Optimizing risk predictive strategies in febrile neutropenic episodes in children and young people undergoing treatment for malignant disease

Dr Robert Stephen Phillips

PhD Thesis

University of York

Health Sciences

January 2014

Abstract

This thesis aimed to investigate the clinical problem of the initial management of febrile neutropenia (FN) in children and young people undergoing treatment for malignant disease, to thoroughly evaluate the existing research, and to collect and synthesise this to quantify the risk of adverse clinical outcomes, through development of develop a new risk prediction model, using individual participant data (IPD). A further aim was to develop methodological approaches to IPD analysis in the development of predictive models, including the graphical display and communication of such information.

The research helped create a global collaboration of 19 research groups (PICNICC) which has shared data on over 5000 episodes of FN. This individual patient data was synthesised using hierarchical logistic regression meta-analysis to develop a new predictive model for MDI, which is robust to internal validation techniques (bootstrapping and leave-one-out cross-validation). The multivariable predictive model derived has six components: Tumour type, temperature, clinical description of being “severely unwell”, and measurements of three elements of the full blood count: haemoglobin concentration, total white cell count and absolute monocyte count. It showed good overall fit (Brier[scaled] 4.5% discordancy), moderate discrimination (AU-ROC 0.736) and good calibration between predicted and actual estimates of the risk of MDI (calibration slope 0.95). A basic implementation of the predictive model has been made ‘live’ at: http://tinyurl.com/PICNICC1

The content of this thesis has directly generated five systematic reviews published in academic journals [1-5], along with a further six peer reviewed papers [6-11]. Further papers are in preparation. This has influenced national [12] and international guidelines [13] on the management of FN in children and young people. We have demonstrated that such a data sharing project is feasible across many different jurisdictions and eras of study; we now need to undertake a series of further projects to evaluate the model and improve the management of paediatric FN worldwide.

Contents

List of Tables

List of Figures

Acknowledgements

Undertaking this thesis would not have been possible without the assistance of my advisory panel, the members of the PICNICC collaboration, and the patients and their parents who have engaged in the original research.

I would also like to thank Suzanne, Lily, Alex and Ted who have received less attention than they should have done in order that this thesis be undertaken. For their patience and continued love I remain hugely grateful.

Declaration

I declare that this is an original work and I am the sole author, with specific technical assistance received with the following elements:

- creating and undertaking the electronic database searches for the original systematic reviews
- double-screening potential references
- checking data extractions
- copy-editing the final work

All other elements of this thesis were completed by myself, including the systematic reviews, Collaborative formation, data collection, cleaning, planning and undertaking the statistical analyses, and preparation of graphical outputs.

I also declare that this work has not previously been submitted at this, or any other University, for any other award.

1. Phillips B, Wade R, Stewart LA, Sutton AJ: Systematic review and meta-analysis of the discriminatory performance of risk prediction rules in febrile neutropaenic episodes in children and young people. *Eur J Cancer* 2010, 46:2950-2964.

2. Phillips R, Wade R, Riley R, Sutton A, Stewart LA: Systematic review and meta-analysis of the value of clinical features to exclude radiographic pneumonia in febrile neutropenic episodes in children and young people. *Journal of Paediatrics and Child Health* 2011;48(8):641-8

3. Phillips RS, Lehrnbecher T, Alexander S, Sung L: Updated systematic review and meta-analysis of the performance of risk prediction rules in children and young people with febrile neutropenia. *PLoS One* 2012, 7:e38300.

4. Phillips RS, Wade R, Lehrnbecher T, Stewart LA, Sutton AJ: Systematic review and meta-analysis of the value of initial biomarkers in predicting adverse outcome in febrile neutropenic episodes in children and young people with cancer. *BMC Med* 2012, 10:6.

5. Haeusler GM, Carlesse F, Phillips RS: An Updated Systematic Review and Meta-Analysis of the Predictive Value of Serum Biomarkers in the Assessment of Fever during Neutropenia in Children with Cancer. *Pediatr Infect Dis J* 2013. 32(10):e390-6

6. Phillips B, Skinner R, M. Lane S, C. Chisholm J: Challenges in the Management of Paediatric Febrile Neutropenia. *Current Pediatric Reviews* 2009, 5:229-233.

7. Phillips B: Clinical decision rules: how to build them. *Archives of disease in childhood - Education & practice edition* 2010, 95:83-87.

8. Phillips B: Clinical decision rules: how to use them. *Archives of disease in childhood - Education & practice edition* 2010, 95:88-92.

