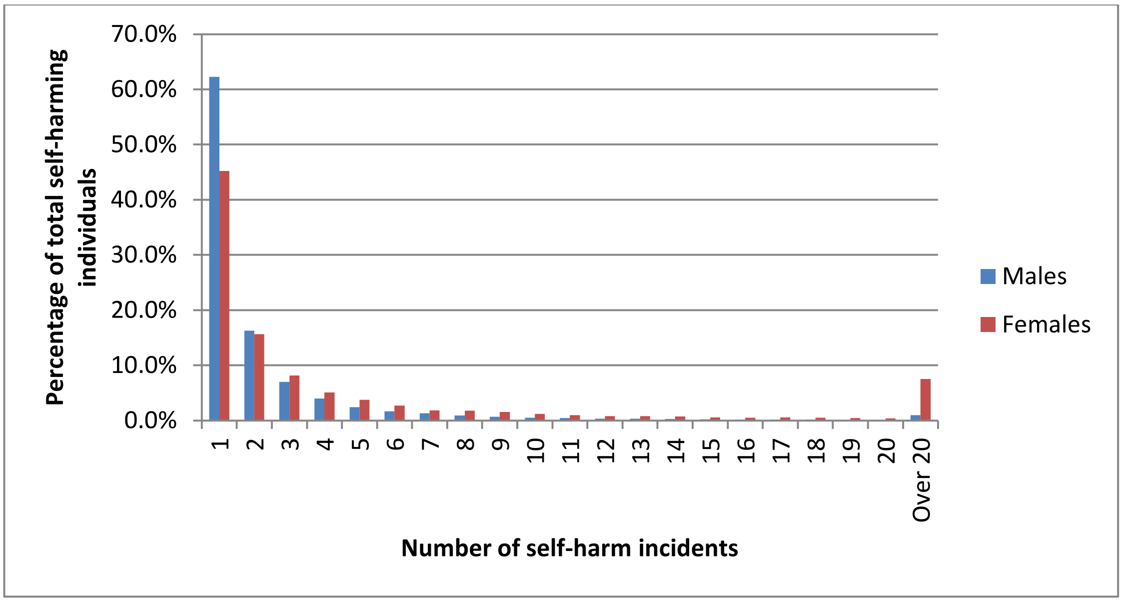

Figure 11. Number of self-harm incidents carried out by each self-harming individual, aggregated across years 2004-2014.

2.3.2 Operational Self-Harm Practices in Prison

2.3.2.1 Assessment, Care in Custody, and Teamwork (ACCT)

Self-harm can present a major challenge and place considerable demand upon prison health care systems, (80) the responsibility for which resides with Primary Care Trusts (PCTs). In 2005 the Prison Service piloted a care-planning system called ACCT (Assessment, Care in Custody, and Teamwork) (4) to improve the care for prisoners at risk of suicide or self-harm, and this was implemented nationally in 2007. Full details are provided elsewhere, (138) but briefly, when an ACCT is active it is referred to as 'open', and any member of prison staff can open an ACCT at any point, thereby identifying a prisoner as being 'at risk'. The ACCT document specifies who opened the ACCT, the date that it was opened and the reason(s) why the prisoner is deemed to be at risk of self-harm or suicide. The ACCT document is designed to ensure prison staff keep a concise record of the prisoner's care, needs and problems. When an ACCT is 'open', prison staff are required to make regular (e.g. random, every 15 min, every 30 min) observations of the prisoner and to provide comments on their subjective and objective mood and behaviour, along with documenting evidence of engagement or communication with the prisoner. The ACCT remains open until the prisoner is no longer perceived to be 'at risk', at which point the ACCT is 'closed', which brings an end to the formal monitoring process. (138)

An important point to mention is that a prisoner only needs to be considered as 'at risk' for an ACCT to be opened, and the reasons for this are variable. Although an ACCT would be opened if a prisoner carried out a self-harm incident, many ACCTs are opened without any incidence of self-harm. The initial ACCT assessment effectively establishes a care pathway system (CAREMAP) for those deemed to be at risk. However, it does not incorporate a standardised diagnostic test to estimate the risk of future self-harm.

2.3.2.2 Reception Screening

Although only 11% of (non-fatal) self-harm incidents occur within the first seven days of reception into prison (see Section 2.3.1), it has been recognised that prisoners are at a heightened risk of suicide upon reception into prison, with a third of all prison suicides taking place in the first seven days. (3)

Due to the increased vulnerability of prisoners during the reception period, all new prisoners are screened upon reception into the prison, using a standardised prison questionnaire which was designed to screen for physical and mental health problems. (139) Although this screening tool is not intended to predict the risk of self-harm or suicide, it does allow for the broad identification of high-risk problems such as self-harm or suicide risk, which may warrant further assessment. If a risk of self-harm or suicide is deemed to be present, this would also trigger the opening of an ACCT document. (138) There is some evidence to suggest that this screening tool can help to identify true cases of psychiatric illness upon entry into prison. (139) This early indication of mental and physical health problems is beneficial to prison staff in terms of prisoner management, but the key issue remains as to whether individuals specifically at risk of self-harm or suicide can be identified at reception into prison. (140) Early recognition of this risk could lead to increased staff awareness and the initiation of appropriate preventative measures being put in place; therefore potentially lowering the rate of self-harm and reducing the demand on the prison healthcare system. (58)

2.3.3 Screening for Self-Harm within Prisons

One way to approach the development of a screening process would be to assess the associated risk factors for self-harm. As seen in Section 2.2.3, there are many risk factors associated with self-harm, and it is necessary for these risk factors to be statistically obtained as clinical intuition is a notoriously error-prone practice of risk assessment. (58, 61) These studies on risk factors are indispensable to broaden our knowledge of self-harm, (58) and they have been used to generate self-harm screening algorithms specifically for prison populations, (80, 141) although these have not been tested prospectively. Also, with regard to the majority of the risk factors that have been identified specifically to self-harm in prisons, a major problem is that there is also conflicting evidence to disregard these same risk factors. (58) This is possibly because a lot of the factors that have been identified as associated with self-harm are non-specific, and are therefore of limited value. (140) When considering the potential of risk factors for suicide, although the prevalence of suicide in prisons is greater than in the general population, it still remains a relatively rare event, meaning that a lot of associated suicide risk factors have a low predictive power. (5, 140)

The evidence to support the routine use of any screening instrument for self-harm in offender populations is limited, and the transferability of any existing self-harm screening instruments is problematic due to the unique environment in which prisoners are accommodated. (5) A review article identified four screening instruments across five studies that have been used to assess for the risk of suicide and self-harm in adult offenders, although three of these instruments were specifically aimed at screening for suicide (or suicide risk) rather than self-harm (or risk of self-harm), and two of the studies used retrospective methodology which may result in non-comparable information between study participants. Additional limited evidence suggests that the Beck Depression Inventory (142) may be predictive of self-harm behaviour among female prisoners, (143) and that the Beck Hopelessness Scale (117) may predictive of self-harm among offenders with mental disorders, (144), but not among female prisoners. (143) A newer scale, Suicide Concerns for Offenders in Prison Environment (SCOPE) (145) has been specifically developed to assess vulnerability to risk of suicide and nonfatal self-harm behaviour in young adult offenders. However, again, this has not been tested with regard to implementation for routine prison use, and although it does demonstrate some evidence for its prospective predictive validity, this was only demonstrated in a female cohort. (143) The limited evidence for the use of screening instruments for self-harm in prisons led Perry et al. to conclude that: "There is a clear need for additional psychometric research on the validity of suicide and self-harm behaviour screening tools in offender populations." (5)

2.4 Research Aims and Hypotheses

In summary, the term 'self-harm' refers to a number of physical behaviours that result in the intentional direct harming of oneself, although the underlying motivations are variable. Self-harm occurs within the general population, but is a lot more prevalent among those with mental health conditions. As the majority of prisoners have mental health conditions and the prison environment causes additional stresses, the result is a very high incidence of self-harm within prisons.

The ACCT process was introduced into prisons to address and manage any prisoners identified as 'at risk' of self-harm, but within the process there is not currently any standardised instrument which is used for assessing the risk of self-harm. It would therefore be helpful to prison staff to introduce a standardised measure during the ACCT assessment, so that limited resources can be focussed to the most appropriate areas.

In response to the perceived need for screening instruments to identify the risk of self-harm among prisoners, a multi-stage prospective study was undertaken to identify potential instruments and determine their predictive validity. The stages included a scoping exercise to identify candidate instruments; a pilot study to test the feasibility of a protocol to implement these instruments in a prison setting; a prospective cohort study to apply the instruments and identify subsequent self-harm over a specified follow-up period, and various psychometric and multivariate analyses to determine the best (if any) predictive instrument, or set of items taken from the instruments.

The working hypothesis of this study is therefore that:

Self-harm within the prison ACCT population can be predicted using a pre-existing screening instrument.

Additionally, the aims of the study are as follows:

- Identify a range of pre-existing screening instruments, capable of potentially predicting self-harm within a prison setting
- Pilot test these instruments to assess their suitability within a prison setting, along with assessing the plausibility of carrying out a research project within a prison setting.
- Carry out a full research project using a refined set of instruments. The study sample will be followed up after a certain amount of time to obtain information as to whether self-harm was carried out during the follow-up period.
- Psychometrically assess and validate all selected instruments within the specific ACCT sample of the study.
- Assess the predictive validity of each of the selected instruments to determine whether they are capable of predicting self-harm.
- If a stand-alone instrument is not predictive, then determine whether any set of demographic or individual instrument items combine to predict self-harm.
- If the instruments are proved to be valid among the ACCT population, then determine whether a model of the influencing factors can be derived to explain the mechanism of self-harm.

3 Pilot Study

This chapter presents the methodological basis of the study, leading to the implementation of the pilot study, which was carried out to ensure that the logistics of the study could be delivered practically. There are a lot of individual risk factors for self-harm, and therefore there are also a large amount of standardised psychometric instruments that may be potentially predictive of self-harm. This chapter presents the process involved with identifying and refining the selection of instruments that would eventually be used in the cohort study. The results of the scoping exercise and pilot study are then presented, along with the implications of these results for the cohort study.

3.1 Assessment and Screening of Self-Harm

Assessment of self-harm can be done via a clinical or psychological professional assessment, via a structured or semi-structured interview or via a series of discrete items that make up a self-report questionnaire. These forms of assessment are usually carried out on a 'lifetime incidence' basis, or following an acute self-harm incident, but they may relate to a more recent, fixed time period (e.g. previous 6 months). (112) The in-depth psychological assessment and structured interviews can reveal a number of elements relating to the self-harm incident, including the type, location and extent of the self-harm, the underlying motivations and personal characteristics that have led to the self-harm event, and whether the individual is likely to self-harm again. The closed-item questionnaires can also extract this information, although these questionnaires vary in length, scope and quality, (146) and often tend to focus on a particular element of the self-harm; for example, the different types of self-harm behaviour that a person has engaged in, (147) or the amount of impulsivity that is present within a person. (148)

Although structured and semi-structured interviews can reveal a lot of information, they are often lengthy and require a lot of resource to carry out. As these assessments are more personal, they are also non-standardised, meaning that they are harder to compare across cases. Another issue is that most of these interviews and questionnaires are designed to be administered specifically after a self-harm event has already taken place, which is not useful when considering a screening process. There are very few measures that adequately screen for suicidal and non-suicidal self-harm, and those that do attempt it often focus specifically on 'suicide attempt'. (112)

3.2 Scoping Methods

There are many questionnaires available to assess and/or screen for self-harm, some of which relate specifically to self–harm behaviours (e.g. The Self-harm Inventory (39)), and some of which relate to other underlying correlates of self-harm such as depression (e.g. The Patient Health Questionnaire (149)). The first stage of the project involved a scoping exercise to systematically identify available instruments that could be used to screen for self-harm. A search was carried out with the SCOPUS database (encompassing MEDLINE, PSYCINFO, CINAHL & EMBASE), using appropriate search terms such as 'Self-harm', 'Self injury', 'Suicide ideation', 'Prison', 'Jail', 'Risk', 'Questionnaire' and 'Screen'. All journal article titles and abstracts were read for any mention of self-harm measurement, scales or instruments. This was followed up with a search of the grey literature (e.g. University theses, commissioning reports etc.) and a related internet search.

Once identified, a range of practical inclusion criteria had to be fulfilled prior to assessing the psychometric properties of the applicable scales according to a standardised protocol.

The practical inclusion criteria included the following:

- The instrument must be able to be administered by generic primary care, prison and research staff that may not have had mental health or clinical training.
- The instrument must be able to be administered orally by staff rather than self-administered (due to low literacy levels among prisoners).
- The instrument must be able to be administered without specialist training specific to the instrument, in line with the circumstances in which it would be administered upon ACCT assessment or prison reception. This is also a practical point with regard to the implementation of the research project.

- The instrument must NOT be specifically designed for administration AFTER a self-harm event (people at risk may or may not have actually carried out a self-harm incident).
- The instrument must be comprised of closed questions with a discrete response format to allow for objectively measured responses and consistency among respondents. This response format also allows for direct psychometric analysis of individual questions and their corresponding response format.
- The instrument must be brief, in line with the circumstances in which it would be administered in a prison environment. Any instruments containing more than 50 individual questions were excluded as inappropriate.
- The instrument must be available for use within the study.

The psychometric criteria that were assessed included:

- Has the instrument been used to directly screen for self-harm?
- Is the instrument directly related to self-harm (or a self-harm correlate)?
- Has the instrument been validated for an offender population?
- Have the psychometric properties of the instrument been assessed?

Each instrument was rated in terms of its practical application and psychometric properties and then a set of potential instruments was taken forward to an expert panel meeting (consisting of two psychometricians, two prison-based clinician researchers, a forensic psychologist, a psychological medicine and healthcare researcher, and a service user, all with relevant experience), in order to reach a consensus on the instruments to be used in the pilot study.

Within the expert panel discussions, the same practical and psychometric criteria were applied to the instruments, along with any further practical information relating to prison policy or existing implementation processes. All comparative strengths and weaknesses of the instruments were considered. The aim was to select an array of scales from the potential set that might have moderately different focus, therefore maintaining a range of different screening criteria that could be tested. Where unanimous consensus could not be reached, disagreements were resolved by majority vote among panel members.

3.3 Scoping Results

Once duplicates were removed, the initial search yielded 955 unique journal article records. Following the title and abstract screening, along with the grey literature and related internet search, 130 unique potential Self-harm or Suicide screening measurement instruments remained. Following the application of the practical and psychometric inclusion criteria, 13 potential screening instruments remained. The majority of the potential scales were removed due to inappropriate administration constraints (i.e. clinician-rated scales), or due to inappropriate or unspecific scale content (i.e. a scale specifically focused on anger or suicide rather than self-harm, without any self-harm component). Potential scales were also removed if they were only to be administered specifically after a self-harm event had occurred, if they were deemed to be too long, or if no further information could be found on the identified scales.

The initial 13 potential screening instruments are summarised in Table 3.

Table 3. Summary of final set of candidate screening instruments

Name of scale	**Short name**	**Reference**	**Designed to measure**
Beck Depression Inventory	BDI	(142)	Depression
Beck Hopelessness Scale	BHS	(117)	Hopelessness
Borderline Symptom List-23	BSL-23	(150)	Symptoms of BPD
Clinical Outcomes in Routine Evaluation System – Outcome Measure	CORE-OM	(151)	Mental Health disorders
Deliberate Self-Harm Inventory	DSHI	(25)	Self-harm behaviours
Depression, Anxiety & Stress Scale	DASS-21	(152)	Depression, Anxiety, Stress
Functional Assessment of Self-Mutilation	FASM	(64, 153)	Self-harm behaviours
Hospital Anxiety and Depression Scale	HADS	(154)	Anxiety, Depression
Patient Health Questionnaire	PHQ-9	(149)	Depression
Prison Screening Questionnaire	PriSnQuest	(155)	Mental Health disorders
Self-harm Inventory	SHI	(39)	Self-harm behaviours
Suicide Concerns for Offenders in Prison Environment	SCOPE	(145)	Suicide risk factors
The Referral Decision Scale	RDS	(156)	Mental Health disorders

Following the discussions of the expert panel, eight instruments remained. The instruments removed at this stage were the RDS, the FASM, the DSHI, the BDI and the HADS. The reasons for their removal at this stage are as follows:

- The RDS is primarily a screening tool for mental health disorders, which was developed for use within the United States criminal justice system. This was discarded in favour of the PriSnQuest, which was developed to perform a similar role within the UK criminal justice system.

- The HADS and BDI are both measures of depression, which is a correlate of self-harm. These measures were left out in favour of the PHQ-9, which contains similar content but is a shorter scale which is already utilised within UK primary healthcare services.
- The DSHI and the FASM are both measures relating to previous self-harm behaviours. These were left out in favour of the SHI, which covers similar content but has favourable psychometric properties. (157)
- The eight remaining instruments (PriSnQuest, SHI, BSL-23, SCOPE, BHS, CORE-OM, DASS-21, PHQ-9) went forward for use in the pilot study. The results of the scoping exercise are summarised in Figure 12.

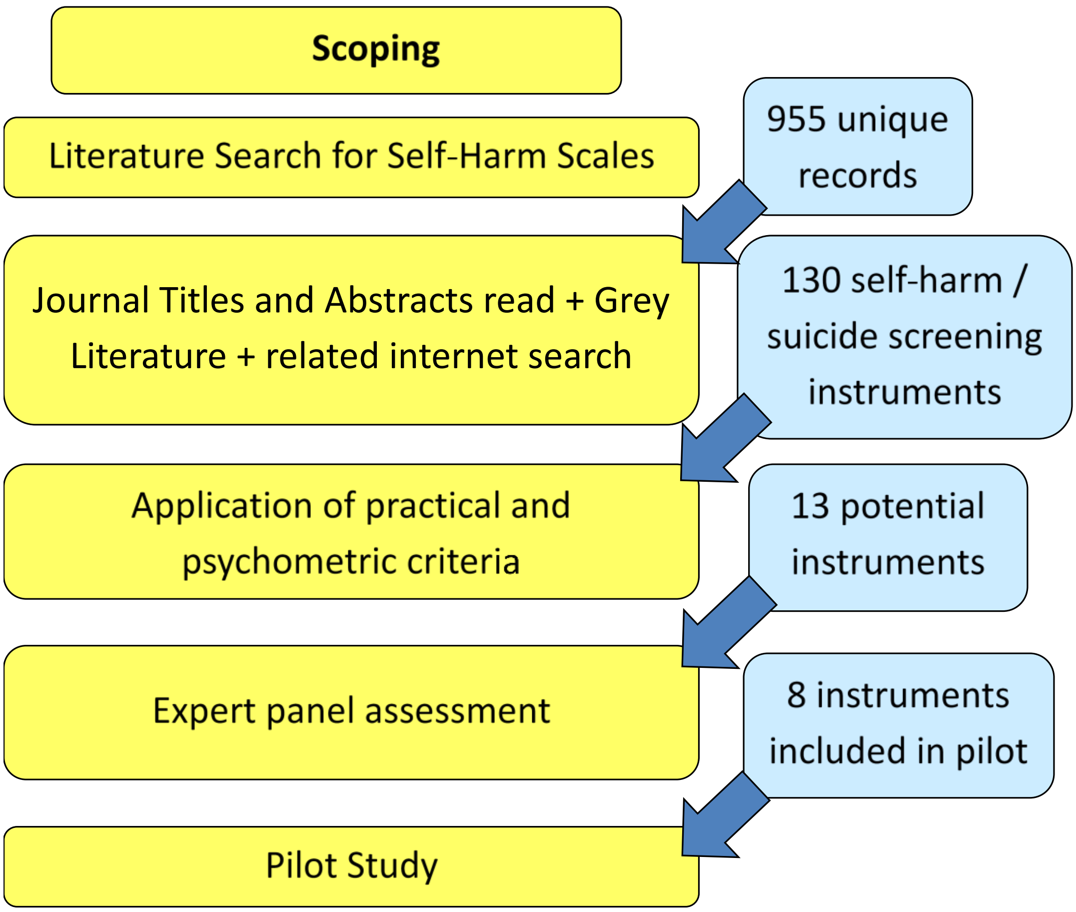

Figure 12. Summary of scoping process, from initial search to final instrument selection

3.4 Pilot Study Methods & Logistics of Process

Following the identification of candidate screening instruments, a pilot study was undertaken in three prisons in Northern England which were collaborating with the Prison and Offender Research in Social Care and Health (PORSCH) network; the male institutions Prisons A & C, along with the female institution (Prison B). The pilot study was undertaken over six weeks to determine several operational aspects of the screening process:

- The operational and safety requirements for introducing a screening procedure, identifying the most appropriate times, locations and implications for staffing (e.g. Prison Officers time for escorting prisoners).
- Evaluating the face validity and acceptability of the chosen screening instruments to prisoners, to assess for problems in their application.
- Discussions with ACCT assessors to see if they foresee, and/or have observed any problems in administration, reliability and validity of the chosen instruments.
- Evaluation of the time taken to administer the questionnaire packs, and to gauge the opinion of the respondents regarding the burden of responding.

Furthermore, the pilot study also served to provide a sample on which to test the follow-up process, which additionally led to the provision of a self-harm incidence estimate for the follow-up period, which can then be utilised for cohort study power calculations. The information gained from the pilot study was to have a direct impact on the final set of instruments selected for inclusion in the cohort study.

To limit the burden of the respondents in the pilot study, a block design with 12 administration patterns was used, meaning that everyone taking part in the study was asked to respond to four of the eight instruments (Table 4). Everyone responded to the DASS-21 and the PHQ-9, along with two of the other six instruments.

Table 4. The block design of the pilot questionnaire packs that were administered

PILOT	Scale								TOTAL
Administration Pattern	CORE-OM	PriSnQuest	BHS	BSL	SHI	SCOPE	PHQ-9	DASS-21	
A	1	1	0	0	0	0	1	1	4
B	0	1	1	0	0	0	1	1	4
C	0	0	1	1	0	0	1	1	4
D	0	0	0	1	1	0	1	1	4
E	0	0	0	0	1	1	1	1	4
F	1	0	0	0	0	1	1	1	4
G	1	0	1	0	0	0	1	1	4
H	0	1	0	1	0	0	1	1	4
I	0	0	1	0	1	0	1	1	4
J	0	0	0	1	0	1	1	1	4
K	1	0	0	0	1	0	1	1	4
L	0	1	0	0	0	1	1	1	4
TOTAL	4	4	4	4	4	4	12	12	

3.4.1 Pilot Study Data Collection

Within the prison system, any incidence of self-harm, or cause for concern that a prisoner may be at risk, triggers the opening of an ACCT Plan. A Unit Manager notifies the Assessor Team and arranges for an Assessor to interview the person at risk within 24 hours. This interview identifies the risk and contributes to the first Case Review. It also presents an opportunity to introduce a standardised diagnostic test, which is currently not present, for the risk of (further) self-harm. Thus, in the three prisons participating in the study, in all cases where an ACCT was opened, the prisoner was approached for inclusion into the pilot study, irrespective of their sentencing status (remand prisoners were also included). If the prisoner consented to inclusion in the study, the pilot questionnaire pack was administered within 72 hours of the opening of the ACCT, where it was safe and sensible to do so. Where it was deemed not safe or inappropriate, the prisoner was excluded from the study. The pilot study recruitment was undertaken over six weeks. All recruitment and data collection was carried out by an experienced on-site prison researcher in two of the prisons, and by members of prison psychology staff in the third prison.

It is acknowledged that this ACCT-based inception cohort were already a pre-selected group considered to be 'at risk' of self-harm. However, given the overall purpose of identifying suitable predictive screening instruments, rather than a prevalence study, together with the practicalities of administering a set of questionnaires within a prison institution, it was deemed unfeasible to screen all prisoners within the scope of this study. It should also be noted that recruitment was based only upon the index ACCT, and subsequent ACCTs by the same individual were discounted, as they were already within the follow-up cohort.

Additionally, when carrying out a project intended to test the predictive properties of potential screening instruments, there is a further potential issue to be considered which may be viewed as a 'risk paradox'. This paradox may present when an individual is identified as being at risk by one (or more) of the instruments that are being assessed. If risk is detected (especially in the case of self-harm risk), then generally something will be done in order to alleviate this risk in the individual. In turn, any element of risk reduction for a given individual may also reduce the probability of the final outcome occurring in the population of interest that was classed as 'high risk' by an instrument. This process may interfere with any attempts to establish the predictive validity of the instruments that are being assessed – the same instrument(s) that identified the risk in the first instance. Although it is acknowledged that this may remain an issue in this study, it is argued that this risk paradox is unlikely to have a major impact on the results, as all study participants are from the prison-ACCT population, and are therefore already classified as being at an increased risk of self-harm.

3.4.2 Pilot Study Follow-up Process

Follow-up was carried out after a period of nine months from the date of questionnaire completion. Follow-up was carried out by checking the prisoner record on the National Offender Management Information System (NOMIS) prison computer record system. The follow-up data that were collected for each study participant included the following:

- Whether the participant had self-harmed during the follow-up period
- The number of self-harm events during the follow-up period
- Dates, descriptions and severity coding of any self-harm events
- The number of ACCTs opened during the follow-up period
- The current prison status and location of the participant, along with corresponding dates of transfer or release
- Whether the index ACCT event was opened due to an actual self-harm event

Each study participant had a valid follow-up time of nine months if they were still within the prison system, or up to the point of their release from their index prison stay. Therefore, the valid follow-up time was variable. If a prisoner had transferred prisons within the follow-up period, all necessary follow-up data were still accessible via the Global Transfer Report on the NOMIS system.

The information available on the NOMIS system was restricted by the quality of the data that were recorded within the database. The NOMIS system contains data that is entered and updated by prison staff, and the information available from an ACCT record or a 'self-harm event alert' is variable, depending on the extent of the information that was entered onto the system.

3.5 Pilot Study Results

There were 75 people recruited to the pilot study; 50 (66.7%) were male, and 22 (29.3%) were female, with three (4%) having missing gender data. The median age was 28 years, ranging from 18 to 62 years (IQR 23-39). Once the data collection routine had been established within each prison, there were no problems reported with the process or logistics of running the pilot study.

3.5.1 Pilot Study Cognitive Debrief

The mean administration time of the questionnaire packs was 37 minutes (SD 11 minutes), but the general consensus from the respondents is that they did not find the interview process burdensome or onerous. Based on participant feedback and the views of the expert panel, a final set of five instruments (from the original eight) were selected for use in the cohort study, and the instruments that were eliminated at this point were the BHS, the SCOPE and the DASS. The BHS was removed as the prisoner respondents found some of the questions confusing. It was also thought that a lot of the questions could be taken out of context when applied within a prison setting. The SCOPE was removed due to a confusing, inconsistent response structure, along with questions that were not applicable to a range of respondents. There were no specific problems found with the DASS, but it was eliminated in favour of the PHQ-9 and the CORE-OM, both of which covered similar content to the DASS, the former already widely used within UK Primary health care.

3.5.2 Instruments Selected for Cohort Study

As a result of the cognitive debrief and the discussions of the expert panel, the five instruments that were selected for the cohort study were the Borderline Symptom List-23, the Clinical Outcomes in Routine Evaluation System – Outcome Measure, the Patient Health Questionnaire-9, the Prison Screening Questionnaire, and the Self-Harm Inventory.

3.5.2.1 Borderline Symptom List-23 (BSL-23)

Please see Appendix A, Questionnaire 3 for a copy of the complete BSL-23 instrument.

The BSL-23 (150) is the short-form version of the Borderline Symptom List, (158) which was developed to reduce patient burden and assessment time. The original Borderline Symptom List (now known as the BSL-95) was developed as a self-reported instrument to quantify typical borderline symptomatology. The full version of the BSL contains 95 items across seven domains: 'self-perception', 'affect regulation', 'self-destruction', 'dysphoria', 'loneliness', 'intrusions' and 'hostility'. The items of the BSL-95 were derived from the criteria of the Diagnostic and Statistical Interview for Borderline Personality Disorder, the opinions of clinical experts and the opinions of borderline patients. The original BSL-95 was developed in Germany among six different samples, and the BSL-23 development was based on a sample of 379 borderline patients, before being further validated in five different samples, including 659 borderline patients. (150) The internal consistency of the BSL-23 was high among all samples, with the Cronbach's alpha value ranging from 0.935-0.969. The test-retest reliability of the BSL-23 (within one week) was also reported as being high ($r = 0.82$; $p < 0.0001$). (150)

The items from the BSL-23 were based on the items from the BSL-95 that had the highest levels of sensitivity to change and the highest ability to discriminate borderline patients from other patient groups. (150, 158) It has 23 items, each with five response categories, scored 0-4. However, the original response categories suggested for the scale items did not pass the initial face-validity tests for the inclusion of the scales; therefore the response categories were adapted for use in the current study.

The original response categories as suggested by the BSL-23 developers are shown in Table 5.

Table 5. Original response categories for BSL-23 items

Response Code	Response Wording
0	Not at all
1	A little
2	Rather
3	Much
4	Very strong

When applied to the items within the scale, the face validity (possibly due to translation issues) of these response categories was questioned by the expert panel (see Section 3.2). The response categories were therefore amended to those displayed in Table 6.

Table 6. Adapted response categories for BSL-23-F items

Response Code	Response Wording
0	Not at all
1	Only occasionally
2	Sometimes
3	Often
4	Most or all the time

It is acknowledged that these revised response category options may affect the properties of the scale. The revised response options reflect a frequency relating to the BSL statements, whereas the original response options were derived to reflect an intensity rating. In order to differentiate the revised BSL-23 from the original, the revised version will be referred to as the BSL-23-F, with the 'F' denoting the frequency element of the response category revision.

The BSL-23 has 23 basic items, with an additional 'overall personal state' question, which is rated on a 0-100% scale.

It also has supplementary items for behaviour assessment. There are 11 of these on the original form, but three of these items were removed for the purposes of the study as they were deemed to be inappropriate for individuals in prison. The three that were removed were as follows:

During the last week:

"I got drunk"

"I took drugs"

"I displayed high risk behaviour by knowingly driving too fast, running around on the roofs of high buildings, balancing on bridges etc."

The supplementary behavioural items were scored as in Table 7 (during the last week).

Table 7. Response categories for BSL-23-Supplement items

Response Code	Response Wording
0	Not at all
1	Once
2	2-3 times
3	4-6 times
4	Daily or more often

3.5.2.2 Clinical Outcomes in Routine Evaluation System – Outcome Measure (CORE-OM)

Please see Appendix A, Questionnaire 1 for a copy of the complete CORE-OM instrument.

The CORE-OM is a 34-item generic measure of psychological distress with a maximum total score of 136, with each individual item scored 0-4 on the same response category structure. (151) The items cover the four domains of subjective well-being (four items), problems/symptoms (12 items), life functioning (12 items), and risk (to self and to others; six items). The CORE-OM was developed in the UK and it has been validated on non-clinical (n=1106) and clinical (n=890) samples. The internal consistency (Cronbach's alpha) ranges from 0.75 – 0.9 among the different domains, and is reported as 0.94 among both clinical and non-clinical samples for the complete item set. Test-retest correlations are reported as 0.9 for the complete item set, and 0.87-0.88 among the individual domains, except the Risk domain which delivered a lower correlation value of 0.64. It is, however, argued that this lower correlation is unsurprising given the situational and reactive nature of the items within this domain. (159)

Within the analysis, the mean item score was generated where less than 10% of items were missing, i.e. at least 31/34 items completed, as per the scale scoring instructions. The CORE-OM comprises four domains, for which the mean item score was generated where there was no more than one item missing within each domain. The non-risk items also form a 28 item subscale, in which the mean item score was generated where less than 10% items were missing i.e. at least 26/28 items were completed.

3.5.2.3 Patient Health Questionnaire-9 (PHQ-9)

Please see Appendix A, Questionnaire 5 for a copy of the complete PHQ-9 instrument.

The PHQ-9 is a nine-item depression scale with 0-3 rating available for each item, resulting in a maximum total score of 27. (149) The items consist of the nine criteria upon which diagnosis of depressive disorders is based, according to the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV). The PHQ-9 was originally developed in the United States for use in primary care, and among this primary care sample (N=3000) the internal consistency (Cronbach's alpha) was 0.89, and the test-retest reliability was reported as 'excellent' (r=0.84). (149)

Depression severity with the PHQ-9 is graded as: 0-4 = minimal; 5-9 = mild; 10-14 = moderate; 15-19 = moderately severe; 20-27 = severe. (149)

Within the analysis, a total score was generated where at least eight out of the nine items were completed, in which the mean item score is imputed for a missing item. Also, the first two items of the PHQ-9 form an initial assessment (known as the PHQ-2), (160) which has a maximum total score of six. The PHQ-2 total score was generated where a response to both items was available.

3.5.2.4 Prison Screening Questionnaire (PriSnQuest)

Please see Appendix A, Questionnaire 2 for a copy of the complete PriSnQuest instrument.

The PriSnQuest is an eight-item instrument with a maximum total score of eight. (155) The PriSnQuest was developed in the UK, building on the development of the Referral Decision Scale (RDS) in the United States. It was developed to screen for mental health problems within the UK criminal justice system. The internal consistency and test-retest reliability of the PriSnQuest have not been reported

Within the analysis, the total score was generated where at least seven out of the eight items were completed, in which the mean item score was imputed for a missing item.

3.5.2.5 Self-Harm Inventory (SHI)

Please see Appendix A, Questionnaire 4 for a copy of the complete SHI instrument.

The SHI is a 22-item questionnaire with a dichotomous response format, resulting in a maximum total score of 22. (39) The items all relate to previous engagement in different self-harm behaviours, and as such, the scale screens for the lifetime prevalence of these behaviours. The scale was initially developed in the United States, among samples taken from mental health and non-mental health settings, as a way of linking self-harm behaviours to a diagnosis of borderline personality disorder. The internal consistency was not reported in the initial development work, but it has subsequently been reported as between 0.8-0.9. (161-163) Additionally, the SHI has been shown to satisfy the requirements of Rasch scaling assumptions among a non-clinical sample. (157)

For the analysis, a total score was generated where less than 10% of items were missing i.e. at least 20/22 items completed. The SHI has demonstrated accuracy in diagnosis of borderline personality disorder (BPD) of 84% at a cut off score of five. (39)

3.5.3 Pilot Study Follow-up

At follow-up, 25 (33.3%) of the prisoners were still housed in the original prison, 28 (37.3%) had been released, 20 (26.7%) had been transferred, and the status of two of them (2.7%) was not known (Table 8).

The mean valid follow-up time was 172 days (SD 100 days). During the follow-up period, 30 (40%) prisoners had carried out at least one self-harm event (Table 9). However, the rate of self-harm varied by prison (Table 10). The number of self-harm events carried out by each individual during follow-up are summarised in Figure 13.

Of those that self-harmed, the median time to the first self-harm event (after the administration of the questionnaires) was 45 days. Importantly, there was only one incidence where the first self-harm event was beyond six months (see Figure 14), and the rate of self-harm did not increase substantially as the follow-up time increased (Table 11). Table 11 also shows the cumulative self-harm rate and the number of prisoners lost to full follow up via release and transfer for various follow up periods. Pilot data suggests a loss to follow up rate of 18.7% at 6 months (11 transferred without data available after transfer and three were missing all follow-up data) and 22.6% at 9 months (14 transferred without data available after transfer and three were missing all follow up data).

Table 8. Status of prisoner at follow-up

Follow-up status	N (%)
Still in original prison	25 (33.3%)
Released	28 (37.3%)
Transferred	20 (26.7%)
Missing status at follow-up	2 (2.7%)
Total	75 (100%)

Table 9. Summary statistics for follow-up time, and time to first self-harm event

		Follow-up time (in days)	Time to first self-harm event (days)
Study population			
Valid N		72	30
Missing N		3	45
Number of days			
Mean		171.65	64.8
Median		216.5	45
Range (min, max)		306 (1, 307)	233 (1,234)
Percentile	25	73.25	18.75
	50	216.5	45
	75	253	106.75

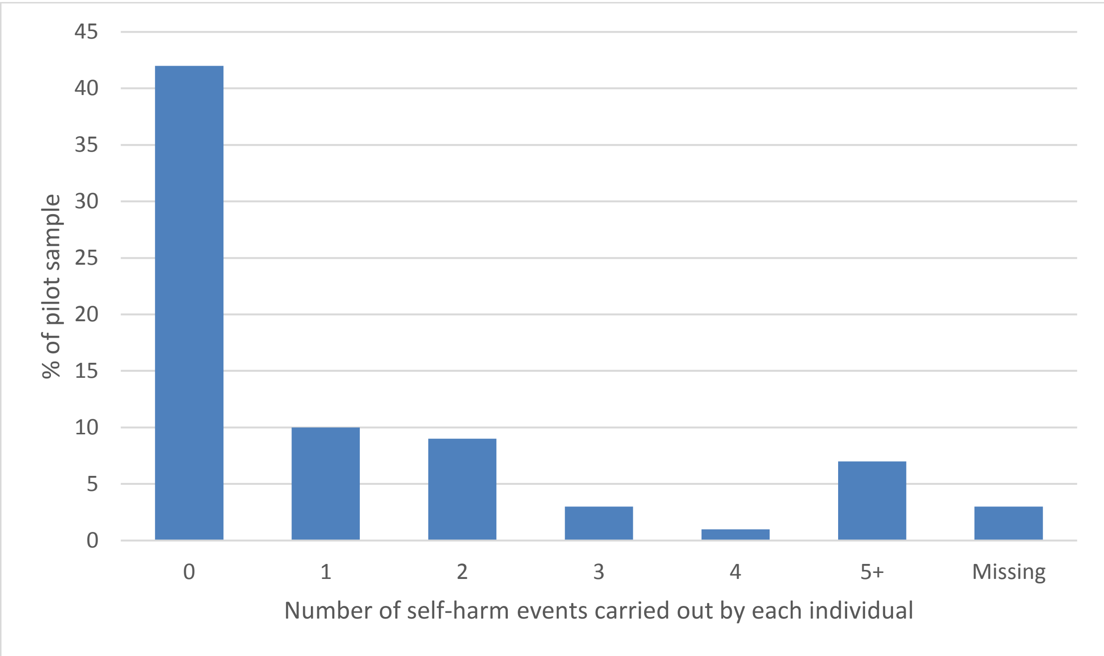

Figure 13. The amount of self-harm events carried out by each individual during follow-up, presented as a percentage of the full pilot sample (n=75)

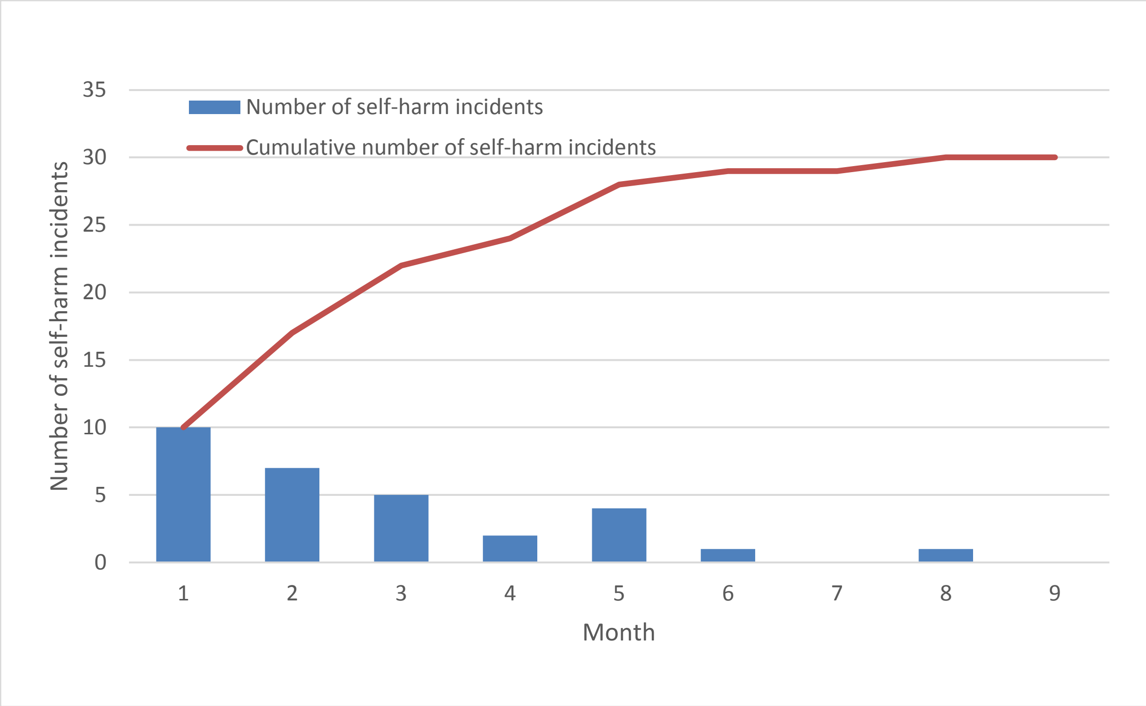

Figure 14. Time to incidence of first self-harm event during pilot study follow-up

Table 10. Self-harm rate by prison

	Prison			
Self-harm	**A N (%)**	**B N (%)**	**C N (%)**	**Total N (%)**
No	10 (47.6%)	8 (34.8%)	24 (77.4%)	42 (56.0%)
Yes	11 (52.4%)	12 (52.2%)	7 (22.6%)	30 (40.0%)*
Missing	0 (0.0%)	3 (13.0%)	0 (0.0%)	3 (4.0%)
Total	**21 (100%)**	**23 (100%)**	**31 (100%)**	**75 (100%)**

*95% CI for overall self-harm rate: (28.9%, 51.1%)

Table 11. Cumulative self-harm rate and loss to follow-up rate due to release and transfer by follow-up time point

Follow up to:	Self-Harm rate N (%)	Released or Transferred with no further follow up N (%)	Loss to follow up – Transferred with no further follow up* N (%)
5 months	28 (37.3%)	28 (37.3%)	13 (17.3%)
6 months	29 (38.7%)	31 (41.3%)	14 (18.7%)
7 months	29 (38.7%)	36 (48.0%)	16 (21.3%)
8 months	30 (40.0%)	39 (52.0%)	17 (22.6%)
9 months	30 (40.0%)	42 (56.0%)	17 (22.6%)

*Includes the additional 3 prisoners missing all follow up data

3.6 Implications for Cohort Study

3.6.1 Process

The pilot study was designed to inform the cohort study, and a number of implications were forthcoming. Firstly, the data collection process and study logistics worked well, so it was agreed amongst the study team that the process for the main cohort study would remain largely unchanged. However, researchers reported difficulty in trying to conduct all interviews within 72 hours of the index ACCT being opened; therefore some potential recruits were missed during the pilot study. This was due to two reasons, the first of which was the logistics of the researcher actually being able to contact the prisoner within this timeframe. The second reason was the unstable, unsafe or vulnerable state of some prisoners within the first 72 hours of the ACCT being opened, which precluded them being approached for inclusion. To address this situation, the timeframe was changed from '72 hours of the ACCT being opened', to 'within two weeks of a prisoner being on an active ACCT'. This was done in order to maximise study recruitment, and it would also allow for the potential inclusion of people who are on a long-term ACCT (some ACCTs never get closed).

Additionally, due to the results of the time to first self-harm event witnessed in the pilot study, the active follow-up period in the cohort study was reduced from nine to six months. Decreasing the follow-up time maximises potential recruitment time for the study, whilst maintaining the opportunity to capture the vast majority of self-harm events (of those who self-harmed within the pilot study, 29/30 (96.7%) self-harmed within six months of the interview).

3.6.2 Initial Sample Size

The original protocol sample size required approximately 1400 prisoners to be recruited into the study. These would all be administered a small set of questionnaires in an overlapping block design. It was originally anticipated that a total of four screening instruments would be administered, and that each prisoner that consented to take part in the study would only respond to two screening instruments, in order to minimise the responder burden. Therefore, a scale administration block design was utilised, in which there were six combinations of two-scale administrations (see Table 12).

The initial sample size was primarily determined by the need to compare the Area under the Curve (AUC) between each pair of self-harm screening instruments. A secondary requirement was to achieve the relevant degree of precision required by the psychometric analysis (Mokken and Rasch Analysis).

A previous prison audit revealed that approximately 20% of inmates are assigned an ACCT in any given year. Other work has shown that up to a quarter of females could self-harm during their current prison term. (33, 164) Thus, assuming a 20% prevalence rate for self-harm, it was estimated that a sample of 405 prisoners would be required to achieve 80% power to detect a difference of 0.1 between a diagnostic test with a receiver operating characteristic (ROC) area under the curve (AUC) of 0.8 and another diagnostic test with an AUC of 0.9, using a two-sided z-test with a 5% significance level. This calculation was based on discrete (rating scale) responses, and assumed: similar levels of variation for responses in prisoners with and without self-harm for both diagnostic tests (i.e. the ratio of the standard deviation of responses of prisoners with self-harm to those without was 1.0 for both diagnostic tests); and a correlation between the two diagnostic tests for both the prisoners with and without self-harm of 0.6 [PASS 2008 (NCSS, LLC, Kaysville, UT, USA)].

Given that an ACCT is an indicator in itself of potential risk for self-harm, it was thought that the self-harm prevalence in this group may be substantially higher than the general estimated level of 20%. As such, the above quoted sample size of 405 would then have sufficient power to detect smaller differences between the AUC of any two diagnostic tests. Consequently, for the comparison of each pair of instruments, a sample size of 405 was required. Given the original block design of six two-instrument combinations, a sample of 840 would provide 420 prisoners who could be compared on any pair of screening instruments (Table 12). With a degree of uncertainty surrounding the follow-up rate that would be achieved, a conservative estimate led to deliberate over-sampling of approximately 70%, meaning that the initial aim was to assess approximately 1400 prisoners. This would allow for recruitment of 840 participants with sufficient follow-up information available for a reliable AUC analysis.

Table 12. Two-scale combinations and administration numbers

Two-instrument combinations			
Combination	Instrument 1	Instrument 2	n
1	**A**	**B**	140
2	**A**	**C**	140
3	**A**	**D**	140
4	**B**	**C**	140
5	**B**	**D**	140
6	**C**	**D**	140
		Total n	**840**
n of Scale A completed			420
n of Scale B completed			420
n of Scale C completed			420
n of Scale D completed			420

For the Rasch analysis that will be carried out, sample size is primarily concerned with the degree of precision of the item calibrations for any given scale. A sample size of 400 respondents for any given instrument would estimate the item difficulty calibrations to within ± 0.3 logits, with a significance level of 0.01. This is the minimum practical level of stability expected for most variables. (165)

3.6.3 Sample Size Re-estimates

The pilot study brought about several changes to the protocol, including the estimated sample size required for the study.

As shown in Table 11 and Figure 14 the rate of self-harm did not increase substantially as the follow-up time increased beyond six months, suggesting that follow-up time could be restricted to a six month period in order to maximize the recruitment period in the cohort study.

The original sample size was inflated by approximately 70% to allow for a final sample with sufficient follow-up time for a reliable AUC analysis. However, after further consideration it was agreed that as the focus of the study was self-harm during the follow-up period post-ACCT or the time to release (whichever was sooner), prisoners who were released prior to the end of the follow-up period would not be considered lost to follow-up, assuming that full data would be available for them during their time in prison post ACCT. A prisoner would therefore only be considered lost to follow-up if they were transferred prior to the end of the follow-up period with no available information after their transfer date, or if no follow-up data are available at all. Given the loss to follow-up rates observed in the pilot study (Table 11) in which a loss to follow-up rate of 18.7% was observed at six months and 22.6% at nine months, it was agreed that a loss to follow-up rate of 20% at six months could be assumed for the cohort study.

The original sample size estimates assumed a self-harm rate of 20%. However, the overall self-harm rate observed during the pilot study was 40% with an overall 95% confidence interval of 28.9% to 51.1%. The proportion of prisoners recruited from each prison in the cohort study is expected to be similar to that in the pilot. However, considerably lower rates were observed in Prison C compared to prisons A and B. It was also planned that the follow-up period in the cohort study would reduce from nine to six months. Thus, when considering the sample size re-estimates, an expected self-harm rate of approximately 30% was considered appropriate, based on the lower limit of the 95% confidence interval in order to limit the deviation from the prior assumption of 20%.

Given the results of the pilot study, the sample size for the AUC analysis was re-estimated assuming a self-harm prevalence rate of 30% and loss to follow-up rate of

20% at six months. These estimates provided a sample size of 359 prisoners to provide 80% power to detect a difference of 0.1 between the AUC for two diagnostic tests at the 5% significance level. Similarly, 475 prisoners would provide 90% power to detect such a difference (Table 13). As per the original sample size assumptions it was assumed that: the detection of a difference of 0.1 between the AUC for two diagnostic tests would involve one test with an AUC of 0.8 and the other with an AUC of 0.9; similar levels of variation for responses in prisoners with and without self-harm for both diagnostic tests (i.e. the ratio of the standard deviation of responses of prisoners with self-harm to those without was 1.0 for both diagnostic tests); and that the correlation between the two diagnostic tests for both the prisoners with and without self-harm was 0.6.

Table 13. Sample size requirements for AUC analysis under levels of power

Self-harm prevalence	30%	30%
Power	80%	90%
N - Sample size (number who Self-Harm)	287(86)	380 (114)
N - Sample size accounting for 20% loss to follow up (number who Self-Harm)	359 (108)	475 (143)

3.7 Summary of Pilot Study and Implications for Cohort Study

The pilot study showed that it was possible to administer a set of screening instruments in a prison setting. It also showed that the prisoners themselves were happy to spend time in an interview setting, and were able to answer questions from a broad range of instruments. Approximately 60% of the pilot study participants were still within the prison system at the time of follow-up, and the loss to follow-up rate at nine months was found to be 22.6%. The self-harm rate was found to be 40%, with the vast majority of these events occurring within six months of the baseline interview.

Given these findings, the block randomisation of instruments was abandoned, and all prisoners were to be administered all of the chosen instruments at the same time, built into a single questionnaire pack (Appendix A). Using a conservative rate of 30% for self-harm and a six month follow-up period with a 20% loss to follow-up rate, it was calculated that 359-475 cases would be sufficient to give 80%-90% power, respectively, for the AUC analysis. This sample size would also, as before, be sufficient for the Rasch analysis. The same prisons involved with the pilot study would be used for the cohort study.

4 Methods

There are a number of different methodologies used within this thesis, and this chapter provides a description of each of the relevant methodologies. Firstly, the cohort study is described in terms of the setting and participants, including the data collection process. An introduction to measurement and psychometrics is provided, including information regarding the confirmatory factor analysis (CFA) and Mokken scaling methodologies, and a more detailed description of Rasch analysis, which was the predominant methodology utilised for the purpose of instrument validation. The methodologies involved in ascribing the predictive validity of each instrument are then described, along with the process used to identify individually predictive items, and how these may be combined into a predictive item set using a logistic regression. Finally, an introduction to structural equation modelling (SEM) is provided, which is a methodology that can be used to explore the complex interaction between traits (variables), in terms of the causal pathway to the final outcome of whether or not self-harm occurred.

4.1 Cohort Study Protocol

4.1.1 Process

The process for the cohort study was the same as in the pilot study, as described in Section 3.4, with the changes described in Sections 3.6 and 3.7.

Briefly, all prisoners who have an ACCT opened were approached to participate in the study. After being given a study information sheet (Appendix B) to consider, all potential participants were asked to provide written, informed consent to take part in the study (Appendix C). Although prisoners were recruited in their prison setting there was, in practice, a variable amount of time available for considering the study information sheet. Following consent, each participant was administered a questionnaire pack featuring five standardised instruments and a range of other sociodemographic and sentencing information which was thought relevant to the study (Appendix A). All questions were read aloud by one of the research team, and responses were directly recorded in the questionnaire pack by the researcher.

Follow-up was carried out by checking the prisoner record on the National Offender Management Information System (NOMIS) prison computer record system, to primarily determine whether the participant had self-harmed during the follow-up period. Each study participant had a valid follow-up time of six months if they were still within the prison system, or up to the point of their release from their index prison stay.

In an unanticipated change from the pilot study, the follow-up process was forcibly amended following a change to the prison NOMIS computer system. In the time period between the pilot study follow-up being carried out and the cohort study follow-up being carried out, a nationwide system change of the NOMIS computer system was implemented, with the result being that the Global Transfer Report was no longer available.

During the pilot study follow-up, if a prisoner was still housed within the original institution, or had been released, then the required follow-up information was available on the NOMIS system. If a prisoner was still within the prison system but had been transferred to a different establishment, the required follow-up information was available from the Global Transfer Report section of the NOMIS system. As the Global Transfer Report had been removed from the NOMIS system for the cohort study follow-up, the required follow-up information was no longer directly available for the transferred prisoners.