9. Phillips B, Stewart LA, Sutton AJ: 'Cross hairs' plots for diagnostic meta-analysis. *Research Synthesis Methods* 2010, **1:**308-315.

10. Phillips R, Sutton A, Riley R, Chisholm J, Picton S, Stewart L : Predicting infectious complications in neutropenic children and young people with cancer (IPD Protocol). *Systematic Reviews* 2012, 1:8.

11. Phillips B, Ranasinghe N, Stewart LA: Ethical and regulatory considerations in the use of individual participant data for studies of disease prediction. *Archives of Disease in Childhood* 2013: 98(7):567-8

12. Phillips R, Hancock B, Graham J, Bromham N, Jin H, Berendse S: Prevention and management of neutropenic sepsis in patients with cancer: summary of NICE guidance. *BMJ* 2012, 345:e5368.

13. Lehrnbecher T, Phillips R, Alexander S, Alvaro F, Carlesse F, Fisher B, Hakim H, Santolaya M, Castagnola E, Davis BL, Dupuis LL, Gibson F, Groll AH, Gaur A, Gupta A, Kebudi R, Petrilli S, Steinbach WJ, Villarroel M, Zaoutis T, Sung L: Guideline for the management of fever and neutropenia in children with cancer and/or undergoing hematopoietic stem-cell transplantation. *J Clin Oncol* 2012, 30**:**4427-4438.

Chapter 1: Introduction to the problem of fever in the immunocompromised host

The treatment of childhood cancer is one of the great success stories of healthcare in the late 20th century.[1] In Europe, children with a malignant disease have an excellent chance of survival, with overall survival rates approaching 75%.[2] Of those who do not survive, the cause of death is most often directly related to the malignancy, but one in twenty-five children with cancer will die due to complications of therapy: one in six of all deaths.[3-4] One important cause of death is infection, frequently presenting as the occurrence of fever with neutropenia.[5-6]

Clinical background

In the mid-1960s, it was noted that adults with a severe reduction in numbers of neutrophils (neutropenia) following chemotherapy were at very high risk of serious infection, and that early aggressive treatment with broad spectrum antibiotics could save lives.[7-8] The clinical phenomenon of neutropenia with fever is known by a variety of synonymous phrases, but frequently as 'febrile neutropenia' or 'neutropenic sepsis'.

The pathophysiology of infection in a child following the administration of chemotherapy or radiotherapy is complex. Deficiencies occur in innate and adaptive immunity, with changes in cellular and non-cellular elements of the defences against infection. There are also marked differences between individuals [9-10] in their response to infection.

Anatomical defences are compromised by the effects of chemotherapy and radiotherapy disrupting the mucous membranes of the gut and integrity of the skin; side effects are experienced as mucositis and dermatitis. This enables colonisation and invasion of bacteria into the blood stream or local infection of the tissue. Foreign bodies such as central venous catheters are frequently inserted, which provide a potential site for bacterial colonisation. The use of cytotoxic agents alters the host intestinal flora, and chemical barriers (such as gastric acidity) can be reduced by supportive care medications.

The production of inflammatory and antimicrobial proteins can be reduced by anti-cancer treatments [9], dampening any coordinated antibacterial response of the complement cascade and inflammatory pathway. Cellular components of the innate phagocytic system

are reduced in number and may be impaired in function. These include the identifiable levels of circulating neutrophils in peripheral blood and the less visible tissue macrophages and natural killer cells.

Chemotherapy also affects adaptive, acquired B- and T-cell mediated immunity. The proliferation rate of these cells is reduced by chemotherapy, depleting circulating cells and humoral antibody levels. Some malignancies (leukaemia and lymphoma) are of white cell origin, and while the absolute numbers of cells may be high, they are functionally incompetent and this increases the risk of infection.[11]

Clinically based risk stratification

The identification of recipients of chemotherapy at high risk of infection, and definition of the entity of febrile neutropenia, led to the rapid use of early assessment and aggressive treatment with broad spectrum intravenous antibiotics until the neutrophil count had recovered. Though no randomised trials evaluated this, the dramatic fall in mortality rates (from 30% to around 1% in Western Europe) was convincing.[5] The next logical step was to explore whether the duration of antibiotics could be safely shortened, and antibiotics discontinued despite neutropenia. This was shown to be possible.[12] From here, the next milestone in refining therapy is to define a subset of patients at low risk of infection, to reduce the intensity of treatment in that group, and facilitate rational intensification of anti-infective treatment in the high-risk group.