An amended protocol was therefore implemented to obtain the required follow-up information for transferred prisoners. The amended protocol involved the identification of the establishment to where the prisoner had been transferred, and then making direct contact with the relevant establishment to obtain the required follow-up information. This approach required the cooperation of the prison governors in the study institutions to provide a letter of reference for the prison-based researchers. It also required the cooperation and goodwill of prison staff within the institutions where transferred study participants were housed at the time of follow-up.

This unforeseen amendment made the follow-up process more difficult and time consuming, although the relevant follow-up information was still eventually obtained for the vast majority of cases.

4.1.2 Ethics

Ethical approval was granted by the National Research Ethics Committee, the Ministry of Justice, with local approval from each local NHS R&D office. The University of Leeds was the sponsor for the study. The Project Steering Committee consisted of the Chief Investigator, an independent chair, and an independent member. The Study Management Group (SMG) consisted of the Chief Investigator, co-applicants, research staff, and a service-user representative.

4.2 Measurement and Psychometrics

Measurement is an imperative feature of scientific research, and it is therefore important that variables and outcomes are quantified in the most appropriate manner. Some things, such as distance and mass, can be measured directly using calibrated measuring instruments. Other things, however, cannot be measured directly, and we are therefore reliant on observing certain manifestations of these traits in order to quantify them. These are known as latent traits, and they are commonplace in healthcare research, with examples including depression, anxiety, pain, independence and function.

The basic idea with any measurement is to place a person upon a continuum of whatever is being measured. In this way, the level of a trait that a person exhibits can be quantified and persons can be compared upon the continuum. This is depicted simply in Figure 15, where the trait in question ranges from a low level through to a high level, although this may be reversed depending upon the nature of the scale and the scoring direction.

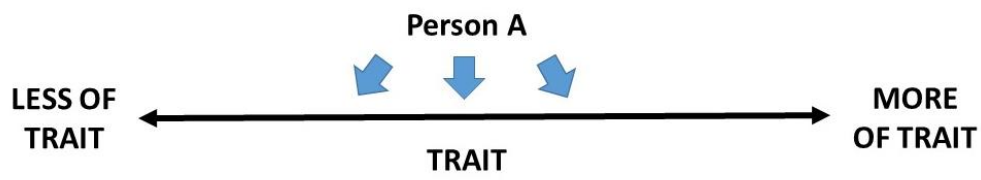

Figure 15. A measurement continuum

When traits can be measured directly (as in the physical sciences) using calibrated measuring instruments, all measures are taken in standardised units (e.g. centimetres, Newtons, degrees Celsius etc.) on either an interval or a ratio scale. Interval and ratio scales are equivalent in terms of the standardisation of the fixed unit, but they differ in that a ratio scale always has an implied absolute zero point (e.g. centimetres), whereas an interval scale has an arbitrary zero value, which is applied as a matter of convention or convenience (e.g. degrees Celsius). (166) With regard to the fixed unit, this means that the ‘distance’ between two consecutive unit points on the scale is always uniform and consistent across the entire range of the scale. Where latent traits are concerned, score values (from questionnaires or patient reported outcome measures, for example) are assigned to represent the magnitude of the trait that is present within a person, but the measurement units from these scales are not (by their nature) standardised. This means that the raw scores from these types of observations are, at best, only ever ordinal in nature, and that the ‘distance’ between two consecutive unit points on a scale is not uniform and consistent across the entire range of the scale. Figure 16 can be seen to depict representations of an ordinal scale and an interval scale (or a ratio scale, if the zero is absolute, as it would be here) on opposing sides of the ruler, where the difference between the measurement units can be observed.

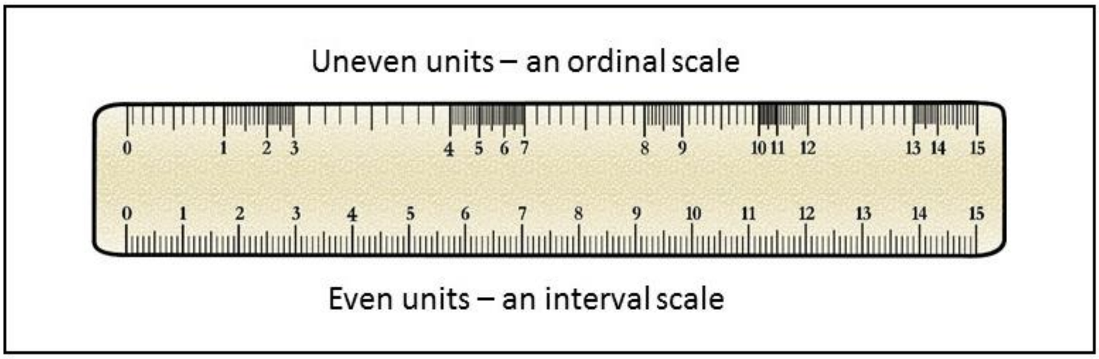

Figure 16. Difference between ordinal and interval scaling properties (provided courtesy of A. Tennant)

Additionally, where measures of latent traits are concerned, there are a series of underlying assumptions that should be tested to ascertain whether the rating scale measure is actually functioning as is intended. The study of these methods for measuring latent (psychological and non-psychological) variables is known as psychometrics, and the purpose of a psychometric analysis is to determine whether a rating scale is functioning correctly; therefore establishing the extent to which a quantitative conceptualisation of a trait has been successfully operationalised (167, 168).

The cohort study incorporated five standardised questionnaires into a single questionnaire pack, along with other socio-demographic and sentencing information which was thought to be relevant, and the primary aim of the study was to see if any of the selected extant questionnaires are predictive of self-harm over a six-month time period. All of these pre-existing questionnaires have been developed to measure a particular thing, but none were specifically designed to predict future self-harm events. Nevertheless, a critical element of the measurement process is to determine whether the selected scales are operating properly within the reference frame of the study, specifically the on-ACCT prison population from which the sample was derived. A psychometric analysis will determine whether these questionnaires hold certain properties which are consistent with quality measurement, and test whether it is valid to sum the scale items into a total score that represents the latent trait under consideration. These desirable psychometric properties qualities include reliability (whether it measures consistently), validity (whether it measures what it intends to measure), and unidimensionality (whether it measures a single construct) and these are described more fully elsewhere (168-170).

The field of psychometrics is broadly partitioned into two areas: traditional psychometrics, which concerns classical test theory (CTT); and modern test theory (MTT), which concerns the more modern and advanced psychometrics methodologies of item response theory (IRT) and Rasch measurement theory (RMT).

4.2.1 Classical Test Theory

Classical test theory was largely developed through the work of Charles Spearman at the turn of the 20th century, (171) and the concept is based around the notion that a person's observed score on a set of items is made up of their 'true score' plus an error value. This is represented by the following formula:

$O = T + E$

Where O represents the observed score, T represents the true (real) score of the individual, and E represents the (measurement) error. The score that is observed is therefore is mixture of relevant information and error. (172) The premise is that a set of items can be summed without any weighting or standardisation to produce a total score, (173) and this forms the basis of well-known psychometric methodologies such as factor analysis. (172) Classical test theory remains popular due to its familiarity and accessibility among prominent statistics packages, (172) and it is recognised that it provides a useful model that has served scale developers for many years. (168, 174) However, CTT also has a number of limitations. (168, 172) Perhaps the most important of these limitations is that the actual values of the true score (T) and the error score (E) cannot be determined, and therefore the assumptions underpinning the theory cannot be tested. (168) A further limitation is that the scores used within CTT analyses are treated as interval measures when they are, at best, only ordinal counts. (168, 172) Additionally, properties of CTT-derived items and scales are sample-dependent, results obtained for samples are scale-dependent, and the standard error of measurement for a person on a scale is assumed to be a constant value, regardless of that person's score on the scale. (168) Although the methodology is widely used, it has been recognised that as the theory of CTT cannot be verified in data; therefore the analysis of total scores as true measurements is problematic. (168)

4.2.2 Modern Test Theory

The origins of modern test theory (MTT) are based on the work of Louis Thurstone, (175) who was the first person to attempt to apply a strong theoretical and mathematical basis to rating scales. (168) Thurstone recognised key attributes that should hold when measuring latent variables, and these requirements included:

- Rating scale items should define a continuum and should be located across the continuum as markers of different levels of the construct of interest
- Rating scales should only measure one clearly defined single aspect
- Rating scales should measure that entity on an equal-interval (interval level) scale
- The scale should transcend the group measured, in that the scale estimates should not be sample-dependent, and the sample estimates should not be scale dependent

(168, 175)

These principles form the foundations of modern test theory, and these are distinctly different from the basis of CTT. Whereas CTT applies successive integers to successive rating scores and treats them as interval measurement, in MTT (both IRT and RMT) some modifications are applied to this process. (176) In MTT, the focus is shifted to the relationship between a person's unobservable measurement on an underlying continuum, and that person's probability of responding to a specific response category within a rating scale. (168) Within this framework, the focus is shifted from the total scale score level (as in CTT), to the item score level. The main differences from CTT are that the models used in MTT are probabilistic in nature, response categories are not assumed to be the same size, and the number of response categories is explicitly finite. (176) Because of these differences, the assumptions of the underlying models can be formally tested, thus forming a stronger basis for measurement.

However, within modern test theory, there is still a further separation into item response theory (IRT) and Rasch measurement theory (RMT). Although these two approaches share the similarities outlined above, they also diverge around a key epistemological point. These two paradigms have been termed as experimental measurement (RMT) and statistical modelling (IRT) as these terms characterise the key feature of the analysis approach. (176, 177) Within the statistical modelling paradigm (where IRT sits), the process involves selecting a model which best represents the observed data, and the justification for the model selection is the empirical evidence of its suitability to the observed data. (176, 178) Within the experimental measurement paradigm (where RMT sits), the emphasis is placed on the primacy of the mathematical model, where the observed data are analysed to check whether they meet *a priori* specifications. (168, 176) This approach is taken not because it describes any particular set of data, but because the inherent properties of the (Rasch) models provide the optimum criterion for fundamental measurement, thus allowing for invariant comparisons within a specified frame of reference. (168, 176)

A further complication between IRT and RMT is that the Rasch model is mathematically equivalent to the one-parameter logistic model that is used within IRT, so the Rasch model is often viewed as a special case within the IRT field. Although this equivalence exists, the fundamental difference between IRT and RMT is the approach to the research agenda. (168, 177)

Measures of latent variables that are provided by rating scales or questionnaires normally produce ordinal scaled scores, where respondents are ranked in order of magnitude of the construct being measured. However, where data are shown to satisfy the requirements of the Rasch measurement model (179), a distinct advantage is that these scale scores can then be transformed into interval scaled measurement where increments in score are of equal units (179, 180). Determining if this is the case, the process of Rasch analysis tests if data accord with model expectations, and provide further diagnostics as to the issues that may attributed to an item. Thus, Rasch measurement theory offers detailed diagnostic information on the way that scales work, and provides a unified framework to perform the inclusive analysis.

Consequently, for all candidate screening instruments going forward into the cohort study, elements of both classical and modern test characteristics are reported. These include an assessment of unidimensionality through a Confirmatory Factor Analysis (CFA); an assessment of ordinal scaling through a Mokken analysis, and an assessment of interval scaling and other associated properties through Rasch analysis.

4.3 Confirmatory Factor Analysis (CFA)

A fundamental assumption of test theory is that a set of items should measure just one attribute or dimension; else the score is not interpretable. (170, 181, 182). This unidimensionality is an assumption where a set of items are to be summated to provide a total score. As the name would suggest, confirmatory factor analysis (CFA) is a confirmatory procedure which makes it possible to test whether the hypothesized factor structure of a questionnaire (based either on empirical data or on theory) is supported by actual data. (183) Confirmatory factor analysis may be in the form of a single set of items (questions) which purport to measure a single domain, or it may seek to confirm that a larger set of items map onto a specific number of pre-specified domains. Consequently, analysis of the dimensional structure of the candidate screening tools which were chosen for the current study represents the foundation of the psychometric analysis, as all further stages have the assumption of unidimensionality. CFA is undertaken with the MPlus package (184), and is based upon a polychoric correlation matrix for polytomous variables, and a tetrachoric correlation matrix for dichotomous variables. The polychoric and tetrachoric correlation coefficients are a measure of association for ordinal variables, which rest upon an assumption of an underlying joint continuous distribution. These correlation matrices therefore provide the most appropriate basis of carrying out a CFA upon an ordinal item set, as the basis of any other CFA assumes that scale scores represent interval level (parametric) data.

While a strict CFA interpretation would require uncorrelated errors between scale items, it is quite common in health-related scales (e.g. depression) to find items which are linked in some fashion such that errors should be correlated. Although this would be indicative of a certain dependency between items, sometimes this reflects certain nuances of a construct that may still be important for clinical decision making. Although these items may breach the assumption of local independence, simply discarding these items is not always appropriate due to the clinically relevant information that they contain. The correlation of errors will therefore be allowed within the CFA, as this takes account of any between-item dependency which may be present.

Several fit statistics will be used to determine if the CFA is satisfactory. The primary measure is the Chi-square statistic where a non-significant value indicates that the data conform to expectations. (185) Supplementary fit statistics include the Root Mean Square Error of Approximation (RMSEA) where a value of <0.08 would be considered sufficient. A Tucker-Lewis Index (TLI) and Comparative Fit index (CFI) value of >0.95 would also support the proposed data structure.

Given these fit parameters, scales can be graded as summarised in Table 14, indicating the degree of support for unidimensionality:

Table 14. CFA fit parameter guidelines

Quality of Support	Chi-Square	RMSEA	TLI	CFI
Strong	>0.05	<0.08	≥0.95	≥0.95
Medium	>0.01	<0.08	≥0.90	≥0.90
Weak	<0.01	<0.08	≥0.90	≥0.90

If a scale provides sufficient evidence of conforming to a CFA, this suggests that the items are unidimensional to a degree, and therefore that it is a valid procedure to obtain a total score from the set of items within the scale. Although this does not confirm any particular level of measurement, it does provide a basis for the validity of any cut points which are stated for the relevant scales.

4.4 Mokken Scaling

Mokken scale analysis is used for scaling items and measuring respondents on an ordinal scale. (186, 187) It is a nonparametric probabilistic version of Guttman scaling (188), and it is used similarly to other techniques for data reduction that allow for the unidimensional measurement of latent variables. Building on the basis of the CFA, the stochastic cumulative scaling model offered by this approach is ideally suited when the intention is to score an underlying latent trait by simple addition of the item response values. (189) It has been shown to have a number of advantages over some other measurement models; for example, it includes an item parameter that shows how items differ in their distribution, it is probabilistic rather than deterministic, and it can be applied in situations in which latent variables must be operationalized with only a small number of indicators. (190)

The process has a number of assumptions which are to be found in most non-parametric and parametric (e.g. Rasch model) Item Response Theory (IRT) Models. These are unidimensionality, local dependence and monotonicity (the probability of affirming an item increases as the underlying level of the construct increases). As with Guttman scaling, model violation is crucial to interpretation, and this revolves around the relative responses upon items (e.g. If a>b and b>c, then it always follows that a>c) and the difference between what is expected within the model, and what is observed within the data. The scalability of the scale is measured by Loevinger's coefficient H, which compares the actual Guttman errors to the expected number of errors if the items would be unrelated. (191) In practice this reflects the amount of discrimination of an item where, for example, very low values of H would indicate poor discrimination (a flat item response function). Consequently, many computer programmes adopt a minimum requirement of H > 0.3 for item selection. The MSP module (192) of STATA 13 (193) was used to carry out the Mokken analysis, and the interpretation of levels of scaling based upon H values are reported in Table 15.

Table 15. Interpretation of Loevinger's H values

H_{ij} Value	Interpretation
< 0.3	poor/no scalability
0.3-0.4	useful but weak scalability
0.4-0.5	medium scalability
>0.5	good scalability

As with the CFA, the use of Mokken scaling in the current study is designed to provide information to support the summation of a set of items to provide an ordinal scale, thus validating any associated cut points. As it has the assumption of unidimensionality, this analysis follows the CFA of the candidate scales.

Given the double monotone homogeneity of the procedure, which orders both persons and items, it can also be considered a prelude to Rasch analysis, in that failure to satisfy Mokken scaling criteria would indicate that it would be unlikely to satisfy Rasch model assumptions. Alternatively, if a Rasch analysis were carried out first, then Mokken scaling may be seen as an appropriate model to use if the item set does not conform to the strict requirements of a Rasch model. This would determine whether the item set was unidimensional and ordinal, but failed to satisfy the assumptions necessary to obtain interval level measurement.

Despite Mokken scaling generally being seen as a useful process, some concerns have been expressed about its merits. It has been argued that Loevinger's H is not a measure of monotone homogeneity, and that it is not sample independent. (194) In practice, these two aspects are satisfied by only the Rasch model.

4.5 Rasch Analysis

While Mokken scaling offers a test to see if a set of items form an ordinal scale, fit of the data to the Rasch measurement model tests to see if the data satisfy the requirements of a quantitative structure, so providing interval scale measurement. (179, 195) Essentially, the Rasch model provides the optimum criterion for fundamental measurement. (168, 180)

Whereas ordinal scales should be comprised of a unidimensional item set with a specific hierarchical (difficulty) ordering, an item set which satisfies the Rasch model also delivers the properties of additivity and specific objectivity. The additivity property relates to the standardised common unit within the Rasch framework (the logit) which is not present within an ordinal measure (see Figure 16). With regard to specific objectivity, this means that the (logit) location estimates of the people being measured do not depend on the sampling distribution of the items, and, likewise, the location of the items does not depend on the sampling distribution of the persons, although a mis-targeted distribution of persons and items will have an impact upon the precision and error surrounding these estimates. To phrase this slightly differently, this means that the relative locations of any two people does not depend on the items to which they responded, and that the relative locations of any two items does not depend on the people from whom the item estimates are generated. (168, 177)

A distinct advantage of scales which satisfy both Rasch and Mokken scale model assumptions is that the items form a unidimensional scale where the raw score is a sufficient statistic, meaning that the total score on the scale gives you an estimate of the person's ability at the ordinal level, and that you do not require any additional information. (196) However, unlike the raw score from a Mokken scale, the raw score from a Rasch scale can be transformed to interval scaling, meaning that mathematical operations such as change scores and parametric statistical techniques can be carried out (given appropriate distributional properties).

Although the Rasch model is mathematically equivalent to a one-parameter logistic model within the field of item response theory (IRT), it has been argued that the practical application of the Rasch model is fundamentally and epistemologically different from the application within an IRT framework. (177) Whereas the IRT paradigm seeks to explain variance within the data by applying different extensions of the model, in the Rasch paradigm the primacy of the model is emphasised due to the inherent measurement properties of the model. (168) In short, the Rasch model provides a practical definition of measurement; the model is fixed and the data are explored to investigate deviations from the model. In other IRT models, the data remains fixed and a range of models are explored to best describe the data. (168, 177, 197, 198)

4.5.1 The Rasch Measurement Model

The Rasch model is a unidimensional measurement model which is based on a probabilistic form of Guttman scaling. It asserts that the 'easier' the question, the more likely it will be answered 'correctly' (or affirmed). The context of this is variable, depending upon the nature of the trait that is being measured. The original work of Rasch (179) was set in the field of attainment, and therefore most of the early work was carried out in the education domain, in the field of assessment and attainment. In the context of education, the levels of a trait as represented by an item or a person are commonly referred to as 'item difficulty' and 'person ability', and the principles behind the model are fairly straightforward to interpret. Within this reference frame, the model asserts that the more able a person, the more likely they will answer an item correctly compared to a less able person. The model assumes that the probability that a person will affirm an item is a logistic function of the difference between the person's ability [β] and the difficulty of the question [δ], and only a function of that difference. (199)

$$P_{ni} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}}$$

Where p_{ni} is the probability that person *n* will affirm item *i* (or answer the item 'correctly'), β is person ability parameter, and δ is the item difficulty parameter.

From this, the expected pattern of responses to a set of questions is determined given the estimated β and δ. When the observed response pattern does not deviate too much from the expected response pattern, then the questions constitute a true Rasch scale.

One of the main advantages of the Rasch model is that the item difficulty and person ability parameters are derived independently, and therefore the item analysis is not dependent upon the sample from which it was taken (specific objectivity). (200, 201)

The Rasch model is a unidimensional model, and thus it assumes that any set of items within an analysis is purported to be unidimensional. As previously suggested, the purpose of a Rasch analysis is not to sort out a set of items into a varying number of underlying factors, but to assess the relationship between a set of items that are supposed to be unidimensional. Any substantive anomalies to the assumed underlying probabilistic relationship within the item set will be highlighted by the analysis process, which allows for a better understanding of the trait that is being measured. (177, 197)

4.5.2 Rasch Analysis Methodology

A Rasch analysis examines the extent to which the observed data (persons' actual responses to scale items) are concurrent with ('fit') predictions of those responses from the Rasch model. Thus, the difference between expected and observed scores indicates the degree to which rigorous measurement is achieved.

The RUMM2030 computer software (202) was used to carry out all Rasch analytic procedures, and in brief, Rasch analysis will be used to evaluate: overall scale fit and reliability; individual item fit; item response threshold ordering; targeting; response dependency; differential item functioning (item bias); and unidimensionality.

4.5.3 Item Fit

To determine how well each question fits the Rasch model, and so contributes to defining a single dimension, a set of fit statistics are used. These statistics include overall fit statistics as well as fit statistics for individual questions. Statistics indicating fit to the model test how far the observed data match the model expectation. An item-trait interaction statistic, reported as a Chi-Square, reflects the property of invariance across the trait. A significant Chi-Square value indicates that the hierarchical ordering of the items is variable across the trait, indicating a lack of the desired scale invariance. In addition, individual item-fit statistics are presented, both as residuals (a summation of individual person and item deviations, standardised into a z-score) and as a chi square statistic (deviation from the model by trait-level ordered groups of people, known as class intervals). Most standardised item-fit residuals should fall between the conventionally accepted ranges of +/- 2.5. (198, 200)

Misfit of an item indicates a lack of the expected probabilistic relationship between the item and other items in the scale, and largely manifests in three main ways: (a) over-discrimination, (b) under-discrimination, (c) no pattern in the data. These are illustrated in Figure 17.

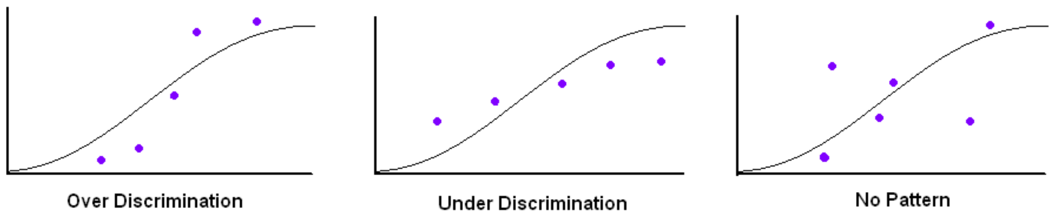

Figure 17. Discrimination patterns

Each plot shows an Item Characteristic Curve for one question. The grey curve represents the Rasch model expected pattern of the responses, and the purple dots represent the actual average response of groups of people of the same overall level of ability (class intervals). The x-axis on the figure represents both the level of ability of the persons and the difficulty of the question in logits, (low-high). The y-axis represents the response score.

Over-discrimination: This type of discrimination is not desirable since it indicates a more deterministic response pattern than is expected. This often occurs when an item is redundant or dependent, which could artificially inflate the scale score. Over-discrimination is characterised by a high negative Fit Residual statistic on an item.

Under-discrimination: This pattern is extremely undesirable as it indicates that the response to the item is not being influenced by (i.e. is not related to) the trait that is being measured by the rest of the items. Under-discrimination is characterised by a high positive Fit Residual statistic on an item.

No Pattern: This pattern is also undesirable. The trait under consideration is not influencing the response to this question; therefore meaning that the responses to the question are largely random.

4.5.4 Response Category Ordering

When responding to, or scoring an item, the response categories should progress in a logical progressive manner in terms of their ordering. i.e. the individual response categories should progress from a response representing 'less' to 'more' of the trait in question, or vice versa. Following this logic, the thresholds (crossover points) between adjacent response categories should also progress in the same logical manner, with reference to the underlying trait that is being measured. (203)

As progress is made along the underlying trait continuum, each response category should emerge as *the most likely* response at some point, and this pattern can be seen in Figure 18.

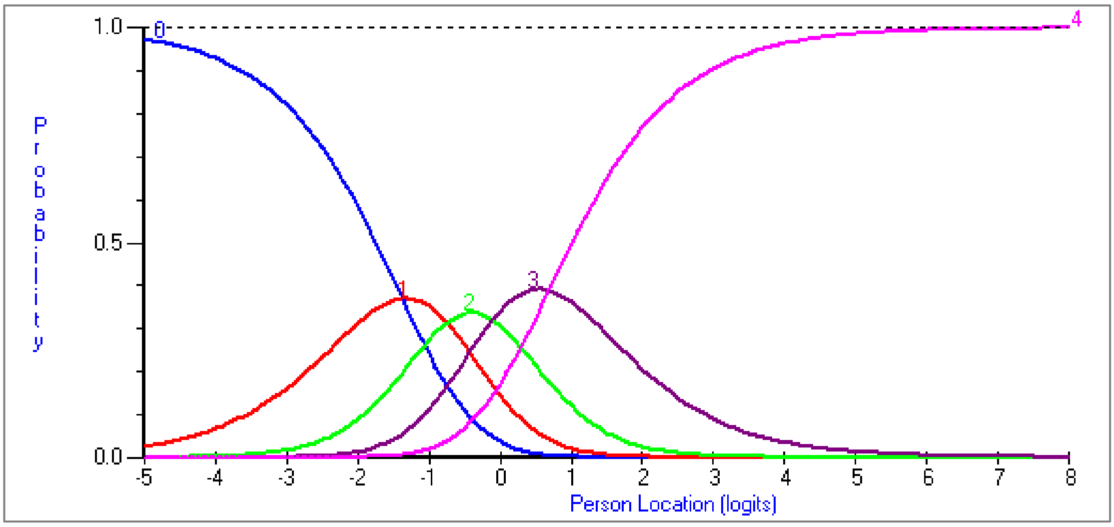

Figure 18. Example of an item with ordered response categories

If the response categories do not work as intended, then this logical structure is not seen. One or more response categories will never emerge as *the most likely* response, and the ordering of the thresholds becomes disordered, or reversed, as can be seen in Figure 19.

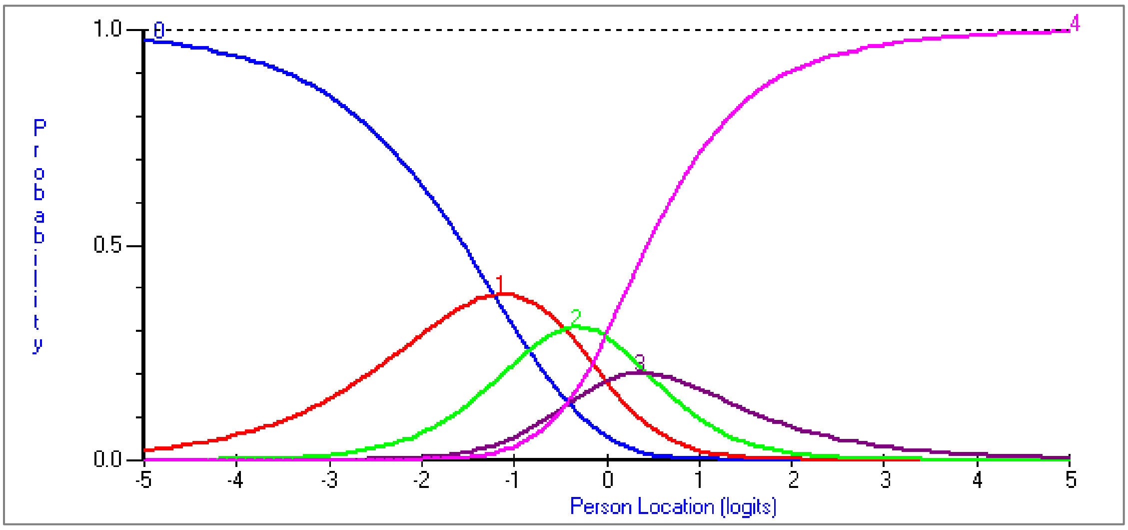

Figure 19. Example of an item with disordered response categories

As can be seen in Figure 19, at no point along the continuum is response 3 the most likely response. Also, the threshold between response categories 3 & 4 falls before the threshold between response categories 2 & 3 on the underlying continuum. This is illogical in measurement terms.

4.5.4.1 Rescoring

Where response options were found not to work as intended across the whole item set, a generic rescore was considered. This is a post-hoc adjustment of the original response categories which treats two (or more) adjacent response categories as equivalent. It is necessary to do this as the disordering of the original response categories implies that the respondents (i.e. the prisoners in this case) do not distinguish between the presented response categories, meaning that the intended discrete, ordered response category structure is not working in the way that it was originally designed. When rescoring, it is logical for this to be guided by the content and wording of each response category. It is often possible to see where the confusion may arise (where response options are similar or overlap), and linking these response options back to the observed threshold patterns helps to inform rescore options. (197, 198, 201)

4.5.5 Person Fit

In the same way that items can misfit the model, certain individual response patterns may show up as misfitting if they are unexpected or contain too much dependence. Most standardised person Fit Residuals should fall between +/- 2.5, with a high positive residual value indicating an unexpected response pattern, and a high negative residual value indicating a certain dependency or predictability within the response pattern. (198, 200)

4.5.6 Response Dependency

Response dependency occurs when the response to one item has a direct influence on the response to another item, over and above the level that is explained by the common trait. Response dependency is indicated by a correlation in the Residuals between items. A fixed criterion residual positive correlation value of 0.2 was used to indicate dependency in this study. (204) However, it has recently been shown that the criterion response dependency value is actually relative, and that a response dependency between items is generally indicated with a positive correlation of 0.2 above the average residual correlation of the complete item set. (205) If dependency is present, then the nature of the probabilistic relationship between items does not hold, as a more deterministic pattern is present. This can affect the apparent unidimensionality of a scale, as well as artificially inflating the overall scale score and the reliability statistics.

4.5.7 Differential Item Functioning (DIF)

Within the framework of Rasch measurement, the scale should work in the same way, irrespective of which group is being assessed. Thus, the probability of a person affirming an item – at a given level of ability – should be the same for younger or older persons, males and females, and so on. This type of analysis is given the name Differential Item Functioning (DIF). The basis of the DIF approach lies in the proportion of individuals at the *same* ability level who correctly answer a question. Under the requirement that the ability under consideration is unidimensional, if the question measures the same ability across groups then, except for random variations, the same response pattern is found irrespective of the nature of the group for whom a function is plotted. Questions that do not yield the same pattern for two or more groups display DIF and are violating the requirement of invariance. (197, 198)

In graphical terms, when you plot the results for different groups, they should fall roughly on top of each other, as in the example shown in Figure 20. For an item displaying DIF, a degree of separation will be apparent between the plots, as shown in Figure 21. The formal test of whether an item is displaying DIF is done via an ANOVA, although the graphical output is also available for each item.

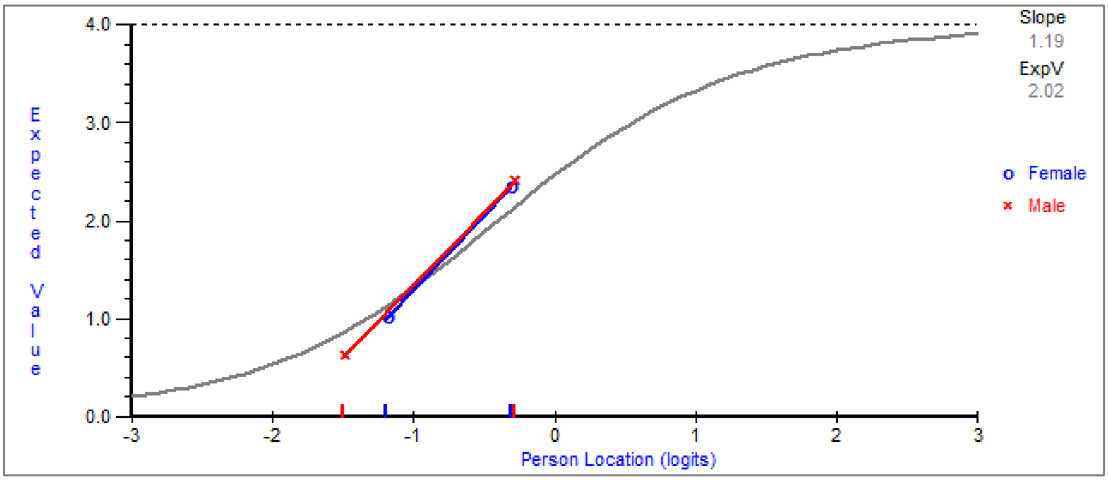

Figure 20. Example of item with no DIF

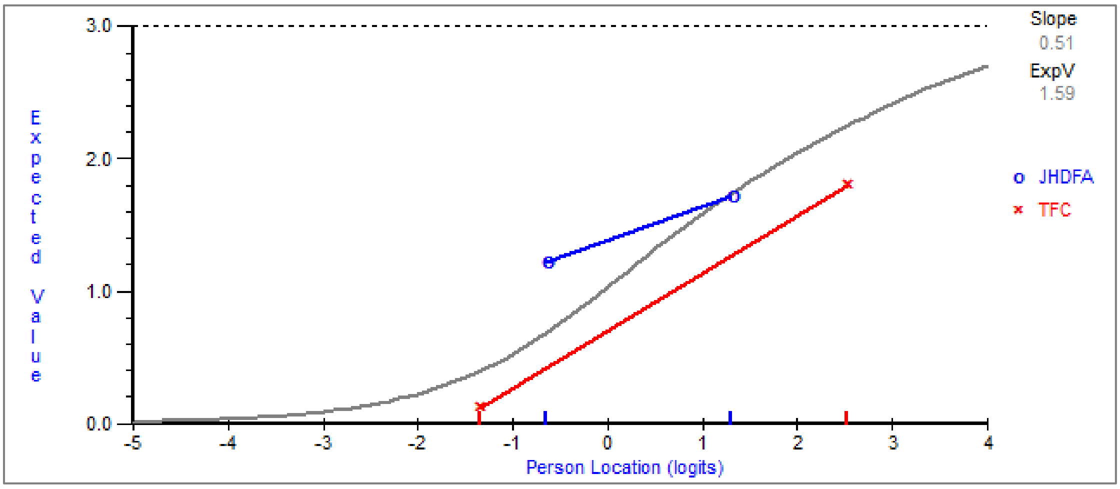

Figure 21. Example of item displaying DIF

Within this study, the DIF variables tested for group invariance were: Prison, Gender, Age Group (up to 29 Vs 30 plus), Remand Status (on remand Vs sentenced), Age left full-time education (less than age 16 Vs 16 years plus) and Religion (whether prisoner stated that they practiced a religion or not). It should be noted that if group sizes become too small, then this is likely to mean that the statistical DIF tests are underpowered, and it was due to this fact that no test of ethnicity DIF was carried out.

4.5.8 Unidimensionality

The Rasch model is a unidimensional measurement model. Unidimensionality is a key concept in the measurement of anything, in that only one thing should ever be measured at one time. (182, 206)

Rasch analysis is a confirmatory procedure, which has the assumption that all of the items belong to a single unidimensional structure. The analysis process will then highlight anomalies within this assumed unidimensional framework. Additionally, a post-hoc test of unidimensionality is also available, following the recommendations by Smith. (207) Within this procedure, independent sets of items are used to generate two estimates for every individual, which are then compared by a t-test.

If any conceptual grouping of items is present within a measure (e.g. into similar, but perhaps overlapping domains), then unidimensionality assessment should be made on this conceptual basis. (182) If no conceptual grouping is present within an item set, then the unidimensionality can be assessed by looking for patterns in the residuals. This process will identify the sets of items within the scale that are the 'most different' from each other. Two separate person ability (location) estimates are then generated based on these two subsets of items. If the scale is unidimensional, then there should be no difference between the estimates that have been generated from the two subsets of items. The two separate person ability estimates are compared for each individual in the analysis via a series of t-tests. The number of significant t-tests should be below 5% of the sample tested (or the lower bound of the binomial confidence interval for proportions should be less than 5%) in order to confirm the unidimensionality of the scale. (182, 197, 198)

4.5.9 Targeting

Targeting is the name given to the relative person and item distributions when they are placed upon the same underlying scale. Ideally, the item and person locations should have an approximately equal distribution, as this is when the information from the scale and population is maximised. An example of a well-targeted scale is given in Figure 22.

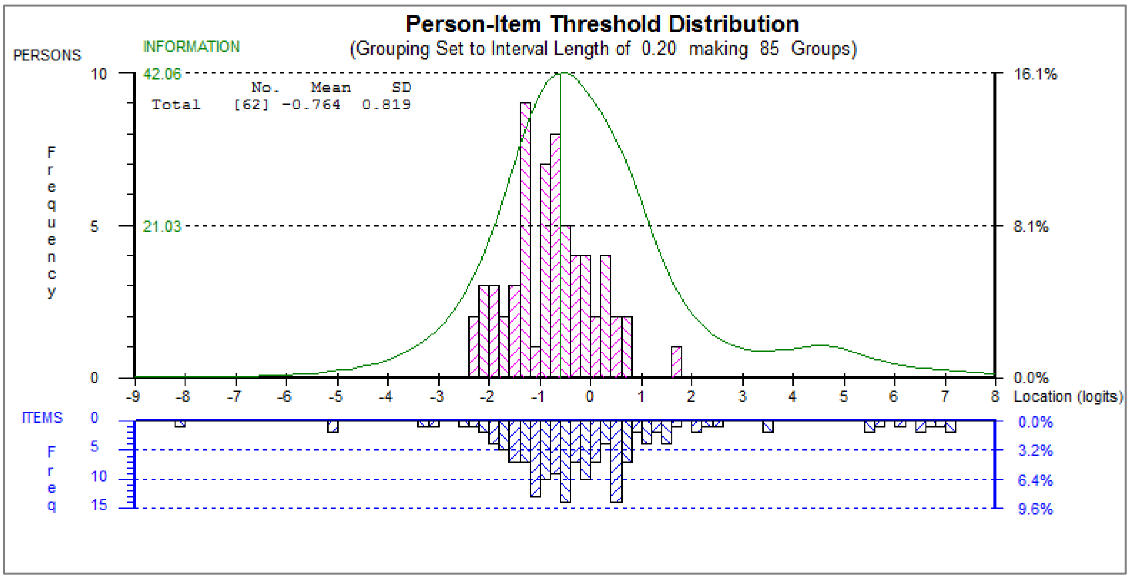

Figure 22. Example of a person-item threshold distribution, otherwise known as a targeting plot

Individual persons that score minimally or maximally upon the scale (traditionally, at the floor or ceiling) are known as extreme cases. These persons fall outside the measurement range of the scale in question, and are therefore removed from the calibration equations, although logit location estimates are extrapolated for these persons.

Included in the targeting plots are the test information curves (the green lines). These curves show the distribution of the information that is obtained from the item set. This should ideally be a single-peaked curve, which would typically be based around the mid-point of the scale. If the distribution of the person-sample generally falls under this information curve, then the item set is well-targeted to the population, and the full measurement continuum is represented. When this is the case, a high person-separation reliability index will also generally be found.

In the case of a poorly-targeted scale, the person-sample distribution will be skewed to the left or right of the information curve, meaning that the information extraction of the scale is not optimal. If this type of distribution is observed, and the intention is to measure this population across the full range of the trait that is being measured, then it suggests that it would be beneficial to introduce additional items to the item set. These additional items would have to be developed in order to specifically target people at either the floor or the ceiling of the scale, depending on which way the skew is observed and the scoring direction of the items. If the skew is showing that a lot of people are scoring maximally, then new items would need to be introduced that were more difficult to endorse. If the skew is showing that a lot of people are scoring minimally, then new items would need to be introduced that were easier to endorse. In both scenarios, this would introduce more separation (and information) at the skewed end of the scale, which would also have the effect of shifting the skew and the information curves to a more central, better targeted position.

This type of skewed distribution is often found among screening tools, meaning that measurement across the continuum is not optimal. However, the primary function of a screening tool is to separate a population into risk groups at a criterion cut-point, rather than to measure individuals incrementally across a continuum, so this type of skewed distribution would perhaps be advantageous in this context. (208)

4.5.10 Rasch Sample Size Considerations

The purposes for carrying out a Rasch analysis are variable, and therefore the appropriate sample size will also vary depending upon the purpose. However, published guidelines suggest that a sample size of approximately 250 cases will give 99% confidence that item calibrations will be stable to within half a logit (even under the condition of poor targeting), and that a sample size of 500 cases is robust. (165)

4.5.11 Rasch Missing Data Considerations

Rasch analysis is not dependent on imputing missing data, as the estimates are only generated on the data that are present. Therefore the whole dataset can be used without eliminating participants or imputing response values, and this will not affect the model or the derived fit statistics, although standard errors of measurement will be larger where missing data are present.

4.5.12 Rasch Analytic Strategy

Although the same indicators were assessed throughout the analytical process, the Rasch analytic strategy was progressed in alternative ways, with each strategy offering a different resolution to any issues that may appear within a set of items. These alternative strategies are summarised in the form of Resolutions A & B:

4.5.12.1 Resolution A

Where misfit anomalies were found, attempts were made to account for the misfit that had been highlighted. In the case of response dependency, where the apparent dependency has a conceptual basis, this can be accounted for by subtesting the related items. This effectively groups the dependent items into one 'testlet', meaning that the total raw score derived from the items does not change, but the dependent relationship between the items has been eliminated. (209)

In the case of DIF, an 'item-split' can be carried out which affectively creates a new item specific to each selected factor grouping. For example, if an item displays a DIF-by-gender, then to split this item by gender would result in two new items – one specific to males and one specific to females. Split items remain anchored to the common set of items, but the logit location (item difficulty estimate) will be independent for each split-item.

These amendments are post-hoc adjustments of the apparent misfit, which will account for the effects of the misfit within the constraints of a particular analysis. Therefore, the person logit estimates will be comparable within this particular analysis whilst maintaining as many of the original scale items as possible. However, it should be pointed out that these post-hoc adjustments do not account for the problems that are inherent to a scale when applied to this particular population.

Resolution A sought to maintain as many original scale items as possible by making the appropriate amendments to account for response dependency and DIF. Where amendments could not be made to account for the source of misfit, individual items were removed from the item set.

4.5.12.2 Resolution B

A second approach was to remove misfitting items iteratively, to try and obtain a set of items which satisfied all fit parameters. When all individual misfit anomalies had been removed, this provides a 'pure' item set on which to base comparable person estimates. When adequate fit statistics were displayed by the 'pure' item set, the removed items were individually re-introduced back into the pure set to see whether the original source of misfit was still apparent. If the source of misfit was still present within the refined item set, then the item would again be removed. If, however, the original source of misfit was no longer apparent, then the item would be marked for re-introduction back into the final item set.

Resolution B sought to find a set of items, free from any form of significant individual or collective misfit, which act together to form a unidimensional scale.

4.6 Area Under the Curve (AUC) Analysis

The premise of screening was introduced in Section 2.1, where it was stated that in an ideal situation a screening test would correctly identify and classify all true cases and all individuals who are not at risk. When evaluating a screening test, certain terms are used to describe particular features of the test; these are sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

Sensitivity refers to the ability of a test to correctly identify true positives (i.e. those that DO have the health condition of interest). Alternatively, in the context of this study; the proportion (or %) of people who DO self-harm, whom the test correctly identifies as SELF-HARMING. (12, 210)

Specificity refers to the ability of a test to correctly identify true negatives (i.e. those that DO NOT have the health condition of interest). Alternatively, in the context of this study; the proportion (or %) of people who DO NOT self-harm, whom the test correctly identifies as NOT SELF-HARMING. (12, 210)

Both sensitivity and specificity are characteristics of the specific test, and they are independent of the population that is being subjected to the test. (12)

Positive predictive value (PPV) refers to the proportion (or %) of people that the test identifies as having the condition, who do truly have it. Alternatively, in the context of this study it answers the question of 'how likely is it that this prisoner will self-harm given that the test says that they will self-harm?' (12, 210)

Negative predictive value (NPV) refers to the proportion (or %) of people that the test identifies as not having the condition, who do truly do not have it. Alternatively, in the context of this study it answers the question of 'how likely is it that this prisoner will not self-harm given that the test result says that they will not self-harm?' (12, 210)

The positive and negative predictive values are characteristics of the test that are variable depending on the prevalence of the condition in the population of interest. (12)

The sensitivity, specificity, PPV and NPV are all calculated using a classic two-by-two table, as depicted in Table 16.

Table 16. Results of a screening test

		Self-Harm		
		Yes	No	Total
Result of screening test	Will Self-Harm	a	b	a + b
	Will not Self-Harm	c	d	c + d
	Total	a + c	b + d	a + b + c + d

a = True positive: the individual does self-harm and the test says that they will self-harm.

b = False positive: the individual does not self-harm but the test says that they will self-harm.

c = False negative: the individual does self-harm but the test says that they will not self-harm.

d = True negative: the individual does not self-harm and the test says that they will not self-harm.

Sensitivity = $\frac{a}{a+c}$

Specificity = $\frac{d}{b+d}$

Positive Predictive Value = $\frac{a}{a+b}$

Negative Predictive Value = $\frac{d}{c+d}$

(211)

The evaluation of any screening instrument tends to involve a trade-off between sensitivity and specificity, which is done by manipulating the cut-off score that is used to identify a case. (5) In the context of this study, the accuracy of a predictive test depends on how well the test separates the group subsequently self-harming from those who do not self-harm. The optimum cut-off point of a test can be determined by plotting each cut-off point onto a receiver operating characteristic (ROC) curve; which is a plot of sensitivity (the true positive rate) on the vertical axis, against 1 minus the specificity (the false positive rate) on the horizontal axis. (210) The optimal cut-off point is the point which maximises the area under the curve (AUC), which is the point which lies the closest to the top left corner of the curve (Figure 23). In practice, the AUC value is calculated for each cut-off point and the optimal cut-off point is taken as the point at which the AUC value is maximised. (210) An AUC value of 1 represents a perfect test; an AUC of 0.5 represents a worthless test (represented by the dotted line on Figure 23. A rule of thumb to the magnitude of the AUC is summarised in Table 17. (212) All AUC analyses were carried out using SPSS (version 21) software package. (213)

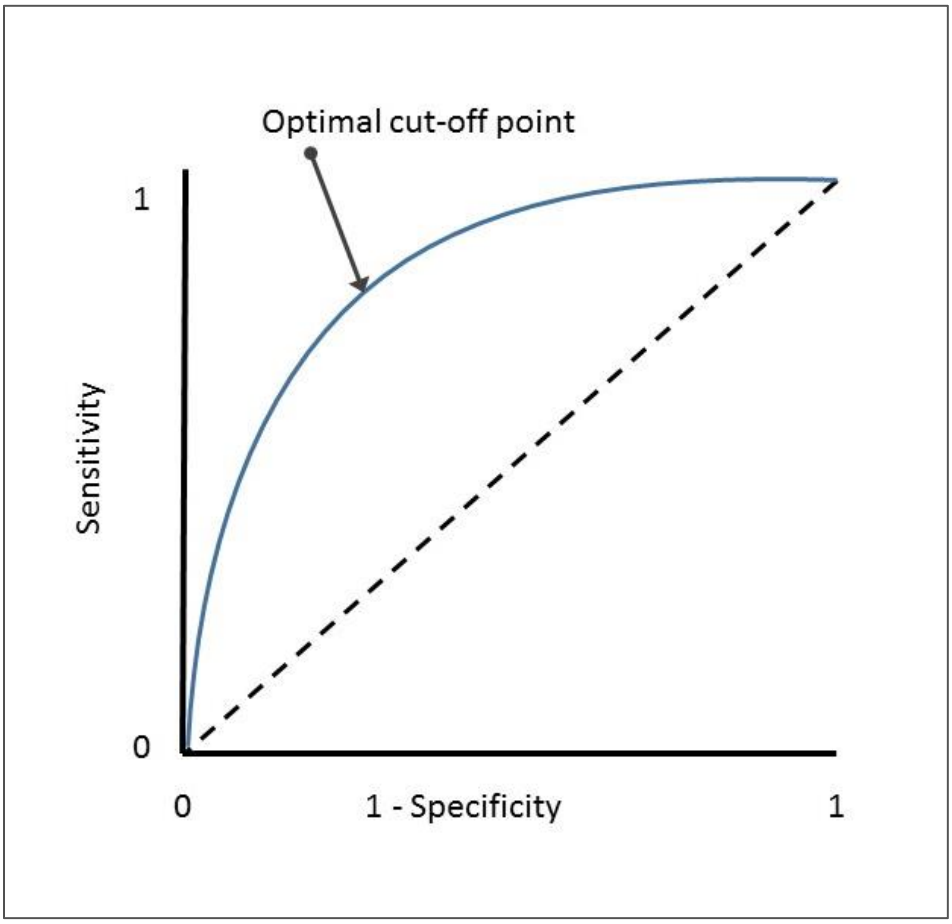

Figure 23. A receiver operating characteristic (ROC) curve

Table 17. Interpretation of AUC values

AUC Value	Judgment
.90-1	Excellent
.80-.90	Good
.70-.80	Fair
.60-.70	Poor
.50-.60	Fail

4.7 Predictive Item Set

The 105 items in the candidate instruments also formed an item pool of potential risk indicators, together with the other items regarding socio-demographic and sentencing criteria which were present in the questionnaire pack.

In the case that none of the selected complete instruments operate sufficiently to adequately predict the occurrence of self-harm with the ACCT population, then it was planned that all of the individual items would be assessed in terms of their individual predictive value. This exploratory analysis will identify items which could be considered to combine to form a predictive algorithm. From this analysis, those items that were individually associated with future SH at p=0.10 (as indicated by crosstab chi-square tests) were refined through a logistic regression procedure in order to identify the algorithm that maximised the predictive AUC value. Within this process, all items that were individually identified as potential predictors were entered into a backwards stepwise binary logistic regression, under a likelihood-ratio removal process (p removal 0.1), as recommended by Field. (214) All statistical analyses were carried out using SPSS 21. (213)

4.8 Structural Equation Modelling (SEM)

As described in Section 2.2, self-harm is a complex issue, and it can be the result of a wide range of influencing factors, which can be both internal and external in nature. The complex interaction of these factors has produced many explanatory models of the resultant self-harm behaviour (see Section 2.2.2). The primary aim of the study is to assess whether any pre-existing, standardised scale is predictive of self-harm, but the resulting data set from this study allows for the investigation into an explanatory model of self-harm among prisoners that are on an ACCT. Although this will not help to achieve the goal of introducing a standardised risk assessment tool into the ACCT process, if an explanatory model can be confirmed, then this could be informative for care pathways and future interventional studies.

4.8.1 Overview of Structural Equation Modelling

Structural Equation Modelling (SEM) builds on the foundations of correlation models and multiple regression approaches, to offer a more complete view of how a number of variables interact with each other with regard to a focal relationship.

4.8.1.1 Correlation

In order to explore a relationship between two variables, a correlation analysis between the two variables can be performed. This will provide an indicator of the strength of the association between the variables. However, this type of analysis is restricted in that there is no direction (causality) attached to the association. Although a correlation matrix may be created in order to assess the pairwise associations between a number of variables, only one association (between two variables) can be assessed at any one time, meaning that more complex multi-variate relationships cannot be evaluated through this methodology. (214)

4.8.1.2 Regression

To evaluate the impact of a number of independent variables on a single dependent variable, regression analysis can be performed. This can be a linear regression if the variables are continuous in nature, or a logistic regression if the dependent variable is categorical. This type of analysis is very useful in ascertaining the impact that each individual independent variable has upon the dependent variable, when all variables are considered together in the model. Additionally, regression analysis provides an index (R^2) of the amount of the dependent variable that is explained by the combined impact of the independent variables. However, a restriction of regression analysis is that the direction of the relationship is fixed upon the dependent variable in the model, of which there can only be one. In this sense, what is modelled is the direct impact of each independent variable upon the dependent variable, but not the impact of each of the independent variables upon each other. (214)

4.8.1.3 Structural Equation Modelling (SEM)

Structural Equation Modelling is a statistical methodology that is used for specifying and estimating models of linear relationships among variables. (215) It can therefore be used to assess the directional relationships between a number of independent and dependent variables. A key component of SEM is that it takes a confirmatory approach to the analysis of a structural theory; therefore providing a comprehensive method for the testing and quantification of substantive theories. (216, 217) Within an SEM framework, any directly observed variables can be utilised as they would be in other regression models. However, unlike other regression models, latent (unobserved) variables can also be incorporated into SEM through a measurement model component that also assesses the individual indicators (items) of the implied latent construct. Another element of SEM that differs from other (more traditional) multivariate procedures is that measurement error is incorporated into the models, where it can be assessed and taken into account. An additional benefit of SEM is that it is the only easily applied methodology that also models the indirect effects of relationships, along with the direct effects. (216)

SEM consists of two main elements: the Factor Analysis (measurement) element, which concerns the individual items that represent an underlying latent variable; and the Path Analysis element, which concerns the interaction and relationships between the independent and dependent (observed or latent) variables in the model. A Full Structural Model would include both of these elements, but this study is primarily concerned with the path analysis element of SEM, as the measurement element has already been dealt with through the Rasch analysis. All latent variables (from the instruments included in the study) have been assessed, amended as necessary, and converted into interval-equivalent measures through the Rasch analytic process.