In adult oncology practice, a large international prospective study of 1,139 patients was undertaken and produced a scoring system to identify patients at low risk of serious medical complications during febrile neutropenia.[13] The factors included: an outpatient presentation; a solid (compared with haematological) malignancy; “young age” (defined in this setting as <60 years); no chronic obstructive airways disease; mild or absent symptoms of infection; normal blood pressure; and absence of dehydration. This system supported earlier work which identified many of the factors as important predictors and has been used as the basis for outpatient management of fever in low-risk neutropenic adult patients.[14] Despite this, there is evidence of minimal uptake of this approach in routine adult practice in the UK.[15]

The adult system did not include children in its derivation, and is of very limited applicability in this group: age is not a discriminator, and chronic airways disease is extremely rare. It is frequently said in paediatric medicine that “Children are not little adults”; when it relates to the use of this scoring system, it appears to be very relevant. [16]

Assessing the risk of adverse outcome of each episode of febrile neutropenia has been undertaken by different groups, with many creating a clinical decision rule (CDR) which purports to allow clinicians to accurately judge risk and treat appropriately. A clinical decision rule is a clinical tool designed to be used at the bedside to assist clinical decision making [17]; the methods used in creating and assessing such rules are explored further later in this chapter.

The existing CDRs in paediatric oncology patients have proposed varied criteria for risk stratification, involving various combinations of bone marrow suppression indicators [18-22], maximum body temperature[18-19], cancer type[23] and the presence of clinically severe illness.[11, 19, 21, 24-25] However, the currently published models differ in describing different numbers of important predictor variables (e.g. Baorto [12] identified one, Klaassen [26] described two, Ammann [19] used five), and different specific variables (bone marrow suppression may be indicated by monocyte count [12], leukocyte count [19] or platelet count [22]).

Biomarker based approaches

Another element of initial risk stratification of significant interest is the use of serum markers of inflammation and infection derived from blood tests taken on admission to predict similar outcomes.

Studies have been undertaken which explore the predictive ability of specific serum markers, for example C-reactive protein (CRP), procalcitonin (PCT), Interleukin-6 (IL6) or Interleukin-8 (IL8).[27-30] These markers have been demonstrated to have some discriminatory ability in other fields of paediatrics (e.g. CRP has been used in septic arthritis, PCT in neonatal sepsis and IL6 in meningitis). In paediatric febrile neutropenia (FNP), the reports have had few patients, episodes and definite outcomes. Drawing these reports together, and synthesising their results, could add greatly to our understanding of their potential clinical usefulness, indicating which markers may be pertinent to examine in newly developed clinical decision rules.

C-reactive protein (CRP) is a substance that was initially discovered in the 1930s by Tillett and Francis [31], its name arising from the observation of a reaction between an unidentified protein and the C-polysaccharide from *Streptococcus pneumoniae*. It was soon discovered to be endogenously generated from within the liver, and its ubiquity in inflammatory conditions recognised soon after. CRP is generated with other acute-phase reactants between four and six hours following tissue damage, and seems to reach a peak around one to three days following injury.[32] Routine measurement of CRP within one hour of admission is possible in most technologically advanced countries, and in some places near-patient testing (for example, in a primary care centre) is possible.

The more recently discovered procalcitonin (PCT) is a pro-hormone of calcitonin, a hormone associated with calcium homeostasis. Pathophysiological studies have shown PCT has a rapid rise in new-onset sepsis, and falls rapidly with the administration of antibiotics and mirrors the clinical course of critically ill patients. Commercial assays are available in many hospitals, and can produce results within two hours.

Interleukin 6 (IL6) is one of a series of cytokines which are released by immune system cells and drive the active and organised process of inflammation. IL6 is released from macrophages and monocyte-lineage cells, and can be measured in the blood rapidly after tissue damage. Interleukin 8 (IL8) is a related protein, first identified in the early 1980s as a neutrophil chemoattractant. It too is produced by a wide range of inflammatory cells, and is rapidly produced following tissue damage.

Although all these markers have a sensible pathophysiological basis for their ability to distinguish between people who do and do not have an incipient severe infection, many other factors are involved with their production. Some factors may reduce the production of these markers. In malignancy, the generative potential of the liver may be compromised, the immune system is suppressed and functions inadequately, potentially reducing the number of cells which could produce cytokines. In contrast, the tissue damage of malignant infiltration may produce an inflammatory response, triggering release of the marker proteins. The toxic effects of chemotherapy may reduce the ability of the kidneys and liver to clear metabolites including these markers. Therefore, just detecting the presence of an inflammatory response, for example to a mild rhinovirus (common cold) infection, does not equate with severe septic shock. The effect of how these elements alter with age, and how the range of different malignancies that affect children as compared with adults vary also need to be taken into account.