4.8.1.4 Mediation

SEM assesses the relationships between variables in a model, and some of these variables within the model may act as mediators. In general terms, a variable is said to function as a mediator to the extent that it accounts for the relation between the predictor and the dependent variable. (218) Mediators often represent properties of a person that transform the predictor variables in some way. (218)

In simple terms, when considering a third variable (labelled C in Figure 24) in the relationship between an independent (predictor) variable (A) and a dependent (outcome) variable (B), a mediator variable is one that can be said to explain the relationship between the two other variables. A variable functions as a mediator when variations in the levels of an independent variable account for variations in the assumed mediator (Path y), and then the variations in this mediator account for variations in the dependent variable (Path z). In the case of full mediation, then when these two relationships are controlled, a previously significant relation between the independent and dependent variables (Path x) will no longer be significant. (218) In other words, although initially it may appear that A directly influences B, in reality A directly influences C (the mediator variable), and, in turn, C directly influences B. In the case of full mediation, when paths y and z are in place, A actually has no (significant) direct on influence on B, even though that relationship was initially thought to be present.

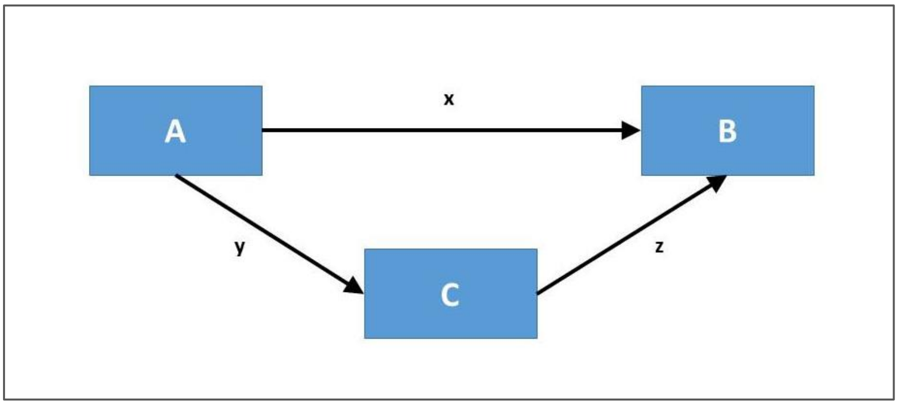

Figure 24. Path diagram of a mediating relationship

Although the terms are often used interchangeably, mediation should not be confused with moderation, which represents a different kind of relationship. In general terms, a moderator is a variable (either qualitative or quantitative) that affects the direction and/or strength of the relationship between an independent or predictor variable and a dependent or criterion variable. (218) Moderation implies that the causal relation between two variables changes as a function of the moderator variable. Moderator variables always function as independent variables, whereas mediating events shift roles from effects to causes, depending on the focus of the analysis.

In simple terms, when considering a third variable (C) in the relationship between an independent (predictor) variable (A) and a dependent (outcome) variable (B), a moderator variable is one that influences the strength of a relationship between two other variables (Figure 25), and a mediator variable is one that explains the relationship between the two other variables (Figure 24).

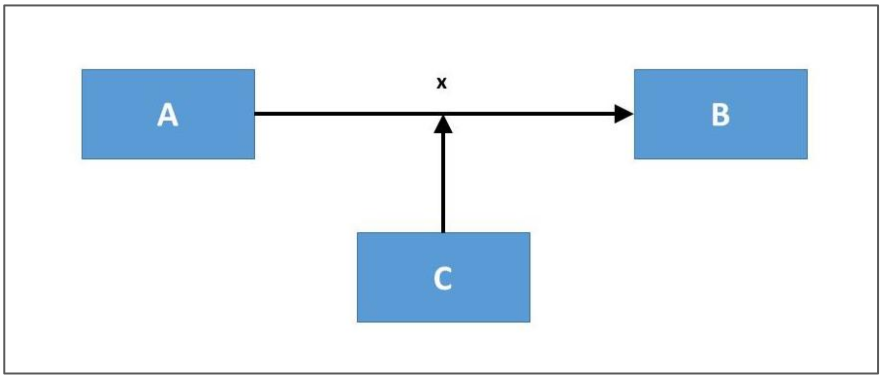

Figure 25. Path diagram of a moderating relationship

All path analysis relationships are summarised are summarised in Figure 26, which has been adapted from Wang et al. (219) It can be seen that if two independent variables both have direct effects on the outcome variable, but are correlated with each other, then this would be seen as a confounding relationship as it is not known which independent variable is affecting the outcome variable.

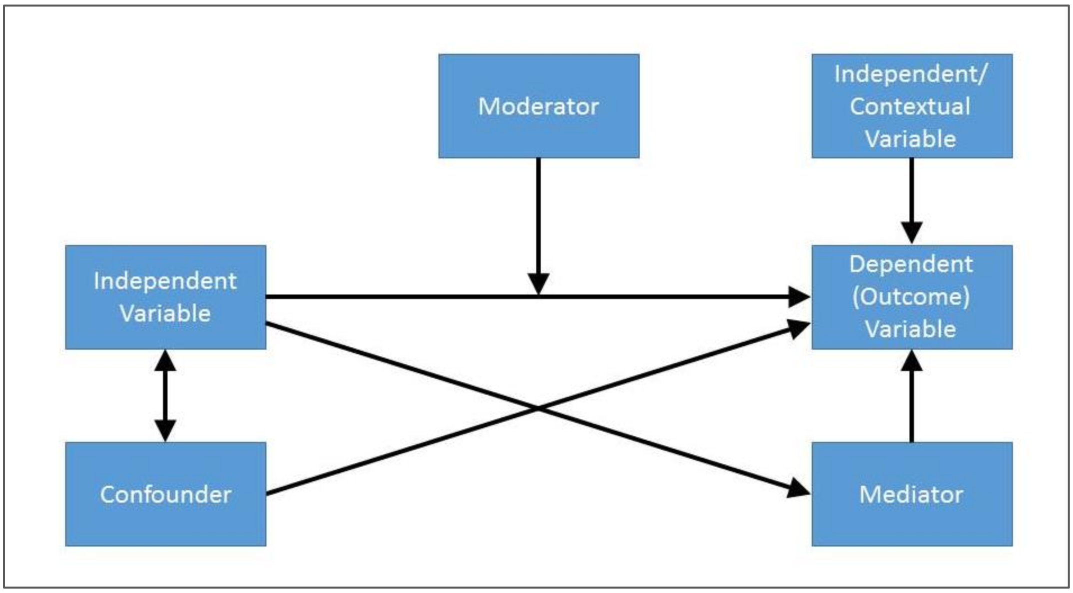

Figure 26. Path analysis relationships between variables

4.8.2 SEM Graphical Conventions

Although structural equation models can be specified non-graphically, it is usual for them to be represented graphically, thus providing an easier way to visualise the relationships and interaction between the variables. The models are graphically portrayed using configurations of four geometric symbols: an oval (or circle), which represents a latent variable; a rectangle (or square), which represents a directly observed variable; a single-headed arrow, which represent the impact of one variable upon another; and a double-headed arrow, which represents covariance or correlations between pairs of variables. (216)

4.8.3 SEM Process

As previously stated, SEM is considered to be a confirmatory procedure, but this covers three different scenarios. (220) A *strictly confirmatory* strategy would involve a theory-led model to be proposed, data to be collected, and then the hypothesised model to be tested. In this scenario, the proposed model is accepted or rejected without any further modification. An *alternative models* strategy involves the proposition of several alternative theory-led models, which are each tested against a single data set. The most appropriate model is then selected as the one which best represents the data, again, with no further modification. A *model-generating* strategy involves the testing of a theoretically derived model, which is then rejected on the basis of poor fit to the sample data. Following this, the researcher proceeds in an exploratory way to modify and re-estimate the model, with the primary focus being to eradicate the sources of misfit to specify a model which best characterises the data. Model re-specification can be either theory or data-driven, with the ultimate objective being to find a model that makes conceptual sense, as well as being statistically well-fitting. (220)

Although it is not the purest form of SEM, it has been identified (216) that the model-generating strategy is the most commonly used, with modification specification searches becoming a common practice within the process of SEM. (217) However, it must be emphasised that any model modification must also be justified on a conceptual basis, and be consistent with pre-existing theories or results of previous research. (217)

4.8.4 Fit Indices

Within SEM there are a number of statistics relating to model testing and evaluating fit, but no decision on goodness-of-fit should be based on one single indicator, regardless of how favourable that single indicator may be. (217) Each indicator of fit represents something slightly different, and therefore any decision to reject or retain a model should be based on multiple fit indices. (217)

The fit indices used in the SEM component of this study were the Chi-square goodness-of-fit statistic, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and the Weighted Root Mean Square Residual (WRMR). These are summarised in Table 18.

Table 18. Summary of fit indices used in structural equation modelling

Statistic	Abbreviation	Interpretation	Ideal Fit Criteria	Acceptable Fit Criteria
Chi-square fit statistic	χ^2	Tests the null hypothesis that the specified model fits the corresponding population matrix perfectly. χ^2 p-value should be non-significant to indicate model fit	p>0.05	
Root Mean Square Error of Approximation	RMSEA	Measures the extent to which the specified model is an approximation of the analysed data.	RMSEA<0.05	RMSEA<0.08
Comparative Fit Index	CFI	Compares the specified model with the null model assumption of no relationship between variables	CFI>0.95	CFI>0.90
Tucker Lewis Index	TLI	Tests the difference between the chi-square value of the specified model with the chi-square value of the null (no relationship) model, adjusted for model complexity	TLI>0.95	TLI>0.90
Weighted Root Mean Square Residual	WRMR	Alternative fit index recommended for models with categorical observed variables	WRMR<0.9	WRMR<1.0

4.8.4.1 Chi-Square

The Chi-square (χ^2) fit statistic tests the null hypothesis that the specified model fits the corresponding population matrix perfectly. The χ^2 p-value should be non-significant to indicate model fit. This is an inferential fit index, meaning the result is generalisable to the studied population. However, the χ^2 index and its p-value are vulnerable to variations in sample size, meaning that as a stand-alone statistic it cannot be fully trusted as a means of model evaluation. (217)

4.8.4.2 Root Mean Square Error of Approximation

The Root Mean Square Error of Approximation (RMSEA) offers an alternative perspective to the χ^2 test. Whereas the χ^2 test starts from the default position of assuming that the model is perfect and tests whether there is any deviation from this, the RMSEA assumes that no model is 'true', and therefore this test evaluates the extent to which the model fails to fit the data. The RMSEA is robust for large sample sizes, but can be misleading in the case of very small sample sizes. It has been suggested that an RMSEA value of less than 0.05 is indicative of good fit, and that values of 0.08 represents a reasonable approximation to the data that have been analysed. (221)

4.8.4.3 Comparative Fit Index

The Comparative Fit Index (CFI) compares the specified model with the null model assumption of ‘no relationship’ between variables. (222) The CFI reflects the ratio of improvement when moving from the null model to the specified model, and CFI values close to 1 are generally considered to indicate well-fitting models. However, there are no strict norms for the CFI, and therefore there are no fixed upper or lower indicators of models that could be deemed as ‘good’ or ‘poor’ respectively. It has been suggested that models provide a good approximation of the data if they demonstrate a CFI in the mid-nineties or above. (217)

4.8.4.4 Tucker-Lewis Index

The Tucker-Lewis Index (TLI) is also known as the non-normed fit index (NNFI). Again, the NNFI is based on the comparison between the specified model and the null model assumption of ‘no relationship’ between variables. (223) This is computed through the comparison of the chi-square value of the null model and the chi-square value of the specified model, which gives an idea of how good the specified model is, compared to the 'worst' (null) model. This is an adjusted index, which takes the complexity of the model into account through the associated degrees of freedom of the specified model. As with the CFI, there are no strict norms for the TLI, and it has again been suggested that models provide a good approximation of the data if they demonstrate a TLI in the mid-nineties or above. (224)

4.8.4.5 Weighted Root Mean Square Residual

The weighted root mean square residual (WRMR) is an alternative fit index that has been recommended for models that contain categorical observed variables. WRMR measures the weighted average differences between the sample and population variances and covariances. (225) It has been suggested that a WRMR value of less than or equal to 1.0 is an indication of good fit. (226)

4.8.5 SEM Analysis Strategy

For all SEM analyses, the MPlus 7 (184) software was used, due to the fact that binary outcome variables (self-harm Vs no self-harm) are accommodated within this program. Additionally, a polychoric correlation matrix is used as the basis for any models containing ordinal variables. Although this is the most appropriate basis for models of this type, it is not available in all SEM software packages.

Prior to entry into MPlus 7, all variables entered into the SEM procedure were tested for multicollinearity using SPSS 21, (213) and all independent variables satisfied the requirement of having a variance inflation factor (VIF) value below 10, and of having a tolerance (1/VIF) value greater than 0.2. (214)

The data were then read into the SEM software, with variables stated as categorical where appropriate. When utilising the outcomes from the study instruments (CORE-OM subscales, BSL-23-F, PHQ-9, PriSnQuest, SHI), the Rasch-transformed 'Resolution B' (See Section 4.5.12.2) values were used as single-value outcomes to represent the latent constructs. Resolution B is derived from a 'pure' set of items, meaning that no additional modelling is necessary to extract a valid summed score. This option was used in order to try and maintain the practicality of the outcomes. As Rasch analysis provides individual person estimates even when missing data points are present, using the Rasch-transformed values maximises the sample size (n) that can be utilised within the SEM as cases will not be removed from the analysis on the basis of missing data.

Despite the strength of the conceptual basis of an initial underlying model, it is unlikely that a good model fit will be obtained at the first time of asking, and therefore SEM becomes an iterative process. A number of different models may be formulated and tested individually to see whether they are accepted or rejected. Alternatively or additionally, models may be amended iteratively, based on adjustments that are suggested through the tests-of-fit and modification indices provided by the software. Any modifications to the model that are made on this basis should also make conceptual sense. The newly specified model is then retested and the modification process may be repeated until a parsimonious model is achieved.

5 Results of Cohort Study

This chapter presents the descriptive results of the cohort study, including details of study recruitment, participant characteristics and the self-harm incidence that was observed within the study cohort.

5.1 Recruitment

The cohort study began recruitment in May 2011, and concluded in May 2012, followed by the six month follow-up, which meant that the study data collection ranged from May 2011 until the end of November 2012. Prisoners recruited to the pilot study were not included in the cohort study sample. Three prisons were included in the study, and the basic descriptive information of these prisons is presented in Table 19.

Table 19. Basic descriptive information of recruiting prisons

Prison	Gender	Era built	Category	Capacity	Wings	Location	Type
A	Male	Early 19th century	B	1001	7	North of England	Remand
B	Female	Mid-20th century	Closed	336	8	North of England	Regional
C	Male	Mid-19th century	B	1212	6	North of England	Local

A flow chart of the total recruitment is given in Figure 27. During the recruitment period 590 prisoners were eligible for inclusion, of which 452 (76.6%) consented (see Table 20). Two prisoners subsequently withdrew making the baseline sample 450. Recruitment rate was similar across prisons, ranging from 70.7%-79.0%.

Table 20. Participation consent rate, presented for individual prisons

	Prison A	Prison B	Prison C	Total
Approached	135	164	291	590
Refused participation	29	48	61	138
Consented	106	116	230	452
% Consented	78.5%	70.7%	79.0%	76.6%
Withdrew from study	1	1	0	2
Total included	105	115	230	450

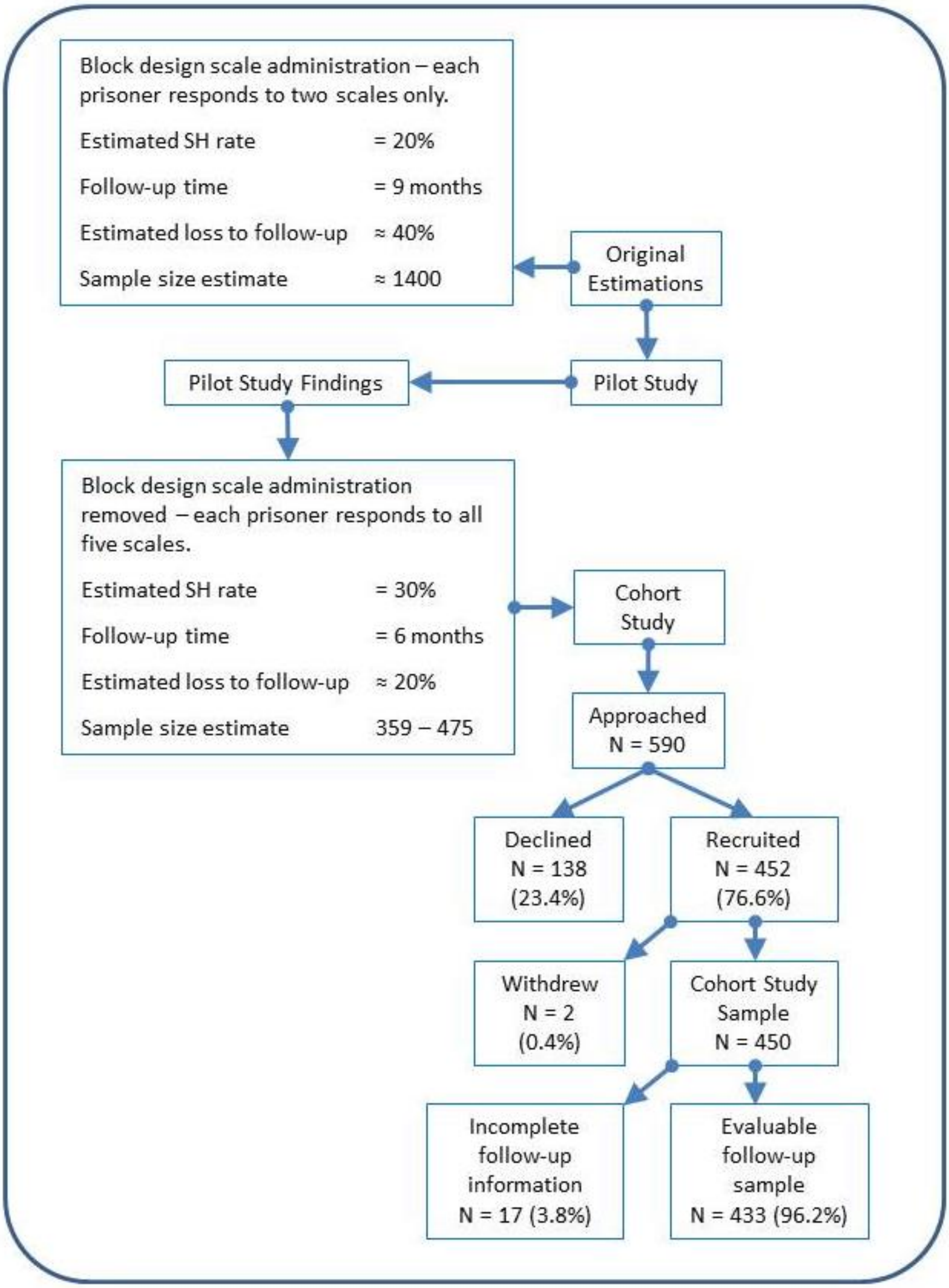

Figure 27. Flowchart of study sample recruitment

Of the 450 participants recruited, full evaluable follow-up information was available for 433 (96.2%). Of the 17 records with incomplete follow-up data, 1 participant was lost to follow-up, 12 participants had inaccessible records and 4 participants had full follow-up but with inconclusive information in their records.

5.2 Study Participant Characteristics

The mean age of the 450 participants consenting to the study was 31.2 years, which did not vary across the three prisons (see Table 21). On average, they left full time education at 15 years old, with over two-fifths leaving without qualifications of any sort. However, this varied by prison, with twice as many without qualifications in one male prison compared with the other. Almost half of participants (49.4%) had children, but only one-in-seven (14.3%) reported receiving a visit during the past seven days.

As an indicator of the prevalence of mental health issues within the sample, the three mental health screening items from the prison reception screening tool (139) were summed into a composite item. Only 4.9% (n=22) of the sample responded negatively to all three questions, with 94.4% (n=425) affirming one or more of the items (1=16.2%, 2=32.2%, 3=46%).

Table 21. Demographic and sentence characteristics of participants recruited. Significance displayed across prisons

Characteristic	Prison A	Prison B	Prison C	Total	Significance*	N
Mean Age (Years)	31.2	29.6	32	31.2	0.102	450
Age leaving full-time education	15.3	15.5	15.3	15.3	0.896	440
% without any qualifications	26.7	36.8	55.3	43.8	<0.001	447
Have Children (%)	51.4	44.3	51.1	49.4	0.447	449
Received visit in last 7 days (%)	15.2	14.8	13.6	14.3	0.858	448
% on remand	56.2	22.6	52.2	45.6	<0.001	245
Of those sentenced **- Tariff in months**	53.8	44.6	32.1	41	0.394	225
- Served	9.8	17.2	14.8	14.7	0.388	239
N	105	115	230	450		

The prisons differed in their function, with the male prisons also being remand facilities. Consequently the proportion on remand differed considerably with just over half the participants on remand in the male prisons, compared with just over one fifth (22.6%) in the female prison. The average tariff of those sentenced was 41 months, of which 14.7 months had been served.

The median time to interview from initiation of the ACCT was 6 days (see Table 22). This differed between the male (A & C) and female (B) prisons, with females being interviewed somewhat later, with a median of 8 days compared to 5 days in the male prisons.

Table 22. Number of days between index ACCT being opened and interview being carried out

	Prison A	Prison B	Prison C	Total
N	105	115	230	450
Mean	6.07	8.96	4.96	6.24
Std. Deviation	3.693	5.287	3.201	4.268
Median	5	8	5	6
IQR	3 - 9	6 - 12	3 - 7	3 - 8
Minimum	1	0	0	0
Maximum	16	30	18	30

5.3 Follow-up Time

The time included in the follow-up period was variable, with the aim being to complete a 6-month follow-up period. In some cases this was not possible due to the prisoner being released, but in some cases the records allowed for a longer follow-up time. Where a longer follow-up was possible, the information has been included for the full follow-up period. However, for the predictive element of the study, the follow-up period was restricted to 198 days (6.5 months). Only one person reported their first self-harm event after this cut-off point. During follow-up, 126 people actually carried out a self-harm event, but only 125 of these were within the valid time frame.

5.4 Incidence of Self-Harm

Recorded across all prisons, the basic self-harm incidence during follow-up was 29.1%, although this varied between a minimum of 16.67% (Prison A) and 33.33% (Prison B). During the follow up period a total of 423 self-harm events were reported, based upon 126 individuals, followed up for 66,789 prisoner days. This gives an 'event incidence' of 6.33 per 1,000 prisoner days of those who had been placed upon an ACCT, or 'prisoner incidence' of 1.84 per 1,000 days. For example, if 20% of the current prison establishment had previously been on an ACCT, then with a prison housing 1000 inmates, one self-harm act per day could be expected. However, this is only the average from the current study, and it is noticeable that this varies considerably by gender (see Table 23), and to a lesser extent, between prisons. Thus, the event incidence rate in the female prison is much higher, at 15.83 per 1000 prisoner days, as opposed to the male event average of 4.02 per 1000. Looking at persons rather than events, there is a clear gradient across prisons, increasing from the person incidence of 1.26 per 1000 prisoner days in the male Prison A, through to the much higher person incidence of 2.83 in the female prison B. There is also a marked difference here between males and females, with the male incidence at 1.66 against the female rate of 2.83.

Table 23. Incidence of self-harm in follow-up separated by prison and gender

	Prison A	Prison B (Female)	Prison C	Total	Male Prisons
N	105	115	230	450	335
N with valid follow up	102	111	220	433	322
Total number of self-harm events reported during follow-up	50	207	166	423	216
Total number of prisoner follow-up days	13470	13074	40245	66789	53715
Event incidence per 1000 prisoner-days	3.71	15.83	4.12	6.33	4.02
Total number of people with self-harm events reported during follow-up	17	37	72	126	89
Person self-harm incidence per 1000 prisoner-days	1.26	2.83	1.79	1.89	1.66
Self-Harm event/person ratio	2.94	5.59	2.31	3.36	2.43
Basic self-harm incidence during follow-up	16.67%	33.33%	32.73%	29.1%	27.64%

The ratio of persons to events is also different across prisons, with the female ratio more than twice that of the male ratio. The frequency of events is shown in more detail in Figure 28.

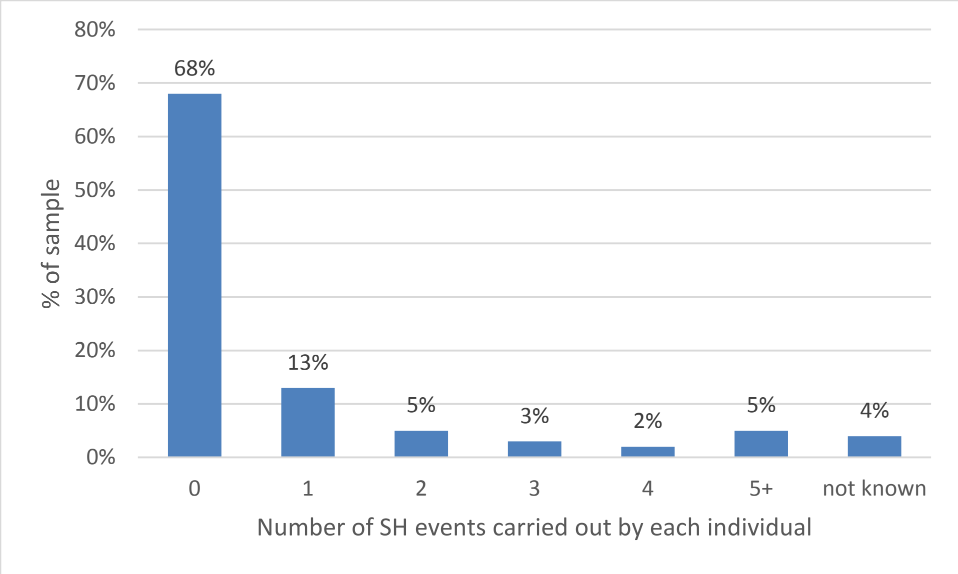

Figure 28. The amount of self-harm events carried out by each individual during follow-up, presented as a percentage of the full cohort study sample (n=450)

The median time to a first follow-up self-harm event was 37 days, with a range of 0-190 days for those within the valid follow-up time (n=125), and 0-245 days for all those recorded with self-harm (n=126) (see Figure 29). The conditional probability of an ACCT Index self-harm event, given previously reported self-harm was 0.33; of subsequent self-harm (i.e. during follow up) given reported previous self-harm was 0.28; and of subsequent self-harm, given a known self-harm ACCT Index event, was 0.47. Additional detail about the nature of the first self-harm events in follow-up is given in Tables 24 and 25.

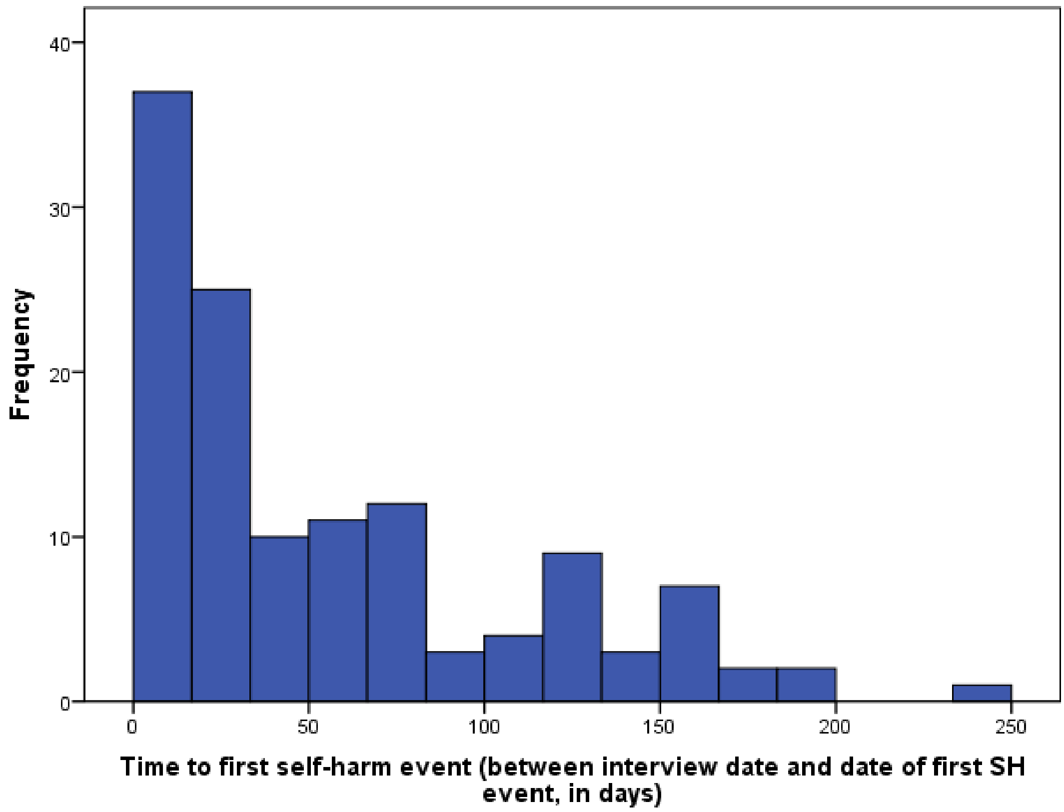

Figure 29.Time (in days) to first self-harm event (of those that self-harmed)

Table 24. Severity details of the first self-harm events during follow-up

Severity of first self-harm event	n (% of total)
Self-harm that was near lethal with intent to die	3 (2.4%)
Self-harm that was near lethal without intent to die	6 (4.8%)
Major (required medical attention at an off-site hospital)	3 (2.4%)
Moderate (required medical attention on-site)	38 (30.2%)
Minor (superficial)	55 (43.7%)
Not known	21 (16.7%)
Total	126 (100.0%)

Table 25. Types of first self-harm events reported during follow-up

Type of first self-harm event	n (% of total)
Cutting	64 (50.8%)
Unspecified Self-Harm	30 (23.8%)
Attempted Hanging / Ligatures / Self-Strangulation	8 (6.3%)
Self-Poisoning	8 (6.3%)
Scratching	2 (1.6%)
Self-Suffocation	2 (1.6%)
Hunger Strike	2 (1.6%)
Opening Old Wounds	2 (1.6%)
Punching things (wall/door etc.)	2 (1.6%)
Swallowed razor blade	2 (1.6%)
Head banging	1 (0.8%)
Biting self	1 (0.8%)
Burning self	1 (0.8%)
Setting fire to own cell	1 (0.8%)
Total	126 (100.0%)

5.5 Personal Characteristics Associated with Self-Harm

As was described in Section 2.2.3, various characteristics may be considered as potential risks or mediating factors for self-harm, and data regarding some of these characteristics was collected during the study. Here, it was reported that 42.2% of people practiced a religion, although this was much more predominant in one of the male prisons (prison C) (see Table 26). Over a third of the participants reported being homeless in the 12 months prior to prison, and 57.9% reported seeing a psychiatrist outside of prison. Almost three-quarters (74.4%) reported receiving medication for mental health problems, and results were almost equivalent for alcohol dependency and drug dependency, with around a third of participants affirming each. Almost four-in-five (78%) reported that they had self-harmed outside of prison, and over three-in-five (61.7%) within prison. Females were much more likely to have self-harmed in prison, but not so outside of prison. Just over four-in-five (82.1%) were recruited from their first ACCT during their current stay in prison, but females were much less likely than males to be on their first ACCT.

Table 26. Potential risk factors for self-harm. Significance reported across prison

Characteristic	Prison A	Prison B	Prison C	Total	Significance*	N
% who practice a religion	30.5	35.7	50.9	42.2	0.001	450
% Homeless in 12 months prior to prison	31.4	34.8	37.6	35.4	0.692	449
% Seen psychiatrist outside prison	62.5	60	54.8	57.9	0.369	447
% Received mental health medication	68.6	81.7	73.4	74.4	0.072	449
% Dependent upon alcohol	29.8	26.3	36.7	32.4	0.125	447
% Dependent upon drugs	29.5	31.3	35.4	33	0.52	449
% who have self-harmed outside prison	83.8	83.5	72.5	78	0.017	449
% who have self-harmed within prison	59	78.3	54.6	61.7	<0.001	449
First time on ACCT in current Tariff	82.7	60.5	92.6	82.1	<0.001	447

* F-test for continuous variables; Chi-Square for proportions

Given the frequency of reported previous self-harm was so high, it could be instructive to examine the type and number of behaviours that have been previously engaged. The Self-Harm Inventory is comprised of a series of different self-harm behaviours, where respondents report on their lifetime history of engaging in each specific behaviour. The affirmation of these behaviours range from 'Tortured self with self-defeating thoughts', reported by four-in-five (79.7%), through to 'Abused laxatives to hurt self' reported by just 5.2% (most of whom were female). Over three-quarters (77.9%) reported that they had attempted suicide sometime in the past, which showed a significant difference across prisons. One in five reported a suicide attempt within the last week (taken from the BSL Supplementary items), but this did not show any difference across prisons. More than half of the behaviours showed a significant difference in reported frequency across prisons, many of which (e.g. engaged in sexually abusive relationships), but not all, related to gender differences. On average, participants who had self-harmed reported nine behaviours, but there were significant differences in the number of behaviours reported, and the patterns of those behaviours. These previous self-harm behaviours are reported across the full sample (n=450) in Table 27, and for the group that carried out self-harm during follow-up (n=126) in Table 28.

Table 27. Reported lifetime self-harm behaviours across full sample (n=450). Significance reported across prison

		% affirming behaviour in lifetime				
SHI Item	**Self-Harm Behaviour**	**Prison A**	**Prison B**	**Prison C**	**Total**	**Significance***
20	Tortured self with self-defeating thoughts	68.3	76.5	86.5	79.7	0.001
18	Attempted Suicide	85.1	82.6	72.1	77.9	0.012
1	Overdosed	71.3	85.2	65.6	72.0	0.001
2	Cut self on Purpose	75.2	78.3	65.9	71.3	0.036
6	Abused Alcohol	68.3	68.7	65.5	67.0	n.s.
5	Banged Head on Purpose	67.3	59.1	50.4	56.6	0.014
13	Abused prescription medication	54.5	54.8	47.5	51.0	n.s.
21	Starved self to hurt self	38.6	53.9	43.9	45.3	n.s.
10	Made medical situations worse	27.2	36.5	55.2	44.0	<0.001
4	Hit Self	35.6	47	39.7	40.7	n.s.
9	Prevented wounds from healing	51.5	47	31.2	40.0	0.001
15	Engaged in emotionally abusive relationships	35.6	65.2	23.2	37.2	<0.001
7	Driven recklessly on Purpose	37.6	12.2	33.5	28.9	<0.001
11	Been promiscuous	37.6	24.3	27	28.8	n.s.
8	Scratched Self on Purpose	27.7	42.6	21	28.2	<0.001
17	Lost Job on Purpose	25.7	14.8	25.1	22.6	n.s.
3	Burned Self on Purpose	20.8	24.3	19.2	20.9	n.s.
14	Distanced yourself from God	15.8	12.2	22.9	18.5	0.042
12	Set relationship to be rejected	18.8	20.9	16.3	18.1	n.s.
19	Exercised an injury on purpose	14.9	13.9	19	16.7	n.s.
16	Engaged in sexually abusive relationships	5.9	26.1	1.9	9.3	<0.001
22	Abused laxatives to hurt self	0	16.5	1.8	5.2	<0.001

*Chi-Square

Table 28. Reported lifetime self-harm behaviours across participants that self-harmed in follow-up (n=126). Significance across prison

		% affirming behaviour in lifetime				
SHI Item	Self-Harm Behaviour	Prison A	Prison B	Prison C	Total	Significance*
2	Cut self on purpose	100.0	91.9	80.9	86.7	n.s.
20	Tortured self with self-defeating thoughts	66.7	83.8	84.1	81.8	n.s.
18	Attempted suicide	80.0	83.8	79.7	81.0	n.s.
1	Overdosed	66.7	91.9	68.1	75.2	0.019
6	Abused alcohol	60.0	70.3	58.0	62.0	n.s.
5	Banged head on purpose	66.7	67.6	52.2	58.7	n.s.
13	Abused prescription medication	60.0	62.2	50.7	55.4	n.s.
21	Starved self to hurt self	60.0	67.6	42.6	52.5	0.042
10	Made medical situations worse	20.0	51.4	58.0	51.2	0.029
9	Prevented wounds from healing	60.0	62.2	39.1	48.8	0.05
4	Hit self	40.0	51.4	43.5	45.5	n.s.
8	Scratched self on purpose	40.0	59.5	21.7	35.5	0.001
15	Engaged in emotionally abusive relationships	46.7	64.9	16.2	35.0	<0.001
11	Been promiscuous	40.0	32.4	23.5	28.3	n.s.
3	Burned self on purpose	33.3	37.8	20.3	27.3	n.s.
7	Driven recklessly on purpose	33.3	16.2	26.1	24.0	n.s.
19	Exercised an injury on purpose	26.7	21.6	24.6	24.0	n.s.
12	Set up relationship to be rejected	33.3	18.9	22.4	22.7	n.s.
17	Lost job on purpose	20.0	13.5	27.5	22.3	n.s.
14	Distanced yourself from God	0.0	18.9	26.1	20.7	n.s.
16	Engaged in sexually abusive relationships	6.7	24.3	1.5	9.4	0.001
22	Abused laxatives to hurt self	0.0	21.6	2.9	8.3	0.002

*Chi-Square

A range of additional baseline and follow-up characteristics regarding the sample are summarised in Appendix D.

6 Psychometric Analysis of Scales

Various methodologies were used to assess the psychometric properties of the separate five instruments that were utilised within the cohort study. The results of these psychometric analyses are presented in this chapter. This includes information regarding the basic psychometric details for each of the instruments, followed by the CFA and Mokken scaling results. A more-detailed breakdown of the Rasch analysis is presented for each of the instruments, which includes information relating to each of the instruments at both the scale level and the individual item level. Furthermore, the results of two separate Rasch-based resolution approaches are presented for each instrument, where these were appropriate.

6.1 Basic Characteristics of Utilised Instruments

The basic characteristics of the five instruments used in the study are summarised in Table 29. Compliance at the scale level was good; the PriSnQuest recorded the lowest proportion of cases with complete data (91.6%). In terms of individual item compliance, this was also good across all items. The mean individual item completion rate was 98.3% (SD 0.91%) across all items. The lowest individual item completion rate was 95.1% (22 non-responses) for item 16 of the SHI: ‘Engaged in sexually abusive relationships’.

The high compliance rate across all scales and individual items would suggest that there is no evidence of responder burnout. Participants were free to stop the questionnaire administration at any point in the process, but very few of them did so, meaning that complete data were present in almost all cases.

Table 29. Basic descriptive and compliance statistics for the five utilised instruments

Characteristic	BSL-23-F	CORE-OM	PriSnQuest	PHQ-9	SHI
Number of items in scale	23	34	8	9	22
Number of response categories for each item	5	5	2	4	2
Response category scoring for scale items	0-4	0-4	0-1	0-3	0-1
Total scale scoring range	0-92	0-136	0-8	0-27	0-22
Number of cases with missing scale data	22	24	38	13	31
Percentage of cases with complete data (n=450)	95.10%	94.70%	91.60%	97.10%	93.10%
Number of cases with missing evaluable scale data	14	6	16	9	12
Percentage of cases with evaluable scores (according to scale instructions)	96.90%	98.70%	96.40%	98.00%	97.30%
Median	50	77	5	19	9
IQR	35-65	60-90	4-6	13.5-23	6-12
Range	0-92	7-122	0-8	0-27	0-22
Internal Consistency Reliability α	0.93	0.9	0.63	0.82	0.78

The medians and inter-quartile ranges (IQR) of all of the instruments are reported in Table 30 for the complete sample across all three prisons. These statistics are based on the evaluable scores for each instrument, as per the scoring instructions for the individual scales. Note that some scales have low reliability in this setting.

Table 30. Basic descriptive statistics for the five instruments across the prisons

			BSL-23-F	CORE-OM	PriSnQuest	PHQ-9	SHI
Prison	A	Median	51	77	5	18	8
		IQR	34.3-65.0	57.0-90.0	3-6	13-21	6.0-11.5
		Range	0-92	12-122	0-8	0-27	0-17
	B	Median	49	74.5	5	17	10
		IQR	34.5-63.0	54.8-86.0	4-6	12-22	7-13
		Range	0-92	12-116	0-8	1-27	0-22
	C	Median	52	79	5	20	8
		IQR	36.0-65.0	62.9-75.0	3.43-6	14-24	5-11
		Range	6-92	7-118	0-8	0-27	0-20

6.2 BSL-23-F

Refer to Appendix A, Questionnaire 3 for a copy of the complete BSL-23-F instrument.

6.2.1 Confirmatory Factor Analysis and Mokken Scaling Analysis

The results of the CFA and Mokken scaling analysis are summarised in Table 31. Weak support was offered for the unidimensionality of the BSL-23-F once error correlations had been added to the model (to account for inter-item dependency), and the item set satisfied Mokken scaling criteria without the removal of any items. This suggests that although some dependency is present between items, the BSL-23-F item set forms an ordinal scale that is robust enough (within this setting) for a scale cut point to be valid for use in an AUC analysis.

The BSL Supplement displayed more favourable results, with the CFA showing strong support for unidimensionality once error correlations had been added to the model. Additionally, the item set satisfied Mokken scaling criteria without the removal of any items, suggesting that the item set forms an ordinal scale, even though the item set is only provided to elicit further information regarding specific self-harming behaviours.

Table 31. CFA and Mokken scaling statistics for the BSL-23-F and BSL Supplement

	Borderline Symptom List-23-F		BSL Supplement	
	Original structure	With Correlated errors	Original structure	With Correlated errors
Original Number of scale items	23	23	8	8
CFA Results				
Chi-Square (df)	1043 (230)	400 (205)	44 (20)	28 (19)
P value	<0.0001	<0.0001	0.0014	0.0934
RMSEA	0.089	0.046	0.053	0.032
CFI	0.928	0.983	0.891	0.962
TLI	0.92	0.979	0.848	0.944
Quality of support	None	Weak	None	Strong
Mokken Results				
Number of items removed to form final scale	0	-	0	-
Loevinger's H	0.57	-	0.71	-
Interpretation of Scalability	Good		Good	

6.2.2 Rasch Analysis

Initial analysis of the BSL-23-F revealed that the items in the scale failed to meet Rasch model expectations (see Table 33 - initial). Individual item fit revealed evidence of a number of problematic items displaying fit parameters outside of the normally expected and accepted range. Additionally, ALL items displayed disordered thresholds, meaning that the response categories were not functioning as intended. At this initial stage, only two items displayed DIF at the Bonferroni-adjusted level. Item 13 (I suffered from shame) displayed DIF-by-age group and Item 16 (Criticism had a devastating effect on me) displayed DIF by both prison and gender, although the prison DIF is likely to just be an interactive manifestation of the gender DIF that is present.

Rescore

As the response options were not working as intended across the whole item set, and the observed response patterns were similar for most items, a generic rescore was implemented (see Table 32).

Table 32. Generic rescore implemented across all BSL-23-F items

Original Response Code	Response Wording	Rescored Response Code
0	Not at all	0
1	Only occasionally	1
2	Sometimes	1
3	Often	2
4	Most or all the time	2

This rescore also has the follow-on effect of reducing the total scale score. Originally the scale would be scored 0-92, but with the rescore in place the total scale score is contracted to 0-46.

Following the generic recode, all items displayed ordered categories except Item 15 (I suffered from voices and noises from inside or outside my head).

The summary fit statistics at this stage are presented in Table 33 (rescored), along with the plot of relative item threshold difficulties and person abilities (the targeting plot - Figure 30).

Sources of individual item misfit at this stage are summarised in Table 34.

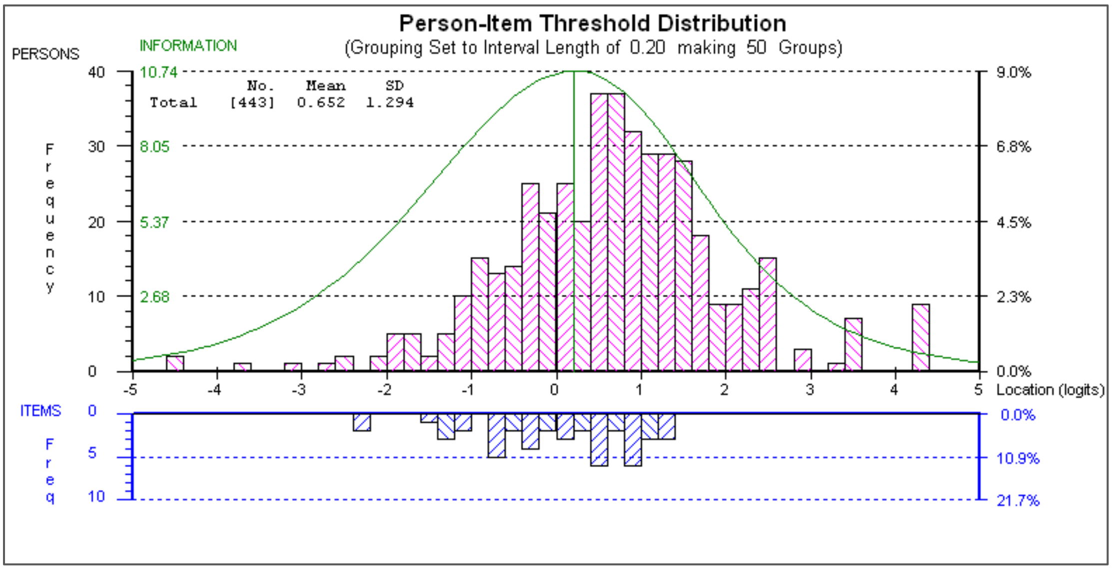

Figure 30. Targeting plot for the BSL-23-F following rescoring

Table 33. Summary Rasch fit statistics for the BSL-23-F

* = not sufficient power in t-test procedure (i.e. < 10 thresholds used to generate comparative estimates)

		Item Location		Person Location		Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI		Alpha	Unidimensionality T-Tests (CI)			
Analysis	CInt	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Value	df	p	with extrms	NO extrms		number of signif. tests	out of:	%	lower bound 95% CI
Initial	7	0	0.39	0.15	0.9	0.77	3	-0.12	1.7	479	138	0	0.918	0.921	0.932	37	436	8.49%	6.40%
Rescored	7	0	0.69	0.65	1.29	0.05	2.7	-0.13	1.3	416.7	138	0	0.895	0.893	0.918	39	431	9.05%	7.00%
Resolution A	7	0	0.77	0.56	1.19	0.25	1.3	-0.18	1.1	116.9	84	0.01	0.852	0.843	0.853	22	424	5.19%	3.10%
Resolution B	7	0	0.73	0.9	1.38	0.33	1.4	-0.18	1.1	99.07	78	0.05	0.823	0.807	0.875	18	414	4.35%	-
Resolution B2	7	0	0.75	0.83	1.38	0.33	1.3	-0.18	1.1	96.08	84	0.17	0.838	0.823	0.882	19	417	4.56%	-
Supplement Rescored	4	0	1.36	-1.62	1.04	0.09	2.0	-0.10	0.4	67.11	24	0	0.018	-0.27	0.486	0	336	0%*	-

Table 34. Summary of individual sources of misfit within the BSL-23-F item set, following a generic recode (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

Even when Item 15 is rescored in an alternative rescore pattern to resolve the disordered thresholds, the reported misfit is still present.

6.2.2.1 Scale Refinement

6.2.2.1.1 Resolution A

Following the generic rescore, Resolution A was reached following the removal of five items (Item 3: 'I was absent-minded and unable to remember what I was actually doing', Item 6: 'I didn't trust other people', Item 15: 'I suffered from voices and noises from inside or outside my head', Item 22: 'I felt as if I was far away from myself', and Item 23: 'I felt worthless'). Additionally, subtests (testlets) were created from Items 1 & 2 ('It was hard for me to concentrate' & 'I felt helpless'), Items 7, 11 & 12 ('I didn't believe in my right to live', 'I hated myself' & 'I wanted to punish myself'), and Items 4, 13 and 21('I felt disgust', 'I suffered from shame' & 'I felt disgusted by myself'). Also, Item 16 ('Criticism had a devastating effect on me') was split for DIF-by-gender.

The summary fit statistics at this stage are presented in Table 33 (Resolution A).

6.2.2.1.2 Resolution B

Following the generic rescore, Resolution B was reached following the removal of 10 items (Item 3: 'I was absent-minded and unable to remember what I was actually doing', Item 6: 'I didn't trust other people', Item 10: 'I had images that I was very much afraid of', Item 11: 'I hated myself', Item 12: 'I wanted to punish myself', Item 15: 'I suffered from voices and noises from inside or outside my head', Item 16: 'Criticism had a devastating effect on me', Item 18: 'The idea of death had a certain fascination for me', Item 21: 'I felt disgusted by myself', and Item 23: 'I felt worthless').

The summary fit statistics at this stage are presented in Table 33 (Resolution B).

The summary fit statistics are also presented at the stage prior to Item 16 ('Criticism had a devastating effect on me') being removed for DIF-by-gender. See Table 33 (Resolution B2).

6.2.2.2 BSL Supplement

The 8 items of the supplement were also looked at as a separate scale.

All thresholds were disordered with category probability response patterns tending towards a dichotomous structure. All items were therefore dichotomised, which resulted in an extremely low person separation index (0.02), along with other unfavourable fit statistics. See Table 33 (Supplement Rescored) for the BSL Supplement summary fit statistics at this stage. This analysis was not progressed due to the lack of power in the tests of fit, as indicated by the low person separation index.

6.3 CORE-OM

Refer to Appendix A, Questionnaire 1 for a copy of the complete CORE-OM instrument.

6.3.1 Confirmatory Factor Analysis and Mokken Scaling Analysis

The CORE-OM is made up of 34-items that can be assessed as a single total score, or broken down into separate domains of Well-being (4 items), Problems or Symptoms (12 items), Functioning (12 items), and Risk (6 items). The CORE-OM is also commonly summed with the Risk domain excluded (CORE minus Risk). Additionally, the short-form 10-item screening tool, the CORE-10, is embedded within the larger 34-item scale. The overall single total score of the CORE-OM was not initially supported by the single-factor CFA, but a weak level of support was displayed once errors had been correlated (to account for the dependency within the item set). The CORE-OM subscales, including the CORE-10, all displayed moderate support once errors had been correlated, with the exception of the Well Being subscale, which displayed strong support of its unidimensionality, without the addition of any error correlations to the model.

For the Mokken scaling, the individual subscale domains of 'Well Being', 'Problems' and 'Risk' displayed moderate scalability, with some item-removal modifications made to the domain item sets. The functioning subscale is more problematic, splitting into two smaller scales; one with weak scaling properties and one with moderate scaling properties. The CORE-10 also initially failed, requiring removal of three items to satisfy moderate scaling criteria. The results of the CFA and Mokken scaling analysis are summarised in Table 35.