Reduced intensity treatment

Based on the identification of a low-risk group, small randomized controlled trials (RCTs) [33-37] of the use of reduced intensity and/or duration of antibiotic therapy have been undertaken in children. The intention is to improve quality of life by reducing hospitalisation, and reducing unnecessary health care costs. These trials have been too small to produce definitive conclusions about the efficacy or safety of the approaches undertaken. The underlying assumption (that parents would prefer an outpatient or home-based approach) has also been called into question.[38]

Preventing febrile neutropenia

Reducing unnecessary hospitalisation, exposure to antibiotics and costs associated with febrile neutropenia has been one approach to improving the management of children who present with infectious complications during cancer therapy. Another strand of research has been attempting to reduce the risk of febrile neutropenia occurring.

An ideal approach to reducing the risk of adverse outcomes of febrile neutropenia would be to prevent neutropenia secondary to chemotherapy and/or to prevent infection in neutropenic patients. Colony-stimulating factors (CSFs) were introduced into clinical trials 20 years ago and are now used widely in both adults and children. They expand the pool of circulating neutrophils by stimulating proliferation and hastening maturation of myeloid progenitor cells in bone marrow, and are used successfully in the treatment of chronic and cyclical neutropenia.[39] Clinical experience suggests that the prescription of granulocyte-CSF (G-CSF) in the oncology setting does not have the same dramatic benefits. An extensive systematic review of 148 RCTs (19 were exclusively in paediatric populations and 12 included both adults and children) with a total of 16,839 cycles of treatment assessed the effects of G-CSF.[40] The synthesis demonstrated no effect of G-CSF on mortality (relative risk 0.95, 95% CI 0.84 to 1.08). There was a small reduction in the number of episodes of febrile neutropenia (relative risk 0.71 95% CI 0.63 to 0.80) and a small effect of duration of hospitalisation (mean difference -2.4 days, 95% CI -3.3 to -1.1) and use of parenteral antibiotics (mean difference -1.8 days, 95% CI -2.5 to -1.1). Subgroup and meta-regression analyses, and the use of Baysian approaches incorporating further data [41] showed these results to be consistent regardless of the type of malignancy or age of patients. If CSFs were oral, cheap and had no side effects then these moderate benefits might be considered

useful: unfortunately, none of these attributes are true for CSFs and their use is not routinely recommended.

An alternative approach to prevention would be to use antibiotic prophylaxis. While this has been losing favour in a number of areas (e.g. urinary tract infection [42], recurrent tonsillitis [43], dental procedures in those with cardiac defects [44]) there is convincing evidence for their use in preventing surgical site infections. [45] A systematic review published in 2005 [46] examined 109 RCTs that evaluated the use of prophylactic antibacterial drugs in cancer patients. This showed that across a broad spectrum of ages, malignancies and methods of administration, prophylactic antibacterials given through periods of neutropenia conferred a survival advantage. The use of quinolone antibiotics reduced overall mortality with a relative risk 0.66 (95% CI 0.55 to 0.81).[46] However, direct clinical application of these data is hindered by substantial variation in the protocols for administration of prophylaxis, making selection of a particular approach difficult.

There is a significant concern that the widespread use of antibiotics may engender resistance, but data from these trials do not suggest the emergence of this problem.[47-48] Cohort studies of the routine use of prophylaxis do suggest an increase in rates of resistance of colonising organisms, and that these rates fall with discontinuing prophylaxis. However, these same data show higher mortality rates in patients when prophylactic antibiotics are not used, despite that fact that there are lower rates of 'resistant' organisms cultured. The balance of community resistance against individual protection is clearly difficult, but seems to favour the use of prophylaxis. These factors probably account for the caution about current widespread use of prophylactic antibiotics: the ongoing challenge is to support their judicious introduction, perhaps in selected patient groups, with close microbiological surveillance.

Vaccination is the most effective anti-infective prophylaxis that is used in the world today [49] and high levels of herd immunity against vaccine-preventable diseases are the best protection for children with cancer. In the setting of acute treatment for malignancy, there is limited evidence for the use of vaccination since most of the serious infections that occur are not vaccine-preventable. Influenza vaccination while on low-intensity therapy (e.g. maintenance treatment for acute lymphoblastic leukaemia)[50] is probably effective, and conjugate Haemophilus influenzae type B (Hib) and pneumococcal vaccines may reduce the risk of invasive infection.[51] Live vaccines remain potentially lethal if given during immunosuppressive therapy, but the use of attenuated varicella vaccine has been

extensively studied during Acute Lymphoblastic Leukemia (ALL) maintenance therapy by groups in Japan and the USA with few adverse events and apparent protective responses. [52]

Environmental prevention, for example the use of face masks and gowns, 'clean' diets and water supplies, and excluding patients from large public gatherings (such as football matches, shopping centres and cinemas) had very limited evidence for efficacy.

Conclusion

Despite the generally good success in treating children with cancer, with overall survival rates approaching 75% [2], one in twenty-five children with cancer will die due to the complications of therapy: this is one in six of all deaths.[3-4] One important cause of death remains infection, frequently presenting as the occurrence of fever with neutropenia.[5-6] The traditional approach to such patients is to admit them to hospital and treat with prolonged courses of intravenous antibiotics until both fever and neutropenia have resolved.

Current practice in paediatric oncology with respect to the risk stratified approaches in febrile neutropenia is variable, both nationally [53] and internationally.[51, 54-55] Some centres use a risk-stratified, reduced intensity approach, others treat all children with aggressive antibiotic therapy Calls for collaborative trials have been made [56-58] but little progress made. The essential problems with research in this area are common in much of paediatric practice, rare conditions with small numbers of cases, and limited collaboration in primary studies. This clinical decision problem is the classic area where systematic review, with meta-analysis, may be able to draw together numerous studies and reach more powerful conclusions than any single study could. The output of clear risk stratification product of this work will inform practice and future therapeutic RCTs.

Chapter 2: Introduction to the methodologies employed in prognostic and diagnostic predictions in medical decision making

Background

Clinical decision rules (CDR) are clinical tools designed to be used at the bedside to assist decision making [17] and are generally either diagnostic or prognostic. Well known examples from adult medicine include the Ottawa ankle rules [59] and Wells DVT rule [60], both using clinical features to predict in particular the absence of an ankle fracture (and so avoid the need for radiography) or the absence of a DVT (and so avoid unnecessary anticoagulation). In paediatrics, such rules have been developed for the prediction of good outcomes from septic arthritis in the limping child [61], the identification of infants at low risk of serious bacterial infection [62], meningitis [63] or radiographic pneumonia [64], but are not generally or widely used.

Clinical decision rules are developed by an initial derivation study that creates the rule. This should be followed by further studies determining their discriminatory validity (do they actually tell the difference between the groups of affected and unaffected) and predictive accuracy (do they predict at the same sorts of proportions of individuals as they were created to do). Such validations can occur at different times, but within the same institution (temporal validation), in different physical locations but with similar clinical settings (geographical validation) and across different clinical settings (domain validation), for example in both tertiary specialised paediatric oncology centres and secondary care hospitals [65]. The final step should be to demonstrate their efficacy in routine practice with multi-site randomised controlled trials [66].

An ideal CDR for the management of febrile neutropenia would predict the risk of adverse outcomes from data collected at or soon after presentation. To this extent, it is 'prognostic' as a prediction of the course of disease [67]. However, this data may well be practically used in two different ways: to decide if the risk was 'low enough' to allow outpatient management, and at the opposite end of the risk scale, to consider the need for increasingly close observation and more aggressive management. The 'low risk' decision collapses into a dichotomy that can be considered 'diagnostic' ("is this a low-risk episode or not?") and such patients discharged for out-patient therapy. The patients at higher risk do not have such a clear difference in potential management options. There are no effective truly prophylactic

measures to prevent septic shock in this group, so the degree of risk generates a heightened degree of concern and observation, but does not require a dichotomous decision to be made. This 'high risk' information is not clearly a 'decision problem' and continuous, 'prognostic' information may be more useful. Such hypotheses about the nature and use of the information are ripe for testing.

It is worth noting that the description and decision of what constitutes 'low risk' is a matter of debate. It reflects both a desire to know if a patient has a significant infection which needs a specific therapy, and an understanding of the likelihood of a fatal or near-fatal outcome. In some settings this may collapse into the same information: the diagnosis of a systemic fungal infection is associated with an extremely high chance of death [68], and the diagnosis needs little further by way of prognostic information. With other infections, for example infection with coagulase-negative *Staphylococci* [69], serious adverse outcomes are rare. The setting of this threshold of 'low enough' risk appears to vary between healthcare professionals and families, and between healthcare professionals themselves [38] , and requires further study.