Table 35. CFA and Mokken scaling statistics for the CORE-OM

	Overall CORE-OM		Well-Being Domain	Problems Domain		Functioning Domain		Risk Domain		CORE-10	
	Original structure	With Correlated errors	Original structure	Original structure	With Correlated errors	Original structure	With Correlated errors	Original structure	With Correlated errors	Original structure	With Correlated errors
Original Number of scale items	34	34	4	12	12	12	12	6	6	10	10
CFA Results											
Chi-Square (df)	1854 (521)	929 (490)	1.546 (2)	170 (54)	76 (48)	405 (54)	79 (46)	36 (9)	16 (8)	122 (35)	50 (30)
P value	<0.0001	<0.0001	0.4617	<0.0001	0.0059	<0.0001	0.0019	<0.0001	0.0425	<0.0001	0.0138
RMSEA	0.076	0.045	0	0.07	0.037	0.122	0.04	0.083	0.048	0.074	0.038
CFI	0.856	0.952	1	0.938	0.985	0.831	0.984	0.885	0.966	0.959	0.991
TLI	0.845	0.946	1	0.925	0.98	0.794	0.977	0.809	0.937	0.947	0.986
Quality of support	None	Weak	Strong	None	Moderate	None	Moderate	None	Moderate	None	Moderate
Mokken Results											
						Functioning Set 1	Functioning Set 2				
Number of items removed to form final scale	-	-	1	2	-	8	8	1	-	3	-
Loevinger's H	-	-	0.42	0.42	-	0.42	0.36	0.5	-	0.41	-
Interpretation of Scalability	-	-	Medium	Medium	-	Medium	Useful but Weak	Medium/ Good	-	Medium	-

6.3.2 Rasch Analysis

The CORE-OM can be assessed as a single total score (representing a single dimension), or as separate domains, along with the CORE-10 short-form. It is postulated that the 4 domains all contribute to a higher order construct, but, prior to this being formed, it holds that each individual domain should function independently. Firstly, the results of the complete CORE-OM will be presented, followed by the independent domains and the CORE-10.

6.3.2.1 The CORE-OM complete scale

Initial analysis of the CORE-OM revealed the scale to be problematic in terms of fit to the Rasch model. The summary fit statistics at this stage are presented in Table 37 (initial). Individual item fit revealed evidence of a number of problematic items displaying fit parameters outside of the normally expected and accepted range. Additionally, the observed response patterns for the items were very similar to those observed for the BSL-23-F as all items initially displayed disordered thresholds, meaning that the response categories were not functioning as intended.

Rescore

As the response options were not working as intended across the whole item set, and the observed response patterns were similar for most items, a generic rescore was again implemented, although this rescore was different for regular scored items and reverse scored items. This recode is presented in Table 36.

Table 36. Generic rescore implemented across CORE-OM items

Original Response Code	Original Reversed Response Code	Response Wording	Rescored Response Code	Rescored Reversed Response Code
0	4	Not at all	0	2
1	3	Only occasionally	1	1
2	2	Sometimes	1	1
3	1	Often	2	0
4	0	Most or all the time	2	0

This rescore also has the follow-on effect of reducing the total scale score. Originally the scale would be scored 0-136, but with the rescore in place the total scale score is contracted to 0-68.

Following the generic recode, 28 items displayed ordered thresholds, but six items still displayed disordered thresholds. Despite the remaining disorder, this response structure was maintained across the item set.

The summary fit statistics at this stage are presented in Table 37 (rescored), along with the targeting plot (Figure 31).

Sources of individual item misfit at this stage are summarised in Table 38.

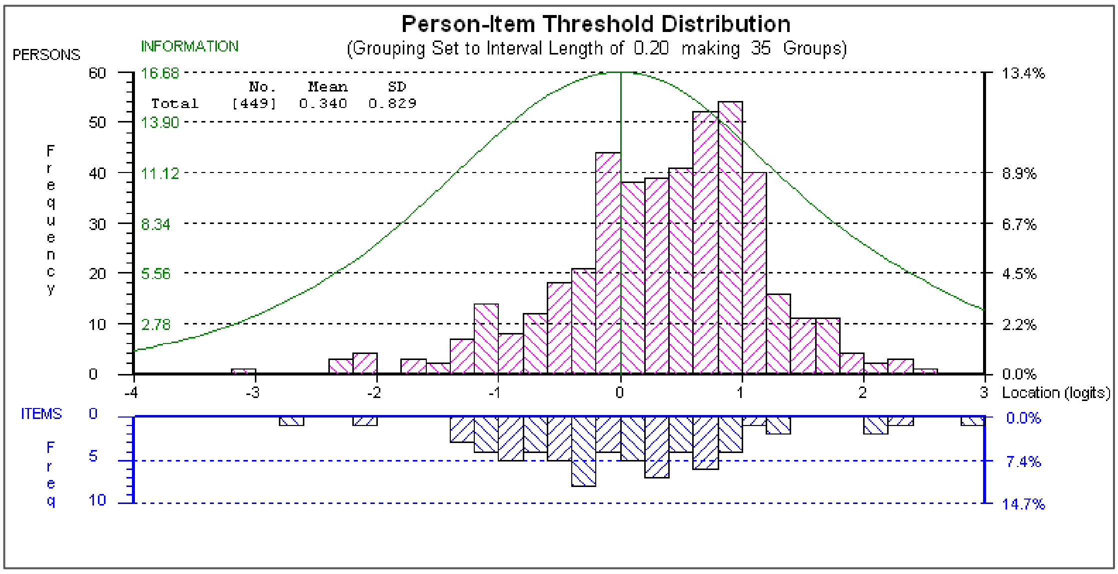

Figure 31. Targeting plot for the CORE-OM following rescoring

Table 37. Summary Rasch fit statistics for the CORE-OM, separate CORE domains and CORE-10

		Item Location		Person Location		Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI			Unidimensionality T-Tests (CI)			
Analysis	CInt	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Value	df	p	with extrms	NO extrms	Alpha	number of signif. tests	out of:	%	lower bound 95% CI
CORE-OM initial	8	0	0.48	0.09	0.48	0.83	3.47	-0.02	1.43	1074	238	0	0.9	0.9	0.9	77	448	17.19%	15.20%
CORE-OM rescored	8	0	0.84	0.34	0.83	0.38	3.26	-0.07	1.23	849.7	238	0	0.89	0.89	0.9	59	448	13.17%	11.20%
Bifactor resolution	5	0	0.69	0.15	0.71	0.41	0.91	-0.32	0.97	27.22	28	0.507	0.856	0.856	-	20	445	4.49%	-
CORE-OM Resolution B	7	0	0.83	0.32	0.94	0.28	1.11	-0.15	0.98	104.04	102	0.425	0.815	0.81	0.83	25	448	5.58%	3.60%
Well-being initial	8	0	0.18	0.45	0.71	0.55	1.7	-0.31	1.2	80.3	28	0	0.43	0.3	0.58	0	448	0.00%*	-
Well-being rescored	5	0	0.26	0.68	1.02	0.99	2.03	-0.29	1.36	63.2	16	0	0.26	0.03	0.54	0	383	0.00%*	-
Prob/Symp initial	8	0	0.33	0.37	0.76	0.82	2.64	-0.17	1.36	221.8	84	0	0.82	0.81	0.85	26	441	5.90%	3.90%
Prob/Symp rescored	7	0	0.59	1.01	1.17	0.02	2.19	-0.12	0.95	154.1	72	0	0.75	0.73	0.83	12	425	2.82%	-
Prob/Symp Resolution A	5	0	0.47	0.96	1.21	0.03	1.48	-0.23	1.11	44.7	32	0.067	0.686	0.64	0.8	7	410	1.71%*	-
Prob/Symp Resolution B	4	0	0.62	1.06	1.21	0.02	1.15	-0.21	1.1	43.03	24	0.0099	0.652	0.594	0.78	9	407	2.21%*	-

* = not sufficient power in t-test procedure (i.e. < 10 thresholds used to generate comparative estimates)

		Item Location		Person Location		Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI			Unidimensionality T-Tests (CI)			
Analysis	CInt	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Value	df	p	with extrms	NO extrms	Alpha	number of signif. tests	out of:	%	lower bound 95% CI
Func initial	7	0	0.38	0.07	0.5	1.05	2.59	-0.08	1.34	201	72	0	0.75	0.74	0.74	37	448	8.26%	6.20%
Func rescored	7	0	0.43	0.12	0.86	0.86	2.71	-0.07	1.34	223.95	72	0	72	69	0.78	36	444	8.11%	6.10%
Func Resolution A	6	0	0.47	0.23	0.99	0.52	1.17	-0.23	1.52	61.94	50	0.1198	0.72	0.671	0.74	19	437	4.35%	-
Func Resolution B	6	0	0.44	0.05	1.03	0.72	1.26	-0.19	1.31	62.96	45	0.0396	0.708	0.638	0.74	12	431	2.78%*	-
Risk1 initial	7	0	0.68	-0.61	0.86	-0.38	1.08	-0.32	0.65	129.2	36	0	0.65	0.6	0.73	23	404	5.69%*	3.60%
Risk2 rescored	7	0	1.34	-0.7	1.38	-0.4	1.11	-0.34	0.69	94.8	36	0	0.65	0.53	0.72	27	401	6.73%*	4.60%
CORE-10 Initial	6	0	0.4	0.3	0.65	0.71	2.61	-0.16	1.16	195.97	50	0	0.764	0.751	0.79	22	446	4.93%	-
CORE-10 Rescored	7	0	0.74	0.73	1.04	0.12	2.6	-0.14	0.92	170.75	60	0	0.71	0.693	0.77	11	442	2.49%	-
CORE-10 Resolution	7	0	0.83	0.78	1.1	0.11	1.24	-0.18	0.85	69.2	48	0.024	0.659	0.612	0.73	5	434	1.15%*	-

* = not sufficient power in t-test procedure (i.e. < 10 thresholds used to generate comparative estimates

Table 38. Summary of individual sources of misfit within the CORE-OM item set, following a generic recode (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										
33										
34										

Although the items with disordered thresholds can be recoded in an alternative rescore pattern to resolve the disordered thresholds, the reported misfit is still present.

6.3.2.1.1 Scale Refinement

6.3.2.1.1.1 Bifactor Resolution

As the CORE-OM has four underlying domains, a bifactor resolution was sought. A bifactor analysis treats each independent domain as a testlet item, and the analysis is based on the shared component of the domains, with the unique component excluded. (209, 227)

The items displaying as clear underdiscriminating measurement anomalies within each domain were removed prior to the formation of the domain subtests (testlets). This meant that Items 2, 8 and 30 were removed from the Problems/Symptoms domain, and Items 3 and 8 were removed from the Functioning domain.

The initial domain grouping revealed various DIF issues. The final bifactor resolution involved splitting the Well-being domain for DIF-by-Gender, and splitting the Risk and Functioning domains for DIF-by-Age group.

The summary fit statistics for the final bifactor resolution are presented in Table 37 (Bifactor resolution).

6.3.2.1.1.2 Resolution B

Following the generic rescore, Resolution B was reached following the removal of 17 items.

The summary fit statistics at this stage are presented in Table 37 (CORE-OM Resolution B).

The removed items, along with the reason for removal are summarised in Table 39.

Table 39. Items removed from CORE-OM to reach Resolution B

Misfit parameter	Items removed
Underdiscrimination	3 / 8 / 19 / 21 / 31 / 34
Overdiscrimination	2 / 9 / 17 / 23
Response Dependence (Corr>0.2)	9 / 22 / 24 / 28 / 32 / 33 / 34
Prison DIF	
Gender DIF	14
Age DIF	34
Religion DIF	

6.3.2.1.1.3 Removed Items

An additional analysis was run on the removed items to see whether they formed an alternative unidimensional item set. However, this item set displayed a high degree of misfitting parameters, both collectively and on an individual item basis.

6.3.2.2 Individual CORE Domains

The initial summary statistics for each domain can be found in Table 37 (headed Well-being initial, Prob/Symp initial, Func initial and Risk initial). All domains displayed the same threshold disordering as was present in the CORE-OM; therefore the same generic rescoring pattern was applied to each individual domain (see Table 36). The summary statistics for each domain following the generic recode can be found in Table 37 (headed Well-being rescored, Prob/Symp rescored, Func rescored and Risk rescored), and the sources of individual item misfit at this stage are summarised in Table 40.

6.3.2.2.1 Well-being & Risk Domains

Following rescoring, the Well-being and Risk domains still displayed a large degree of misfit from a number of sources. Due to the limited number of items within these domains, along with the apparent misfit of various forms, neither resolution was reached for either domain. This means that these subscales did not conform to the strict requirements of Rasch scaling, but they may still conform to ordinal scale requirements, or have use as a series of single-indicator items. This does not preclude the domains being used as part of a bifactor analysis, but as independent domains these item sets fail to conform to the expectations of Rasch analysis.

6.3.2.2.2 Problems/Symptoms Domain

After the application of the generic recode, Resolution A was reached following the removal of Items 2, 8 and 20, and subtesting Items 23 & 27 to account for the response dependency between the items.

The summary fit statistics at this stage are presented in Table 37 (Prob/Symp Resolution A).

Following the generic rescore, Resolution B was reached following the removal of Items 2, 8, 20 and 23. The summary fit statistics at this stage are presented in Table 37 (Prob/Symp Resolution B).

6.3.2.2.3 Functioning Domain

After the application of the generic recode, Resolution A was reached following the removal of Items 3 and 19, and subtesting items 25 & 33 to account for the response dependency between the items. Additionally, Item 1 was split for DIF-by-Gender.

The summary fit statistics at this stage are presented in Table 37 (Func Resolution A).

Following the generic rescore, Resolution B was reached following the removal of Items 1, 3 and 19. The summary fit statistics at this stage are presented in Table 37 (Func Resolution B).

Table 40. Summary of individual sources of misfit within the CORE domains, following a generic recode (misfit sources highlighted in red)

	Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
Well-being	4										
	14										
	17										
	31										
Problems/Symptoms	2										
	5										
	8										
	11										
	13										
	15										
	18										
	20										
	23										
	27										
	28										
	30										
Functioning	1										
	3										
	7										
	10										
	12										
	19										
	21										
	25										
	26										
	29										
	32										
	33										
Risk	6										
	9										
	16										
	22										
	24										
	34										

6.3.2.2.4 CORE-10

The initial summary statistics for the CORE-10 short form can be found in Table 37 (headed CORE-10 initial). All CORE-10 items displayed the same threshold disordering as was present in the CORE-OM; therefore the same generic rescoring pattern was applied (see Table 36). The summary statistics for the CORE-10 following the generic recode can be found in Table 37 (headed CORE-10 rescored), and the sources of individual item misfit at this stage are summarised in Table 41.

After the application of the generic recode, both Resolutions A and B were reached following the removal of Items 3 and 23.

The summary fit statistics at this stage are presented in Table 37 (CORE-10 Resolution).

Table 41. Summary of individual sources of misfit within the CORE-10, following a generic recode (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
2										
3										
7										
10										
15										
16										
18										
23										
27										
28										

6.4 PHQ-9

Refer to Appendix A, Questionnaire 5 for a copy of the complete PHQ-9 instrument.

6.4.1 Confirmatory Factor Analysis and Mokken Scaling Analysis

The results of the CFA and Mokken scaling analysis are summarised in Table 42. Weak support was offered for the unidimensionality of the PHQ-9 when error correlations had been added to the model, and the item set satisfied Mokken scaling criteria without the removal of any items. This suggests that some dependency is present between items, but the PHQ-9 item set forms an ordinal scale that is robust enough (within this setting) for a scale cut point to be valid for use in an AUC analysis.

Table 42. CFA and Mokken scaling statistics for the PHQ-9

	PHQ-9	
	Original structure	With Correlated errors
Original Number of scale items	9	9
CFA Results		
Chi-Square (df)	142 (27)	52 (22)
P value	<0.0001	0.0003
RMSEA	0.098	0.056
CFI	0.941	0.984
TLI	0.921	0.974
Quality of support	None	Weak
Mokken Results		
Number of items removed to form final scale	0	-
Loevinger's H	0.66	-
Interpretation of Scalability	Good	

6.4.2 Rasch Analysis

Initial analysis of the PHQ-9 showed that the scale failed to satisfy Rasch model expectations. The summary fit statistics at this stage are presented in Table 44 (initial). Individual item fit revealed evidence of relatively few problematic items displaying fit parameters outside of the normally expected and accepted range.

However, all items except one (Item 4: ‘Feeling tired or having little energy’) displayed disordered thresholds, meaning that the response categories were not functioning as intended.

Rescore

As the response options were not working as intended across almost the whole item set, and the observed response patterns were similar for most items, a generic rescore was implemented.

Table 43. Generic rescore implemented across all PHQ-9 items

Original Response Code	Response Wording	Rescored Response Code
0	Not at all	0
1	Several Days	1
2	More than half the days	1
3	Nearly every day	2

This rescore also has the follow-on effect of reducing the total scale score. Originally the scale would be scored 0-27, but with the rescore in place the total scale score is contracted to 0-18.

Following the generic recode, all items displayed ordered categories.

The summary fit statistics at this stage are presented in Table 44 (rescored), along with the targeting plot (Figure 32).

Sources of individual item misfit at this stage are summarised in Table 45.

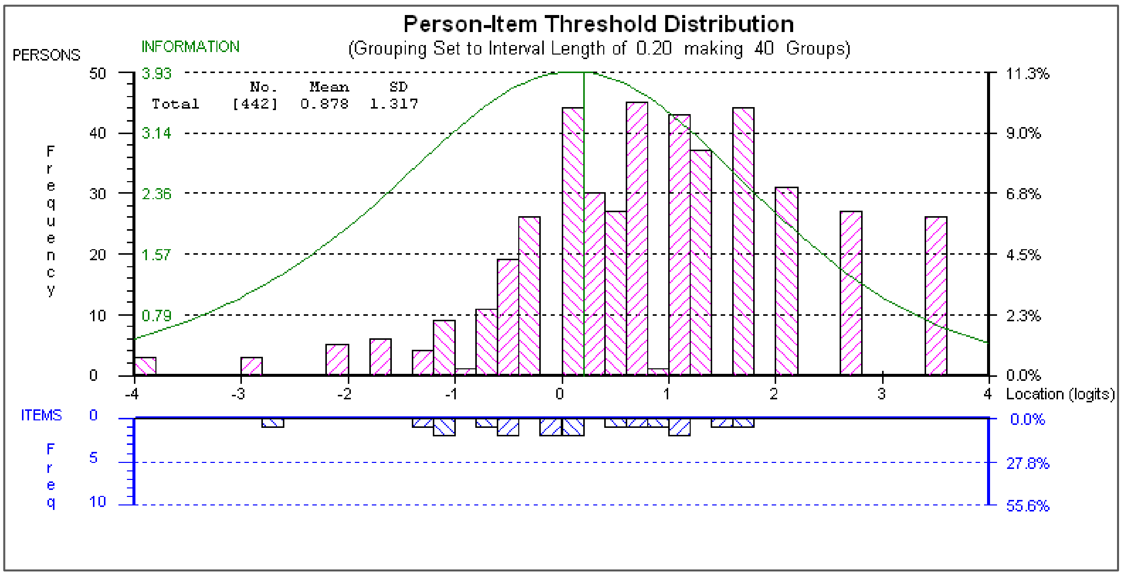

Figure 32. Targeting plot for the PHQ-9 following rescoring

Table 44. Summary Rasch fit statistics for the PHQ-9

		Item Location		Person Location		Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI		Alpha	Unidimensionality T-Tests (CI)			
Analysis	CInt	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Value	df	p	with extrms	NO extrms		number of signif. tests	out of:	%	lower bound 95% CI
PHQ-9 initial	7	0	0.44	0.6	0.96	0.28	1.4	-0.2	1.15	90.2	54	0.0015	0.746	0.717	0.82	14	412	3.40%	-
PHQ-9 rescored	6	0	0.64	0.88	1.32	0.32	1.28	-0.27	1.3	70.18	45	0.0095	0.743	0.702	0.81	13	412	3.16%*	-
PHQ-9 Resolution A	6	0	0.55	0.8	1.26	0.31	1.13	-0.28	1.27	46.44	40	0.224	0.732	0.69	0.79	13	412	3.16%*	-
PHQ-9 Resolution B	6	0	0.51	0.73	1.24	0.41	0.95	-0.29	1.32	50.75	40	0.119	0.7	0.649	0.78	8	411	1.95%*	-

* = not sufficient power in t-test procedure (i.e. < 10 thresholds used to generate comparative estimates)

Table 45. Summary of individual sources of misfit within the PHQ-9 item set, following a generic recode (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										
9										

6.4.2.1 Scale Refinement

6.4.2.1.1 Resolution A

It can be seen in Table 45 that following the generic rescore there are very few sources of underlying misfit to amend in order to reach Resolution A. Despite no response dependency being apparent at a residual correlation of 0.2, a lower level dependency was present between Items 1 and 2. This dependency also holds on a conceptual level, as Items 1 and 2 are the two ‘summary’ items that make up the PHQ-2 short form.

Resolution A was reached following the subtesting of Items 1 and 2 into a testlet to account for underlying conceptual dependency.

The summary fit statistics at this stage are presented in Table 44 (Resolution A).

6.4.2.1.2 Resolution B

Following the generic rescore, Resolution B was reached following the removal of Item 2.

The summary fit statistics at this stage are presented in Table 44 (Resolution B).

6.5 PriSnQuest

Refer to Appendix A, Questionnaire 2 for a copy of the complete PriSnQuest instrument.

6.5.1 Confirmatory Factor Analysis and Mokken Scaling Analysis

The results of the CFA and Mokken scaling analysis are summarised in Table 46. Strong support was offered for the unidimensionality of the PriSnQuest once error correlations had been added to the model, and the item set satisfied Mokken scaling criteria to a medium level without the removal of any items. This suggests that although some dependency is present between items, the PriSnQuest item set forms an ordinal scale that is robust enough (within this setting) for a scale cut point to be valid for use in an AUC analysis.

Table 46. CFA and Mokken scaling statistics for the PriSnQuest

	PriSnQuest	
	Original structure	With Correlated errors
Original Number of scale items	8	8
CFA Results		
Chi-Square (df)	126 (20)	26 (17)
P value	<0.0001	0.0714
RMSEA	0.109	0.035
CFI	0.909	0.992
TLI	0.872	0.987
Quality of support	None	Strong
Mokken Results		
Number of items removed to form final scale	0	-
Loevinger's H	0.48	-
Interpretation of Scalability	Medium	-

6.5.2 Rasch Analysis

Initial analysis of the PriSnQuest showed the scale to be problematic in terms of fit to the Rasch model. The summary fit statistics at this stage are presented in Table 47 (initial), along with the initial targeting plot Figure 33. Individual item fit revealed evidence of some items displaying fit parameters outside of the normally expected and accepted range, but the individual item misfit did not suggest the same level of misfit as was found in the overall scale fit statistics.

As the PriSnQuest items are all dichotomously scored, there is no opportunity for item thresholds to be disordered as each item only has a single measurement threshold. Therefore no rescoring is necessary, or possible, among the PriSnQuest items. The sources of individual item misfit at this stage are summarised in Table 48.

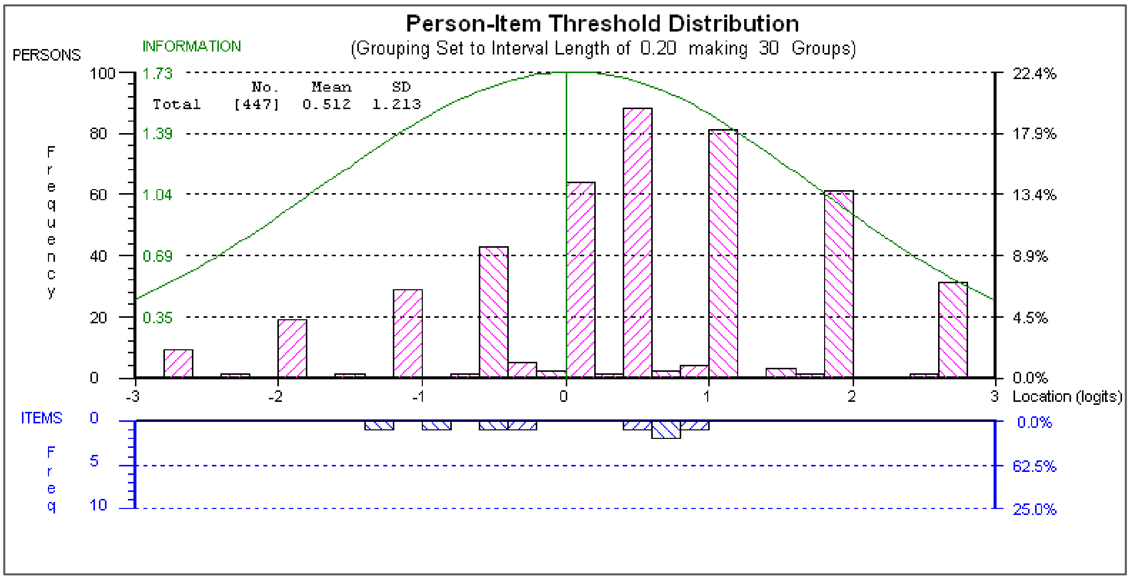

Figure 33. Targeting plot for the PriSnQuest

Table 47. Summary Rasch fit statistics for the PriSnQuest

		Item Location		Person Location		Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI		Alpha	Unidimensionality T-Tests (CI)			
Analysis	CInt	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Value	df	p	with extrms	NO extrms		number of signif. tests	out of:	%	lower bound 95% CI
PriSnQuest initial	6	0	0.84	0.51	1.21	0.41	1.61	-0.12	0.84	82	40	0	0.44	0.26	0.63	8	405	1.98%*	-
Prisn subtest	6	0	0.67	0.31	1.1	0.84	1.36	-0.03	0.74	70.9	35	0.0003	0.36	0.16	0.58	2	404	0.5%*	-
Prisn Male	5	0	0.87	0.5	1.24	0.26	1.3	-0.12	0.83	54.41	32	0.008	0.45	0.28	0.65	5	300	1.67%*	-
Prisn Female	3	0	0.84	0.55	1.19	0.2	0.94	-0.12	0.87	21.66	16	0.155	0.43	0.23	0.6	0		0%*	-

* = not sufficient power in t-test procedure (i.e. < 10 thresholds used to generate comparative estimates)

Table 48. Summary of individual sources of misfit within the PriSnQuest (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										

6.5.2.1 Scale Refinement

At this initial stage, the main anomaly seemed to be the sizeable response dependency that was apparent between Item 4: ‘Have you recently felt that life isn’t worth living?’, and Item 5: ‘Have you recently found yourself wishing you were dead and away from it all?’ (Residual correlation = 0.505). This apparent dependency was accounted for through subtesting the affected items, and the summary statistics following this amendment are presented in Table 47 (Prisn subtest).

Despite accounting for this item dependency, the PriSnQuest appeared similar to the CORE Well-being and Risk domains, in that the PriSnQuest has a limited number of items within the scale, and even after accounting for various forms of apparent misfit, neither Resolution A or Resolution B was applicable to this set of items. Again, this means this scale did not conform to the strict requirements of Rasch scaling, but this does not preclude it conforming to ordinal scale requirements, or having use as a screening tool or a series of single-indicator items.

However, the PriSnQuest appeared to function differently in male and female populations, suggested by the Gender DIF that is apparent in the initial analysis. Consequently, it may be useful to treat the PriSnQuest as a different scale among male and female ACCT populations. The summary statistics of the initial PriSnQuest for the separate male and female samples are presented in Table 47 (Headed Prisn Male and Prisn Female). Although the fit of the scale to the model is weak for males, it does appear that a separate gender-based solution is more appropriate.

6.6 Self-Harm Inventory

Refer to Appendix A, Questionnaire 4 for a copy of the complete Self-Harm Inventory (SHI) instrument.

6.6.1 Confirmatory Factor Analysis and Mokken Scaling Analysis

The results of the CFA and Mokken scaling analysis are summarised in Table 49. Weak support was offered for the unidimensionality of the SHI once error correlations had been added to the model, but an excellent level of Mokken scalability was displayed without the removal of any items. This again suggests that some dependency is present between some items, but that the SHI item set forms an ordinal scale that is robust enough (within this setting) for a scale cut point to be valid for use in an AUC analysis.

Table 49. CFA and Mokken scaling statistics for the SHI

	SHI	
	Original structure	With Correlated errors
Original Number of scale items	22	22
CFA Results		
Chi-Square (df)	1924 (231)	277 (198)
P value	<0.0000	0.0002
RMSEA	0.053	0.03
CFI	0.846	0.953
TLI	0.83	0.946
Quality of support	None	Weak
Mokken Results		
Number of items removed to form final scale	0	-
Loevinger's H	0.91	-
Interpretation of Scalability	Good	-

6.6.2 Rasch Analysis

Initial analysis of the SHI revealed a few individual elements of misfit, but the overall scale did not appear to be too problematic in terms of fit to the Rasch model. Individual item analysis, however, revealed evidence of some items displaying fit parameters outside of the normally expected and accepted range. The majority of this misfit was attributable to DIF parameters – mainly in the form of Gender DIF, but Prison DIF (unrelated to the Gender DIF) was also present.

As the SHI items are all dichotomously scored, there is no opportunity for item thresholds to be disordered as each item only has a single measurement threshold. Therefore no rescoring is necessary, or possible, among the SHI items.

The initial summary statistics for the SHI can be found in Table 50 (initial), along with the targeting plot (Figure 34). The sources of individual item misfit at this stage are summarised in Table 51.

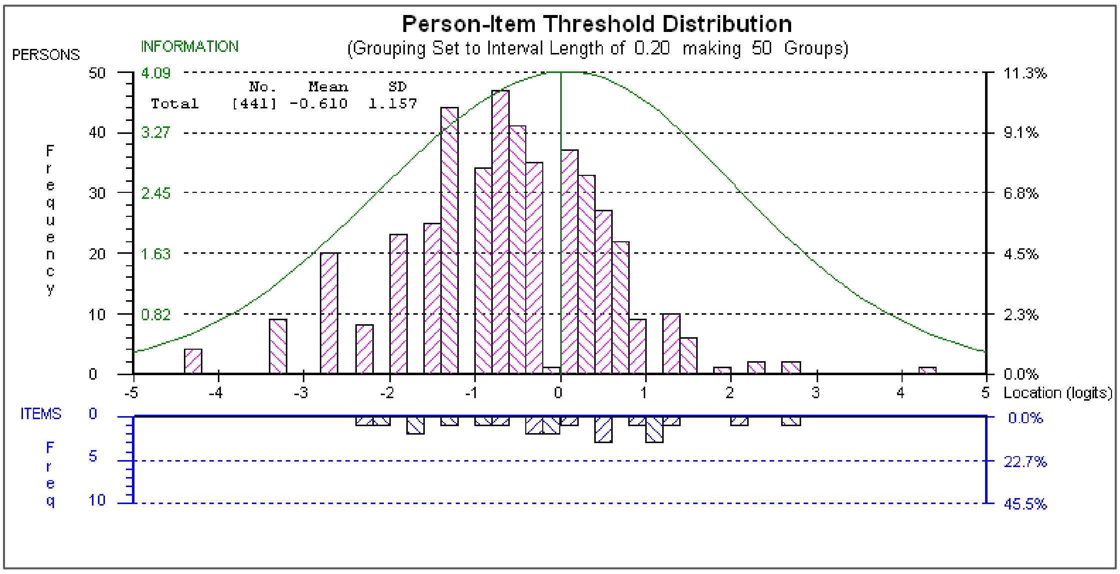

Figure 34. Targeting plot for the SHI

Table 50. Summary Rasch fit statistics for the SHI

		Item Location		Person Location		Item Fit Residual		Person Fit Residual		Chi Square Interaction			PSI		Alpha	Unidimensionality T-Tests (CI)			
Analysis	CInt	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Value	df	p	with extrms	NO extrms		number of signif. tests	out of:	%	lower bound 95% CI
SHI initial	6	0	1.32	-0.61	1.16	-0.1	1.25	-0.19	0.86	141.5	110	0.023	0.76	0.74	0.78	36	435	8.28%	6.20%
SHI Resolution A	6	0	1.58	-0.81	1.13	-0.1	1.05	-0.18	0.66	160	140	0.113	0.75	0.74	0.76	29	435	6.67%	4.60%
SHI Resolution B	6	0	1.13	-0.41	1.23	-0.04	1.07	-0.19	0.87	67.61	65	0.3881	0.651	0.61	0.71	7	427	1.64%*	-
SHI Male	6	0	1.63	-0.8	1.16	-0.07	1.07	-0.16	0.56	128.38	110	0.111	0.753	0.748	0.78	16	322	4.97%	-
SHI Female	2	0	1.41	-0.37	1.22	-0.15	0.65	-0.18	0.91	17.91	22	0.7116	0.781	0.744	0.78	11	113	9.73%	5.70%

* = not sufficient power in t-test procedure (i.e. < 10 thresholds used to generate comparative estimates)

Table 51. Summary of individual sources of misfit within the SHI (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

At this initial stage, there was response dependency that was apparent between Item 1: 'Overdosed?', and Item 18: 'Attempted suicide?' (residual correlation = 0.347), along with a lower level dependency apparent between Item 11: 'Been promiscuous?' and Item 12: 'Set yourself up in a relationship to be rejected?'.

However, the majority of the misfit was attributable to DIF parameters – mainly in the form of Gender DIF, but Prison DIF (unrelated to the Gender DIF) was also present. Religion DIF was also present for Item 14: 'Distanced yourself from God as punishment?'. This was the only Religion DIF present across any of the scales.

6.6.2.1 Scale Refinement

6.6.2.1.1 Resolution A

Resolution A was reached by subtesting Items 1 and 18 together, and Items 11 and 12 together, into separate testlets to account for the apparent response dependency. Additionally, a number of items were sequentially split to account for the apparent DIF. Items 7, 8, 15, 16 and 22 were split for DIF-by-gender, Items 10 and 20 were split for DIF-by-Prison, with only Prison C separated, and Item 14 was split for DIF-by-Gender.

The summary fit statistics at this stage are presented in Table 50 (Resolution A).

6.6.2.1.2 Resolution B

Resolution B was reached following the sequential removal of 9 items; all of which were presenting with some form of DIF. Items 1, 7, 8, 10, 14, 15, 16, 20 and 22 were removed in order to create a set of items which was free from any form of misfit.

The summary fit statistics at this stage are presented in Table 50 (Resolution B).

6.6.2.1.3 Gender Separation

The large amount of Gender DIF that is apparent in the initial SHI analysis suggests that the SHI is functioning differently for males and females. It may therefore be beneficial to treat the SHI as a different scale among male and female ACCT populations. The summary statistics of the initial SHI analysis for the separate male and female samples are presented in Table 50 (labelled SHI Male and SHI Female, respectively). An example of an item displaying a clear Gender DIF is presented in Figure 35.

The sources of individual item misfit for the separate Male and Female samples are also summarised in Table 52 (Males) and Table 53 (Females).

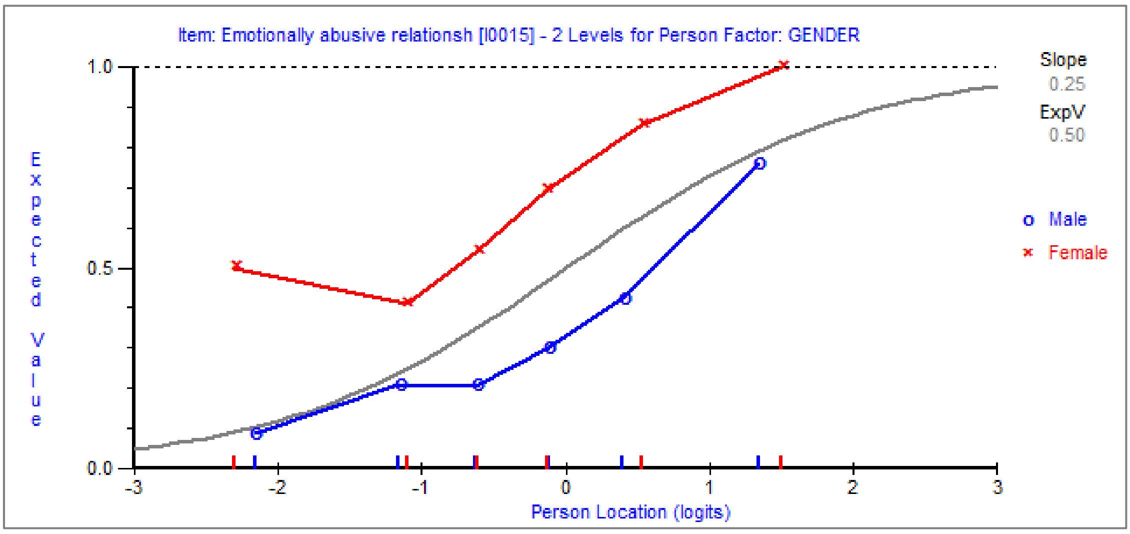

Figure 35. An example of an item ('emotionally abusive relationships') displaying Gender DIF, with females obtaining a higher affirmation rate at all levels of the underlying trait

Table 52. Summary of individual sources of misfit within the SHI – Male sample only (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

Table 53. Summary of individual sources of misfit within the SHI – Female sample only (misfit sources highlighted in red)

Item	Disordered Thresholds	Fit Res>2.5	Fit Res<-2.5	Misfitting Chi-square	Misfitting F-Stat	Prison DIF	Gender DIF	Age DIF	Religion DIF	Response Dependence (Corr>0.2)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

6.7 Summary of Psychometric Properties of Instruments

Through the Confirmatory Factor Analysis, all of the five candidate instruments displayed some level of evidence for the unidimensionality assumption, although this mostly involved some degree of error-correlation to account for apparent dependency within the sets of items. Additionally, although the overall CORE-OM showed no scalability, all of the other instruments displayed at least a moderate level of Mokken criteria scalability, without the need for any scale modifications to be made. The individual CORE domains did display a moderate degree of scalability, but this required some degree of modification in all cases.

Therefore, with the exception of the CORE-OM, it seems that these instruments can be used in their original format within the prison ACCT population to provide ordinal estimates of their respective constructs (i.e. with respect to the magnitude). The CORE-OM in its various subscale forms, however, will require some modification to support internal construct validity in this setting.

The Rasch model is more demanding with regard to its quest for quantitative structure, and this is reflected where data from the instruments are fitted to the model. In their original form, none of the selected instruments completely satisfy all of the requirements of the Rasch model. However, with some refinement, most of the instruments contain a set of items which conform to Rasch model expectations, although the analysis and refinement capabilities are rather limited for the shorter instruments and subscales. Although it is not necessary for an item set to conform to Rasch model measurement standards for the purpose of carrying out an Area Under the Curve analysis, the Rasch analytic process is useful in determining the item-level details and structure regarding the entire instrument. Additionally, it offers the opportunity to refine an item set to a level where interval level measurement can be achieved.

The psychometric analyses that were carried out provide details regarding the internal construct validity of the instruments, meaning that they assess whether all of the items within a scale are measuring the same thing in a reliable way. However, what this does not provide is evidence that the instruments measure what they intend to measure, which is provided by a sound conceptual basis and evidence of external construct validity that should be part of the scale development process.

7 Predictive Analysis

This chapter reports on the results of the predictive validity element of the study. This includes the results of the area-under-the-curve (AUC) analysis for each of the instruments that were used within the cohort study, along with the corresponding receiver-operating-characteristic (ROC) curves where appropriate. Additionally, this chapter reports the results from the exploratory analysis of individual predictors of self-harm, and how a logistic regression can be used to combine these individual items in order to create a screening algorithm instrument that could potentially be used as part of the self-harm monitoring process.

7.1 Area Under the Curve Analysis

The Area Under the Curve (AUC) analysis was run on all of the instruments (and subscales) to assess the predictive capabilities of each scale, in terms of the final outcome of whether a prisoner carried out a self-harm event during the follow-up period.

7.1.1 AUC on Original Total Instrument Scores

The AUC results for all original total instrument scores are summarised in Table 54. An AUC of 1 represents a scale that can perfectly discriminate between prisoners who will and will not self-harm, and an AUC of 0.5 represents as scale giving a 50:50 chance of correctly discriminating between prisoners who will and will not self-harm. Where the AUC is significantly different from the null hypothesis assuming an AUC 0.5, the Receiver Operating Characteristic (ROC) curves are presented.

The only instrument scores which offered a significant predictive value on the original total score were the PriSnQuest and the Self-Harm Inventory. The corresponding ROC curves for these instruments are presented in Figure 36 and Figure 37.

Table 54. Summary of AUC analysis for all instrument and subscale scores, with original scoring applied (significant results are highlighted in green)

Instrument	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
BSL-23-F Total Score	0.524	0.031	0.443	0.463	0.585
Average BSL-23-F	0.529	0.031	0.353	0.468	0.59
CORE - Total OM	0.52	0.032	0.525	0.458	0.583
Average CORE - Total OM	0.52	0.031	0.515	0.459	0.581
CORE-Well-Being	0.491	0.032	0.779	0.429	0.554
Average CORE Well-Being	0.492	0.031	0.802	0.431	0.554
CORE-Problems	0.501	0.031	0.971	0.44	0.562
Average CORE-Problems	0.501	0.031	0.967	0.441	0.562
CORE-Functioning	0.517	0.031	0.583	0.457	0.578
Average CORE-Functioning	0.522	0.03	0.486	0.462	0.581
CORE-Risk	0.543	0.031	0.162	0.481	0.605
Average CORE-Risk	0.543	0.031	0.163	0.481	0.604
CORE - non Risk	0.504	0.032	0.89	0.442	0.567
Average CORE - non Risk	0.508	0.031	0.796	0.447	0.569
CORE-10	0.496	0.03	0.889	0.436	0.555
Average CORE-10 score	0.491	0.03	0.773	0.432	0.55
PHQ-9 Total Score	0.503	0.031	0.928	0.443	0.563
PHQ-2 Total Score	0.509	0.031	0.762	0.449	0.57
PriSnQuest Total score	0.565	0.03	0.038	0.506	0.624
SHI Total Score	0.566	0.031	0.035	0.506	0.626

a. Under the nonparametric assumption b. Null hypothesis: true area = 0.5

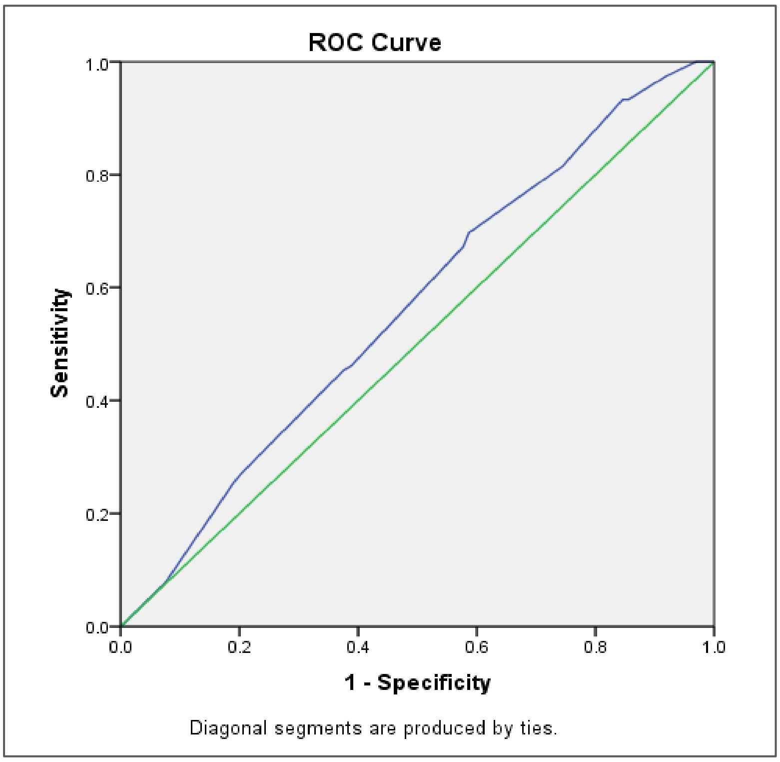

Figure 36. ROC curve of the PriSnQuest

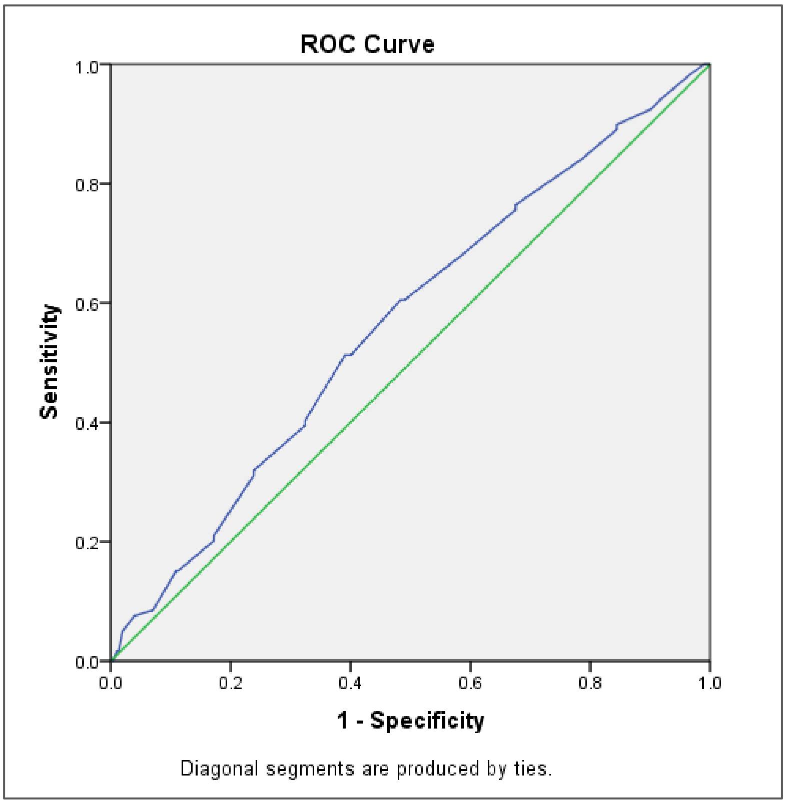

Figure 37. ROC curve of the Self-Harm Inventory

7.1.2 AUC on Rasch-refined Instrument Scores

Additionally, the AUC analysis was run for the optimal resolution resulting from the Rasch analysis for each scale and subscale (see Table 55). The logit estimates for each person were converted back into an equivalent raw score for the items (and scoring parameters) which comprised the final item set. Resolution B was used in the majority of instances, but where this was not available, the rescored scale analysis was used. If no rescore was applicable, then the conversion was based on the initial analysis (as per Table 55). The PHQ-9 also offers a Resolution A to utilise, and the PriSnQuest was separated into gender specific conversions, as suggested by the Rasch analysis (see Section 6.5.2).

Table 55. Summary of AUC analysis for all instrument and subscale Rasch converted scores (significant results are highlighted in green)

Instrument	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
BSL-23-F Resolution B conversion 0-28	0.507	0.031	0.831	0.447	0.567
CORE-OM Resolution B conversion 0-34	0.527	0.031	0.387	0.466	0.587
CORE Well Being rescored conversion 0-8	0.511	0.031	0.725	0.45	0.572
CORE Problems Resolution B conversion 0-16	0.506	0.031	0.833	0.447	0.566
CORE Functioning Resolution B conversion 0-18	0.531	0.03	0.319	0.472	0.59
CORE Risk rescored conversion 0-12	0.54	0.031	0.194	0.479	0.601
CORE Non-Risk Resolution B conversion 0-30	0.525	0.031	0.412	0.465	0.585
CORE10 Resolution B conversion 0-16	0.493	0.03	0.814	0.434	0.551
PHQ9 Resolution A conversion 0-18	0.508	0.031	0.809	0.447	0.568
PHQ9 Resolution B conversion 0-16	0.511	0.031	0.732	0.45	0.571
PHQ 2 Location conversion 0-6	0.511	0.031	0.719	0.451	0.572
PriSnQuest initial conversion 0-8	0.567	0.03	0.03	0.508	0.626
PriSnQuest male subtest conversion 0-8	0.58	0.036	0.028	0.51	0.65
PriSnQuest female subtest conversion 0-8	0.53	0.057	0.606	0.418	0.642
SHI Resolution B conversion 0-13	0.581	0.03	0.009	0.521	0.641

a. Under the nonparametric assumption b. Null hypothesis: true area = 0.5

Again, the only scale scores which offered a significant predictive value were the PriSnQuest (initial and male specific resolution) and the Self-Harm Inventory. The corresponding ROC curves are presented in Figure 38, Figure 39, and Figure 40.

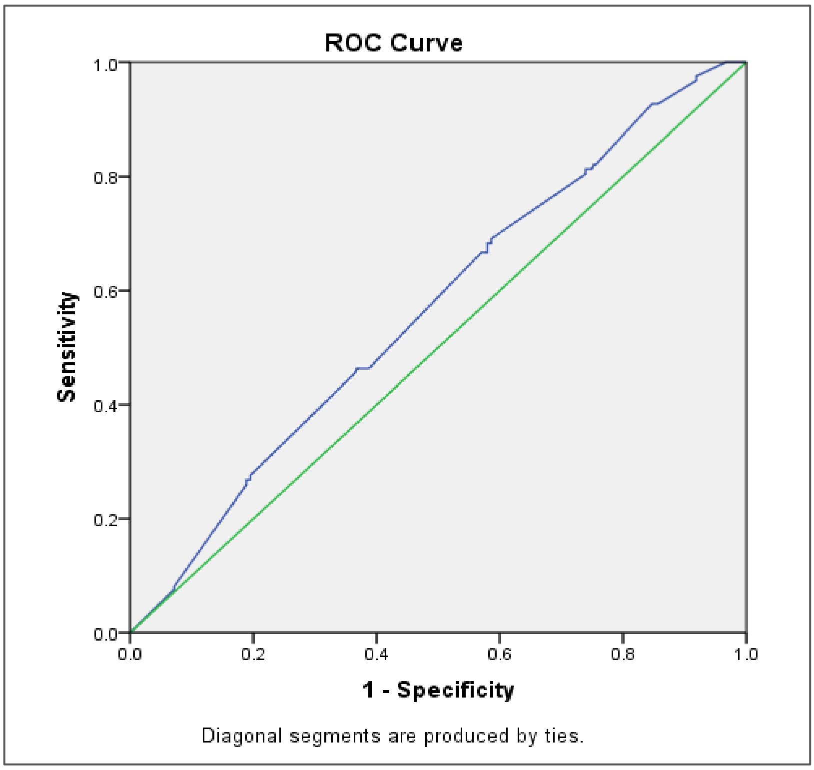

Figure 38. ROC curve of the PriSnQuest initial, converted from Rasch estimates

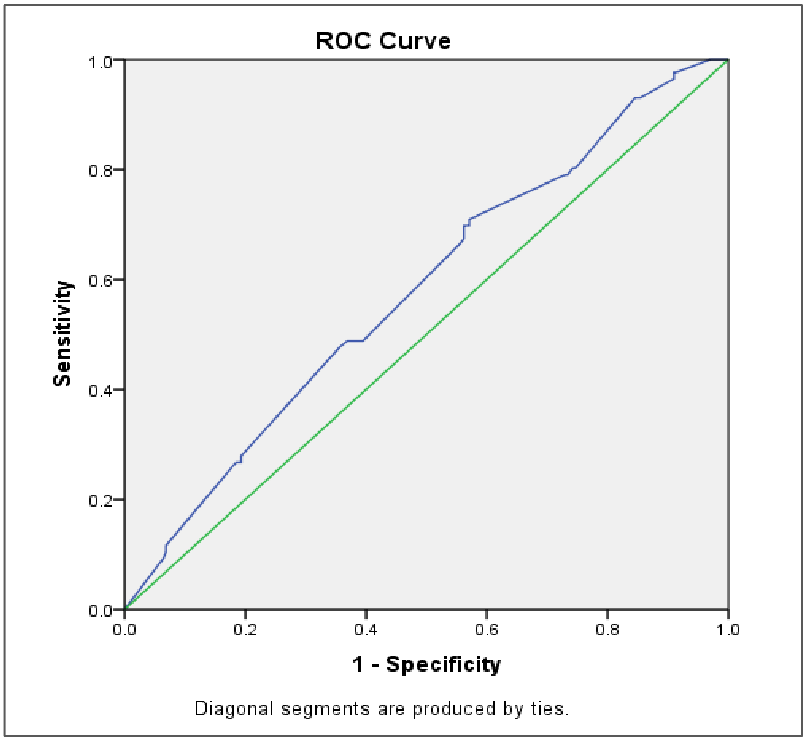

Figure 39. ROC curve of the PriSnQuest (male specific), converted from Rasch estimates

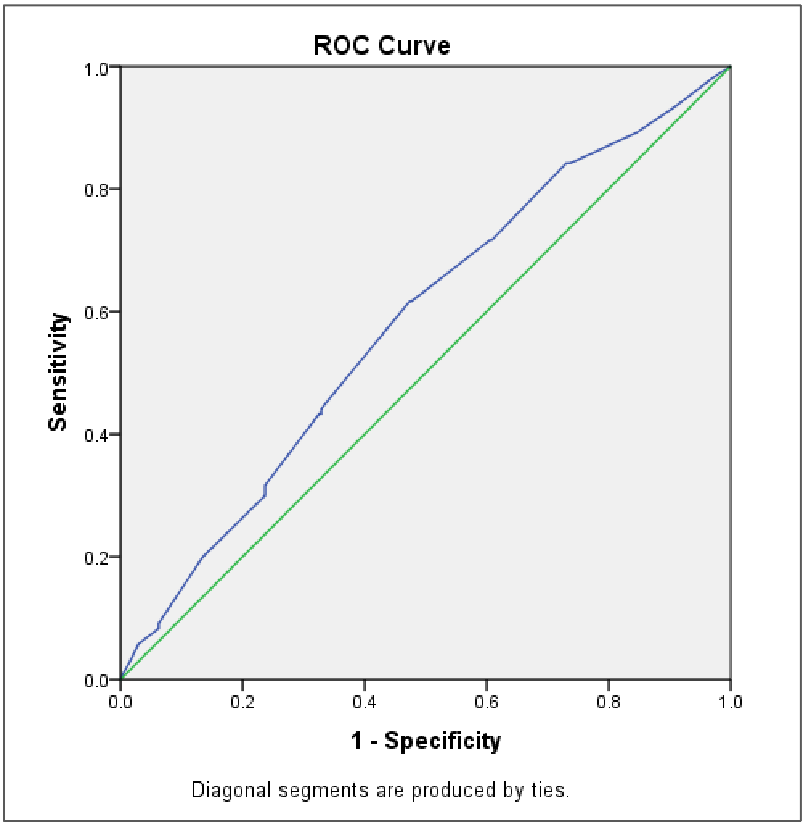

Figure 40. ROC curve of the Self-Harm Inventory, converted from Rasch estimates

7.1.3 Gender-specific AUC

Based on indications in the literature, (137, 228) along with the indications provided within the Rasch analysis, the AUC analysis was repeated on a gender specific basis to assess whether gender-specific results differed from the results of the collated (not gender-specific) analysis.

7.1.3.1 Male-specific AUC

The male-specific AUC results for the original scoring across all instruments are summarised in Table 56.

Table 56. Summary of male-specific AUC analysis for all scale and subscale scores (significant results are highlighted in green)

Instrument	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
BSL-23-F Total Score	0.541	0.037	0.273	0.469	0.614
Average BSL-23-F	0.545	0.037	0.229	0.472	0.617
CORE - Total OM	0.542	0.037	0.268	0.47	0.614
Average CORE - Total OM	0.541	0.036	0.266	0.471	0.611
CORE-Well-Being	0.486	0.038	0.711	0.413	0.56
Average CORE Well-Being	0.488	0.037	0.738	0.415	0.561
CORE-Problems	0.54	0.037	0.282	0.469	0.612
Average CORE-Problems	0.538	0.036	0.294	0.468	0.609
CORE-Functioning	0.524	0.036	0.518	0.454	0.594
Average CORE-Functioning	0.532	0.035	0.385	0.463	0.601
CORE-Risk	0.56	0.037	0.098	0.488	0.633
Average CORE-Risk	0.56	0.037	0.099	0.488	0.632
CORE - non Risk	0.527	0.037	0.48	0.454	0.599
Average CORE - non Risk	0.531	0.036	0.397	0.461	0.601
CORE-10	0.515	0.035	0.682	0.446	0.585
Average CORE-10 score	0.509	0.035	0.801	0.441	0.578
PHQ-9 Total Score	0.543	0.036	0.243	0.473	0.614
PHQ-2 Total Score	0.536	0.036	0.33	0.465	0.607
PriSnQuest Total score	0.577	0.036	0.04	0.506	0.647
SHI Total Score	0.517	0.038	0.656	0.443	0.59

a. Under the nonparametric assumption b. Null hypothesis: true area = 0.5

With the original scoring applied, the only instrument which offered a significant predictive value among the male sample was the PriSnQuest. The corresponding ROC curve is presented in Figure 41.

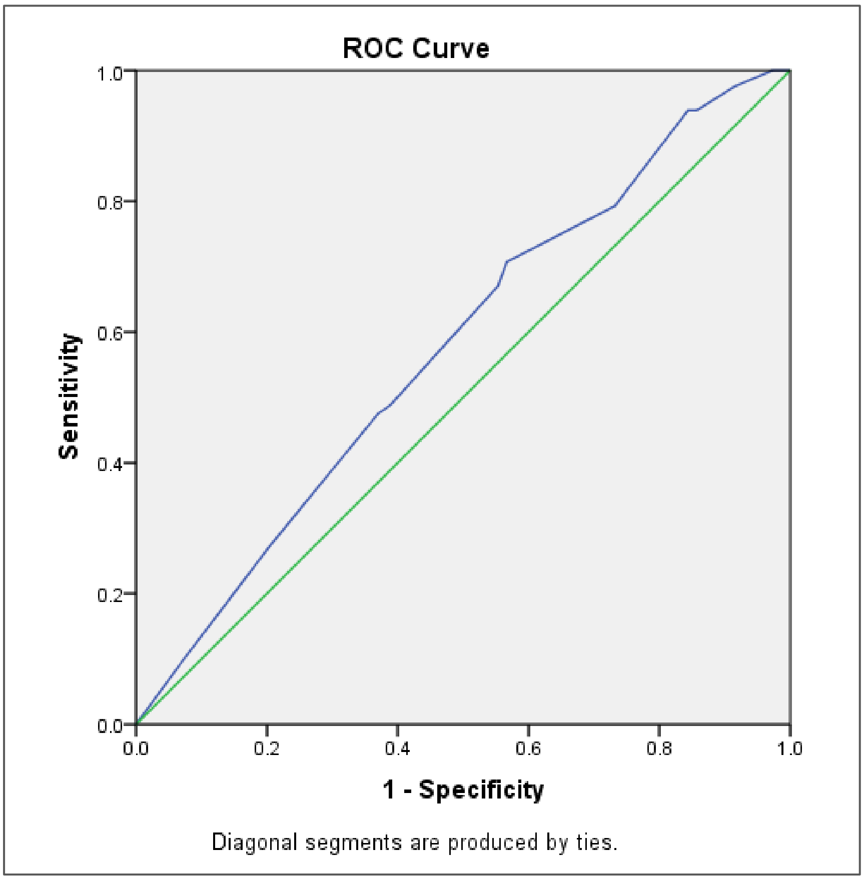

Figure 41. ROC curve of the PriSnQuest (original scoring) for males

Again, the analysis was repeated among the male-specific sample, utilising the Rasch-transformed scores for the instruments in the AUC analysis.

The male-specific Rasch-transformed AUC results for all scales are summarised in Table 57.

Table 57. Summary of AUC analysis for all instrument and subscale Rasch converted scores for males (significant results are highlighted in green)

				Asymptotic 95% Confidence Interval	
Instrument	Area	Std. Error[a]	Asymptotic Sig.[b]	Lower Bound	Upper Bound
BSL-23-F Resolution B conversion 0-28	0.518	0.036	0.618	0.448	0.589
CORE-OM Resolution B conversion 0-34	0.546	0.035	0.21	0.476	0.615
CORE-OM Resolution B conversion 0-34	0.546	0.035	0.21	0.476	0.615
CORE Well Being rescored conversion 0-8	0.51	0.037	0.78	0.438	0.582
CORE Problems Resolution B conversion 0-16	0.529	0.035	0.418	0.46	0.599
CORE Functioning Resolution B conversion 0-18	0.547	0.034	0.194	0.48	0.615
CORE Risk rescored conversion 0-12	0.556	0.036	0.122	0.485	0.628
CORE Non-Risk Resolution B conversion 0-30	0.552	0.035	0.151	0.484	0.621
CORE10 Resolution B conversion 0-16	0.515	0.035	0.681	0.447	0.583
PHQ9 Resolution A conversion 0-18	0.545	0.036	0.225	0.474	0.616
PHQ9 Resolution B conversion 0-16	0.546	0.036	0.209	0.476	0.617
PHQ 2 Location conversion 0-6	0.538	0.036	0.307	0.466	0.609
PriSnQuest initial conversion 0-8	0.579	0.036	0.031	0.509	0.648
PriSnQuest male subtest conversion 0-8	0.58	0.036	0.028	0.51	0.65
SHI Resolution B conversion 0-13	0.549	0.038	0.19	0.475	0.622

a. Under the nonparametric assumption b. Null hypothesis: true area = 0.5

The only instrument scores which offered a significant predictive value for males were the two alternative conversions of the PriSnQuest (initial and male-specific resolution). The corresponding ROC curves are presented in Figure 42 and Figure 43.

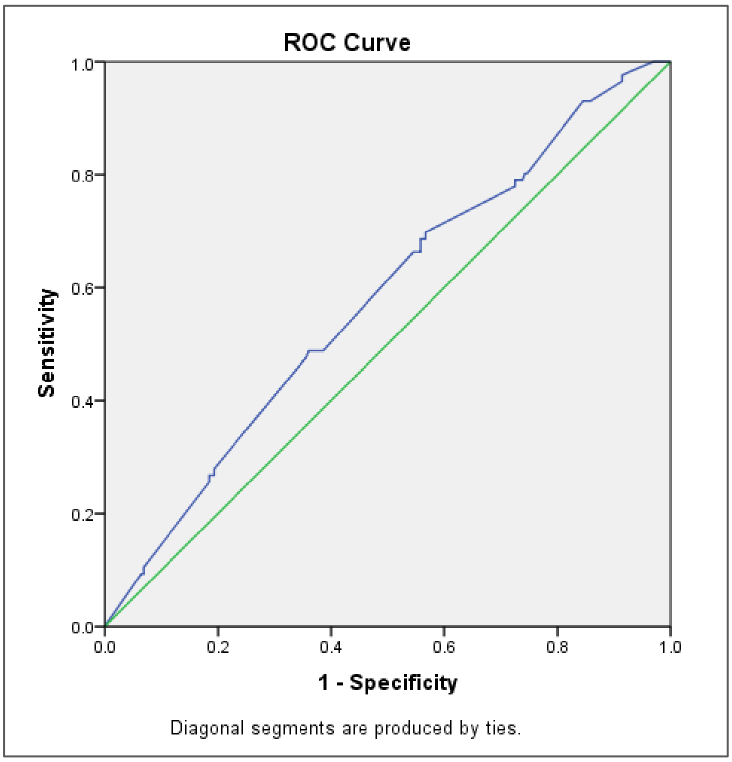

Figure 42. ROC curve of the PriSnQuest initial for males, converted from Rasch estimates

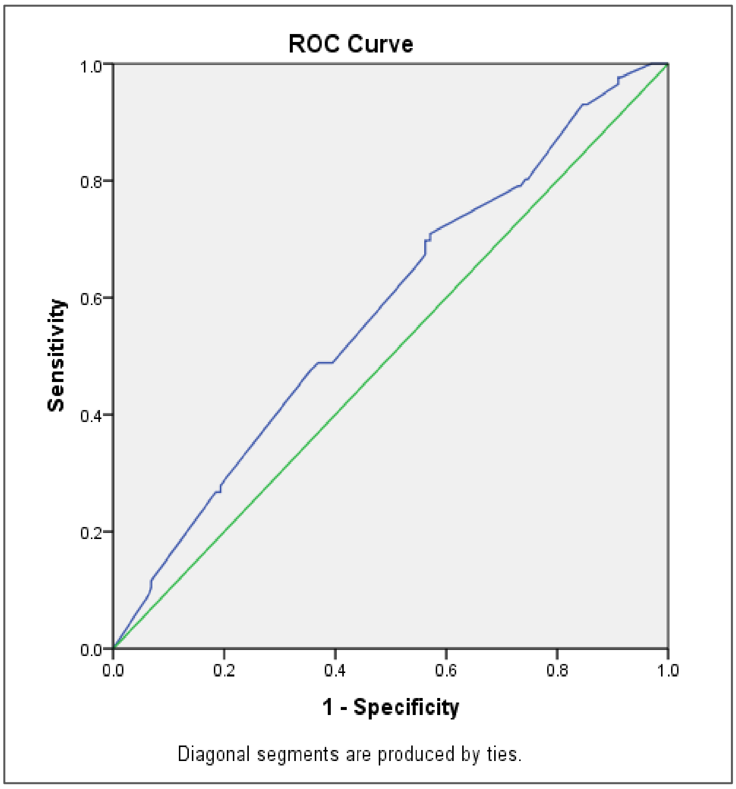

Figure 43. ROC curve of the PriSnQuest (male specific conversion) for males, converted from Rasch estimates

7.1.3.2 Female-specific AUC

The female-specific AUC results for the original scoring across all instruments are summarised in Table 58.

Table 58. Summary of female-specific AUC analysis for all scales and subscale scores (significant results are highlighted in green)

Instrument	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
BSL-23-F Total Score	0.483	0.058	0.773	0.369	0.597
Average BSL-23-F	0.494	0.058	0.92	0.381	0.607
CORE - Total OM	0.474	0.062	0.653	0.352	0.595
Average CORE - Total OM	0.473	0.062	0.641	0.351	0.594
CORE-Well-Being	0.499	0.059	0.98	0.382	0.615
Average CORE Well-Being	0.499	0.059	0.98	0.382	0.615
CORE-Problems	0.416	0.058	0.151	0.302	0.53
Average CORE-Problems	0.416	0.058	0.151	0.302	0.53
CORE-Functioning	0.504	0.06	0.947	0.386	0.622
Average CORE-Functioning	0.499	0.06	0.99	0.381	0.617
CORE-Risk	0.511	0.059	0.854	0.395	0.627
Average CORE-Risk	0.511	0.059	0.854	0.395	0.627
CORE - non Risk	0.458	0.061	0.471	0.337	0.578
Average CORE - non Risk	0.456	0.061	0.455	0.336	0.577
CORE-10	0.456	0.058	0.457	0.343	0.57
Average CORE-10 score	0.453	0.058	0.418	0.339	0.566
PHQ-9 Total Score	0.417	0.057	0.154	0.305	0.528
PHQ-2 Total Score	0.466	0.058	0.563	0.353	0.579
PriSnQuest Total score	0.53	0.057	0.606	0.418	0.642
SHI Total Score	0.671	0.051	0.003	0.57	0.771

a. Under the nonparametric assumption b. Null hypothesis: true area = 0.5

The only instrument score which offered a significant predictive value among the female sample was the SHI. The corresponding ROC curve is presented in Figure 44.

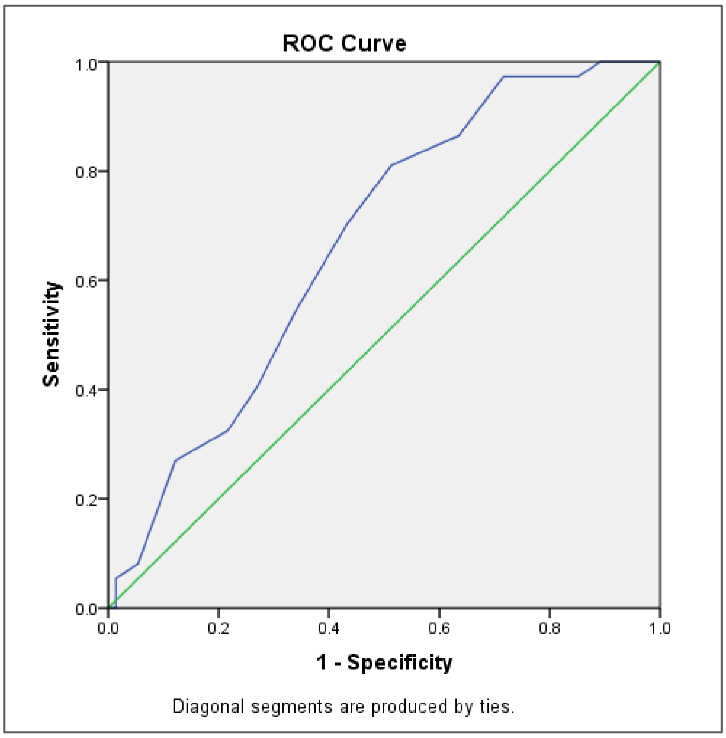

Figure 44. ROC curve of the SHI for females

This analysis was repeated among the female-specific sample, utilising the Rasch-transformed scores for the instruments in the AUC analysis.

The female-specific Rasch-transformed AUC results for all scales are summarised in Table 59.

Table 59. Summary of AUC analysis for all scales and subscale Rasch converted scores for females (significant results are highlighted in green)

Instrument	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
BSL-23-F Resolution B conversion 0-28	0.493	0.058	0.898	0.379	0.606
CORE-OM Resolution B conversion 0-34	0.476	0.061	0.687	0.356	0.597
CORE Well Being rescored conversion 0-8	0.511	0.059	0.856	0.394	0.627
CORE Problems Resolution B conversion 0-16	0.455	0.06	0.436	0.338	0.571
CORE Functioning Resolution B conversion 0-18	0.492	0.059	0.893	0.376	0.609
CORE Risk rescored conversion 0-12	0.514	0.059	0.805	0.399	0.63
CORE Non-Risk Resolution B conversion 0-30	0.456	0.06	0.453	0.338	0.575
CORE10 Resolution B conversion 0-16	0.446	0.059	0.359	0.331	0.562
PHQ9 Resolution A conversion 0-18	0.422	0.057	0.18	0.31	0.533
PHQ9 Resolution B conversion 0-16	0.427	0.057	0.21	0.314	0.539
PHQ 2 Location conversion 0-6	0.466	0.058	0.561	0.353	0.579
PriSnQuest initial conversion 0-8	0.53	0.057	0.606	0.418	0.642
PriSnQuest female subtest conversion 0-8	0.53	0.057	0.606	0.418	0.642
SHI Resolution B conversion 0-13	0.654	0.052	0.009	0.552	0.756

a. Under the nonparametric assumption b. Null hypothesis: true area = 0.5

As with the original instrument scoring, the only scale score which offered a significant predictive value for females was the conversion of the SHI Resolution B. The corresponding ROC curve is presented in Figure 45.

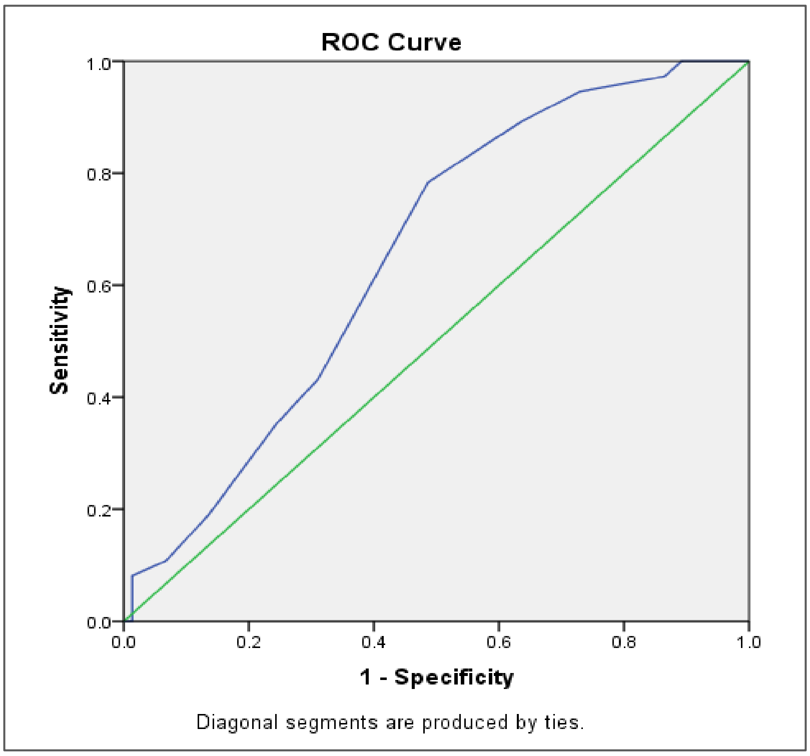

Figure 45. ROC curve of the SHI Resolution B for females, converted from Rasch estimates

7.1.4 AUC Analysis Summary

The primary aim of the study was to determine whether any pre-existing instruments could predict self-harm among an ACCT population. From the candidate instruments that were selected for the study, none of these performed the task adequately enough to be considered a useful aid for prison staff to utilise as part of a standardised ACCT process. Although two scales demonstrated an AUC significantly different from 0.5, all scales failed to have any meaningful predictive value. The utilisation of Rasch-transformed scores within the AUC analysis had very little effect on the results, although the Rasch analytic process revealed useful diagnostic information regarding the items within the instruments, and highlighted the gender biases that were present within the SHI, which informed the AUC analysis in terms of gender separation. Although the results of the AUC analysis were improved to an extent when gender separation was considered, there were still no meaningful predictive values observed among the instruments. Across all analyses, the highest observed AUC value was 0.671 (for the female-specific SHI), and even this should not be considered as anything greater than a 'poor' level of discrimination.

7.2 Self-Harm Predictive Algorithm

The failure of the candidate screening instruments to predict future self-harm, while disappointing, was not entirely unexpected. Each of the scales may contain some items that do discriminate with regard to predicting self-harm, but they also probably contain many items that do not discriminate for self-harm. Due to this dilution of the discriminating questions, the instrument total scores on which the AUC analysis is based may be compromised with respect to predicting self-harm. It was the hope that one of the extant, standardised instruments would be predictive of self-harm, but it was recognised that this may not be the case. It was therefore anticipated that it may be necessary to examine the potential of individual items as predictors, with one of the study aims stating; *"If a stand-alone instrument is not predictive, then determine whether any set of demographic or individual instrument items combine to predict self-harm"*.

7.2.1 Predictive Assessment of Individual Items

The candidate instruments contain 105 individual items, and these form an item pool of potential risk indicators together with other socio-demographic and sentencing criteria (e.g. a prisoner's education level, or remand status etc.). The psychometric and AUC analysis is also informative in that there seems to be a gender difference in the way in which instruments (and their constituent items) worked, and therefore this assessment of individual predictor items was carried out on a gender-specific basis.

Therefore, to refine the item set down to an appropriate number to enter into a logistic regression, all items were assessed for their capacity to potentially indicate risk of self-harm. This exploratory analysis was achieved by running individual item crosstab chi-square tests, with all items individually associated with future self-harm at p=0.10 progressing on to the logistic regression. To account for the small number of self-harm cases within individual response category groups, all items with multi-category response options were dichotomised into categories that represented a 'complete absence' and 'some presence' of either a sign or symptom. The individual items that were statistically significantly associated with future self-harm at p=0.05 are included in Table 60.

Table 60. Items and other indicators associated ($p<0.05$) with future self-harm, separated by gender

Male Risk Factors							
Variable (Odds Ratios refer to affirmation of variable)	p-value	OR	CI	Sensitivity	Specificity	PPV	NPV
Prisoner has no qualifications (no qualifications = 1)	0.001	2.335	1.409-3.872	62.10%	58.80%	36.00%	80.60%
BSL1. In the last week it was hard for me to concentrate	0.017	>100	not calculated	100%	6.50%	27.80%	100%
BSL S8. During the last week I had uncontrollable sexual encounters of which I was later ashamed or which made me angry	0.028	8.475	0.869-82.65	3.60%	99.60%	75%	73.90%
Ever Self-Harmed in prison	0	3.423	1.967-5.958	75.90%	52.10%	37.10%	85.30%
Index ACCT due to Self-Harm?	0	3.42	1.986-6.836	71.40%	59.60%	46.70%	80.80%
SHI2. Have you ever cut yourself on purpose?	0	3.075	1.604-5.894	84.10%	36.70%	32.20%	86.60%
PQuest2. In the past year have you been taking longer over the things you do?	0.024	2	1.086-3.685	80.70%	32.30%	29.90%	82.40%
Ever received medication for mental health problems	0.023	1.981	1.091-3.596	80.50%	32.50%	30.70%	81.70%
SHI19. Have you ever exercised an injury on purpose?	0.045	1.858	1.007-3.427	25.30%	84.60%	37.50%	75.60%
PQuest1. In the past year have you previously seen a psychiatrist?	0.018	1.821	1.105-3.001	54.70%	60.20%	33.80%	78.10%
Acquisitive Crime (Burglary, Robbery, Theft)	0.043	1.712	1.015-2.89	38.80%	73.00%	34.40%	76.60%
Male Protective Factors							
Age left full time education (16+ = 1)	0.034	0.578	0.348-0.962	37.60%	48.90%	21.30%	68.10%
SHI6. Have you ever Abused alcohol?	0.028	0.559	0.332-0.941	57.80%	28.90%	22.90%	65.30%
Dependent on alcohol	0.013	0.497	0.284-0.867	24.10%	60.90%	18.80%	68.30%
CORE19. Over the last week I have felt warmth or affection for someone (with scoring reversed) 1=Less than all the time	0.003	0.476	0.288-0.786	49.40%	32.80%	21.60%	63.30%

Female Risk Factors							
Variable (Odds Ratios refer to affirmation of variable)	p-value	OR	CI	Sensitivity	Specificity	PPV	NPV
Life or indefinite sentence	0	8.4	2.479-28.46	32.40%	94.60%	75%	73.70%
SHI2. Have you ever cut yourself on purpose?	0.01	4.795	1.331-17.269	91.90%	29.70%	39.60%	88%
PHQ-9-7. Over the last 2 weeks - Trouble concentrating on things, such as reading the newspaper or watching television	0.04	4.449	0.96-20.619	94.60%	20.30%	37.20%	88.20%
PQuest8. In the past year have you recently heard voices saying a few words or sentences when there was no one around to account for this?	0.001	4.19	1.768-9.928	73%	60.80%	48.20%	81.80%
CORE 25. Over the last week I have felt criticised by other people	0.003	3.544	1.501-8.366	73%	56.80%	45.80%	80.80%
SHI3. Have you ever Burned yourself on purpose?	0.011	3.145	1.269-7.793	37.80%	83.80%	53.80%	72.90%
Ever Self-Harmed in prison	0.05	3.056	0.96-9.723	89.20%	27%	37.90%	83.30%
Index ACCT due to Self-Harm?	0.017	2.9	1.189-7.084	63.30%	62.70%	43.20%	79.20%
SHI8. Have you ever Scratched yourself on purpose?	0.015	2.708	1.203-6.096	59.50%	64.90%	45.80%	76.20%
SHI10.Have you ever made medical situations worse on purpose (e.g. skipped medication)?	0.017	2.664	1.174-6.044	51.40%	71.60%	47.50%	74.60%
BSL S4. During the last week I had episodes of binge eating	0.031	2.609	1.079-6.309	37.80%	81.10%	50%	72.30%
SHI21. Have you ever starved yourself to hurt yourself?	0.022	2.588	1.132-5.918	67.60%	55.40%	43.10%	77.40%
SHI9. Have you ever prevented wounds from healing?	0.032	2.41	1.071-5.419	62.20%	59.50%	43.40%	75.90%
BSL 15. Over the last week I suffered from voices and noises from inside or outside my head	0.032	2.41	1.071-5.419	62.20%	59.50%	43.40%	75.90%
Female Protective Factor							
First time on an ACCT?	0	0.224	0.096-0.523	36.10%	28.40%	19.70%	47.70%

7.2.2 Logistic Regression

Following the exploratory analysis into which individual items could be considered for a predictive algorithm, all items which were individually significant at p=0.10 were tested for tolerance and multicollinearity, and then entered into a backwards stepwise binary logistic regression, under a likelihood-ratio removal process (p removal 0.1). (214) This was carried out separately for the male and female samples.

Following the initial analysis run, a composite item of 'Prison self-harm history' was created from three individual items: 'Have you ever self-harmed in prison?', 'Was the prisoner's index ACCT due to self-harm?', and Item 1 of the BSL supplement 'During the last week I hurt myself by cutting, burning, strangling, head banging etc.'. This grouped the prisoners into three categories: those that had never self-harmed in prison; those that had self-harmed in prison, but not recently (not within the previous two weeks); and those that had self-harmed in prison recently (within the previous two weeks). The composite item was significantly predictive for the male sample, so it was used instead of the constituent items. It was not significantly predictive for the female sample, so the individual items were retained.

Additionally at this point, the male sample statistical analysis was switched from SPSS to STATA 13 (193), as STATA offered the opportunity to apply a Firth adjustment (229) following the discovery of complete separation within the data set, which can occur when the (self-harm) event numbers are limited. Where complete separation occurs within the data, the maximum likelihood values of the logistic regression cannot be estimated, and the Firth adjustment allows for the convergence of finite estimates; therefore reducing the bias within the analysis. (230)

The final models contained 11 independent variables for men (see Table 61) and seven independent variables for women (see Table 62). Both models were statistically significant, (male model: χ^2 (df 12, N = 301) = 47.57, $p < 0.001$; and female model: χ^2 (df 7, N = 94) = 53.46, $p < 0.001$) indicating that the models were able to distinguish between prisoners who went on to carry out a self-harm event in the follow-up, and those who did not. Seven of the 11 independent variables in the male model, and five of the seven independent variables in the female model made a unique statistically significant contribution to the final models.

Table 61. Logistic Regression predicting likelihood of self-harm during follow-up for males (STATA output)

Variable	B	S.E.	z	Sig.	95% C.I.	
					Lower	Upper
Do you have any qualifications? (yes=0, no=1)	1.122977	0.31065	3.61	0	0.514114	1.731839
Have you accessed healthcare during this prison stay?	-1.14773	0.442399	-2.59	0.009	-2.01481	-0.28064
In the past year, have you previously seen a psychiatrist?	0.660485	0.308926	2.14	0.033	0.055001	1.26597
Have you ever cut yourself on purpose?	0.785021	0.396684	1.98	0.048	0.007535	1.562508
Have you ever abused alcohol?	-1.06009	0.331151	-3.2	0.001	-1.70914	-0.41105
Have you ever driven recklessly on purpose?	-0.6994	0.32734	-2.14	0.033	-1.34098	-0.05783
Have you ever intentionally exercised an injury to hurt yourself?	0.670756	0.370553	1.81	0.07	-0.05551	1.397025
In the last week have you felt warmth or affection for someone?	-0.532	0.304642	-1.75	0.081	-1.12909	0.065083
In the last week, have you thought that you are to blame for your problems and difficulties	1.029772	0.58036	1.77	0.076	-0.10771	2.167257
In the last week, has it been hard for you to concentrate?	1.998831	1.518229	2.84*	0.092**	-0.97684	4.974505
Prisoners self-harm history in prison: ('no prison self-harm history' is reference category)			11.36*	0.003***		
self-harmed, but not recently	0.922775	0.557201	1.66	0.098	-0.16932	2.014869
self-harmed recently	1.526448	0.468689	3.26	0.001	0.607833	2.445062
Constant	-4.36395	1.648187	-2.65	0.008	-7.59434	-1.13357

* Chi-square value ** Firth-adjusted p-value ***Overall significance of categorical item

Table 62. Logistic Regression predicting likelihood of self-harm during follow-up for females (SPSS output)

Variable	B	S.E.	Wald	df	Sig.	Odds Ratio	95% C.I. for Odds Ratio	
							Lower	Upper
Life or indeterminate sentence?	2.016	1.143	3.112	1	0.078	7.506	0.799	70.488
Has prisoner had ANY sort of correspondence during stay (yes or no)?	3.698	1.985	3.471	1	0.062	40.351	0.825	1973.135
Have you ever seen a psychiatrist outside prison?	1.453	0.739	3.867	1	0.049	4.274	1.005	18.183
Is this the first time in this sentence that you have been put on an ACCT?	-2.027	0.762	7.086	1	0.008	0.132	0.03	0.586
Have you ever intentionally scratched yourself on purpose?	2.362	0.74	10.2	1	0.001	10.617	2.491	45.252
During the last week I had episodes of binge eating.	2.714	0.867	9.806	1	0.002	15.096	2.761	82.544
During the last week I took medication that had not been prescribed or if had been prescribed, I took more than the prescribed dose.	2.213	0.878	6.349	1	0.012	9.139	1.635	51.093
Constant	-7.022	2.4	8.563	1	0.003	0.001		

For each prisoner on an ACCT, a risk score can be calculated by multiplying each variable with the regression coefficient of the prediction model. To create a more easily applicable prediction rule, regression coefficients were rounded to half points and then doubled to form simple summative indices of complete numbers. (231) This was again done separately for males and females. The receiver operating characteristic (ROC) curve for the male prediction model is displayed as Figure 46, and the ROC curve for the female prediction model is displayed as Figure 47. When maximising the Kappa value in the agreement between the prediction model and the outcome of self-harm, corresponding AUC values are 0.81 for males and 0.867 for females. The properties of the gender-specific predictive models are summarised in Table 63.

Table 63. Properties of the gender-specific predictive models

Predictive Algorithm	AUC	Sensitivity	Specificity	PPV	NPV	Correctly Classified
Male	0.81	55%	85.5%	57.9%	84%	77.4%
Female	0.867	71.4%	93.1%	83.3%	87%	86%

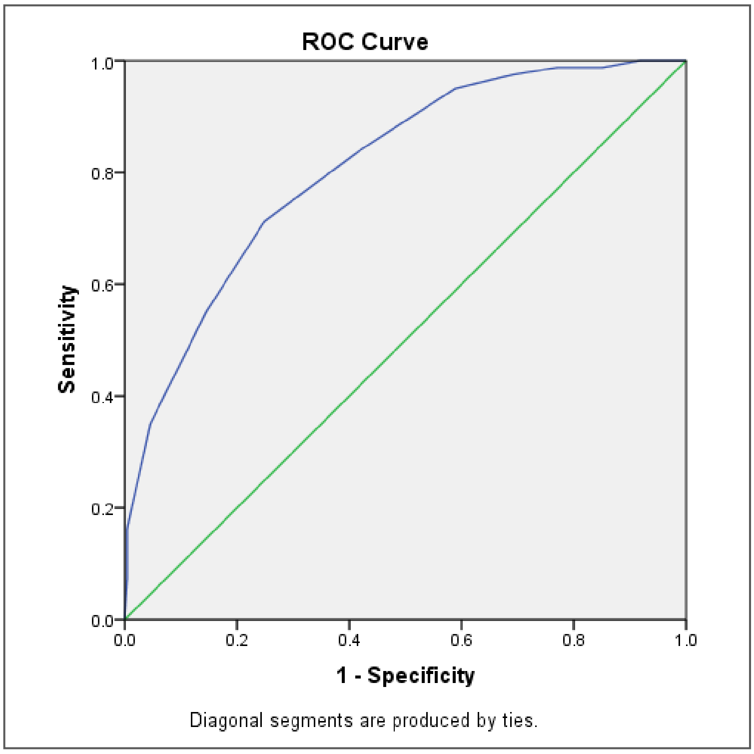

Figure 46. ROC curve for male predictive risk model (AUC = 0.81)

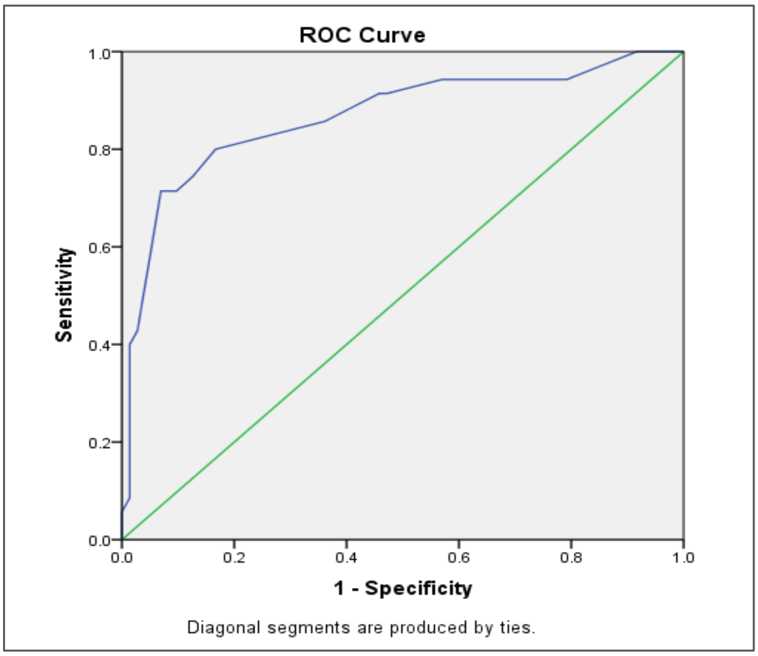

Figure 47. ROC curve for female predictive risk model (AUC = 0.867)

For these values that are presented, it should be noted that specificity and sensitivity are properties of the instrument, whereas PPV and NPV differ by the self-harm prevalence rate within a given population. As the self-harm rate varied by prison, the PPV and NPV will therefore differ across institutions, although this will only apply to the male institutions as the female institution was considered separately.

For the sensitivity and specificity values obtained within the male prisons, where the self-harm prevalence rate is lower (i.e. Prison A), the PPV will also be lower, but the NPV will be higher. This means that there will be a higher proportion of false positive results of the screening test, but a lower proportion of false negatives.

Where the self-harm prevalence rate is higher (i.e. Prison C), the PPV will also be higher, but the NPV will be lower. This means that there will be a lower proportion of false positive results of the screening test, but a higher proportion of false negatives.

In order to put the sensitivity and specificity values of these prediction models into context, and to demonstrate the absolute numbers that would be found, a range of example scenarios are presented that replicate the institutional findings within this study. In each scenario, the values relate to the inferred screening of one thousand prisoners. These examples are presented for comparison in Table 64. Refer back to Table 16 (Section 4.6) for a further description of the values that are presented.

Example 1 represents the male predictive model, as applied to the complete male sample in the study, with an overall self-harm rate of 28%.

Example 2 represents the male predictive model, as applied to the Prison A male sample in the study, with an overall self-harm rate of 17%.

Example 3 represents the male predictive model, as applied to the Prison C male sample in the study, with an overall self-harm rate of 33%.

Example 4 represents the null model (a random 50:50 chance) as applied to the complete male sample in the study, with an overall self-harm rate of 28%.

Example 5 represents the female predictive model, as applied to the complete female sample in the study, with an overall self-harm rate of 33%.

Example 6 represents the null model (a random 50:50 chance) as applied to the complete female sample in the study, with an overall self-harm rate of 33%.

Table 64. Comparison of identified case numbers across a range of scenarios for n=1000 inferred screening administrations

Example	1	2	3	4	5	6
Self-Harm Rate	28%	17%	33%	28%	33%	33%
Sensitivity	55%	55%	55%	50%	71%	50%
Specificity	86%	86%	86%	50%	93%	50%
PPV	60%	44%	65%	28%	83%	33%
NPV	83%	90%	79%	72%	87%	67%
True positives (a)	154	94	182	140	234	165
False positives (b)	104	120	97	360	47	335
False negatives (c)	126	76	148	140	96	165
True negatives (d)	616	710	573	360	623	335
Total n	1000	1000	1000	1000	1000	1000

These values are presented at the single cut-point where sensitivity and specificity were maximised. Although it can be seen that the risk classification of the prediction models (Examples 1,2,3 and 5) is far superior to the null models (Examples 4 and 6), none of the individual cut values presents a compelling argument for the use of the screening algorithms, as there are still a lot of false positives and false negatives identified. However, with the introduction of variable cut points for specific purposes, the number of false positives or negatives could be minimised.

By examining crosstabs of different cut points relative to the sensitivity and specificity achieved, it is possible to create a low-medium-high risk classification for the risk of self-harm. A 'low' risk classification seeks to maximise the sensitivity of the prediction model, meaning that among those that do self-harm, their identification is maximised. This provides a low cut-point, above which true positive identification is maximised. This cannot be used as single cut point as it also maximises the amount of false positives, but it is useful as it minimises the false negatives identified (i.e. anyone below the cut point value is highly unlikely to self-harm). A 'high' risk classification seeks to maximise the specificity of the prediction model, meaning that among those that do not self-harm, their identification is maximised. This provides a high cut-point, below which true negative identification is maximised. This cannot be used as single cut point as it also maximises the amount of false negatives, but it is useful as it minimises the false positives identified (i.e. anyone above the cut point value is highly likely to self-harm). In order to exemplify why these should not be taken as single cut points, four further examples are presented in Table 65 which replicate the cut points for the male and female screening models across 1000 inferred screening administrations.

Table 65. Replication of predictive model cut points across n=1000 inferred screening administrations

	Male		Female	
Cut point	**Low**	**High**	**Low**	**High**
Self-Harm Rate	28%	28%	33%	33%
Sensitivity	100%	16%	100%	40%
Specificity	8%	100%	8%	99%
PPV	30%	94%	35%	94%
NPV	100%	75%	100%	77%
True positives (a)	280	45	330	132
False positives (b)	661	3	614	9
False negatives (c)	0	235	0	198
True negatives (d)	59	717	56	661
Total n	1000	1000	1000	1000

For males, a cut point of 1 maximises the sensitivity of the prediction model at 100%, also producing an NPV of 100% (specificity 8.1%, PPV 28.3%). A cut point of 10 maximises the specificity of the prediction model at 99.5%, producing a PPV of 92.9% (sensitivity 16.3%, NPV 76.7%).

For females, a cut point of 2 maximises the sensitivity of the prediction model at 100%, also producing an NPV of 100% (specificity 8.3%, PPV 34.7%). A cut point of 16 results in the specificity of the prediction model being 98.6%, also producing a PPV of 93.3% (sensitivity 40.0%, NPV 77.2%).

When all individuals are classified (post-hoc) within these risk categories, both genders have a minimal level of self-harm among those categorised as low risk, and those classified as high risk subsequently self-harmed in almost 75% of the male cases, and almost 90% of the female cases. This categorisation by level of risk could contribute to identifying appropriate care pathways and, given the strength of the negative tests, may facilitate sign-off from the ACCT. It is plausible that the respective gender-specific item sets, which resulted from the logistic regression, could form single page clinical decision aids which could be administered by any prison staff within a few minutes.

Table 66. Levels of risk for self-harm from gender-specific predictive algorithm

Predictive algorithm results expressed as level of risk	% Males who SH	% Females who SH
Low (Males: <2, Females: <3)	0	0
Medium (Males: 2<9, Females: 3<15)	21.2	23.8
High (Males: 10+, Females: 16+)	73.7	88.2

When considering the use of any screening instrument or algorithm, the potential costs have to be measured against the potential gains, which is variable depending on the trade-off between sensitivity and specificity. In order to maximise sensitivity and capture all true cases, there will also be an associated increase in false positives identified. If a very large amount of false positives are identified, this would swamp the system, and the added cost and resource of monitoring all of these cases would not be sustainable. However, this would have to be considered alongside any potential reduction in self-harm or suicide events, where it has been estimated that each self-harm episode costs an average of £809 to treat in a hospital setting, (232) and that each completed suicide could cost between £1.7 million and £3.2 million. (233) If the amount of false positives could be limited to a manageable amount, then the implementation of a screening instrument could be beneficial both in real terms (less self-harm) as well as financially.

7.3 Summary of Predictive Analysis

The primary aim of the study was to determine whether any pre-existing instruments could predict self-harm among an ACCT population, and an AUC analysis was carried out on five candidate instruments to determine whether they could be considered useful for prison staff to incorporate into a standardised ACCT process. Although two scales demonstrated an AUC significantly different from 0.5, all scales failed to have any meaningful predictive value, and the utilisation of Rasch-transformed scores within the AUC analysis had very little effect. When gender separation was considered, the results of the AUC analysis were improved to an extent, but there were still no meaningful predictive values observed among the instruments.

As none of the candidate instruments were usefully predictive, the pool of items was assessed for individual items that may be predictive of self-harm. These potentially useful items were then entered into a gender-specific logistic regression analysis, which resulted in the production of gender-specific predictive algorithms that were statistically significant in predicting future self-harm. However, as this analysis was carried out post-hoc, although it is plausible that these item sets could be useful, their direct predictive capacity and operational functionality remains unknown.

8 Explanatory Model of Self-Harm

This chapter explores possible explanatory mechanisms that may contribute to the final outcome of self-harm. Although none of the pre-existing instruments used in the study were predictive of self-harm, they were all shown to display some level of measurement validity within the specific ACCT population that was under consideration. All of these instruments were originally developed to measure different constructs, and a structural equation modelling approach offers the opportunity to assess the conceptual relationship between some of these constructs in terms of their impact on self-harm being carried out. An explanatory model of self-harm in the ACCT population may be useful as it could potentially contribute to the understanding of the process that leads to self-harm. This, in turn, could then help to inform care pathways and targeted interventions.

8.1 Model Development

The initial model of an SEM analysis should be conceptually derived, based on pre-existing models and empirical evidence that are presented in the research literature. Although SEM is technically a confirmatory approach, there is an exploratory element of the process that may reveal the relationships between a number of variables.

Based on the indications in the literature, (137, 140) and supported by the differing self-harm incidence and predictive analysis results of this study, it was decided that separate models for males and females should be pursued. Although the conceptual model basis is equivalent for males and females, it was thought that the varying relationships and pathways would manifest differently for males and females, and that a gender-equivalent model may be difficult to achieve.

8.2 Model Basis

The conceptual basis of the tested models was framed on Joiner's Interpersonal-Psychological Theory of Suicidal Behaviour (IPTSB), (76, 234) which was developed further, by Ireland & York, (78) who proposed an Integrated Model of Self-Injurious Activity (see Figure 48).

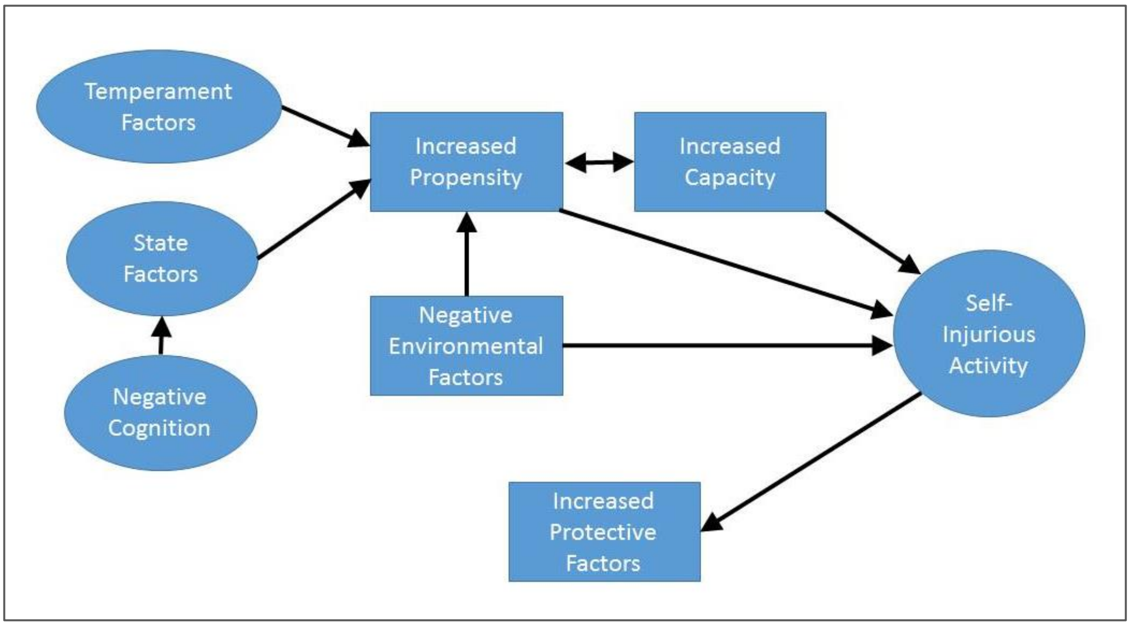

Figure 48. Integrated model of self-injurious activity (Ireland & York, 2012)

This model suggests that self-harm is largely predicted by the central facets of an 'increased propensity' and 'increased capacity' to carry out self-harm. An 'increased capacity' refers to the engagement in a range of self-damaging behaviours, with an increased history, an increased number of different self-harm behaviours, and an increased involvement in general risk-taking behaviour representing the concept. This view is supported by the finding that the number of *different* self-harm behaviours is the single best predictor of future suicide. (54)

An 'increased propensity' refers both to factors of temperament such as personality type and coping strategies, and factors of current state such as psychological distress.(78) Self-harm is a (failed) coping-mechanism that is used to deal with excess stress. At times of increased stress, this is when people that self-harm are at their most vulnerable and are therefore more likely to self-harm. The 'propensity' of a prisoner to self-harm is therefore likely to be impacted by the level of current psychological distress that is experienced, along with the ability of the prisoner to apply a coping strategy to alleviate any increased stress.

Within the integrated model, the 'negative environmental factors' element is reflected by the context of the prison setting, with this being influenced by a person's sense of ‘thwarted belongingness’ (76) through being in prison. The context of the prison setting also has the practical influence of affecting the availability of different methods of self-harm.

A number of simplified variations on Ireland & York’s model (78) were therefore used as the conceptual basis for model testing, focused around the key elements of increased capacity and increased propensity for self-harm. Although an initial model could be more conceptually complete, along with having many inclusions as risk-factor or protective-factor contributors (see Section 2.2.3), it was decided that it would be technically beneficial to derive an initial model that was as simple as possible, as added complexity may add complications to the specification and interpretation of the model.

Within the path models that were specified, the Self-Harm Inventory score was used as the sole indicator of increased capacity, as it provides a count of different self-harm behaviours and risk-taking activities that a person has previously engaged in. This signifies the extent to which previous self-harm behaviour has escalated, and a capacity to self-harm has increased. The ‘increased propensity’ within the models is represented by a number of different variables, which were specified to interact in different ways through the various models that were tested. This propensity element variably included indicators of: depression (as represented by the PHQ-9), coping (as represented by the ‘functioning’ subscale of the CORE-OM), and borderline personality disorder symptomology (as represented by the BSL-23-F). Additionally, in some of the male models that were tested, the PriSnQuest was used as an overall measure of self-harm propensity, as this includes individual risk factors relating to background, depression and borderline symptoms. Within the male sample, the PriSnQuest was also shown to be the best single instrument at predicting the risk of self-harm in follow-up. The CORE-Risk subscale was also used in some of the male models, as an indicator of the current propensity to engage in high-risk behaviours associated with self-harm.

Although it has been previously recognised as an independent risk factor, the variable of prisoner age was included in many of the models as a technical addition in order for the specified models to have the necessary degrees of freedom for the model to run.

Table 67 includes all abbreviations that were used in the SEM path analysis to indicate the different variables and relationships that were specified within the different models.

Table 67. Abbreviations used in path analysis model specification

Abbreviation	Meaning
BORDERLINE	BSL-23-F scale score
COPING	CORE-OM Functioning subscale score
DEP	PHQ-9 scale score
ON	Regressed on
PQUEST	PriSnQuest scale score
RISK	CORE-OM Risk subscale score
SHFU	Self-harm in follow-up
SHI	Self-Harm Inventory scale score
WITH	Correlated with

8.3 Female Model

The female models were derived first for two main reasons. Firstly, the overall rate of self-harm was greater for females than it was for males, and it was thought that an explanatory model may be more straightforward where the outcome incidence was greater. Secondly, within the predictive analysis of the cohort study, for females, the SHI was identified as the best single instrument at predicting the risk of self-harm in follow-up. Self-harm capacity (as measured by the SHI) is a key component of the conceptual model, (76, 78) and therefore the female models were based around the focal relationship of self-harm capacity being the key indicator of self-harm in follow-up.

8.3.1 Results

The models that were tested are summarised in Table 68, along with the corresponding fit indices of each specified model. It can be seen that models F5-F8 all offer support for the specified models, with all individual relationships being significant, and no further relationships suggested by the modification indices.

The models F5-F8 are all presented as individual schematic diagrams in Figure 49 to Figure 52.

Table 68. Female path analysis model fit

			Chi-square			RMSEA						
Model	**Descriptor/Modification**	**N**	**Value**	**df**	**p**	**Estimate**	**90% C.I.**	**90% C.I.**	**Probability RMSEA <=0.05**	**CFI**	**TLI**	**WRMR**
F1	SHFU ON COPING, SHI. COPING WITH DEP	115	15.4	3	0.0015	0.19	0.103	0.288	0.006	0.707	0.413	1.147
F2	SHFU ON DEP, SHI. COPING WITH DEP	115	12.9	3	0.005	0.169	0.082	0.269	0.016	0.766	0.532	1.057
F3	SHFU ON COPING, SHI. COPING WITH DEP. COPING ON SHI. SHI ON AGE	115	11.82	5	0.037	0.109	0.024	0.191	0.099	0.843	0.686	0.789
F4	SHFU ON DEP, SHI. COPING WITH DEP. DEP ON SHI. SHI ON AGE	115	11.06	5	0.0502	0.103	0	0.185	0.123	0.86	0.721	0.774
F5	SHFU ON SHI.SHI ON COPING, AGE. COPING WITH DEP	115	7.47	6	0.2797	0.046	0	0.136	0.45	0.966	0.944	0.622
F6	SHFU ON SHI.SHI ON DEP, AGE. COPING WITH DEP	115	6.437	6	0.376	0.025	0	0.126	0.55	0.99	0.983	0.565
F7	SHFU ON SHI.SHI ON COPING, AGE. COPING WITH BORDERLINE	115	3.748	6	0.71	0	0	0.091	0.827	1	1.056	0.395
F8	SHFU ON SHI.SHI ON BORDERLINE, AGE. COPING WITH BORDERLINE	115	3.035	6	0.8044	0	0	0.077	0.889	1	1.074	0.336
Guide	**Ideal Fit Criteria**				**p>0.05**	**<0.05**				**>0.95**	**>0.95**	**<0.90**

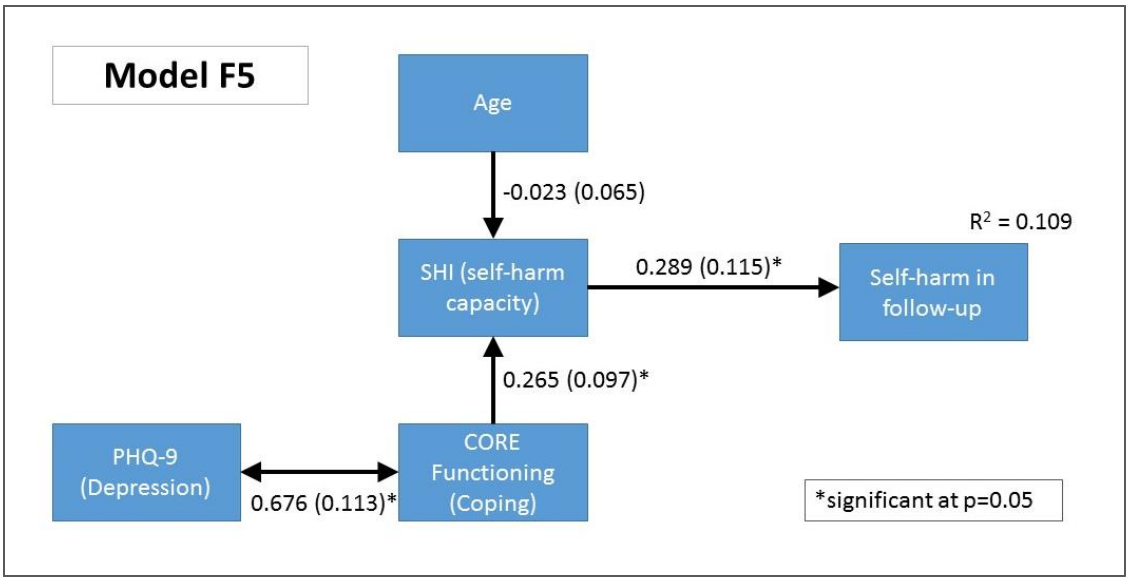

Figure 49. Schematic diagram of female model F5

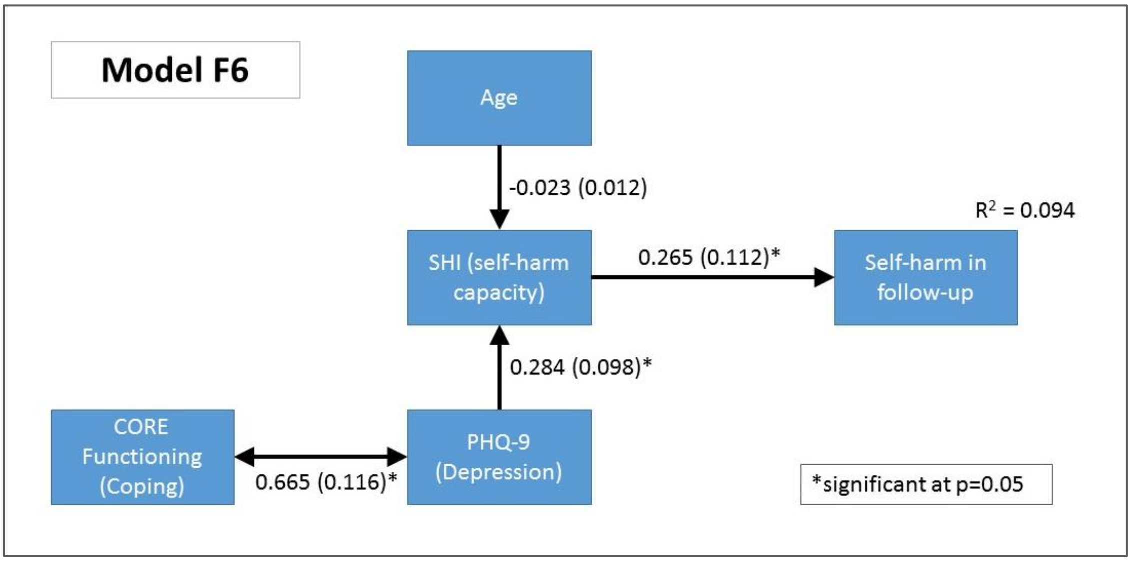

Figure 50. Schematic diagram of female model F6

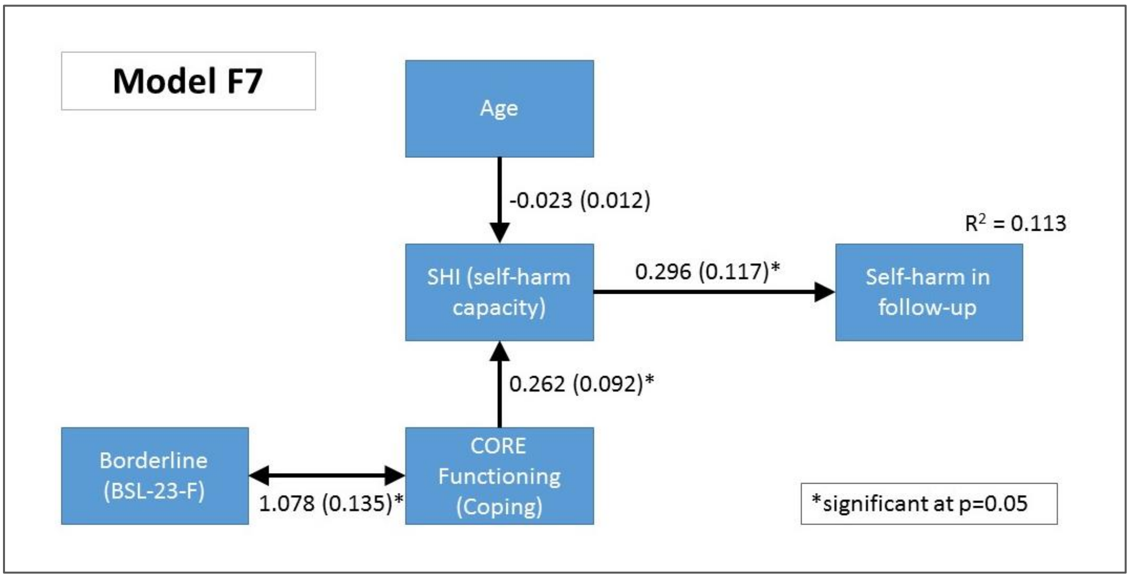

Figure 51. Schematic diagram of female model F7

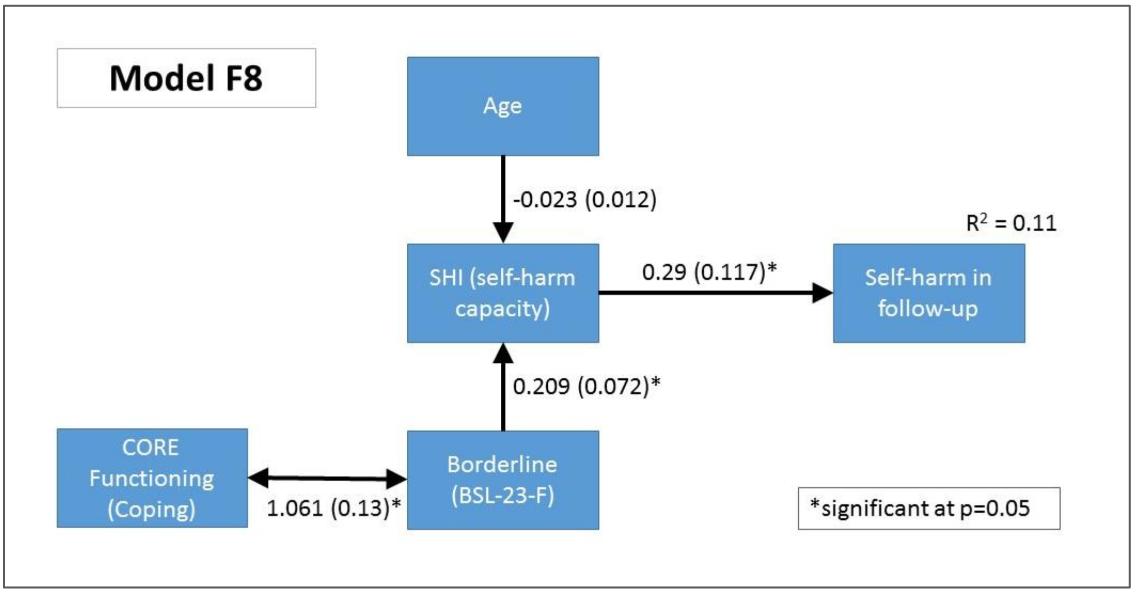

Figure 52. Schematic diagram of female model F8

8.3.2 Interpretation

All models were tested as potential pathways of the relationship between the capacity to self-harm and the propensity to self-harm, to see if any one of these eight models (F1-F8) would offer a valid solution. It would appear that models F5-F8 all offer potentially useful models with regard to explaining the pathway to self-harm when on an ACCT. These models are all very similar, and as they all seem to be supported by the model fit statistics it seems that some indicators relating to the 'propensity' element may be interchangeable. Crucially, the 'capacity' element remains steady across all of the working models, and this consistently appears to influence the self-harm study outcome, with the likelihood of carrying out self-harm increasing as the self-harm capacity (SHI score) increases. This finding offers support to Joiner's Interpersonal-Psychological Theory of Suicidal Behaviour (76) along with Ireland & York's Integrated Model of Self-Injurious Activity. (78) This also ties in with the work of Latimer et al., (235) who proposed an item bank of self-harm behaviours that represents an increasing scale of self-harm capacity.

Interestingly, the propensity to self-harm seems to be fully mediated by the capacity to self-harm within these models. The fully-mediated relationships specified in models F5-F8 are supported by the model fit results, whereas the partially-mediated relationships specified in models F1-F4 were not supported, either by the total model results or the direct relationships between the propensity indicators and the self-harm outcome variable. This full-mediation suggests that perhaps a critical threshold of 'capacity' needs to be reached before any self-harm is carried out. Alternatively, it may be that those with a decreased capacity actually did carry out self-harm, but that this was self-managed and therefore not picked up by the prison recording system. Either way, the interaction between propensity and capacity appears to be supported.

As noted, the propensity indicator of coping, as represented by the CORE-OM function subscale, seemed to be highly correlated with both the depression and borderline symptom variables, which appeared to be interchangeable in terms of their position within the model. This also makes sense conceptually, as a higher level of depression will generally reduce the ability to cope with a situation, and the lower the ability to cope with a situation, then generally the higher the resulting depression will become.

However, despite the fit-statistics and significant individual relationships suggesting that models F5-F8 are all well-fitting models, it should also be noted that the R^2 values for each of these models is fairly limited. For models F5-F8, the R^2 values vary from 0.094 (F6) to 0.113 (F7), meaning that as a maximum, model F7 explains 11.3% of the variance in the dependent variable (self-harm in follow-up). Therefore, although these models may be potentially useful, this highlights that a more proficient explanatory model may also be more complex. Alternatively, it may highlight the restrictions of trying to identify a generalisable model within the complex and highly individualised process of self-harm.

8.4 Male Model

It was postulated that the conceptual model would remain similar for both males and females, but that it would manifest in a slightly different way across genders, meaning that a common model may be more difficult to achieve. Nevertheless, the same eight initial models (as specified in F1-F8) were also tested separately for males (labelled as M1-M8).

Although the model fit results offered support for models M5 and M8 (see Table 69), the focal relationship of self-harm capacity (SHI score) directly influencing the self-harm outcome variable was not significant in either model (in M5: $\beta = .103$, SE = .067, $p = .121$; in M8: $\beta = .092$, SE = .068, $p = .175$). See Figure 53 and Figure 54 for the individual path results of models M5 and M8, respectively.

As these models were not fully supported, a number of additional models were also tested. These were largely based around the focal relationship of the PriSnQuest directly influencing the self-harm outcome variable, as the PriSnQuest was the best single instrument at predicting the risk of self-harm in follow-up.

8.4.1 Results

The models that were tested are summarised in Table 69, along with the corresponding fit indices of each specified model. It can be seen that a number of the models appear to be supported by the model fit statistics (M5, M8, M16, M18 – M21). However, only models M16 and M21 appeared to be fully supported, with all individual relationships being significant, and no further relationships suggested by the modification indices.

The models M16 and M21 are presented as individual schematic diagrams in Figure 55 and Figure 56, respectively.

Table 69. Male path analysis model fit

			Chi-square			RMSEA						
Model	**Descriptor/Modification**	**N**	**Value**	**df**	**p**	**Estimate**	**90% CI**	**90% CI**	**Probability RMSEA <=0.05**	**CFI**	**TLI**	**WRMR**
M1	SHFU ON COPING, SHI. COPING WITH DEP.	326	41.3	3	0	0.198	0.147	0.254	0	0.612	0.224	1.894
M2	SHFU ON DEP, SHI. COPING WITH DEP.	326	40.8	3	0	0.197	0.146	0.252	0	0.617	0.234	1.893
M3	SHFU ON COPING, SHI. COPING WITH DEP. COPING ON SHI. SHI ON AGE.	335	16.95	5	0.0046	0.084	0.043	0.13	0.083	0.89	0.78	0.963
M4	SHFU ON DEP, SHI. COPING WITH DEP. DEP ON SHI. SHI ON AGE.	335	36.3	5	0	0.137	0.097	0.18	0	0.712	0.424	1.432
M5	SHFU ON SHI.SHI ON COPING, AGE. COPING WITH DEP.	335	8.963	6	0.1757	0.038	0	0.087	0.589	0.973	0.955	0.694
M6	SHFU ON SHI.SHI ON DEP, AGE. COPING WITH DEP.	335	17.134	6	0.0088	0.074	0.034	0.117	0.139	0.898	0.829	0.973
M7	SHFU ON SHI.SHI ON COPING, AGE. COPING WITH BORDERLINE.	335	13.46	6	0.0363	0.061	0.014	0.105	0.29	0.961	0.935	0.842
M8	SHFU ON SHI.SHI ON BORDERLINE, AGE. COPING WITH BORDERLINE.	335	7.072	6	0.3142	0.023	0	0.077	0.733	0.994	0.991	0.585
M9	SHFU ON BORDERLINE. COPING WITH BORDERLINE. BORDERLINE ON SHI. SHI ON AGE.	335	40.219	6	0	0.13	0.094	0.17	0	0.822	0.703	1.462
M10	SHFU ON PQUEST.PQUEST ON SHI. SHI ON BORDERLINE, AGE. COPING WITH BORDERLINE.	335	47.368	10	0	0.106	0.076	0.137	0.001	0.876	0.813	1.227
Guide	**Ideal Fit Criteria**				**p>0.05**	**<0.05**				**>0.95**	**>0.95**	**<0.90**

			Chi-square			RMSEA						
Model	**Descriptor/Modification**	**N**	**Value**	**df**	**p**	**Estimate**	**90% CI**	**90% CI**	**Probability RMSEA <=0.05**	**CFI**	**TLI**	**WRMR**
M11	SHFU ON RISK. RISK ON SHI. SHI ON BORDERLINE, AGE. COPING WITH BORDERLINE.	335	97.04	10	0	0.16	0.133	0.191	0	0.739	0.608	1.766
M12	SHFU ON BORDERLINE, SHI. BORDERLINE ON PQUEST.BORDERLINE WITH COPING. SHI WITH RISK.	332	325.32	9	0	0.325	0.296	0.356	0	0.376	-0.04	3.738
M13	SHFU ON PQUEST, COPING, SHI. COPING ON PQUEST.	325	5	1	0.025	0.111	0.032	0.216	0.091	0.898	0.492	0.792
M14	SHFU ON PQUEST, AGE. PQUEST ON SHI. SHI WITH COPING.	335	28.36	6	0.0001	0.105	0.068	0.146	0.009	0.83	0.716	1.221
M15	SHFU ON PQUEST, AGE. PQUEST WITH SHI, COPING. SHI WITH COPING.	335	11.12	5	0.049	0.06	0.004	0.109	0.302	0.953	0.907	0.776
M16	SHFU ON PQUEST. PQUEST WITH SHI, COPING. SHI WITH COPING. SHI ON AGE.	335	5.64	4	0.228	0.035	0	0.095	0.578	0.988	0.969	0.56
M17	SHFU ON PQUEST. PQUEST WITH SHI, COPING. SHI WITH COPING. SHI ON AGE, BORDERLINE.	328	473.21	7	0	0.451	0.417	0.486	0	0.141	-0.72	4.78
M18	SHFU ON PQUEST. PQUEST ON SHI, COPING, BORDERLINE. COPING ON BORDERLINE, SHI. SHI ON AGE, BORDERLINE.	328	7.234	6	0.2997	0.025	0	0.079	0.715	0.998	0.995	0.558
M19	SHFU ON PQUEST. PQUEST ON SHI, COPING, BORDERLINE. COPING ON BORDERLINE, SHI ON AGE, BORDERLINE.	328	8.835	7	0.2647	0.028	0	0.077	0.712	0.997	0.993	0.623
M20	SHFU ON PQUEST. PQUEST ON SHI, BORDERLINE. COPING ON BORDERLINE. SHI ON AGE, BORDERLINE, COPING.	328	9.92	7	0.1934	0.036	0	0.082	0.636	0.995	0.989	0.648
M21	SHFU ON PQUEST. PQUEST ON SHI, BORDERLINE. SHI ON AGE, BORDERLINE.	328	5.146	4	0.2727	0.03	0	0.093	0.621	0.994	0.987	0.566
Guide	**Ideal Fit Criteria**				**p>0.05**	**<0.05**				**>0.95**	**>0.95**	**<0.90**

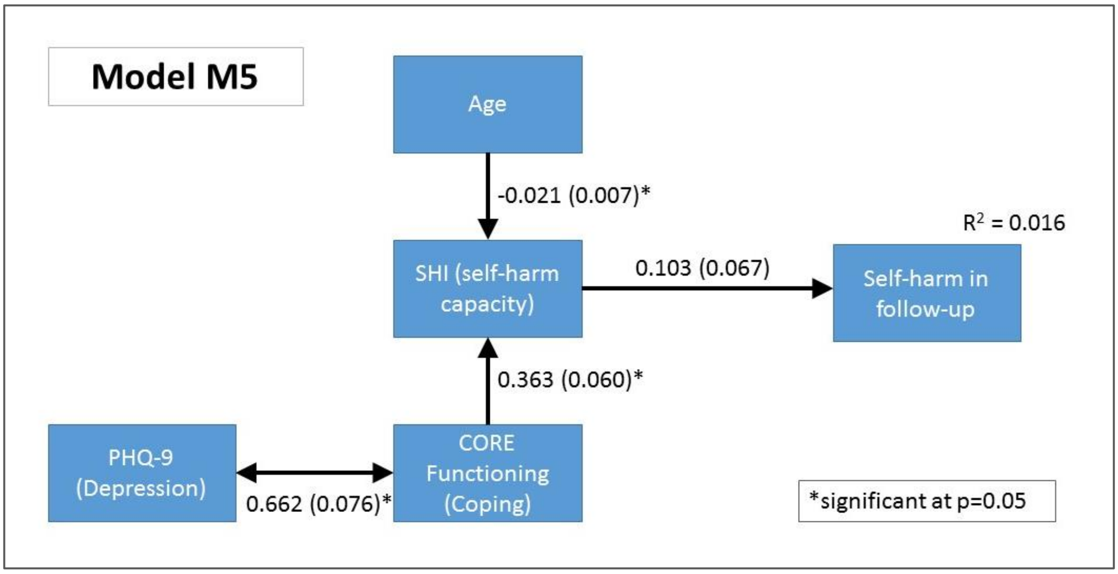

Figure 53. Schematic diagram of male model M5

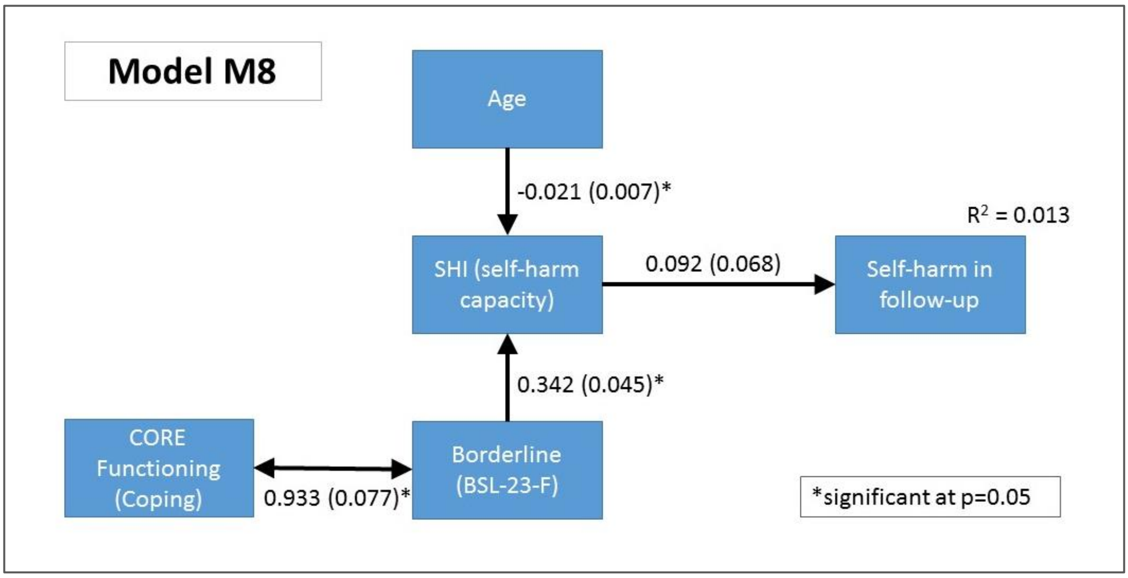

Figure 54. Schematic diagram of male model M8

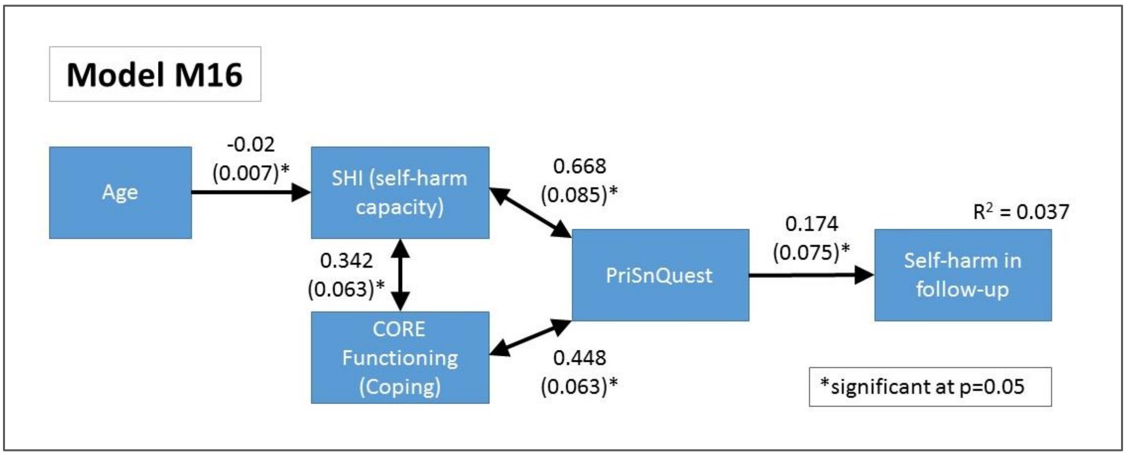

Figure 55. Schematic diagram of male model M16

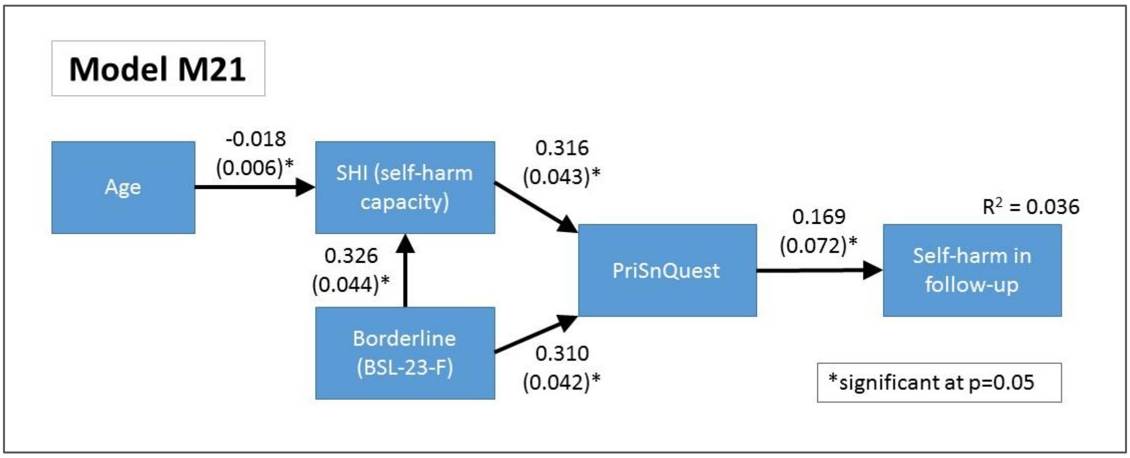

Figure 56. Schematic diagram of male model M21

8.4.2 Interpretation

Within the male models, the 'capacity' element of self-harm was shown not to have a direct significant impact on self-harm in follow-up; therefore the conceptual model seems to manifest slightly differently for males and females. Although 'capacity' is still included as a key component of the male model, the 'propensity' element provides the focal relationship. This is provided in this instance by the PriSnQuest, which was the best overall single indicator of self-harm in the male sample.

The M16 model (see Figure 55) shows that the SHI, the PriSnQuest and the CORE Functioning subscale (coping) are all significantly associated with each other, with the direct impact on self-harm provided by the PriSnQuest. The PriSnQuest and the CORE Functioning (coping) subscale both represent different elements of 'propensity', and therefore this model does not deviate too far from the conceptual relationship proposed by Ireland & York. (78) Only the PriSnQuest has a direct impact on the final self-harm outcome, but self-harm capacity (SHI) and coping (CORE-Functioning) may also have an indirect impact through the association with the PriSnQuest. This model fits well, and confirms the associational relationships between self-harm current-state risk factors (PriSnQuest), self-harm capacity, and coping. This model alludes to a complex relationship between different contributing factors of self-harm, suggesting that the pathway to self-harm is not straightforward, even when only a limited number of variables are examined.

As in the M16 model, the M21 model (see Figure 56) also shows that the PriSnQuest is the only variable with a direct impact on self-harm during follow-up. However, in this model the PriSnQuest is directly influenced by the self-harm capacity (SHI), and also by borderline symptomology (BSL-23-F), both directly and indirectly (partially mediated through self-harm capacity). Again, this model colludes with Ireland & York's model, where self-harm capacity and propensity are associated, and are influenced by the underlying temperament of a person. Although borderline symptomology could be viewed as an indicator of self-harm propensity in itself, it could also represent the underlying 'temperament' of a person as specified in the Integrated Model of Self-Injurious Activity (see Figure 48).

Despite the overall concordance of male models M16 and M21 to the Integrated Model of Self-Injurious Activity, it can be seen that the pathway to self-harm does manifest slightly differently to the female models F5-F8. With regard to the relationship between self-harm capacity and propensity, the male models suggest that the impact of self-harm capacity upon self-harm in follow-up is fully mediated through self-harm propensity, whereas the female models suggest that the relationship is the other way around. If this is true, then this could help to explain the differing pathways to self-harm exhibited by males and females.

However, despite the fit-statistics and significant individual relationships suggesting that models M16 and M21 are well-fitting models, it should again be noted that the R^2 values for each of these models is even more limited than it was for the female models. Both models M16 and M21 exhibit R^2 values lower than 0.04, meaning that less than 4% of the variance in the dependent variable (self-harm in follow-up) is explained. Therefore, although these models may be potentially useful, this again highlights that a more complete explanatory model is also likely to be more complex. Additionally, this may also be lower than in the female models due to the reduced overall self-harm rate within the male sample.

8.5 A Joint Model, Moderated by Gender

The broad conceptual model basis was common across both males and females, although it was thought that this would manifest slightly differently across genders. Through running a number of common models across genders (F/M1-F/M8), it was seen that models F/M5 and F/M8 were both supported by the model fit statistics. These models appear to work across both genders, although the focal relationship of self-harm capacity (SHI score) directly influencing the self-harm outcome variable was different across genders, being a significant relationship for females, but not for males. These findings appear to suggest that it may be appropriate to introduce gender as a moderator on the focal relationship of the model, with all the other specified relationships within the model held constant across gender groups. The representation of this can be seen in Figure 57 and Figure 58, where gender is specified as a moderator which only impacts upon the focal relationship of the models.

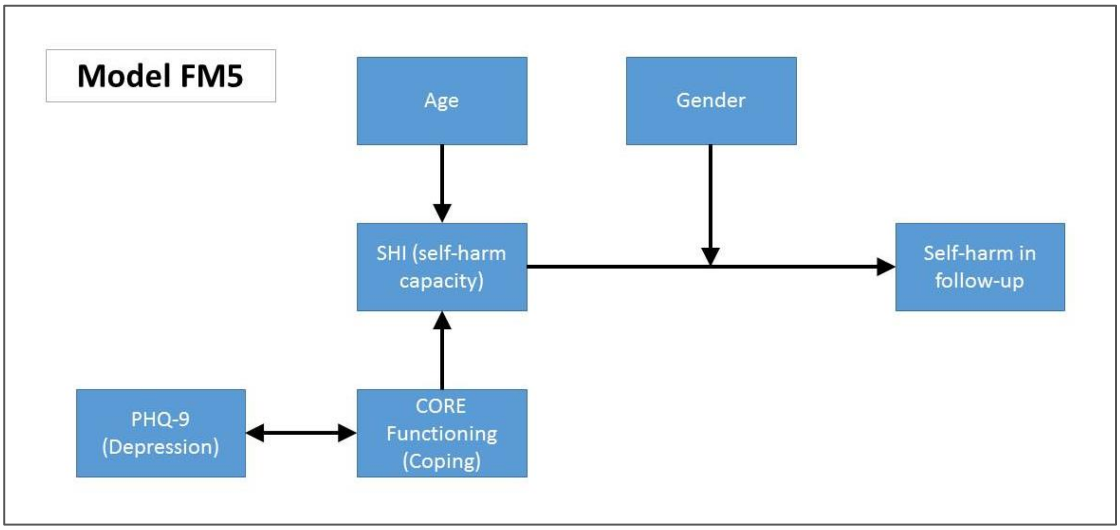

Figure 57. Schematic diagram of joint model FM5, with gender specified as an effect moderator

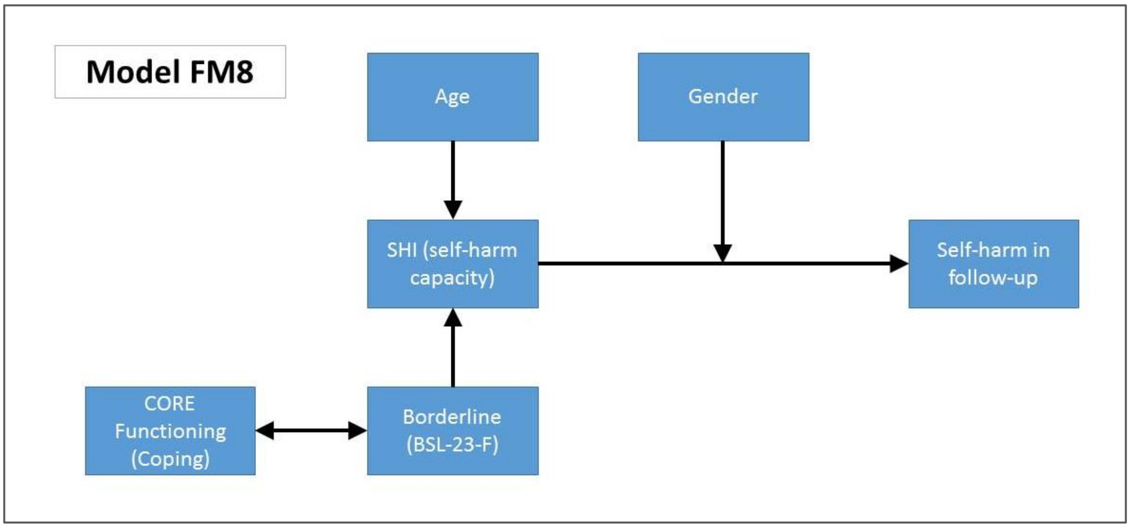

Figure 58. Schematic diagram of joint model FM8, with gender specified as an effect moderator

In order to test the common structure across gender groups, each model was run three times. Firstly, the model was applied across the full sample (n=450), with no gender grouping variable introduced (labelled as FM5 and FM8 within Table 70). Secondly, the gender grouping variable was introduced, and all of the individual relationships in the model were constrained to be equal across groups *except* for the focal relationship between SHI and the self-harm outcome variable, which was allowed to differ across gender groups. This specification therefore treats gender as a moderator on the focal relationship (labelled as FM5a and FM8a within Table 70). Thirdly, the gender grouping variable was retained, but in this case *all* of the individual relationships in the model were constrained to be equal across groups (labelled as FMb and FM8b within Table 70). Using the DIFFTEST function within MPlus, this specification allows us to formally test (using a chi-square test) whether there is any difference between the 'a' and the 'b' models, thus determining whether the moderator effect is significant.

8.5.1 Results

The results are summarised in Table 70, along with the corresponding fit indices of each specified model. The model path results are reported separately for each version of the models FM5 and FM8. These are presented in Tables 71 to 76.

It can be seen that all three versions of each model (FM5 and FM8) are supported by the model fit statistics, and within the common models (FM5, FM5b, FM8 and FM8b) all individual relationships are significant (see Table 71, Table 73, Table 74 and Table 76). However, the gender-moderated models (FM5a and FM8a) are perhaps more interesting. All of the constrained individual relationships are significant, but the gender-moderated focal relationships are only significant for females, and not for males (see Tables 72 and 75).

Despite this apparent gender moderation, the Chi-Square test of model difference determined that there was no difference between the gender-moderated model (the 'a' model) and the fully-constrained model (the 'b' model) in either case (FM5 or FM8).

Table 70. Gender-moderated path analysis model fit

			Chi-square			RMSEA						
Model	**Descriptor/Modification**	**N**	**Value**	**df**	**p**	**Estimate**	**90% C.I.**	**90% C.I.**	**Probability RMSEA <=0.05**	**CFI**	**TLI**	**WRMR**
FM5	SHFU ON SHI.SHI ON COPING, AGE. COPING WITH DEP.	450	8.926	6	0.1778	0.033	0	0.075	0.699	0.981	0.968	0.693
FM5a	As above, with group variable added. All relationships constrained to be equal across groups, EXCEPT SHFU on SHI.	450	16.64	15	0.3409	0.022	0	0.069	0.797	0.989	0.986	0.959
FM5b	As above, with group variable added. All relationships constrained to be equal across groups to test for model difference with FM5a	450	18.453	16	0.2981	0.026	0	0.069	0.777	0.984	0.98	1.019
DIFFTEST	**Chi-square test for model difference between FM5a & FM5b**		**1.824**	**1**	**0.1768**							
FM8	SHFU ON SHI.SHI ON BORDERLINE, AGE. COPING WITH BORDERLINE.	450	7.16	6	0.3063	0.021	0	0.067	0.813	0.996	0.993	0.597
FM8a	As above, with group variable added. All relationships constrained to be equal across groups, EXCEPT SHFU on SHI.	450	12.891	15	0.6107	0	0	0.055	0.928	1	1.011	0.807
FM8b	As above, with group variable added. All relationships constrained to be equal across groups to test for model difference with FM8a	450	14.904	16	0.5317	0	0	0.058	0.906	1	1.005	0.886
DIFFTEST	**Chi-square test for model difference between FM8a & FM8b**		**1.898**	**1**	**0.1683**							
Guide	**Ideal Fit Criteria**				**p>0.05**	**<0.05**				**>0.95**	**>0.95**	**<0.90**

Table 71. Individual path results for joint model FM5

Variable	Relationship	Variable	Beta Estimate	S.E.	p-value	R^2
SHI	→	SHFU	0.150	0.057	0.009	0.033
COPING	→	SHI	0.331	0.053	0.000	
AGE	→	SHI	-0.022	0.006	0.000	
COPING	↔	DEP	0.674	0.064	0.000	

Table 72. Individual path results for joint model FM5a, with unconstrained focal relationship across gender groups

	Variable	Relationship	Variable	Beta Estimate	S.E.	p-value	R^2
FEMALE	SHI	→	SHFU	0.277	0.115	0.016	0.101
MALE	SHI	→	SHFU	0.102	0.066	0.123	0.016
	COPING	→	SHI	0.334	0.051	0.000	
	AGE	→	SHI	-0.021	0.006	0.000	
	COPING	↔	DEP	0.667	0.063	0.000	

Table 73. Individual path results for joint model FM5b, with all relationships constrained to be equal across gender groups

Variable	Relationship	Variable	Beta Estimate	S.E.	p-value	R^2
SHI	→	SHFU	0.149	0.058	0.010	0.032
COPING	→	SHI	0.339	0.050	0.000	
AGE	→	SHI	-0.021	0.006	0.000	
COPING	↔	DEP	0.669	0.063	0.000	

Table 74. Individual path results for joint model FM8

Variable	Relationship	Variable	Beta Estimate	S.E.	p-value	R^2
SHI	→	SHFU	0.138	0.058	0.018	0.028
BORDERLINE	→	SHI	0.299	0.040	0.000	
AGE	→	SHI	-0.022	0.006	0.000	
COPING	↔	BORDERLINE	0.974	0.067	0.000	

Table 75. Individual path results for joint model FM8a, with unconstrained focal relationship across gender groups

	Variable	Relationship	Variable	Beta Estimate	S.E.	p-value	R^2
FEMALE	SHI	→	SHFU	0.272	0.116	0.019	0.098
MALE	SHI	→	SHFU	0.092	0.068	0.174	0.013
	BORDERLINE	→	SHI	0.304	0.039	0.000	
	AGE	→	SHI	-0.021	0.006	0.000	
	COPING	↔	BORDERLINE	0.965	0.066	0.000	

Table 76. Individual path results for joint model FM8b, with all relationships constrained to be equal across gender groups

Variable	Relationship	Variable	Beta Estimate	S.E.	p-value	R^2
SHI	→	SHFU	0.139	0.059	0.018	0.028
BORDERLINE	→	SHI	0.306	0.038	0.000	
AGE	→	SHI	-0.021	0.006	0.000	
COPING	↔	BORDERLINE	0.965	0.066	0.000	

8.5.2 Interpretation

Both of the models, FM5 and FM8, appear to work well for both males and females together when specified as a single-group analysis. For each of these models, when a gender moderator is included on the focal relationship, this shows that despite the relationship being significant for females and non-significant for males, that the difference between groups (the moderator effect) is actually non-significant, meaning that the path models are valid across both gender groups.

To look further into this, the difference between the male and female beta-estimates in model FM5 can be quantified as 0.175, and the 95% confidence intervals of this difference are -0.085 to 0.435. Similarly, the difference between the male and female beta-estimates in model FM8 is 0.180, and the 95% confidence intervals of this difference are -0.084 to 0.444.

For both models, we therefore cannot reject the null hypothesis that the female and male versions are equivalent. As this null hypothesis cannot be rejected, we must conclude that the same model applies for both groups. However, it could be argued that the effects *may* actually be different across groups, and the size of this difference is quantified by the confidence intervals. So, for both models, there may be no moderator effect, or it may be 0.44.

It should be noted that across all of the models (female, male and joint), the age variable was only included as a technical addition in order for the models to remain identified. However, the relationship between age and SHI was often shown to be significant. Across all models, the relationship consistently suggests that as age increases, the SHI ('capacity') decreases. As the SHI is a lifetime self-harm behaviour count, then obviously this relationship cannot hold as a causal relationship. However, this does suggest that at the particular cross-section taken within this sample, younger people tend to have a higher capacity (measured by SHI score), although this is very slight in real terms. Regardless of the significance, this relationship is retained within the analysis in order for the technical identification of the models.

Again, it should be noted that the R^2 values for each of these models is limited, and it can be seen how this value varies between males and females in Tables 72 and 75. For the joint models FM5 and FM8, the R^2 values are 0.032 and 0.028 respectively, meaning that both models only explain around 3% of the variance in the dependent variable (self-harm in follow-up); therefore again highlighting a restriction with these models.

8.6 Summary of Path Analysis

Valid path models of self-harm have been found separately for males, females and jointly. It was hypothesised that the pathway to self-harm would manifest differently for males and females, and this was supported through the separate gender-specific models that were derived. However, the joint analysis that included gender as a moderating factor revealed that the joint models were equivalent across genders, despite the focal relationship being non-significant for males.

Supporting Ireland & York's Integrated Model of Self-Injurious Activity, (78) the relationships that have been observed between the variables are both conceptually and statistically sound and are perhaps more intuitive than the results of the logistic regression analysis. This owes to the unique benefits of the structural equation modelling process, where the direct and indirect relationships of variables can be assessed simultaneously, not only with regard to a single dependent final outcome variable, but also with regard to each other. These relationships are masked in the logistic regression analysis, as this process only considers the direct relationship of each individual item on the final outcome of self-harm in follow-up.

9 Discussion and Summary

This final chapter presents a discussion of the results in relevance to the existing body of literature, along with the practical implications of the results within the prison setting. The limitations of the different elements of the thesis studies are also presented, along with recommendations for further study. This chapter concludes with a concise summary of the research that is presented within this thesis.

9.1 Summary of Findings and Overview

This thesis has investigated whether it is possible to predict self-harm among an adult offender population within a prison setting, using pre-existing standardised instruments that could be used for screening purposes. Firstly, a scoping exercise and pilot study were carried out in order to identify and refine a selection of instruments that had the potential to predict self-harm in this setting. A large prospective cohort study was then undertaken across three prisons in order to assess the predictive properties of the selected instruments within a specified follow-up time. Additionally, each instrument was psychometrically assessed using several different methodologies, in order to provide a rigorous assessment of the validity of each instrument within the specific population used in the study.

As none of the standardised instruments proved to be usefully predictive, an extension to the predictive work was carried out, where each individual instrument item was assessed for its predictive properties alongside a range of demographic items. A logistic regression then led to the identification of gender-specific item sets, which can form predictive algorithms which appear to potentially have useful clinical application. Finally, a path analysis was carried out within a structural equation modelling framework, in order to explore the factors which may influence self-harm through direct and indirect pathways.

The hypothesis of this thesis as stated in chapter 2 was **"self-harm within the prison ACCT population can be predicted using a pre-existing screening instrument."** Although the study obviously did not test all pre-existing instruments that are currently available, the work contained within this thesis fails to support the hypothesis.

Although the hypothesis is not supported, all of the study aims (as outlined in Chapter 2) were met. A number of interesting findings were observed, which could be important in informing and influencing future research in the area. The major findings will be presented and discussed in this chapter, along with the limitations of the research and recommendations for future study. Firstly, to date, this is the largest prospective cohort study of its type, producing accurate figures of self-harm incidence within the specific ACCT population that was studied. Secondly, the psychometric assessment revealed that most of the instruments displayed a degree of validity within the study setting, but that each instrument had its own specific limitations. Thirdly, none of the selected instruments displayed meaningful properties of predicting self-harm. However, a set of gender-specific individual items was identified that appeared to be usefully predictive. Finally, a path analysis supported the view that capacity and propensity are integral elements to the self-harm pathway, although the relationship may be slightly different for males and females. For females, self-harm propensity was fully-mediated by self-harm capacity, but this relationship was reversed for males, where self-harm capacity was fully-mediated by self-harm propensity.

9.2 Self-Harm Rates

The basic self-harm incidence during the six-month follow-up was 29.1%, although this value was variable across prison and gender. Moreover, the 'event incidence' among study participants was 6.33 per 1,000 prisoner days, and 'prisoner incidence' was 1.84 per 1,000 days. This varies considerably by gender, with the event incidence rate in the female prison (15.83 per 1000 prisoner days) being much higher than the male event average (4.02 per 1000 prisoner days). Looking at persons rather than events, there is still a marked difference between genders, with the male incidence at 1.66 per 1000 prisoner days against the female rate of 2.83.

This overall rate of 29.1% would appear to relate very closely to the 30% that has been previously reported by Brooker et al. (132) However, this value refers to whether any self-harm occurs during the full term of incarceration, whereas the value in this study relates to the specific follow-up period.

It is important to note that the overall basic self-harm incidence during the six-month follow-up (of 29.1%) refers to new incidents of self-harm, whereas a lot of the published research refers to the lifetime prevalence of self-harm among specific populations. Of the research listed in Section 2.3.1, the highest recorded prevalence was among Greek male prisoners, at 49.4%. (135) In the current study, the lifetime prevalence of self-harm was (self) reported at 88%. Although this is far higher than all other recorded rates, it should be considered that the study population is specifically targeted as those who are on an ACCT, meaning that they have been identified as being at particular risk of self-harm (or further self-harm), so this is not directly comparable to other general prison populations.

The overall incidence rate recorded for males during follow-up was 27.6%, which is more than three times higher than the self-harm incidence rate of 8.13% recorded among the general male prison population in 2014. (129) This difference in rates would probably be expected, given the difference of study populations. For females, the overall incidence rate recorded during follow-up was 33.3%, which is not markedly higher than the self-harm incidence rate of 28.3% recorded among the general female prison population in 2014, and is actually lower than the peak rate of 37.7% recorded in 2009. (129) This suggests that the ACCT population in female prisons may appear to be quite similar to the more general female prison population in terms of self-harm activity.

In follow-up, among the individuals that self-harm, males recorded an average of 2.4 self-harm incidents per individual and females report an average of 5.6 self-harm incidents per individual. This is comparable to the values recorded in 2014 among the general prison population, where males recorded an average of 2.9 self-harm incidents per individual and females recorded an average of 6.1 self-harm incidents per individual, although the difference in data collection time scales should be taken into consideration. Unfortunately, the data were not available in order to calculate the event incidence per 1000 prisoner-days among the general prison population, in which case the recorded rates would be directly comparable.

The most common type of self-harm recorded in follow-up was cutting (51%), which is the same as the most common type previously reported, although the rate is not as high as the previously reported 75%. (236) However, 24% of the recorded follow-up self-harm events were classified as 'unspecified', so the proportion of cutting events may actually be higher than recorded. The difference in study populations may also be attributable for this divergence, as the specific ACCT population may be more prone to carrying out more different types of self-harm behaviour.

As mentioned in Section 2.2.1, the specific definition of self-harm that is being used in any particular study also needs to be taken into account when results are being compared. In this instance, self-harm was defined and classified as "self-poisoning or self-injury, irrespective of the apparent purpose of the act". This is quite a broad definition, and therefore may be more inclusive of some self-harm events when compared to other studies that have a narrower definition. However, it should also be noted that the active definition of self-harm and the numbers reported within this study do not include indirect, psychological, or unreported self-harm events, meaning that there will be an under-reporting of self-harm incidence compared to when these types of self-harm events are considered.

9.3 Psychometric Properties of Scales

The psychometric analysis of the potential screening instruments showed that four out of the selected five were found to have acceptable psychometric properties. This means that the instruments were acting as valid ordinal scales, and therefore the raw scores act as a sufficient statistic to justify the use of cut points within the AUC analysis. The CORE-OM, however, would require some modification for use in this setting, if it were to be used as a single total score.

Cronbach's alpha levels across all instruments were largely comparable to those previously reported (Table 77), meaning that the reliability recorded among the present sample is largely consistent with what has been reported in other (different) samples.

Table 77. Reported reliability levels of study instruments

Instrument	Alpha in present study	Person Separation Index	Alpha reported in previous study	Reference
BSL-23-F	0.93	0.92	0.94 - 0.97	(150)
CORE-OM	0.9	0.9	0.75 - 0.9	(159)
PriSnQuest	0.63	0.44	-	-
PHQ-9	0.82	0.75	0.89	(149)
SHI	0.78	0.76	0.8 - 0.9	(161-163)

It should be noted, however, that there are certain limitations to the Cronbach's alpha value which should be considered. For example, the Cronbach's alpha value is sometimes seen as a measure of unidimensionality, whereas unidimensionality of an item set is actually an assumption that must hold for the Cronbach's alpha value to be valid. (237) Additionally, alpha does not take the relative distribution of items and persons (targeting) of a scale into consideration, meaning that the alpha value will always be inflated in the case of mis-targeting, and where there are large floor or ceiling effects. The person separation index (PSI) from the Rasch analysis does take the targeting into account, and therefore a difference between the PSI and alpha values (with the PSI reporting a lower value) often indicates that a mismatch in targeting is present, as is the case for the PHQ-9 and PriSnQuest. Also, both the Cronbach's alpha and the PSI values will be artificially inflated where response dependency is present within the item set, (238) which was apparent to varying degrees across all of the study instruments (see Section 6). This means that the full-scale Cronbach's alpha values, as presented here and within other studies, may be misleading as it is likely that the reported reliability levels are over-estimated.

When using the Rasch analytic approach for instrument development, validation and refinement, a notable benefit is that it provides feedback at the individual item level that can then be related back to the underlying conceptual basis of the instrument. If an item is found not to work in the manner that it was intended, this provides the opportunity to amend the items based on the evidence provided from the analysis. For example, it may be beneficial to change the wording of a particular item, or to amend the number of response categories that are presented, or the labels that are attached to them. If the situation allows, the amended items could then be retested within the same population in order to complete the experimental cycle.

This was not the intention of the present study, but certain post-hoc amendments were made within each resolution in order to account for some of the anomalies that were evidenced. For example, the original response categories of the CORE-OM, BSL-23-F and PHQ-9 instruments were shown not to operate within this particular sample. Although this was accounted for through a post-hoc response category rescore, if the intention was to amend the instruments then it would be recommended that less response categories (with appropriate labels) are presented and retested within this population.

A significant advantage of the Rasch analytic approach is that it offers a unified framework in which to investigate multiple elements regarding an instrument as whole, and the relationships between each of its component items. When the instruments were assessed within the Rasch analytic framework, this identified several weaknesses within each. Instruments with a polytomous response structure almost always required rescoring, as the response categories did not appear to work well within this setting. A degree of response dependency and misfit were also present within the instruments, although it is likely that these issues are related. A minimal amount of age and prison DIF was present in some of the instruments, suggesting that the impact was fairly limited, but the presence of Gender DIF was more widespread. Fit to the Rasch model was resolved in most cases, through two resolutions. The first resolution (A) involved retaining as many items and as much information as possible within each instrument by, where appropriate, accounting for the misfit issues by adjusting the model within the Rasch analytic framework. The second resolution involved the iterative removal of misfitting items so that only a 'pure' set remained; this approach may offer practical advantages when considering the function and setting of the intended instruments. In both cases, item deletion was often involved, providing a solution which, although more valid, is not necessarily optimal.

The psychometric analysis results for the CORE-OM were largely in agreement with other work that has assessed its psychometric properties using principal components analysis and Mokken Scaling, (239) and Rasch Analysis, (240) where the complete item set was found to be multidimensional, to have disordered response categories, and to have a large amount of dependency and redundancy within the item set. (239, 240)

Likewise, a Rasch analysis of the PHQ-9 within a different sample (241) also demonstrated similar results to the current analysis. Although the response categories of the items were all ordered, whereas in the present study they were not, there was some consistent misfit among the first two items of the PHQ-9. This apparent misfit is potentially due to a dependency between these two items, which relate to 'depressed mood' and 'anhedonia' respectively. These two items summarise the major symptoms of depression, and perhaps should not be considered alongside the other items of the PHQ-9. (241) Interestingly, these two items make up the PHQ-2, which is the short form of the PHQ-9. (160)

A previous Rasch analysis of the SHI displayed a largely robust structure among a group of University students. (157) Although these properties of the SHI are again consistent with the psychometric properties reported in the current analysis, there are some interesting differences between the populations in terms of the SHI item set. Among both populations, the item set seems to work well. However, the 'difficulty' order (in terms of affirmation rates) is markedly different for a few items. Item 20 of the SHI: 'Self-defeating thoughts', is the 'easiest' item to affirm among both populations. This means that the empirical frequency count is the highest, and that this item represents the lowest point on the self-harm behavioural spectrum among both populations. Likewise, Items 16: 'Sexually abusive relationship' and 22: 'Abused laxatives' are the two most 'difficult' items to affirm among both populations, marking the upper end of the self-harm behavioural spectrum as measured by the SHI. However, the main divergence among the populations comes with Items 18: 'Attempted suicide' and 1: 'Overdosed', which are placed in hierarchical orders 18 and 19 respectively (out of 22, with position 22 being the most difficult) in the University population, but are placed in hierarchical orders 2 and 3 respectively in the current, prison ACCT population. This puts these two items at opposite ends of the behavioural spectrum, depending on the population in question. Given the differences between the study populations, it is understandable why this divergence is apparent. However, it also emphasises the point that measurement instruments should be separately validated and calibrated within each different population of intended use.

Although a Rasch analysis has not previously been carried out on the BSL-23, the French version of the instrument has been psychometrically assessed among BPD patients, including a one-factor confirmatory factor analysis of the instrument. (242) Despite the differences in language, population and response format, the results appeared to be largely similar to those reported here; that is, the instrument did not completely support a single-factor structure as 23 separate items, and this seemed to be due to the dependency that was present in the scale. There was also some consistency in the items that were identified as having dependency, with both analyses reporting this present among Items 7, 11, 12, 21, and 23. (242) Although the BSL-23-F has never been tested in this modified format, with the response categories now reflecting frequency rather than intensity, it is interesting to note the similarities in the analysis results. This offers support to the modification that was made, suggesting that either intensity or frequency responses are relevant.

It would appear as though Rasch analysis has not previously been carried out on the PriSnQuest. Although the PriSnQuest relies on a total score, and should therefore be unidimensional to a degree, it should be considered that the primary purpose of the PriSnQuest is as a screening tool, rather than an outcome measure. (208) It is therefore not necessary for this instrument to satisfy the properties of the Rasch model, as it acts as a binary indicator at a specific cut point, to determine whether a prisoner should receive a more in-depth psychological assessment upon entry into prison.

Despite each of the instruments displaying individual limitations to a varying extent, the majority of the questionnaires were still shown to have a certain level of internal construct validity in a prison setting. As the total scores are valid at an ordinal level, they could therefore be used for screening purposes to identify, for example, depression or borderline symptoms among this population.

Going forward into the predictive validity and SEM stages, the Rasch resolution B of each instrument was used, as this was seen as the most practical. Resolution A takes account of some of the issues within each scale by adjusting the item-modelling within the Rasch framework. This strategy retains more information, but it may not be as practical to use in a day-to-day setting, particularly as the raw score of the item set may no longer be a sufficient statistic if some DIF adjustment has been carried out.

The resolution B retains a pure set of items from each instrument, so in practical terms, any progress or predictive path that was observed from these scores would be easier to implement into practice within a prison setting.

9.3.1 Alternative Psychometric Approach

When assessing the psychometric properties of the selected instruments, the adopted approach was to fit the instrument data to the Rasch model, so that the measurement properties of the instruments could be formally tested against the known criteria of the pre-defined measurement model. An alternative approach could have been to apply a less constrained model, such as a two parameter logistic (2PL) model. (243) The addition of model parameters often offers a better-fitting model that can explain more of the variance within a dataset. However, the measurement properties remain specific to the particular dataset and the raw score is not a sufficient statistic to allow for interval transformation.

Within the context of this study, the Rasch approach was preferred for a number of reasons. Firstly, it was the intention to fully investigate the measurement properties of the instruments under consideration. The only model that offers fully-testable assumptions under a unified framework is the Rasch model. (168) For example, the additional discrimination parameter that is utilised in the 2PL model is also influenced by the inter-dependency between individual items, or by specific items and the trait itself. (244) Without any sort of conceptual investigation, the increased discrimination parameter of an item may be readily accepted as a function of the model, whereas it may truly represent a departure from the assumption of items being statistically independent. (244)

Secondly, the two parameter logistic model does not have the property of specific objectivity, meaning that the item and person calibrations remain specific to the dataset that has been used. This would render the psychometric assessment less generalizable for use outside of the current study, as the discrimination parameters are likely to be variable.

Thirdly, only the Rasch model offers the benefit of the raw score being a sufficient statistic for interval-level measurement transformation. (180) Although an interval-level transformation is not necessary for the predictive risk element of the study, it does provide a robust platform for the explanatory modelling (SEM) element of the study. Additionally, if this can be provided on the basis of a raw-score sufficiency statistic, then this would also provide a pragmatic basis for its implementation and further assessment.

Within the psychometric assessment of each of the instruments, the focus was on the investigation of the items, rather than to explain the variance within the dataset. (176) This approach also provides useful information at the item level, which could be used to inform any future formulation of measurement instruments within this population.

9.4 Predictive Validity

The primary aim of the study was to determine whether any pre-existing instruments could predict self-harm among an ACCT population. The AUC analysis that was carried out on the candidate instruments determined that none of these performed the task adequately enough to be considered a useful aid for prison staff to utilise as part of a standardised ACCT process.

Although the final outcome being studied was different, this result supports the findings of a recent systematic review that investigated the prediction of suicide following self-harm. (245) In this review, as in the present study, none of the individual scales assessed were found to have sufficient evidence to support their use for predictive purposes.

Although it was disappointing that none of the instruments predicted self-harm at follow-up, this was not entirely unexpected, as each of the instruments was made up of individual items with varying levels of discrimination. The entire item set was therefore re-assessed as an item pool of potentially individually predictive items, alongside a range of other background, socio-demographic and sentencing information. Using a logistic regression approach, these potential indicators of risk were combined to create gender-specific risk algorithms, where each indicator was weighted by its adjusted beta value of observing self-harm during follow-up. The resulting predictive algorithms displayed reasonable AUC values of 0.81 for males and 0.867 for females. The corresponding positive and negative predictive values indicate that both of these algorithms are better at screening out risk, rather than screening in risk of self-harm (see Table 63). Additionally, three levels of risk can be identified, with both genders displaying zero subsequent self-harm when categorised as low risk (see Table 66).

As a set of indicator items, it was not expected that these items hold a probabilistic relationship to one another, or that the risk of future self-harm was a latent construct which determined the responses to the various indicators. The purpose of the item set is merely to discriminate between those who went on to self-harm and those that did not, with a predictive usefulness that is greater than would be obtained by chance alone. (208) This risk factor approach has often been used to incorporate individual risk factors into composite scales to assess for the risk of suicide following self-harm, (245) and these are commonly used in clinical practice, with a wide variety of scales being used across different healthcare settings. (246) In a prison setting, Blaauw et al. (141) used this approach for the identification of inmates that carried out suicide.

A similar approach was utilised by Lanes (80) and Barton et al., (247) in order to identify self-harm (self-injurious behaviour) in male prisoners. These studies produced AUC values of 0.89 (80) and 0.91, (247) with 93% (80) and 87% (247) of cases correctly classified, both of which are superior to the values obtained in the present study. However, both of these studies used retrospective data to classify the difference between prisoners with and without a history of self-harm, whereas the current study used prospective data to classify whether self-harm occurred among an ACCT population during an active follow-up period.

An issue with these risk factor item sets, as is the case in the present study, is that although these item sets seem to work statistically, it is likely that the identified items involve an element of capitalisation on chance within the specific dataset that is used. Due to this restriction, it is vital that any of these risk factor items sets are revalidated prospectively. Another major issue with a lot of the scales that have been derived in this way are that they use solely retrospective data, and they are never further validated prospectively, meaning that along with the chance capitalisation, no process of causality can be assumed.

Additionally, the practical implementation of risk factor item sets may be limited for a number of reasons. The identified risk factors are often comparatively common in the populations of interest, (245) meaning that an impractical amount of false negatives would be identified. Another issue with the item set identified in the present study, is that many of the items are static in nature. These static items refer to background and lifetime information which cannot change once the item has been affirmed. For example, for the item 'Have you ever cut yourself on purpose?', then if this has been affirmed then this response is fixed as it cannot be 'undone'. This impracticality was highlighted by Vollm & Dolan, (248) who identified that although these simple check lists may be useful to identify those at risk of self-harm upon prison reception, this risk is not static; therefore risk assessment has to be a continuous process and should not be restricted to reception screening.

Within the final gender-specific item sets that were identified, there are a mixture of items relating to indicator correlates, specific self-harm behaviours and prison-demographic and sentencing information. Within both item sets, there are some items which may seem irregular upon first viewing. For males, one such item may be 'alcohol dependency', which was found to be a protective factor, meaning that if this item were affirmed, then self-harm during follow-up would be less likely. Alcohol dependency is often viewed as a self-harm risk-factor, (29, 68, 86, 105-109) so this finding may be seen as an anomaly. However, within this same data set, this finding was replicated with the SHI item 'have you ever abused alcohol', which also had a negative relationship with outcome self-harm.

In and of itself, this finding may seem a little absurd, as it would seem to suggest that to reduce self-harm in prisons, you should make people dependent on alcohol. However, there are a number of plausible reasons for this discovery. One explanation may be that some people only physically harm themselves when under the influence of excessive alcohol. It is recognised that alcohol may trigger self-harm by increasing impulsivity, impairing judgement and increasing the pain threshold. (86) When within a prison setting, the chances of excessive alcohol intake are reduced; therefore resulting in an apparent protection against self-harm. Also, within prison, those that are considered as alcohol dependent may be offered entry into a support network such as Alcoholics Anonymous (or a similar prison support group, intervention or initiative). This supportive network may then offer an unseen confounding 'protection' against self-harm, which appears as though being dependent on alcohol offers protection against self-harm. One further explanation may be that the abuse of alcohol is the coping strategy that is employed by some people, without further escalating to physical self-harm; therefore resulting in an apparent negative relationship with self-harm in follow-up. This same coping strategy explanation may also hold for the apparent protective item of 'driving dangerously'.

Within the female set, one potentially irregular finding was that 'any form of correspondence during the prison stay' was found to increase the chances of self-harm in follow-up. Again, this may not appear immediately intuitive, as it would be imagined that correspondence would offer a supporting role. However, it may be that this correspondence serves as a reminder of a person's life outside of prison, which, with further reinforcement, may act as a trigger to a self-harm event. Also, although correspondence is generally viewed as a good thing, it may also be used for the delivery of bad news, adverse family events or the breakdown of a relationship, all of which may increase stress levels enough to result in self-harm. Despite these plausible explanations, it remains likely that this finding is the result of a capitalisation on chance within this data set, which is why any risk factor item set must also be further validated prospectively in order to assess its stability and integrity.

Although a few items appeared to offer a protective value in the final male set, the only protective factor in the final female item set was whether the index ACCT within the study was the first time that they had been placed on an ACCT. This may imply that the monitoring and supportive system is more active for people on their first ACCT, or, in support of the observed female path model (see Section 8.3) it may reflect that women on a first-time ACCT are less likely to have reached the required level of 'capacity' that is necessary to carry out self-harm.

9.4.1 Alternative Predictive Approach – Machine Learning

The primary purpose of this study was to assess the predictive capacity of pre-existing, standardised measurement instruments, to see whether any of them were useful in identifying prospective self-harm. As this was shown not to be the case, an experimental prediction element was carried out using a classical logistic regression methodology, in order to try and generate a usefully predictive algorithm from the information available within the collected dataset.

In terms of the generation of this predictive algorithm, an alternative approach would be to explore the possibilities of a machine learning application. Machine learning is a relatively new, but rapidly expanding technical field, which is at the intersection of informatics and statistics. (249) It is closely connected with data science and knowledge discovery, and it can potentially have extremely useful applications in healthcare, where many problems involve dealing with uncertainty. (250)

The fundamental basis of machine learning is the artificial generation of knowledge from experience, (251) which occurs when a system can learn from new information in order to complete a specified task. Machine learning can generally be categorised into two categories: unsupervised learning (predictive learning), where the goal is to use input data to predict an outcome; and supervised learning (descriptive learning), where there is no outcome data, and the goal is to discover patterns in the data. (250) Machine learning is a broad field, and it largely deals with the problem of how to extract features from data in order to solve predictive tasks. (250) As the problem in this study is one of prediction, the application of a machine learning process seems appropriate.

When conventional prediction models are utilised in a health research setting, these are often based on the principle that there are only a relatively small number of important risk factors, and that the successful outcome of the prediction depends on the selection of these key variables. (252) However, many risk factors typically interact with each other in a complicated and generally unknown way, and are therefore eliminated from conventional prediction models meaning that potentially useful indicators may remain unidentified. (252)

Inductive machine learning approaches can account for this complexity, as they are able to discover relevant structural and/or temporal patterns in data, which are often hidden and inaccessible to the human expert. (250) Additionally, these machine learning techniques have been shown to be superior to classical statistical models in predicting health outcomes. (252)

However, a major criticism of fully-automatic machine learning techniques is that they often represent a 'black box', meaning that the algorithms and interactions involved in the prediction remain unknown, thus detracting from their full acceptance. (250) This could certainly be viewed as a limitation when applied in the context of predicting self-harm, as there would no explanatory process presented. If this methodology were applied, it would perhaps hinder the understanding of the nature of self-harm within a prison setting, as the risk indicators and their interactions would not be revealed. This could also limit the development and application of any potential interventions, as it would not be possible to target an intervention at a specific risk indicator. Furthermore, it has been identified that health research and biomedical data sets can often contain a lot of uncertainty and incompleteness, and also that some medical problems are just difficult to solve, which makes the application of fully automated machine learning to be difficult or impossible. (250) There are also additional issues associated with machine learning applications, including that classical supervised machine learning techniques require a large amount of training data, which is often not available in the field of health research. Additionally, especially with fully-automated procedures, there is often a danger of modelling artifacts (i.e. undesired outcomes or errors). (250)

Despite the issues associated with a fully-automated machine learning, it is recognised that in such circumstances as are often found in health research, that some human expert input could greatly enhance the results of the machine learning, thus an interactive or ensemble approach could be highly successful. (250) A distinct advantage of keeping a human agent in the loop, is that they are able to perceive the complete context of a situation in a very short amount of time, which is something that a computer cannot do. (250) It would *not* be recommended that a decision-making process based completely on computer-derived information should replace the human decision-making process within most healthcare applications, but machine learning may help to inform certain choices for the human decision maker. This would especially be the case within the context of self-harm prediction, but the potential benefits and insights that could be provided by machine learning should not be overlooked.

9.5 Path Analysis and Explanatory Models

The path analysis revealed separate models of self-harm for males and females, along with a joint model. Path analysis models supported the view that capacity and propensity are integral elements to the self-harm pathway, (76, 78) although this appeared to manifest slightly differently among males and females. However, when gender was included as a moderating factor on a common model, this revealed that the joint model was equivalent across genders, despite the focal relationship being non-significant for males.

Self-harm is the result of excessive stress that an individual experiences, and this may be triggered by a huge array of potential variables, many of which are likely to be of an acute nature. Although models F5-F8, M16, M21, FM5 and FM8 all displayed good model fit statistics, and seemed to offer a valid representation of the relationship between factors that contribute to self-harm, it is recognised that there will also be additional variables that will contribute towards this process. This restriction is also highlighted by the limited R^2 values that were observed in the models, meaning that very little of the variance in the final outcome of self-harm is actually explained by the models. An ideal model would undoubtedly include other variables that were either not captured, not specified, or that are too individualised to be generalisable. Specifically, the particular variable(s) which create the additional stress that may trigger a self-harm event will be numerous, and likely to be different among individuals, depending on their personal circumstances. Although this particular unspecified stressor was not captured within this study, it would remain difficult to accurately capture (at the group level) whatever stressor this may be.

Additionally, the explanatory variables that are included in the models are all based upon a person's responses at the specific time that the interview was carried out (and the data collected). Some of this information is based on a person's individual history, so therefore the reference point remains relatively fixed. However, other elements that were captured at the point of interview are transient, and are likely to fluctuate throughout the period of follow-up. Unfortunately, due to the way that the study was designed, it is unlikely that any acute stressor variables that may trigger self-harm are captured consistently within the present dataset, although the stressor event may potentially be captured among the participants that had a short time span between the data collection and the first self-harm event in follow-up.

Alongside the non-capture of an acute trigger event, it should also be recognised that the instruments used within the study cover different time periods. For example, the PHQ-9 covers the time period two weeks prior to completion, whereas the SHI covers the lifetime period up to the point of completion. Due to the single data-collection point, and the different time periods covered by the variables in the path analysis, the direct specification of a causal model is difficult to achieve as the chronological order of events is not fully captured within the study data. This fact should be taken into account when viewing the path models, as the temporal relationship is only valid between the points of data collection and follow-up.

Also, as a number of different models were tested, it is recognised that there is an experimental element to the modelling within this study. SEM is supposed to be a confirmatory procedure, and once modification specifications are followed, the analysis enters a more exploratory phase. Although the practicalities of SEM almost always lead to an experimental aspect, (216) it has been pointed out that any results obtained through a modification specification search may be unique to the dataset, and that capitalisation on chance can occur within the provided modification indices. (217) Within the final models that have been stated, there remains the possibility that the apparently well-fitting models are the result of chance fluctuations. (215, 217) Therefore, as with the predictive item sets, the results of the path analysis should be viewed with a degree of caution, and the final models should be cross-validated before any real validity can be claimed. (217)

9.6 Research Findings in Context

It is recognised that a prison population are at increased risk of self-harm, but the reasons for this may be variable. It may not actually be being in prison *per se* that is the reason for the high risk, as it may just be that a prison population is made up of a higher proportion of people with a propensity to self-harm, due to the large proportion of prisoners with mental health issues. (54, 108) The specific ACCT population that were considered within the present study, screened positive for mental issues in 94.4% of cases, highlighting this overlap.

Although the prison setting may not be the root cause of the observed self-harm, it is likely that it contributes through certain stressors that are specific to the prison setting, which may ultimately trigger the self-harm that occurs. It has been observed that the high rates of self-harm behaviour in prisons cannot be addressed without adequate attention being paid to the high rates of psychiatric disorder and vulnerability factors in prisoners, (2) and that steps should be taken to improve the prison experience for those with mental health issues. (128)

Self-harm remains a significant problem in prisons, and the identification of those most at risk would help towards the introduction of timely coping strategies which could be key for the successful management of self-harm within a prison setting, especially as self-harm is associated with a disproportionate utilisation of health resources. (34) Screening measures are important for early recognition of risk, (58, 253) and a standardised measure may also provide legal protection. (58, 254)

This study failed to support any standardised measures for the prediction of self-harm, which has also been the case when using standardised measures to predict suicide following self-harm. (245) It has been warned that the use of these standardised scales, or an over-reliance on the identification of risk factors in clinical practice, may provide false reassurance that could be potentially dangerous. (245)

Following a self-harm event, it is suggested that comprehensive psychosocial assessments of the risks and needs that are specific to the individual should be central to the management of people who have self-harmed. (245) This may be a plausible approach following a self-harm event, or perhaps if a prisoner had been identified as being at high risk of self-harm, but considering the limited resources within the prison system, the use of comprehensive assessment instruments would not be feasible in day-to-day practise, especially when being used for early risk assessment at prison reception. (248)

The gender-specific predictive risk item sets identified in this study may be useful in this regard, as they offer the opportunity to classify three levels differing levels of risk that could be used at reception into prison. If the risk classification was medium or high, then a further in-depth assessment could be carried out, as per the recommendations of Chan. (245) Given the high negative predictive values, the predictive item sets appear to function better at screening out self-harm than screening it in. This could therefore be potentially useful to assist the ‘sign-off’ from an ACCT, if the clinician or ACCT team worker deemed it safe to do so. Although this is not the ideal intention, it could still help to save time and focus the limited resources that are available.

Despite an apparently limited predictive power, the implementation of a screening process that is specific to self-harm could certainly contribute to an increased awareness of self-harm and mental health issues amongst prison staff. It has been identified that 29% of prison staff have not received any ACCT training, and 82% have not received any training in mental health awareness. (255) This is consistent with other reports of a lack of staff training and policy, along with an inconsistency in response to self-harm behaviour. (256) Additionally, in over 20% of suicide cases, non-medical staff had documented signs of suicidality, but no referral or further action was taken. (257) This evidence leads to the critical point that an improvement in staff awareness and attitude, along with further training, are important factors which may help prevent self-harm and suicide in prisons. (138, 140, 258). Although this staff awareness shortfall has been identified and is being addressed, it has been acknowledged that much work remains to be done. (259)

9.7 Limitations of the Study

There are a number of limitations with the study, further to those which have been covered in the individual sections.

With regard to the final predictive item sets that were derived from the logistic regression, it has already been pointed out that should these predictive item sets be implemented as clinical decision aids, then further testing would be *essential* to ensure that the predictive properties remain when the predictive algorithm is applied in new samples. It has also been identified that the practical application of the item sets may be limited due to the static nature of the items within them. If monitoring needs to be carried out on a regular basis, then the items within any decision-making tool should also be dynamic, although it may still be potentially useful to apply a non-dynamic instrument on a singular basis, perhaps on initial reception into prison.

However, if these predictive item sets were to be administered to all prisoners upon reception, then it is acknowledged that the item sets may not have been derived on the most appropriate sample, as this study is specific to the high-risk ACCT population, rather than the general prison population. This reiterates the need to revalidate the predictive validity of the item sets, should they be used in this way. Despite this limitation, it should be remembered that the primary aim of the study was to assess the predictive capabilities of pre-existing instruments among high-risk prisoners; the intention was not to develop a new instrument. The ACCT population were chosen for the study as they were seen as the most vulnerable 'at risk' prisoners, and therefore they are the prisoners where a risk tool would be more applicable. As the majority of ACCTs are closed within 24 hours due to the risk of SH being considered as 'low', the idea was to identify a pre-existing, brief questionnaire that may be helpful in predicting risk, so that staff time and resource could be better focused on those that need more help. The ACCT population were also chosen due to their increased self-harm rate, so that the AUC analysis could be carried out on an appropriate sample, whilst retaining a study that was practical to complete in terms of logistics, staffing and cost. Additionally, the ACCT process also offers a natural home for the establishment of a standardised risk assessment, if a suitable instrument had been identified. It would not be logistically feasible for this study to have been run with all new prisoner admissions, and although self-harm incidence may be higher within prison than in a general population, the self-harm rate among the general prison population would still not have been high enough to power a predictive analysis without recruiting a very large sample.

A further limitation of the study concerns the primary outcome that was used in the study, which was a simple binary outcome of whether or not any self-harm was carried out during the follow-up period. It may be argued that alternative outcomes could have been used, but each of these may also present their own difficulties. For example, the number of self-harm events could have been used as the primary outcome, but this would have made it more difficult to classify risk, as the risk level should still be perceived as high where there is *any* occurrence of a self-harm event. Additionally, a count of numerous minor, superficial, self-harm events may be far less severe than a single suicide attempt. With this in mind, severity may also have been considered as an alternative outcome. However, the intended severity of an incident is not always known *a priori* by the person carrying out the self-harm. For this reason, it has been stated that lethality can be misleading as an indicator of the severity of self-harm, as there is a variable gap between objective and subjective lethality. (16)

It was seen in Chapter 2.2.3 (Predictors of Self-Harm) that there are many factors that have been associated with an increased risk of self-harm, and it is recognised that not all of these factors were represented in the dataset that was collected. For example, there is evidence to suggest that a major contributing factor to the occurrence of self-harm is when an individual has experience of adolescent sexual molestation or physical assault, (260) but this historical information was not collected as part of the study. Information relating to other prison-specific (e.g. shared cell status), general (e.g. recent arguments, emotional trauma), and more specific (e.g. anxiety, impulsivity, psychotic tendencies, neuroticism, extroversion, etc.) potential self-harm indicators is also missing from the dataset. Although this may be seen as a limitation of the study, it would have been impractical to collect information relating to *all* potential self-harm indicators, and again, the primary aim of the study should be taken into consideration.

On a similar theme, it is acknowledged that although none of the pre-existing instruments were usefully predictive, only five instruments were selected for the study. A scoping exercise was performed to identify the most appropriate instruments for the study, as it is not possible test *all* available pre-existing scales that relate to different self-harm risk factors. A number of potential instruments were therefore not selected for the study, but of those identified and considered this was largely due to them being deemed inappropriate in some way. A large amount of instruments did not satisfy the practical criteria that had been specified for a useful application within a prison setting, and a number of other instruments were rejected following the pilot study. The final selection of five instruments all satisfied the criteria that anyone can administer them (with no formal training), they are not specific to being post self-harm, they are brief (<50 items) and understandable (demonstrating good face validity among prisoners).

Additionally, the sample size of the study may not have been powered appropriately, but it is likely that this had little effect on the final outcome. The AUC values that were used for the initial power calculations (0.8 Vs 0.9) turned out to be an optimistic estimate, as the highest AUC observed in the study was 0.671. However, it is likely that the implications of this inflated estimate are limited, as the direct comparison of two competing instruments was not necessary due to the poor performance across *all* instruments. The sample size remained adequate for all psychometric analyses, (165) although it may be considered a limitation of the logistic regression and structural equation modelling, especially among the female sample.

Furthermore, this study only considered whether or not a self-harm event occurred during the follow-up period. However, this follow-up period was variable for each individual, as it was classified as 'six months after the date of questionnaire completion, or up to the point of release from index prison stay (whichever is sooner)'. This follow-up time frame was intended to be variable, as this would also be the case for the practical implementation of any potential screening instrument. An alternative approach to the assessment methodology would have been to consider a time-coded model (i.e. survival analysis), where the variable follow-up time is taken into account.

This survival analysis approach was assessed alongside the utilised methodology within the broader HTA study, (261) and it was found to largely replicate the findings of the non-time coded analysis. Although there was very little difference, in this particular application any major difference in the findings would probably have limited practical relevance. Within the context of this study, the intention was to identify a screening instrument which would be valid within a given time-frame (up to six months). Usually when time-coded models are applied, the intention is to model the time-to-event data, in order to estimate survival time following some clinical episode. (210) Although this approach could potentially be useful if reduction in time-to-self-harm were being investigated (and interventions were being compared), it has limited value when assessing the prospective prediction of a given outcome within a particular time frame.

Other elements of the study may also have had an unknown impact on the outcome.

Firstly, the active recruitment period was amended from 24 hours post-ACCT to two weeks following the initiation of the ACCT. Initially, the intention was to recruit responders within the acute phase of their ACCT. However, this proved to be logistically difficult, especially when prison staff deemed certain prisoners as too vulnerable to approach for recruitment. Although this amendment to the study was necessary, if the intention is to administer a questionnaire as part of a standardised ACCT interview, then it is acknowledged that the difference in the results obtained at an acute and non-acute stage may be variable.

Secondly, the wording of the scoring categories of the BSL-23 was changed, resulting in score categories that are more reflective of frequency rather than intensity of symptoms. It is acknowledged that this could impact upon the responses, and therefore the predictive validity of the instrument, but the face-validity of the original response categories was limited, possibly due to translation issues, so this amendment was justified.

Thirdly, all of the items within the questionnaire pack were verbally delivered to address the issue of low literacy among a prison population. (262) As all responses back to the researcher were also verbal, this may impact upon the responses that were given, due to the prisoner-researcher interaction. For example, prisoner responses may be affected for reasons of posturing, shame, sensitivity, manipulation, or a fear of reprisal. (59)

Lastly, some of the recorded self-harm events may have been carried out for manipulative purposes, and the impact of this is unknown. However, it has been identified that there are not many self-injurious behaviours that are concurrently manipulative, of low lethality, and performed with low or no suicidal intent. (263) Even when they are carried out in order to manipulate, self-injurious behaviours can still be performed with suicidal intent and may result in death or serious injury. (264) With this in mind, along with the practicalities of classification within the prison system, the active definition of self-harm within the study is again seen as the most appropriate, as the intent of any self-harm is immaterial.

9.8 Recommendations for Future Research

Within this study, a logistic regression analysis revealed sets of risk indicator items that were statistically significant in predicting future self-harm. Any item set which is derived in this way should always be re-tested and validated before it is implemented into practice.

Going forward, the specific purpose of any standardised instrument needs to be considered. From a measurement perspective, different properties are required of instruments depending on the purpose, and sometimes these properties are of competing interest. (208) The functionality of the predictive algorithms as monitoring tools may be limited by the static nature of the items, so it is suggested that the specific function, target population and timing of administration are considered for any predictive item set or clinical decision aid. It is therefore recommended that future research could consider two areas where a standardised instrument may be appropriate:

Firstly, a brief screening tool that could be administered to all new receptions into prison, to identify those that are considered particularly vulnerable, or at high risk of self-harm. This would be a singular administration, so the incorporation of static items would not present an issue. A positive identification would lead to a more in-depth psychological assessment, and potentially an individualised care plan. Although there is already a brief screening tool which is utilised upon reception into prison, (139) this is designed to broadly identify physical and mental health issues rather than a specific vulnerability to self-harm, so a brief self-harm-specific screening tool could provide additional useful information at this point.

Secondly, a dynamic self-harm risk assessment instrument that could be used for monitoring purposes. This element may focus more on the present affective state of the individual, which is more likely to fluctuate over time.

It is possible that it may be more appropriate to create a composite model of risk assessment that incorporates both a (more) static element alongside a dynamic element. Potentially, this may correspond with the two areas of self-harm capacity and self-harm propensity that have been identified as key contributors. Although it is recommended that more research should be carried out in this area relating to the explanatory model, it is acknowledged that it may be difficult to create a standardised instrument from this due to the complex and individualised circumstances of each person. This view is supported by Chan et al., (245) who state that the idea of risk assessment as risk prediction is a fallacy, as we are unable to say with any certainty who will and will not go on to have poor outcomes.

Given that it is so difficult to predict self-harm, even with the large amount of apparent risk factors which have been previously identified, it may be beneficial to run a future study that placed the emphasis on investigating the protective factors of self-harm. It has been recognised that there is not much known about protective factors, (58, 247, 265, 266) and perhaps understanding more about these factors would help to improve the validity of risk assessment. If research in this area were successful, then perhaps this would also be more directly transferable into targeting effective interventions.

This study investigated the moderator effects of gender on self-harm within an SEM framework. It is recognised that there are a vast number of potential moderators and mediators which may have an impact on prospective self-harm and could inform interventional practice, and it is recommended that future research be carried out in this area. However, any research in this area should be conceptually derived, with a confirmatory approach applied to the modelling.

The primary aim of this study was to determine whether any pre-existing instruments could predict self-harm among an ACCT population. When this was shown not to be the case, an exploratory logistic regression analysis did reveal predictive algorithms which were statistically significant in predicting future self-harm, although the operational functionality of these may be limited. With regard to maximising the predictive capacity of available data, it would be recommended that a machine learning approach is utilised in order to investigate this. The potential benefits of machine learning are presented in Section 9.4.1, and although it would be recommended that a human should remain in any final decision-making process, the potential of machine learning techniques should be explored in order to aid and optimise the prediction of self-harm.

Additionally, this study only considered the binary primary outcome of whether or not a self-harm event occurred during the follow-up period. An area for potential future research may be to consider different self-harm types and patterns, and for each of these to be considered separately. If these separate types of self-harm were investigated as different outcomes, it may again be beneficial to implement machine learning algorithms. It is likely that separate pathways and indicators will lead to different outcomes, and machine learning methodologies could provide a way to cope with the additional complexity that is being modelled.

Finally, another direction for future research may be to focus on one specific type of self-harm that is more clinically relevant, rather than considering self-harm as a single inclusive outcome. This could be a specific type (e.g. non-severe cutting) that is perhaps more clinically manageable, and where treatment-management or interventions may be more readily introduced. Alternatively, the focus may be on more extreme types of self-harm, where prevention would be more appropriate than a managed-treatment approach.

9.9 Summary

The key findings from this body of work can be summarised as follows:

i. Self-harm is a common occurrence among the prison ACCT population, with 29.1% of the study participants deliberately self-harming during the follow up period, although this varied considerably across gender and participating prisons.
ii. Four of the five selected instruments did display a certain level of psychometric validity among the study population; therefore validating the cut-points for the predictive analysis. However, all instruments required some refinement to meet the strict measurement criteria of the Rasch model.
iii. Of the five pre-existing instruments that were selected for the study, none of these displayed a meaningful predictive validity.
iv. Logistic regression analysis did reveal gender-specific item sets, producing predictive algorithms which were statistically significant in predicting future self-harm; however, the operational functionality of these item sets may be limited.
v. Structural equation modelling revealed an insightful explanatory model of the process that may be involved in the culmination of self-harm in prison. Path analysis models supported the view that capacity and propensity are integral elements to the self-harm pathway, although the complete explanatory model is likely to be more complex.

10 List of References

1. Frühwald S, Frottier P. Suicide in prison. The Lancet. 2005;366(9493):1242-4.
2. Jenkins R, Bhugra D, Meltzer H, Singleton N, Bebbington P, Brugha T, et al. Psychiatric and social aspects of suicidal behaviour in prisons. Psychol Med. 2005 Feb;35(2):257-69. PubMed PMID: 15841683.
3. Shaw J, Baker D, Hunt IM, Moloney A, Appleby L. Suicide by prisoners national clinical survey. The British Journal of Psychiatry. 2004;184(3):263-7.
4. HM Prison Service. Safer custody group. The ACCT approach. Caring for people at risk in prison. London: 2005.
5. Perry AE, Marandos R, Coulton S, Johnson M. Screening tools assessing risk of suicide and self-harm in adult offenders: a systematic review. Int J Offender Ther Comp Criminol. 2010 Oct;54(5):803-28. PubMed PMID: 20223913.
6. Commission on Chronic Illness. Prevention of chronic illness. Chronic Illness in the United States. Cambridge, MA: Harvard University Press; 1957.
7. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organisation; 1968.
8. Burke W, Coughlin SS, Lee NC, Weed DL, Khoury MJ. Application of population screening principles to genetic screening for adult-onset conditions. Genetic testing. 2001;5(3):201-11.
9. Linnane E, Paul A, Parry R. Screening of newborn infants for cholestatic hepatobiliary disease: Does test fulfil screening criteria? BMJ: British Medical Journal. 1999;319(7222):1435.
10. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bulletin of the World Health Organization. 2008;86(4):317-9.
11. Brodersen J. Measuring psychosocial consequences of false-positive screening results-breast cancer as an example: Copenhagen University, Faculty of Health and Medical Sciences, Department of Public Health; 2006.
12. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia, Critical Care & Pain. 2008;8(6):221-3.
13. Favazza AR, Favazza B. Bodies under siege: Self-mutilation in Culture and Psychiatry: Johns Hopkins University Press; 1987.
14. Klonsky ED, Oltmanns TF, Turkheimer E. Deliberate self-harm in a nonclinical population: prevalence and psychological correlates. The American journal of psychiatry. 2003 Aug;160(8):1501-8. PubMed PMID: 12900314. Pubmed Central PMCID: 4362719.
15. McAllister M. Multiple meanings of self harm: A critical review. International journal of mental health nursing. 2003;12(3):177-85.

16. Ougrin D, Zundel T. Defining self-harm. In: Ougrin D, Zundel T, Ng AV, editors. Self-Harm in Young People: A Therapeutic Assessment manual. London: Hodder Arnold; 2009. p. 1-17.
17. Emerson L, E. The case of Miss A: A preliminary report of a psychoanalytic study and treatment of a case of self-mutilation. Psychoanalytic review. 1913;1(1):41-54.
18. Menninger KA. Man against himself. 1938.
19. Kreitman N, Philip AE, Greer S, Bagley C. Parasuicide. The British Journal of Psychiatry. 1969;115(523):746-7.
20. Kreitman N. Parasuicide: John Wiley & Sons; 1977.
21. Beck AT, Resnik HL, Lettieri DJ. The prediction of suicide: Charles Press Publishers Bowie, Maryland; 1974.
22. Beck AT, Beck R, Kovacs M. Classification of suicidal behaviors: I. Quantifying intent and medical lethality. The American journal of psychiatry. 1975.
23. Messer JM, Fremouw WJ. A critical review of explanatory models for self-mutilating behaviors in adolescents. Clin Psychol Rev. 2008 Jan;28(1):162-78. PubMed PMID: 17618024.
24. Favazza AR. The coming of age of self-mutilation. J Nerv Ment Dis. 1998 May;186(5):259-68. PubMed PMID: 9612442.
25. Gratz KL. Measurement of deliberate self-harm: Preliminary data on the deliberate self-harm inventory. Journal of Psychopathology and Behavioral Assessment. 2001 Dec;23(4):253-63. PubMed PMID: WOS:000172934900005. English.
26. Muehlenkamp JJ, Gutierrez PM. An Investigation of Differences between Self-Injurious Behavior and Suicide Attempts in a Sample of Adolescents. Suicide and Life-Threatening Behavior. 2004 Spr;34(1):12-23. PubMed PMID: WOS:000220498700003. English.
27. O'Carroll PW, Berman AL, Maris RW, Moscicki EK, Tanney BL, Silverman MM. Beyond the Tower of Babel: a nomenclature for suicidology. Suicide & life-threatening behavior. 1996 Fall;26(3):237-52. PubMed PMID: 8897663.
28. Silverman MM. Challenges to classifying suicidal ideations, communications, and behaviours. International handbook of suicide prevention research, policy and practice RC O'Connor, S Platt and J Gordon Malden, MA John Wiley & Sons. 2011:9-25.
29. Skegg K. Self-harm. Lancet. 2005 Oct 22-28;366(9495):1471-83. PubMed PMID: 16243093.
30. Favazza AR. Bodies under siege : self-mutilation and body modification in culture and psychiatry 2nd ed. Baltimore: Johns Hopkins University Press; 1989.
31. Duffy DF. Self-injury. Psychiatry. 2006;5(8):263-5.
32. Favazza AR, Rosenthal RJ. Varieties of Pathological Self-mutilation. Behavioural neurology. 1990;3(2):77-85. PubMed PMID: 24487146. Epub 1990/01/01. eng.
33. Self-harm: The Short-term Physical and Psychological Management and Secondary Prevention of Self-harm in Primary and Secondary Care: Quick Reference Guide: National Institute for Clinical Excellence; 2004.

34. Smith HP, Kaminski RJ. Inmate self-injurious behaviors - Distinguishing Characteristics Within a Retrospective Study. Crim Justice Behav. 2010 Jan;37(1):81-96. PubMed PMID: WOS:000272798700005. English.
35. Latimer S, Covic T, Tennant A. Co-calibration of deliberate self harm (DSH) behaviours: towards a common measurement metric. Psychiatry Res. 2012 Nov 30;200(1):26-34. PubMed PMID: 22727708.
36. Turp M. Hidden Self-Harm: Narratives from Psychotherapy. London: Jessica Kingsley Publishers; 2002.
37. Linehan MM, Comtois KA, Brown MZ, Heard HL, Wagner A. Suicide Attempt Self-Injury Interview (SASII): development, reliability, and validity of a scale to assess suicide attempts and intentional self-injury. Psychol Assess. 2006 Sep;18(3):303-12. PubMed PMID: 16953733. Epub 2006/09/07. eng.
38. Walsh BW, Rosen PM. Self Mutilation: Theory, Research, and Treatment. New York: The Guilford Press; 1988.
39. Sansone RA, Wiederman MW, Sansone LA. The Self-Harm Inventory (SHI): development of a scale for identifying self-destructive behaviors and borderline personality disorder. J Clin Psychol. 1998 Nov;54(7):973-83. PubMed PMID: 9811134.
40. Best R. Students who self-harm: a case study of prevalence, awareness and response in an English university. Pastoral Care in Education. 2009 2009/09/01;27(3):165-203.
41. Patton GC, Harris R, Carlin JB, Hibbert M, Coffey C, Schwartz M, et al. Adolescent suicidal behaviours: a population-based study of risk. Psychological medicine. 1997;27(03):715-24.
42. Ross RR, McKay HB. Self-mutilation: Lexington Books Lexington, MA; 1979.
43. Lloyd-Richardson EE, Perrine N, Dierker L, Kelley ML. Characteristics and functions of non-suicidal self-injury in a community sample of adolescents. Psychological medicine. 2007;37(08):1183-92.
44. Freeman DJ, Wilson K, Thigpen J, McGee RK. Assessing intention to die in self-injury behavior. Psychological assessment of suicidal risk. 1974:18-42.
45. Jacobs DG. The Harvard Medical School guide to suicide assessment and intervention: Jossey-Bass; 1999.
46. Maris RW. Assessment and prediction of suicide: Guilford Press; 1992.
47. Brent DA, Perper JA, Goldstein CE, Kolko DJ, Allan MJ, Allman CJ, et al. Risk factors for adolescent suicide: a comparison of adolescent suicide victims with suicidal inpatients. Archives of general psychiatry. 1988;45(6):581-8.
48. Douglas J, Cooper J, Amos T, Webb R, Guthrie E, Appleby L. “Near-fatal” deliberate self-harm: characteristics, prevention and implications for the prevention of suicide. Journal of affective disorders. 2004;79(1):263-8.
49. Brown GK, Henriques GR, Sosdjan D, Beck AT. Suicide intent and accurate expectations of lethality: predictors of medical lethality of suicide attempts. Journal of consulting and clinical psychology. 2004;72(6):1170.
50. Millstein SG, Halpern-Felsher BL. Perceptions of risk and vulnerability. J Adolescent Health. 2002;31(1):10-27.

51. Carter GL, Clover K, Whyte IM, Dawson AH, Este CD. Postcards from the EDge project: randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. Bmj. 2005;331(7520):805.
52. Harrington R, Kerfoot M, Dyer E, McNiven F, Gill J, Harrington V, et al. Randomized trial of a home-based family intervention for children who have deliberately poisoned themselves. Journal of the American Academy of Child & Adolescent Psychiatry. 1998;37(5):512-8.
53. Hawton K, Harriss L, Simkin S, Bale E, Bond A. Self-cutting: patient characteristics compared with self-poisoners. Suicide & life-threatening behavior. 2004 Autumn;34(3):199-208. PubMed PMID: 15385174.
54. Nock MK, Joiner TE, Gordon KH, Lloyd-Richardson E, Prinstein MJ. Non-suicidal self-injury among adolescents: Diagnostic correlates and relation to suicide attempts. Psychiatry Research. 2006 Sep 30;144(1):65-72. PubMed PMID: WOS:000241171500007. English.
55. Glenn CR, Klonsky ED. Prospective prediction of nonsuicidal self-injury: a 1-year longitudinal study in young adults. Behav Ther. 2011 Dec;42(4):751-62. PubMed PMID: 22036002. Pubmed Central PMCID: 4433036.
56. De Leo D, Burgis S, Bertolote JM, Kerkhof AJ, Bille-Brahe U. Definitions of suicidal behavior: Lessons learned from the WHO/EURO Multicentre Study. Crisis. 2006;27(1):4-15.
57. Appelbaum KL, Savageau JA, Trestman RL, Metzner JL, Baillargeon J. A National Survey of Self-Injurious Behavior in American Prisons. Psychiat Serv. 2011 Mar;62(3):285-90. PubMed PMID: WOS:000287845800008. English.
58. Lohner J, Konrad N. Risk factors for self-injurious behaviour in custody: Problems of definition and prediction. International journal of prisoner health. 2007;3(2):135-61.
59. Lohner J, Konrad N. Deliberate self-harm and suicide attempt in custody: distinguishing features in male inmates' self-injurious behavior. International journal of law and psychiatry. 2006 Sep-Oct;29(5):370-85. PubMed PMID: 16782200.
60. Fulwiler C, Forbes C, Santangelo SL, Folstein M. Self-mutilation and suicide attempt: distinguishing features in prisoners. The journal of the American Academy of Psychiatry and the Law. 1997;25(1):69-77. PubMed PMID: 9148884.
61. Haycock J. Manipulation and suicide attempts in jails and prisons. The Psychiatric quarterly. 1989 Spring;60(1):85-98. PubMed PMID: 2813645.
62. Verona E, Sachs-Ericsson N, Joiner TE, Jr. Suicide attempts associated with externalizing psychopathology in an epidemiological sample. The American journal of psychiatry. 2004 Mar;161(3):444-51. PubMed PMID: 14992969.
63. Nock MK. Actions speak louder than words: An elaborated theoretical model of the social functions of self-injury and other harmful behaviors. Applied and Preventive Psychology. 2008;12(4):159-68.
64. Nock MK, Prinstein MJ. A functional approach to the assessment of self-mutilative behavior. J Consult Clin Psychol. 2004 Oct;72(5):885-90. PubMed PMID: 15482046.

65. Linehan MM. One Population or Two? Annals of the New York Academy of Sciences. 1986;487(1):16-33.
66. Stanley B, Winchell R, Molcho A, Simeon D, Stanley M. Suicide and the self-harm continuum: Phenomenological and biochemical evidence. International Review of Psychiatry. 1992;4(2):149-55.
67. Brausch AM, Gutierrez PM. Differences in non-suicidal self-injury and suicide attempts in adolescents. Journal of Youth and Adolescence. 2010;39(3):233-42.
68. Hamza CA, Stewart SL, Willoughby T. Examining the link between nonsuicidal self-injury and suicidal behavior: A review of the literature and an integrated model. Clinical Psychology Review. 2012;32(6):482-95.
69. Morgan HG, Burns-Cox CJ, Pocock H, Pottle S. Deliberate self-harm: clinical and socio-economic characteristics of 368 patients. The British Journal of Psychiatry. 1975;127(6):564-74.
70. Favazza AR, Rosenthal RJ. Diagnostic issues in self-mutilation. Hospital & community psychiatry. 1993 Feb;44(2):134-40. PubMed PMID: 8432496.
71. Owens D, Horrocks J, House A. Fatal and non-fatal repetition of self-harm. Systematic review. The British journal of psychiatry : the journal of mental science. 2002 Sep;181(SEPT.):193-9. PubMed PMID: 12204922.
72. Hawton K, Hall S, Simkin S, Bale L, Bond A, Codd S, et al. Deliberate self-harm in adolescents: a study of characteristics and trends in Oxford, 1990-2000. Journal of child psychology and psychiatry, and allied disciplines. 2003 Nov;44(8):1191-8. PubMed PMID: 14626459.
73. Cooper J, Kapur N, Webb R, Lawlor M, Guthrie E, Mackway-Jones K, et al. Suicide after deliberate self-harm: a 4-year cohort study. The American journal of psychiatry. 2005 Feb;162(2):297-303. PubMed PMID: 15677594.
74. Hawton K, Harriss L. Deliberate self-harm in young people: characteristics and subsequent mortality in a 20-year cohort of patients presenting to hospital. The Journal of clinical psychiatry. 2007 Oct;68(10):1574-83. PubMed PMID: 17960975.
75. Stanley B, Gameroff MJ, Michalsen V, Mann JJ. Are suicide attempters who self-mutilate a unique population? American Journal of Psychiatry. 2001;158(3):427-32.
76. Joiner T. Why people die by suicide. Cambridge, MA: Harvard University Press; 2005.
77. Latimer S, Meade T, Tennant A. Measuring engagement in deliberate self-harm behaviours: psychometric evaluation of six scales. BMC Psychiatry. 2013;13:4. PubMed PMID: 23286337. Pubmed Central PMCID: 3605243.
78. Ireland JL, York C. Exploring application of the Interpersonal-Psychological Theory of Suicidal Behaviour to self-injurious behaviour among women prisoners: Proposing a new model of understanding. International journal of law and psychiatry. 2012 Jan-Feb;35(1):70-6. PubMed PMID: WOS:000300202100010. English.

79. Plener PL, Libal G, Keller F, Fegert J, Muehlenkamp JJ. An international comparison of adolescent non-suicidal self-injury (NSSI) and suicide attempts: Germany and the USA. Psychological medicine. 2009;39(09):1549-58.
80. Lanes E. Identification of risk factors for self-injurious behavior in male prisoners. Journal of forensic sciences. 2009 May;54(3):692-8. PubMed PMID: 19432746.
81. Ministry of Justice. Safety in Custody Statistics Quarterly Bulletin. J anuary to March 2012. England and Wales. 2012.
82. Daniel AE. Preventing suicide in prison: a collaborative responsibility of administrative, custodial, and clinical staff. The journal of the American Academy of Psychiatry and the Law. 2006;34(2):165-75. PubMed PMID: 16844795.
83. Chapman AL, Dixon-Gordon KL. Emotional antecedents and consequences of deliberate self-harm and suicide attempts. Suicide & life-threatening behavior. 2007 Oct;37(5):543-52. PubMed PMID: 17967121.
84. Krasser G, Rossmann P, Zapotoczky HG. Suicide and Auto-aggression, Depression, Hopelessness, Self-communication--Results of a Prospective Study. Archives of Suicide Research. 2003;7(3):237-46.
85. Kemperman I, Russ MJ, Shearin E. Self-injurious behavior and mood regulation in borderline patients. Journal of personality disorders. 1997;11(2):146-57.
86. Butler J, Longhitano C. Self-harm. Medicine. 2008;36(9):455-8.
87. Zila LM, Kiselica MS. Understanding and counseling self-mutilation in female adolescents and young adults. Journal of Counseling & Development. 2001;79(1):46-52.
88. Woods J. Layers of meaning in self-cutting. Journal of Child Psychotherapy. 1988;14(1):51-60.
89. Suyemoto KL, MacDonald ML. Self-cutting in female adolescents. Psychotherapy: Theory, Research, Practice, Training. 1995;32(1):162.
90. Miller F, Bashkin EA. Depersonalization and self-mutilation. The Psychoanalytic Quarterly. 1974.
91. Firestone RW, Seiden RH. Suicide and the continuum of self-destructive behavior. Journal of American College Health. 1990;38(5):207-13.
92. Haines J, Williams CL, Brain KL, Wilson GV. The psychophysiology of self-mutilation. J Abnorm Psychol. 1995 Aug;104(3):471-89. PubMed PMID: 7673571.
93. Suyemoto KL. The functions of self-mutilation. Clinical psychology review. 1998;18(5):531-54.
94. Crouch W, Wright J. Deliberate self-harm at an adolescent unit: A qualitative investigation. Clinical Child Psychology and Psychiatry. 2004;9(2):185-204.
95. Favazza AR. Why patients mutilate themselves. Psychiat Serv. 1989;40(2):137-45.
96. Yates TM. The developmental psychopathology of self-injurious behavior: Compensatory regulation in posttraumatic adaptation. Clinical Psychology Review. 2004;24(1):35-74.

97. Shelton D, Barta B, Trestman R, Wakai S. Biopsychosocial Vulnerability-Stress Modeling for an Incarcerated Population. Journal for Evidence-based Practice in Correctional Health. 2016;1(1):1.
98. Association AP, Association AP. Diagnostic and statistical manual-text revision (DSM-IV-TRim, 2000): American Psychiatric Association; 2000.
99. Stetka BS, Correll CU. A guide to DSM-5. Med scape. 2013;21.
100. Fliege H, Lee J-R, Grimm A, Klapp BF. Risk factors and correlates of deliberate self-harm behavior: A systematic review. Journal of psychosomatic research. 2009;66(6):477-93.
101. Klonsky ED. The functions of deliberate self-injury: a review of the evidence. Clin Psychol Rev. 2007 Mar;27(2):226-39. PubMed PMID: 17014942.
102. Haw C, Hawton K, Houston K, Townsend E. Psychiatric and personality disorders in deliberate self-harm patients. The British journal of psychiatry : the journal of mental science. 2001 Jan;178(1):48-54. PubMed PMID: 11136210.
103. Zlotnick C, Mattia JI, Zimmerman M. Clinical correlates of self-mutilation in a sample of general psychiatric patients. The Journal of nervous and mental disease. 1999;187(5):296-301.
104. Herpertz S, Sass H, Favazza A. Impulsivity in self-mutilative behavior: psychometric and biological findings. Journal of psychiatric research. 1997;31(4):451-65.
105. Kapur N. Self-harm in the general hospital. Psychiatry. 2009;8(6):189-93.
106. Kapur N, Gask L. Introduction to suicide and self-harm. Psychiatry. 2009;8(7):233-6.
107. Hawton K, Saunders KEA, O'Connor RC. Self-harm and suicide in adolescents. The Lancet. 2012;379(9834):2373-82.
108. Meltzer H, Lader D, Corbin T, Singleton N, Jenkins R, Brugha T. Non-fatal suicide behaviour among adults aged 16-74 in Great Britain. London: The Stationery Office, 2002.
109. Williams D, Padmanabhan V. Deliberate self-harm. The Foundation Years. 2009;5(2):72-4.
110. Lilley R, Owens D, Horrocks J, House A, Noble R, Bergen H, et al. Hospital care and repetition following self-harm: multicentre comparison of self-poisoning and self-injury. The British journal of psychiatry : the journal of mental science. 2008 Jun;192(6):440-5. PubMed PMID: 18515895.
111. Sinclair J, Hawton K. Suicide and deliberate self-harm. In: Lloyd G, Guthrie E, editors. Handbook of liaison psychiatry. Cambridge: Cambridge University Press; 2007.
112. Ougrin D, Kyriakopoulos M. Prevalence and Natural History of Self-Harm. Self-Harm in Young People: A Therapeutic Assessment Manual. 2009:19.
113. Cooper J, Kapur N, Dunning J, Guthrie E, Appleby L, Mackway-Jones K. A clinical tool for assessing risk after self-harm. Annals of Emergency Medicine. 2006 Oct;48(4):459-66. PubMed PMID: WOS:000240886200018. English.

114. Cooper J, Kapur N, Mackway-Jones K. A comparison between clinicians' assessment and the Manchester Self-Harm Rule: a cohort study. Emergency Medicine Journal. 2007 Oct 1;24(10):720-1. PubMed PMID: WOS:000249778900012. English.
115. Steeg S, Kapur N, Webb R, Applegate E, Stewart S, Hawton K, et al. The development of a population-level clinical screening tool for self-harm repetition and suicide: the ReACT Self-Harm Rule. Psychological Medicine. 2012;42(11):2383-94.
116. McMillan D, Gilbody S, Beresford E, Neilly L. Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. Psychological medicine. 2007;37(6):769-78.
117. Beck AT, Weissman A, Lester D, Trexler L. The measurement of pessimism: the hopelessness scale. J Consult Clin Psychol. 1974 Dec;42(6):861-5. PubMed PMID: 4436473.
118. Jacobson CM, Gould M. The Epidemiology and Phenomenology of Non-Suicidal Self-Injurious Behavior Among Adolescents: A Critical Review of the Literature. Archives of Suicide Research. 2007;11(2):129-47. PubMed PMID: WOS:000208193200001. English.
119. Briere J, Gil E. Self-mutilation in clinical and general population samples: prevalence, correlates, and functions. The American journal of orthopsychiatry. 1998 Oct;68(4):609-20. PubMed PMID: 9809120.
120. Favazza AR, DeRosear L, Conterio K. Self-mutilation and eating disorders. Suicide & life-threatening behavior. 1989 Winter;19(4):352-61. PubMed PMID: 2609364.
121. Hawton K, Fagg J, Simkin S, Bale E, Bond A. Trends in deliberate self-harm in Oxford, 1985-1995. Implications for clinical services and the prevention of suicide. The British journal of psychiatry : the journal of mental science. 1997 Dec;171(DEC.):556-60. PubMed PMID: 9519096.
122. Hawton K, Bergen H, Casey D, Simkin S, Palmer B, Cooper J, et al. Self-harm in England: a tale of three cities. Multicentre study of self-harm. Soc Psychiatry Psychiatr Epidemiol. 2007 Jul;42(7):513-21. PubMed PMID: 17516016.
123. Murphy E, Dickson S, Donaldson I, al. e. The MaSH Project: self-harm in Manchester, 1 September 2003 to 31 August 2005. Manchester: Manchester Mental Health and Social Care Trust NHS and The University of Manchester, 2007.
124. Swannell SV, Martin GE, Page A, Hasking P, St John NJ. Prevalence of nonsuicidal self-injury in nonclinical samples: Systematic review, meta-analysis and meta-regression. Suicide and Life-Threatening Behavior. 2014;44(3):273-303.
125. Singleton N, Meltzer H, Gatward R. Psychiatric morbidity among prisoners in England and Wales. London: Office for National Statistics, 1998.
126. Fazel S, Danesh J. Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. The lancet. 2002;359(9306):545-50.
127. Bland RC, Newman SC, Thompson AH, Dyck RJ. Psychiatric disorders in the population and in prisoners. International journal of law and psychiatry. 1998;21(3):273-9.
128. Ginn S. Dealing with mental disorder in prisoners. BMJ. 2012;345(7885):e7280. PubMed PMID: 23175711.

129. Ministry of Justice Statistics Bulletin. Safety in Custody Statistics England and Wales. Deaths in prison custody to June 2015. Assaults and Self-harm to March 2015. 2015.
130. Liebling A. Vulnerability and Prison Suicide. Brit J Criminol. 1995 Spr;35(2):173-87. PubMed PMID: WOS:A1995QV91100001. English.
131. Powis B. Offenders' risk of serious harm: a literature review. Offenders and Corrections Unit, Home Office Research, Development and Statistics Directorate., 2002.
132. Brooker C, Repper J, Beverley C, al. e. Mental Health Services and Prisoners: A Review. Sheffield, UK: University of Sheffield, School of Health and Related Research., 2002.
133. Daigle MS, Côté G. Nonfatal Suicide-Related Behavior Among Inmates: Testing for Gender and Type Differences. Suicide and Life-Threatening Behavior. 2006;36(6):670-81.
134. Borrill J, Burnett R, Atkins R, Miller S, Briggs D, Weaver T, et al. Patterns of self-harm and attempted suicide among white and black/mixed race female prisoners. Crim Behav Ment Heal. 2003;13(4):229-40.
135. Sakelliadis EI, Papadodima SA, Sergentanis TN, Giotakos O, Spiliopoulou CA. Self-injurious behavior among Greek male prisoners: Prevalence and risk factors. Eur Psychiat. 2010 Apr;25(3):151-8. PubMed PMID: WOS:000276028200004. English.
136. Safety in custody statistics: https://www.gov.uk/government/collections/safety-in-custody-statistics: Ministry of Justice; [cited 2016 February].
137. Corston J. Corston Report. 2007.
138. Humber N, Hayes A, Senior J, Fahy T, Shaw J. Identifying, monitoring and managing prisoners at risk of self-harm/suicide in England and Wales. J Forensic Psychi Ps. 2011;22(1):22-51. PubMed PMID: WOS:000287503900003. English.
139. Gavin N, Parsons S, Grubin D. Reception screening and mental health needs assessment in a male remand prison. Psychiatric Bulletin. 2003;27(7):251-3.
140. Hawton K, Linsell L, Adeniji T, Sariaslan A, Fazel S. Self-harm in prisons in England and Wales: an epidemiological study of prevalence, risk factors, clustering, and subsequent suicide. The Lancet. 2014;383(9923):1147-54.
141. Blaauw E, Kerkhof AJ, Hayes LM. Demographic, criminal, and psychiatric factors related to inmate suicide. Suicide and Life-Threatening Behavior. 2005;35(1):63-75.
142. Beck AT, Ward C, Mendelson M. Beck depression inventory (BDI). Archives of general psychiatry. 1961;4(6):561-71.
143. Perry AE, Gilbody S. Detecting and predicting self-harm behaviour in prisoners: a prospective psychometric analysis of three instruments. Soc Psychiatry Psychiatr Epidemiol. 2009 Oct;44(10):853-61. PubMed PMID: 19290444.
144. Gray NS, Hill C, McGleish A, Timmons D, MacCulloch MJ, Snowden RJ. Prediction of violence and self-harm in mentally disordered offenders: a prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. J Consult Clin Psychol. 2003 Jun;71(3):443-51. PubMed PMID: 12795569.

145. Perry AE, Olason DT. A new psychometric instrument assessing vulnerability to risk of suicide and self-harm behaviour in offenders: Suicide Concerns for Offenders in Prison Environment (SCOPE). Int J Offender Ther Comp Criminol. 2009 Aug;53(4):385-400. PubMed PMID: 18519741.
146. Borschmann R, Hogg J, Phillips R, Moran P. Measuring self-harm in adults: A systematic review. Eur Psychiat. 2012;27(3):176-80.
147. Sansone RA, Sansone LA. Measuring self-harm behavior with the self-harm inventory. Psychiatry (Edgmont). 2010 Apr;7(4):16-20. PubMed PMID: 20508804. Pubmed Central PMCID: 2877617.
148. Carli V, Jovanovic N, Podlesek A, Roy A, Rihmer Z, Maggi S, et al. The role of impulsivity in self-mutilators, suicide ideators and suicide attempters - a study of 1265 male incarcerated individuals. J Affect Disord. 2010 Jun;123(1-3):116-22. PubMed PMID: 20207420.
149. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001 Sep;16(9):606-13. PubMed PMID: 11556941. Pubmed Central PMCID: 1495268.
150. Bohus M, Kleindienst N, Limberger MF, Stieglitz RD, Domsalla M, Chapman AL, et al. The short version of the Borderline Symptom List (BSL-23): development and initial data on psychometric properties. Psychopathology. 2009;42(1):32-9. PubMed PMID: 19023232.
151. Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, et al. CORE: Clinical Outcomes in Routine Evaluation. Journal of Mental Health. 2000;9(3):247-55.
152. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther. 1995 Mar;33(3):335-43. PubMed PMID: 7726811.
153. Lloyd E, Kelley M, Hope T, editors. Self-mutilation in a community sample of adolescents: Descriptive characteristics and provisional prevalence rates. annual meeting of the Society for Behavioral Medicine, New Orleans, LA; 1997.
154. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983 Jun;67(6):361-70. PubMed PMID: 6880820.
155. Shaw J, Tomenson B, Creed F. A screening questionnaire for the detection of serious mental illness in the criminal justice system. J Forensic Psychi Ps. 2003 Apr;14(1):138-50. PubMed PMID: WOS:000184719000009. English.
156. Teplin LA, Swartz J. Screening for Severe Mental Disorder in Jails - the Development of the Referral Decision Scale. Law and Human Behavior. 1989 Mar;13(1):1-18. PubMed PMID: WOS:A1989U032300001. English.
157. Latimer S, Covic T, Cumming SR, Tennant A. Psychometric analysis of the Self-Harm Inventory using Rasch modelling. Bmc Psychiatry. 2009 Aug 19;9. PubMed PMID: WOS:000269619200001. English.
158. Bohus M, Limberger MF, Frank U, Chapman AL, Kuhler T, Stieglitz RD. Psychometric properties of the Borderline Symptom List (BSL). Psychopathology. 2007;40(2):126-32. PubMed PMID: 17215599.

159. Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE—OM. The British Journal of Psychiatry. 2002;180(1):51-60.
160. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Medical care. 2003;41(11):1284-92.
161. Sansone RA, Butler M, Dakroub H, Pole M. Borderline personality symptomatology and employment disability: A survey among outpatients in an internal medicine clinic. Primary Care Companion to the Journal of Clinical Psychiatry. 2006;8(3):153-7.
162. Sansone RA, Reddington A, Sky K, Wiederman MW. Borderline personality symptomatology and history of domestic violence among women in an internal medicine setting. Violence and Victims. 2007;22(1):120-6.
163. Sansone RA, Songer DA, Sellbom M. The relationship between suicide attempts and low-lethal self-harm behavior among psychiatric inpatients. Journal of Psychiatric Practice. 2006;12(3):148-52.
164. Meltzer H, Jenkins R, Singleton N, Charlton J, Yar M. Non fatal suicide behaviour among prisoners. . London: Office for National Statistics, 1999.
165. Linacre JM. Sample Size and Item Calibration Stability. Rasch Measurement Transactions 1994;7:328.
166. Stevens SS. On the theory of scales of measurement. 1946.
167. Andrich D, Styles I. Final Report on the Psychometric Analysis of the Early Development Instrument EDI) Using the Rasch Model: A Technical Paper Commissioned for the Development of the Australian Early Development Instrument (AEDI): Citeseer; 2004.
168. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technology Assessment. 2009;13(12).
169. Nunnally JC. Psychometric Theory. London: McGraw-Hill; 1967.
170. Thurstone LL. The measurement of social attitudes. Journal of Abnormal and Social Psychology. 1931;26:249-69. PubMed PMID: WOS:000201075000021. English.
171. Spearman C. " General Intelligence," Objectively Determined and Measured. The American Journal of Psychology. 1904;15(2):201-92.
172. DeVellis RF. Classical test theory. Medical care. 2006;44(11):S50-S9.
173. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. Value in Health. 2015;18(1):25-34.
174. Hambleton RK, Swaminathan H. Item response theory: Principles and applications: Springer Science & Business Media; 2013.
175. Thurstone LL. Attitudes can be measured. American journal of Sociology. 1928;33(4):529-54.
176. Andrich D. Rating scales and Rasch measurement. Expert review of pharmacoeconomics & outcomes research. 2011 Oct;11(5):571-85. PubMed PMID: 21958102.

177. Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care. 2004 Jan;42(1 Suppl):I7-16. PubMed PMID: 14707751.
178. Novick MR. The axioms and principal results of classical test theory. Journal of Mathematical Psychology. 1966;3(1):1-18.
179. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institution for Educational Research; 1960.
180. Newby V, Conner G, Grant C, Bunderson C. The Rasch model and additive conjoint measurement. Journal of applied measurement. 2009;10(4):348.
181. Hattie J, Krakowski K, Rogers HJ, Swaminathan H. An assessment of Stout's index of essential unidimensionality. Applied Psychological Measurement. 1996 Mar;20(1):1-14. PubMed PMID: WOS:A1996VC52900001. English.
182. Horton M, Marais I, Christensen KB. Dimensionality. Rasch models in health.137-58.
183. De Vet HC, Adèr HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. Quality of Life Research. 2005;14(5):1203-18.
184. Muthén LK, Muthén BO. Mplus User's Guide: Seventh Edition. Los Angeles, CA: Muthén & Muthén; 1998-2012.
185. Kline RB. Principles and practice of structural equation modelling. 3 ed. New York; London: Guilford Press; 2011.
186. Mokken RJ. The theory and procedure of scale analysis with applications in political research. Mouton, New York, Berlin.: Walter de Gruyter; 1971.
187. Sijtsma K, Molenaar IW. Introduction to nonparametric item response modeling. Thousand Oaks, CA: Sage Publications; 2002.
188. Guttman L. The basis for scalogram analysis. Stouffer et al Measurement and prediction The American Soldier Vol IV. New York: Wiley; 1950.
189. Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. BMC Med Res Methodol. 2012;12:74. PubMed PMID: 22686586. Pubmed Central PMCID: 3464599.
190. van Shuur WH. Mokken scale analysis: Between the Guttman scale and parametric Item response theory. Political Analysis. 2003;11:139-63.
191. Crichton N. Information point Mokken Scale Analysis. Journal of clinical nursing. 1999;8(4):388-.
192. Hardouin J-B. MSP: Stata module to perform the Mokken Scale Procedure. Statistical Software Components. 2013.
193. StataCorp. Stata Statistical Software: Release 13. College Station, TX.: StataCorp LP; 2013.
194. Roskam EE, van den Wollenberg AL, Jansen PGW. The Mokken Scale: A Critical Discussion. Applied Psychological Measurement. 1986 Sep;10(3):265-77. PubMed PMID: WOS:A1986E881300005. English.
195. Luce RD, Tukey JW. Simultaneous Conjoint-Measurement - a New Type of Fundamental Measurement. Journal of Mathematical Psychology. 1964;1(1):1-27. PubMed PMID: WOS:A1964CHD5200001. English.

196. Fisher RA. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London. 1922;Series A(222):309-68.
197. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: An introduction and illustrative example. International Journal of Nursing Studies. 2009 Mar;46(3):380-93. PubMed PMID: WOS:000264776400010. English.
198. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Care and Research. 2007;57(8):1358-62.
199. Wright B, Panchapakesan N. A procedure for sample-free item analysis. Educational and Psychological measurement. 1969;29(1):23-48.
200. Andrich D. Rasch models for measurement. London: Sage Publications; 1988.
201. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). British Journal of Clinical Psychology. 2007 Mar;46(1):1-18. PubMed PMID: WOS:000244295300001. English.
202. Andrich D, Sheridan B, Luo G. RUMM 2030: Rasch Unidimensional Measurement Models (computer software). Perth, Western Australia.: RUMM Laboratory; 2010.
203. Andrich D. An Expanded Derivation of the Threshold Structure of the Polytomous Rasch Model That Dispels Any "Threshold Disorder Controversy". Educational and Psychological Measurement. 2013 Feb;73(1):78-124. PubMed PMID: WOS:000312304600005. English.
204. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical care. 2007;45(5):S22-S31.
205. Christensen KB, Makransky G, Horton M. Critical Values for Yen’s Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Applied Psychological Measurement. 2017;41(3):178-94.
206. Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. Archives of physical medicine and rehabilitation. 1989;70(4):308-12.
207. Smith Jr EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of applied measurement. 2002;3(2):205-31.
208. Kirshner B, Guyatt G. A methodological framework for assessing health indices. Journal of chronic diseases. 1985;38(1):27-36.
209. Andrich D. Cronbach’s alpha in the presence of subscales. International conference on outcome measurement; Bethesda, Maryland2010.
210. Bowers D, House A, Owens DH. Understanding clinical papers. Second ed: John Wiley & Sons; 2010.
211. Hennekens CH, Buring JE, Mayrent SL. Epidemiology in medicine: Boston: Little Brown and Company, 1987; 1987.
212. Metz CE, editor Basic principles of ROC analysis. Seminars in nuclear medicine; 1978: Elsevier.

213. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp; 2012.
214. Field A. Discovering statistics using SPSS. Second ed. London: Sage; 2005.
215. MacCallum RC, Austin JT. Applications of structural equation modeling in psychological research. Annual review of psychology. 2000;51(1):201-26.
216. Byrne BM. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Second ed. New York: Routledge; 2010.
217. Raykov T, Marcoulides GA. A first course in structural equation modeling. Second ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers; 2006.
218. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology. 1986;51(6):1173.
219. Wang PP, Badley EM, Gignac M. Exploring the role of contextual factors in disability models. Disability and rehabilitation. 2006;28(2):135-40.
220. Jöreskog KG. Testing structural equation models. In: Bollen KA, Long JS, editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 294-316.
221. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136-62.
222. Bentler PM. Comparative fit indexes in structural equation models. Psychological bulletin. 1990;107(2):238-46.
223. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological bulletin. 1980;88(3):588-606.
224. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal. 1999;6(1):1-55.
225. Yu C-Y. Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes. Los Angeles: University of California 2002.
226. Yu C-Y, Muthen B, editors. Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. annual meeting of the American Educational Research Association, New Orleans, LA; 2002.
227. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2007;16 Suppl 1(SUPPL. 1):19-31. PubMed PMID: 17479357.
228. Ginn S. Women prisoners. BMJ. 2013;346(7891):e8318. PubMed PMID: 23315254.
229. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27-38.
230. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Statistics in medicine. 2002;21(16):2409-19.

231. van de Stadt L, Witte B, Bos W, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Annals of the rheumatic diseases. 2013;72(12):1920-6.
232. Tsiachristas A, McDaid D, Casey D, Brand F, Leal J, Park A-L, et al. General hospital costs in England of medical and psychiatric care for patients who self-harm: a retrospective analysis. The Lancet Psychiatry. 2017;4(10):759-67.
233. The Howard League for Penal Reform. The cost of prison suicide. Research Briefing. London: 2016.
234. Joiner Jr TE, Van Orden KA. The interpersonal–psychological theory of suicidal behavior indicates specific and crucial psychotherapeutic targets. International Journal of Cognitive Therapy. 2008;1(1):80-9.
235. Latimer S, Meade T, Tennant A. Development of item bank to measure deliberate self-harm behaviours: Facilitating tailored scales and computer adaptive testing for specific research and clinical purposes. Psychiatry research. 2014;217(3):240-7.
236. Jones A. Self-mutilation in prison: A comparison of mutilators and nonmutilators. Crim Justice Behav. 1986;13(3):286-96.
237. Schmitt N. Uses and abuses of coefficient alpha. Psychological assessment. 1996;8(4):350-3.
238. Marais I. Local dependence. In: Christensen KB, Kreiner S, Mesbah M, editors. Rasch models in health. London: ISTE Ltd; 2013. p. 111-30.
239. Bedford A, Watson R, Lyne J, Tibbles J, Davies F, Deary IJ. Mokken scaling and principal components analyses of the CORE-OM in a large clinical sample. Clinical psychology & psychotherapy. 2010 Jan-Feb;17(1):51-62. PubMed PMID: 19728291.
240. Mavranezouli I, Brazier JE, Young TA, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). Quality of Life Research. 2011;20(3):321-33.
241. Horton M, Perry AE. Screening for depression in primary care: a Rasch analysis of the PHQ-9. BJPsych Bull. 2016;40(5):237-43.
242. Nicastro R, Prada P, Kung A-L, Salamin V, Dayer A, Aubry J-M, et al. Psychometric properties of the French borderline symptom list, short form (BSL-23). Borderline personality disorder and emotion dysregulation. 2016;3(4).
243. Birnbaum A. Some latent trait models and their uses in inferring an examinee's ability. In: Novick FMLMR, editor. Statistical theories of mental rest scores. Reading, MA: Addison-Wesley; 1968. p. 397–479.
244. Tuerlinckx F, De Boeck P. The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. Psychological methods. 2001;6(2):181-95.
245. Chan M, Bhatti H, Meader N, Stockton S, Evans J, O'Connor R, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. The British Journal of Psychiatry. 2016;209(4):277-83.
246. Quinlivan L, Cooper J, Steeg S, Davies L, Hawton K, Gunnell D, et al. Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. BMJ open. 2014;4(5):e004732.
247. Barton JJ, Meade T, Cumming S, Samuels A. Predictors of self-harm in male inmates. Journal of Criminal Psychology. 2014;4(1):2-18.

248. Völlm BA, Dolan MC. Self-harm among UK female prisoners: A cross-sectional study. Journal of Forensic Psychiatry and Psychology. 2009;20(5):741-51.
249. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015;349(6245):255-60.
250. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics. 2016;3(2):119-31.
251. Samuel AL. Some studies in machine learning using the game of checkers. IBM Journal of research and development. 1959;3(3):210-29.
252. Song X, Mitnitski A, Cox J, Rockwood K, editors. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Medinfo; 2004.
253. Morgan J, Hawton K. Self-reported suicidal behavior in juvenile offenders in custody: Prevalence and associated factors. Crisis. 2004;25(1):8-11. PubMed PMID: 15384651.
254. O'Leary WD. Custodial suicide: Evolving liability considerations. Psychiatric Quarterly. 1989;60(1):31-71.
255. Ward J, Bailey D. A participatory action research methodology in the management of self-harm in prison. Journal of mental health. 2013;22(4):306-16.
256. Roe-Sepowitz DE. Indicators of self-mutilation: Youth in custody. Florida: The Florida State University; 2006.
257. Fruehwald S, Frottier P, Matschnig T, Eher R. The relevance of suicidal behaviour in jail and prison suicides. European psychiatry : the journal of the Association of European Psychiatrists. 2003 Jun;18(4):161-5. PubMed PMID: 12814848.
258. Saunders KE, Hawton K, Fortune S, Farrell S. Attitudes and knowledge of clinical staff regarding people who self-harm: a systematic review. Journal of Affective Disorders. 2012;139(3):205-16.
259. Forrester A, Slade K. Preventing self-harm and suicide in prisoners: job half done. The Lancet. 2014;383(9923):1109-11.
260. Nock MK, Kessler RC. Prevalence of and risk factors for suicide attempts versus suicide gestures: analysis of the National Comorbidity Survey. J Abnorm Psychol. 2006 Aug;115(3):616-23. PubMed PMID: 16866602.
261. Horton M, Wright N, Dyer W, Wright-Hughes A, Farrin A, Mohammed Z, et al. Assessing the risk of self-harm in an adult offender population: an incidence cohort study. Health Technology Assessment (Winchester, England). 2014;18(64):1.
262. Andersen HS. Mental health in prison populations. A review–with special emphasis on a study of Danish prisoners on remand. Acta Psychiatrica Scandinavica. 2004;110(s424):5-59.
263. Dear GE, Thomson DM, Hills AM. Self-harm in prison - Manipulators can also be suicide attempters. Crim Justice Behav. 2000 Apr;27(2):160-75. PubMed PMID: WOS:000086102300002. English.
264. Andover MS, Schatten HT, Crossman DM, Donovick PJ. Neuropsychological Functioning in Prisoners with and without Self-Injurious Behaviors Implications for the Criminal Justice System. Crim Justice Behav. 2011 Nov;38(11):1103-14. PubMed PMID: WOS:000295082500003. English.

265. Bonner RL. Correctional suicide prevention in the year 2000 and beyond. Suicide and Life-Threatening Behavior. 2000;30(4):370-6.
266. Eidhin MN, Sheehy N, O'Sullivan M, McLeavey B. Perceptions of the environment, suicidal ideation and problem-solving deficits in an offender population. Legal and Criminological Psychology. 2002;7(2):187-201.

11 Appendices

11.1 APPENDIX A: Study Questionnaire

ACCT

Follow-up Study

Date: ☐☐ / ☐☐ /20 ☐☐

Start Time: ☐☐ : ☐☐

Finish Time: ☐☐ : ☐☐

(Please use 24hr clock format)

Date current ACCT opened: ☐☐ / ☐☐ /20 ☐☐

Pack Number: L0250

Prisoner ID:

Background Information Questionnaire

This questionnaire is to be completed by all participants following completion of the consent form and prior to commencing the selected questionnaires.

RESEARCHER—Please read the following statement out loud:

All information provided in this questionnaire will remain entirely anonymous and confidential. Information provided will be used for the purpose of this research only.

Prisoner ID: [ ]

1. Age: How old are you? [ ][ ] **years old**

2. Gender: Male [ ] **Female** [ ]

3. Ethnicity: Which of the following best describes your ethnic group?

RESEARCHER—Please display ethnicity options to respondent

White	
White British	[ ]
White Irish	[ ]
Other white background: Please write here	[ ]

Mixed	
White and Black Caribbean	[ ]
White and Black African	[ ]
White and Asian	[ ]
Mixed other: Please write here	[ ]

Asian or Asian British	
Indian	[ ]
Pakistani	[ ]
Bangladeshi	[ ]
Asian other: Please write here	[ ]

Black or Black British	
Caribbean	[ ]
African	[ ]
Black Other: Please write here	[ ]

Chinese or other Ethnic Group	
Chinese	[ ]
Other ethnic background: Please write here	[ ]

4. Religion: Do you practice a religion? **Yes** ☐ **No** ☐ → **Go to next question**

a. If Yes do you consider yourself to be

RESEARCHER—Please display religion options to respondent

Christian	☐	Hindu	☐	Muslim	☐	7th Day	☐		
Buddhist	☐	Jewish	☐	Sikh	☐	Any other religion	☐	Prefer not to say	☐

5. Education

a. How old were you when you finished your full time education? ☐☐ **Years old**

b. Do you have any of the following qualifications?

RESEARCHER—Please display qualification options to respondent

GCSE	☐	GCE 'O' Level	☐	'A' Level, Highers	☐
City and guilds	☐	Teaching Diploma, HNC	☐	Degree : Details	☐
None of these	☐	Other	☐		

c. Since you have been in prison have you been involved in any kind of education or training?

Yes	☐ →	Please write details here				
No	☐ →	If 'No' please say why?	Not offered any	☐	Not interested	☐
			Other:	☐		
			Details			

6. Family: Do you have any children under 16? **Yes** ☐ **No** ☐ → **Go to next question**

If Yes how many ☐

7. Visits and Correspondence

a. Since you have been in prison how often have friends/family visited you?

☐ Times a week ☐ Times a month ☐ Never ☐ Not been here long enough

b. Have you received a visit in the past 7 days? Yes ☐ No ☐

c. Since you have been in prison how often do you speak with friends/family by phone?

RESEARCHER—Please display correspondence frequency options to respondent

Daily ☐ 1-2 times a week ☐ 3-6 times a week ☐ 1-2 times a fortnight ☐

1-2 times a month ☐ Very rarely ☐ Never ☐ Not been in long enough ☐

d. Since you have been in prison how often do you write letters to your friends/family?

RESEARCHER—Please display correspondence frequency options to respondent

Daily ☐ 1-2 times a week ☐ 3-6 times a week ☐ 1-2 times a fortnight ☐

1-2 times a month ☐ Very rarely ☐ Never ☐ Not been in long enough ☐

e. Since you have been in prison how often do you receive letters from friends/family?

RESEARCHER—Please display correspondence frequency options to respondent

Daily ☐ 1-2 times a week ☐ 3-6 times a week ☐ 1-2 times a fortnight ☐

1-2 times a month ☐ Very rarely ☐ Never ☐ Not been in long enough ☐

8. Prison Experience: Are you in prison because you are:

RESEARCHER—Please display remand options to respondent

a. On remand awaiting trial ☐ } Go to next question

b. On remand awaiting sentencing ☐ }

c. Sentenced ☐

What is length of your current sentence? ☐☐ yrs ☐☐ mths ☐☐ days

☐ Life/Indeterminate

How long have you served? ☐☐ yrs ☐☐ mths ☐☐ days

8d. What offence(s) best describe what you are currently on remand/sentenced for?

RESEARCHER—Please display offence options to respondent

Violence against another person resulting in injury	☐	Drug related offences	☐
Violence against another person resulting in death	☐	Sexual offences	☐
Breach of licence (please detail prior index offence below)	☐	Burglary/theft offences	☐
Other offence (please detail below)	☐	Vehicle related offences	☐

9. Homelessness

a. Were you homeless at any point during the 12 months before you came to prison?

Yes ☐ No ☐

10. Healthcare

a. Have you accessed healthcare during this prison stay?		Yes ☐	No ☐
b. Have you accessed any listener services during this prison stay?		Yes ☐	No ☐
c. Have you ever seen a psychiatrist outside prison?		Yes ☐	No ☐
d. Have you ever received medication for any mental health problems? (answer yes if antidepressants or antipsychotics)		Yes ☐	No ☐
e. Would you consider yourself to be dependent on alcohol?		Yes ☐	No ☐
f. Would you consider yourself to be dependent on drugs?		Yes ☐	No ☐
g. Have you ever tried to harm yourself?	In prison	Yes ☐	No ☐
	outside prison	Yes ☐	No ☐
h. Is this the first time in this sentence you have been put on an ACCT?		Yes ☐	No ☐

→ If No: How long ago was previous ACCT? ____________________

→ How many have you had before this one? ____________________

Questionnaire 1

core
OUTCOME MEASURE

Site ID
letters only | numbers only

Client ID
Therapist ID | numbers only (1) | numbers only (2)

Sub codes
D D / M M / Y Y Y Y
Date form given

Age
Male ☐
Female ☐

Stage Completed
S Screening
R Referral
A Assessment
F First Therapy Session
P Pre-therapy (unspecified)
D During Therapy
L Last Therapy Session
X Follow up 1
Y Follow up 2

Stage
Episode

IMPORTANT - PLEASE READ THIS FIRST
This form has 34 statements about how you have been OVER THE LAST WEEK.
Please read each statement and think how often you felt that way last week.
Then tick the box which is closest to this.
Please use a dark pen (not pencil) and tick clearly within the boxes.

Over the last week

		Not at all	Only Occasionally	Sometimes	Often	Most or all the time	OFFICE USE ONLY
1	I have felt terribly alone and isolated	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	F
2	I have felt tense, anxious or nervous	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	P
3	I have felt I have someone to turn to for support when needed	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0	F
4	I have felt OK about myself	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0	W
5	I have felt totally lacking in energy and enthusiasm	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	P
6	I have been physically violent to others	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	R
7	I have felt able to cope when things go wrong	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0	F
8	I have been troubled by aches, pains or other physical problems	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	P
9	I have thought of hurting myself	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	R
10	Talking to people has felt too much for me	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	F
11	Tension and anxiety have prevented me doing important things	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	P
12	I have been happy with the things I have done	☐ 4	☐ 3	☐ 2	☐ 1	☐ 0	F
13	I have been disturbed by unwanted thoughts and feelings	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	P
14	I have felt like crying	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	W

Please turn over

Survey: 151

Page: 1

Over the last week

	Not at all	Only Occasionally	Sometimes	Often	Most or all the time	OFFICE USE ONLY
15 I have felt panic or terror	0	1	2	3	4	P
16 I made plans to end my life	0	1	2	3	4	R
17 I have felt overwhelmed by my problems	0	1	2	3	4	W
18 I have had difficulty getting to sleep or staying asleep	0	1	2	3	4	P
19 I have felt warmth or affection for someone	4	3	2	1	0	F
20 My problems have been impossible to put to one side	0	1	2	3	4	P
21 I have been able to do most things I needed to	4	3	2	1	0	F
22 I have threatened or intimidated another person	0	1	2	3	4	R
23 I have felt despairing or hopeless	0	1	2	3	4	P
24 I have thought it would be better if I were dead	0	1	2	3	4	R
25 I have felt criticised by other people	0	1	2	3	4	F
26 I have thought I have no friends	0	1	2	3	4	F
27 I have felt unhappy	0	1	2	3	4	P
28 Unwanted images or memories have been distressing me	0	1	2	3	4	P
29 I have been irritable when with other people	0	1	2	3	4	F
30 I have thought I am to blame for my problems and difficulties	0	1	2	3	4	P
31 I have felt optimistic about my future	4	3	2	1	0	W
32 I have achieved the things I wanted to	4	3	2	1	0	F
33 I have felt humiliated or shamed by other people	0	1	2	3	4	F
34 I have hurt myself physically or taken dangerous risks with my health	0	1	2	3	4	R

THANK YOU FOR YOUR TIME IN COMPLETING THIS QUESTIONNAIRE

Total Scores						
Mean Scores (Total score for each dimension divided by number of items completed in that dimension)						
	(W)	(P)	(F)	(R)	All items	All minus R

Questionnaire 2

RESEARCHER– Please turn the response pack to the **Questionnaire 2** set of responses, and then read the following statement out loud to the respondent:

Please listen to each statement and indicate the response that best describes how you have felt in the PAST YEAR:

1. Have you previously seen a psychiatrist?	**Yes** ○	**No** ○
2. Have you been taking longer over the things you do?	**Yes** ○	**No** ○
3. Have you recently been able to enjoy your normal everyday activities?	**Yes** ○	**No** ○
4. Have you recently felt that life isn't worth living?	**Yes** ○	**No** ○
5. Have you recently found yourself wishing you were dead and away from it all?	**Yes** ○	**No** ○
6. Have you recently felt that your thoughts have been directly interfered with, or controlled by another, in a way that people would find hard to believe?	**Yes** ○	**No** ○
7. Have there recently been times when you felt that people were plotting to cause you serious harm or injury?	**Yes** ○	**No** ○
8. Have you recently heard voices saying a few words or sentences when there was no one around to account for this?	**Yes** ○	**No** ○

Questionnaire 3

RESEARCHER – Please turn the response pack to the **Questionnaire 3** set of responses, and then read the following statement out loud to the respondent:

*In the following questionnaire I will read through a set of difficulties and problems which possibly describe you. Please listen to each statement and decide how much you suffered from each problem in the course **OF THE LAST WEEK**. In case you have no feelings at all at the present moment, please answer according to how you think you might have felt. Please answer honestly. <u>All questions refer to **THE LAST WEEK**. If you felt different ways at different times in the week, give a rating for how things were for you on average.</u>*

In the course of the last week…….

1. It was hard for me to concentrate

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

2. I felt helpless

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

3. I was absent-minded and unable to remember what I was actually doing

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

4. I felt disgust

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

5. I thought of hurting myself

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

In the course of the last week.......

6. I didn't trust other people

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

7. I didn't believe in my right to live

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

8. I was lonely

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

9. I experienced stressful inner tension

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

10. I had images that I was very much afraid of

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

11. I hated myself

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

12. I wanted to punish myself

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0 ◯	1 ◯	2 ◯	3 ◯	4 ◯

In the course of the last week.......

13. I suffered from shame

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

14. My mood rapidly cycled in terms of anxiety, anger and depression

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

15. I suffered from voices and noises from inside or outside my head

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

16. Criticism had a devastating effect on me

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

17. I felt vulnerable

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

18. The Idea of death had a certain fascination for me

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

19. Everything seemed senseless to me

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
○	○	○	○	○

In the course of the last week…….

20. I was afraid of losing control

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
◯	◯	◯	◯	◯

21. I felt disgusted by myself

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
◯	◯	◯	◯	◯

22. I felt as if I was far away from myself

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
◯	◯	◯	◯	◯

23. I felt worthless

Not at all	Only Occasionally	Sometimes	Often	Most or all the time
0	1	2	3	4
◯	◯	◯	◯	◯

RESEARCHER – Please turn the response pack to the **Questionnaire 3 – Supplement A** set of responses, and then read Question 24 out loud to the respondent:

24. Now we would like to know in addition, the quality of your *overall* personal state in the course of the last week. 0% means *absolutely down*, 100% means *excellent*. Please state the percentage which comes closest

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

very bad ← → excellent

BSL Supplement Items for Assessing Behaviour

RESEARCHER – Please turn the response pack to the **Questionnaire 3 – Supplement B** set of responses, and then read the following statement out loud to the respondent:

Also, DURING THE LAST WEEK, please select the most appropriate response to indicate how you would respond to the following statements:

During the last week………………..

S1. I hurt myself by cutting, burning, strangling, head banging etc

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4

S2. I told other people that I was going to kill myself

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4

S3. I tried to commit suicide

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4

S4. I had episodes of binge eating

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4

S5. I induced vomiting

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4

During the last week…….

S6. I took medication that had not been prescribed or if had been prescribed, I took more than the prescribed dose

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4
○	○	○	○	○

S7. I had outbreaks of uncontrolled anger or physically attacked others

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4
○	○	○	○	○

S8. I had uncontrollable sexual encounters of which I was later ashamed or which made me angry

Not at all	Once	2-3 Times	4-6 Times	Daily or More Often
0	1	2	3	4
○	○	○	○	○

Questionnaire 4

RESEARCHER – Please turn the response pack to the **Questionnaire 4** set of responses, and then read the following statement out loud to the respondent:

Please respond to the following questions by selecting Yes or No. Select Yes only to those things that you have done intentionally, or on purpose, to hurt yourself.

Have you ever intentionally, or on purpose, done any of the following:

1. Overdosed?

Yes ○ **No** ○ **If YES, number of times** ______

2. Cut yourself on purpose?

Yes ○ **No** ○ **If YES, number of times** ______

3. Burned yourself on purpose?

Yes ○ **No** ○ **If YES, number of times** ______

4. Hit yourself?

Yes ○ **No** ○ **If YES, number of times** ______

5. Banged your head on purpose?

Yes ○ **No** ○ **If YES, number of times** ______

6. Abused alcohol?

Yes ○ **No** ○ **If YES, number of times** ______

7. Driven recklessly on purpose?

Yes ○ **No** ○ **If YES, number of times** ______

8. Scratched yourself on purpose?

Yes ○ **No** ○ **If YES, number of times** ______

9. Prevented wounds from healing?

Yes ○ **No** ○

10.Made medical situations worse on purpose (e.g. skipped medication)?

Yes ○ No ○

11. Been promiscuous (i.e. had many sexual partners)?

Yes ○ No ○ If YES, how many

12. Set yourself up in a relationship to be rejected?

Yes ○ No ○

13. Abused prescription medication?

Yes ○ No ○

14. Distanced yourself from God as punishment?

Yes ○ No ○

15. Engaged in emotionally abusive relationships?

Yes ○ No ○ If YES, number of relationships

16. Engaged in sexually abusive relationships?

Yes ○ No ○ If YES, number of relationships

17. Lost a job on purpose?

Yes ○ No ○ If YES, number of times

18. Attempted suicide?

Yes ○ No ○ If YES, number of times

19. Exercised an injury on purpose?

Yes ○ No ○

20. Tortured yourself with self-defeating thoughts?

Yes ○ No ○

21. Starved yourself to hurt yourself?

Yes ○ No ○

22. Abused laxatives to hurt yourself?

Yes ○ No ○ If YES, number of times

23. Have you engaged in any other self-destructive behaviors not asked about in this inventory?

If so, please describe. (RESEARCHER– Please describe below)

Questionnaire 5

RESEARCHER – Please turn the response pack to the **Questionnaire 5** set of responses, and then read the following statement out loud to the respondent:

OVER THE LAST 2 WEEKS have you been bothered by any of the following problems? – Please select the most appropriate response:

1. Little interest or pleasure in doing things

Not at all	Several days	More than half the days	Nearly every day
0	1	2	3
○	○	○	○

2. Feeling down, depressed or hopeless

Not at all	Several days	More than half the days	Nearly every day
0	1	2	3
○	○	○	○

3. Trouble falling or staying asleep, or sleeping too much

Not at all	Several days	More than half the days	Nearly every day
0	1	2	3
○	○	○	○

4. Feeling tired or having little energy

Not at all	Several days	More than half the days	Nearly every day
0	1	2	3
○	○	○	○

5. Poor appetite or overeating

Not at all	Several days	More than half the days	Nearly every day
0	1	2	3
○	○	○	○

6. Feeling bad about yourself - or that you are a failure or have let yourself or your family down

Not at all	Several days	More than half the days	Nearly every day
0 ○	1 ○	2 ○	3 ○

7. Trouble concentrating on things, such as reading the newspaper or watching television

Not at all	Several days	More than half the days	Nearly every day
0 ○	1 ○	2 ○	3 ○

8. Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual

Not at all	Several days	More than half the days	Nearly every day
0 ○	1 ○	2 ○	3 ○

9. Thoughts that you would be better off dead, or of hurting yourself in some way

Not at all	Several days	More than half the days	Nearly every day
0 ○	1 ○	2 ○	3 ○

11.2 APPENDIX B: Study Information Sheet

Leeds NHS Primary Care Trust | northumbria UNIVERSITY | PORSCH NE | UNIVERSITY OF LEEDS

SCREENING FOR THE RISK OF SELF HARM WITHIN THE PRISON SETTING

PARTICIPANT INFORMATION SHEET

You are being invited to take part in a research study which is looking at potential risk for self harm among prisoners. We are interested in measuring your level of risk at the present time, following your recent ACCT. Before you agree to take part, it is important that you understand why the study is being done and what it will involve. Please read the following information carefully and feel free to ask any questions.

Who is conducting this study?
The study is being conducted by researchers from Leeds University, Northumbria University and NHS Leeds. It is funded by the National Institute of Heath Research under its Health Technology Assessment Programme.

What is the purpose of the study?
The actual or perceived risk of self harm, such as cutting or burning, would lead to an ACCT referral. At the present time the ACCT does not incorporate a standardised diagnostic test for the risk of further self harm. Several such tests are available, and these take the form of a simple questionnaire. However, we do not know which if any of these questionnaires will work properly in a prison setting. The purpose of the study is to find out if one or more of these questionnaires can be used in this setting. The study will begin in November 2010 and end in October 2012.

Why have you been chosen?
You have been chosen because of your recent referral for an ACCT, and all such prisoners will be approached to help during an eight month recruitment period.

Do you have take part?
No. Taking part is entirely voluntary and you are also free to withdraw at any point. If you decide not to take part, this will not affect the help you receive from the prison medical service in any way. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. You are free to withdraw at any point.

What does the study involve?
We would ask you to complete some short questionnaires. These would be about the type of self harm activities that people engage with, and your own feelings at the time. **This whole process should take less than 40 minutes of your time** and you can refuse to answer any questions or participate in any measures that you do not feel happy about.

Once this stage is complete, that will conclude your involvement. Some nine months after completion of the questionnaires, we would check the ACCT records to see what, if any, self harm activities you have undertaken during that period.

Is the research confidential?
Your current GP and the Safer Custody/ACCT Team will be informed that you are taking part in the study; otherwise, your involvement in the study would be completely confidential. If you disclose details of a very serious crime, such as one involving children, or if you tell us that you are feeling suicidal, we would have to report this to an appropriate agency. We would also have to disclose it if you were threaten harm to self, someone else, or to prison security. However, we will not ask about particular crimes and you should not answer any questions you feel unhappy about. Otherwise, everything you say, and what you have reported on the questionnaires will be made anonymous, and your views will be grouped together with those of others so that your identity is hidden. Any information collected from you will only be seen by members of the research team.
The information collected **will not** be held in your prison record or have an impact on your sentence.

What are the benefits of taking part?
There might not be any direct benefit to you from taking part in this study. However, your participation in this research will enable us to identify possible screening questionnaires which will help us to improve our health care service response to these circumstances.

What will happen to the results of the study?
The findings will be reported to a range of people. This will include prisoners, staff working with prisoners, government officials and other researchers. No individual will be identified in any reports or publications arising from this research.
No individual scores or reports will be available from the study, but the final report will be available from the NIHR HTA web site (www.hta.ac.uk) by early 2013.

What will happen if you agree to be involved?
After you have signed a copy of our consent form, you will be given details of where and when the interview will take place.

If you require further details about the research, you can contact:

XXXX through your Healthcare Department.

Thank you for reading this!

v.1.3. 30/03/2011

11.3 APPENDIX C: Study Consent Form

Leeds NHS
Primary Care Trust

northumbria UNIVERSITY

PORSCH NE

UNIVERSITY OF LEEDS

CONSENT FORM

Title of Project: SCREENING FOR THE RISK OF SELF HARM WITHIN THE PRISON SETTING

Name of Researcher: XXXXXXXXX

Contact: (Details Provided)

<u>Please initial boxes</u>

1. I confirm that I have read and understood the information sheet dated 30/03/2011 (version 1.3) for the above study and that I have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time - without giving any reason – and without my medical care or legal rights being affected. ☐

3. I agree that any data collected about me as part of the study may be used, anonymously, in the presentation of the research. ☐

4. I agree that the prison GP and Safer Custody/ACCT Team can be informed of my participation in the research. ☐

5. I understand that if I threaten harm to self, someone else, or to prison security, then someone will be informed of this information. ☐

6. I understand that the information collected **will not** be held in my prison record or have an impact on my sentence. ☐

7. I agree to take part in the above study. ☐

Name in block capitals………………………………………………………………...................

Signed………………………………………………………… Date……………………….

Witnessed (researcher)………………………………… Date……………………...

1 copy to be kept by the participant; 1 copy to be kept by the researcher

PRISONER ID: _____________ **PROJECT ID (From Questionnaire Pack):** _________

STRICTLY CONFIDENTIAL v.1.3. 30/03/2011

11.4 APPENDIX D: Additional Study Participant Characteristics

Variable	Detail	Full Population (n=450)	Evaluable Population (n=433)
Age (years)	Mean (standard deviation)	31.2 (9.89)	31.2 (9.96)
	Median	29	29
	IQR	24-36	24-36
	Range	16-80	16-80
	<30 years	233 (51.8%)	222 (51.3%)
	>=30 years	217 (48.2%)	211 (48.7%)
Prison	A	105 (23.3%)	102 (23.6%)
	B	115 (25.6%)	111 (25.6%)
	C	230 (51.1%)	220 (50.8%)
Gender	Male	335 (74.4%)	322 (74.4%)
	Female	115 (25.6%)	111 (25.6%)
Ethnicity	White (British / Irish / other)	407 (90.4%)	391 (90.3%)
	Other ethnic background	39 (8.7%)	38 (8.8%)
	Missing	4 (0.9%)	4 (0.9%)
Religion	No	260 (57.8%)	254 (58.7%)
	Yes	190 (42.2%)	179 (41.3%)
Children under 16	No	227 (50.4%)	219 (50.6%)
	Yes	222 (49.3%)	213 (49.2%)
	Missing	1 (0.2%)	1 (0.2%)
Age when finished full time education (years)	Number of prisoners	440	424
	Number of patients with missing data	10	9
	Mean (standard deviation)	15.3 (3.49)	15.4 (3.45)
	Median	15	15
	IQR	14-16	14-16
	Range	0-45	0-45
	<16 years	242 (53.8%)	232 (53.6%)
	>=16 years	208 (46.2%)	201 (46.4%)
	Total	450 (100.0%)	433 (100.0%)
Education or training received in prison	No	204 (45.3%)	200 (46.2%)
	Yes	245 (54.4%)	232 (53.6%)
	Missing	1 (0.2%)	1 (0.2%)
Received a visit in the past 7 days	No	382 (84.9%)	368 (85.0%)
	Yes	64 (14.2%)	61 (14.1%)
	Missing	4 (0.9%)	4 (0.9%)
Sentenced	No	203 (45.1%)	198 (45.7%)
	Yes	245 (54.4%)	233 (53.8%)
	Missing	2 (0.4%)	2 (0.5%)
Homeless at any point in the 12 months before coming to prison	No	289 (64.2%)	278 (64.2%)
	Yes	159 (35.3%)	153 (35.3%)
	Missing	2 (0.4%)	2 (0.5%)
Seen a psychiatrist outside prison	No	188 (41.8%)	183 (42.3%)
	Yes	259 (57.6%)	247 (57.0%)
	Missing	3 (0.7%)	3 (0.7%)
Received medication for mental health problems	No	115 (25.6%)	111 (25.6%)
	Yes	334 (74.2%)	321 (74.1%)
	Missing	1 (0.2%)	1 (0.2%)

Variable	Detail	Full Population (n=450)	Evaluable Population (n=433)
Ever self-harmed in prison	No Yes Missing	172 (38.2%) 277 (61.6%) 1 (0.2%)	167 (38.6%) 265 (61.2%) 1 (0.2%)
Ever self-harmed outside prison	No Yes Missing	99 (22.0%) 350 (77.8%) 1 (0.2%)	96 (22.2%) 336 (77.6%) 1 (0.2%)
First ACCT	No Yes Missing	80 (17.8%) 367 (81.6%) 3 (0.7%)	77 (17.8%) 353 (81.5%) 3 (0.7%)
Accessed listener services in prison	No Yes Missing	316 (70.2%) 133 (29.6%) 1 (0.2%)	306 (70.7%) 126 (29.1%) 1 (0.2%)
Dependent on alcohol	No Yes Missing	302 (67.1%) 145 (32.2%) 3 (0.7%)	289 (66.7%) 141 (32.6%) 3 (0.7%)
Dependent on drugs	No Yes Missing	301 (66.9%) 148 (32.9%) 1 (0.2%)	290 (67.0%) 142 (32.8%) 1 (0.2%)
Length of sentence remaining	On remand / < 1 year >=1year	312 (69.3%) 138 (30.7%)	303 (70.0%) 130 (30.0%)
Violent or sexual related offence committed	Violent/sexual offence Other crime	186 (41.3%) 264 (58.7%)	173 (40.0%) 260 (60.0%)
Violent or sexual or drug or theft related offence committed	Violent/sexual/Drug/Burglary offence Other crime	310 (68.9%) 140 (31.1%)	296 (68.4%) 137 (31.6%)
Index ACCT due to self-harm	No Yes Not Known	158 (35.1%) 154 (34.2%) 138 (30.7%)	157 (36.3%) 151 (34.9%) 125 (28.9%)
Days between index ACCT and baseline interview	Number of prisoners Mean (standard deviation) Median IQR Range Total	450 6.2 (4.27) 6 3-8 0-30 450 (100.0%)	433 6.2 (4.22) 6 3-8 0-30 433 (100.0%)
Prison status at follow up	Still in original prison Released Transferred but still in prison Transferred and subsequently released Back in original prison after multiple transfers Back in prison system after release and re-arrest Not known	120 (26.7%) 191 (42.4%) 98 (21.8%) 16 (3.6%) 4 (0.9%) 19 (4.2%) 2 (0.4%)	118 (27.3%) 189 (43.6%) 86 (19.9%) 16 (3.7%) 4 (0.9%) 19 (4.4%) 1 (0.2%)

Variable	Detail	Full Population (n=450)	Evaluable Population (n=433)
Length of follow up by prison status at follow up	Released with less than 6 months follow up	177 (39.3%)	175 (40.4%)
	Released with at least 6 months follow up	49 (10.9%)	49 (11.3%)
	Still in prison with less than 6 months follow up	45 (10.0%)	41 (9.5%)
	Still in prison with at least 6 months follow up	177 (39.3%)	167 (38.6%)
	At least 6 months follow up, prison status not known	1 (0.2%)	1 (0.2%)
	Lost to follow up	1 (0.2%)	
Length of follow up (Months)	Number of prisoners	449	433
	Number of prisoners with missing data	1	0
	Mean (standard deviation)	5.1 (3.16)	5.1 (3.14)
	Median	5.5	5.5
	IQR	2.5-6.9	2.4-6.8
	Range	0-16.4	0-16.4
	Total	450 (100.0%)	433 (100.0%)
Self-harm in follow-up	No	307 (68.2%)	307 (70.9%)
	Yes	126 (28.0%)	126 (29.1%)
	Not Known	17 (3.8%)	
Number of Self-harm events in follow-up	Number of prisoners	433	433
	Number of prisoners with missing data	17	0
	Mean (standard deviation)	1.0 (2.82)	1.0 (2.82)
	Median	0	0
	IQR	0-1	0-1
	Range	0-26	0-26
	Total	450 (100.0%)	433 (100.0%)
Previous SH in and out of prison	No previous SH	54 (12%)	
	Previous SH OUTSIDE prison only	118 (26.2%)	
	Previous SH INSIDE prison only	45 (10%)	
	Previous SH INSIDE & OUTSIDE prison	232 (51.6%)	
	Missing	1 (0.2%)	
Previous SH yes/no	No previous SH	54 (12%)	
	Previous SH (anywhere)	395 (87.8%)	
	Missing	1 (0.2%)	