Taking all of these factors into account, the clinical use of a rule can only be countenanced if it is valid (truthful) and accurate (meaningful). In the setting of managing febrile neutropenia, the ideal output for a CDR is to use data available at the start of the episode to diagnose the patient as either seriously infected or not, and accurately predict their subsequent chance of important morbidity.

CDR and other predictive models - Derivation

A study of risk prediction, including the derivation and validation of clinical decision rules as a subset of risk prediction, requires that precise, accurate and unbiased information is collected so that any relationships discovered between predictors and outcomes are likely to be valid.

The prototype of a study that aims to create a clinical decision rules is one which prospectively collects information about a cohort of patients that present with a given problem (for example, fever in the neutropenic child). This cohort should be from consecutive patients, or a random sample of everyone who had been affected with the

condition. To collect patients by a different method may introduce a significant bias between those collected and those excluded. For example, if only patients who present 'during office hours' are included, this group could be more generally more 'well' than those who are sufficiently unwell to be taken for assessment in the middle of the night [70]. Using all patients obviously avoids this problem. Random selection is less intuitive, but derived from the principle that a truly random selection will accurately reflect the total population sampled.

The information collected for each patient should be the same. Gathering information prospectively appears to improve the likelihood of complete data capture and reduce biases arising from outcome reporting. For example, avoiding cases with a negative outcome (such as death) being more accurately captured and recorded than those with a better outcome [71]. The way that missing data is handled could reduce the efficiency of a study if it doesn't use as much information as it could, leading to unnecessarily wide confidence intervals, or could introduce biases and so incorrect predictive estimates if handled inappropriately.

A further consideration is that patients involved in the study need to be similar to the patients that the studies' results will be applied to. This is necessary because different groups may have different outcomes. For example, patients treated in tertiary clinics at super-specialised hospitals may have different outcomes than those from local hospitals, general practice or the community [72-73]. These differences should not be interpreted as meaning one sample location is 'wrong'; they indicate that the truth varies according to the population or case-mix under consideration, and needs to be made specific for the question asked.

In addition to the test data being collected in the same way for each patient, the outcomes should be assessed similarly, regardless of how the patient has presented (rather than patients who have been assessed at high risk of a problem undergoing a different outcome assessment than those at low risk). An example might be to only undertake chest X-rays (CXR) on children who have crackles or reduced air entry on physical examination when developing CDR for the detection of pneumonia. In this way, the study will tautologically prove the absence of these signs is perfect at ruling-out CXR positive pneumonia as they will never have been diagnostically tested. Such "differential verification" procedures have been shown quantitatively to overestimate the ability of the test to accurately diagnose a disease [74]. Another variation in this theme is a contamination of the reference standard with the test result: if the definition of pneumonia is "radiographic findings of pneumonia with an

appropriate clinical pattern", then children without classic signs or symptoms of pneumonia can't have the diagnosis, and therefore signs and symptoms of pneumonia cannot have anything less than near-perfect sensitivity.

The need for outcomes to be assessed blind to (without knowledge of) prognostic or diagnostic data is theoretically important, as there may be a tendency among clinicians to attribute different outcomes on the basis of 'clinical likelihood'. The exaggeration of effect has been demonstrated in studies of therapeutics [75], but has not been clearly demonstrated in studies of diagnosis [74] and prognosis [76]. This may reflect a true difference, or merely that there are fewer data available from prognostic and diagnostic studies.

Any good CDR should be based on outcomes which are important to patients and clinicians, and studies should be of sufficient duration for the outcome to become apparent. Some studies report outcomes too early; if new events are still likely to occur, this poorly reflects the true predictive value of a potential marker[77]

Data analysis presents a further series of challenges. There are various approaches to creating a CDR including the use of regression models, classification and regression trees (CART), neural networks and Bayesian networks. No clear superiority for one technique has been demonstrated [78], but it has been shown that different approaches can produce different results from the same data set [79]. This highlights an acknowledged difficulty with model building – that differing techniques may reach different conclusions from the same information.

The assumptions underlying these models are that the data collected is a true representation of the population of interest, that data have been collected accurately and that the various predictors can be combined simply, with different weightings of different elements.

The functional form of the predictors needs to be accurately assessed. For most model building techniques, the initial assumption is that the predictors and outcome have a linear relationship. If this is not the case, a transformation or non-linear form should be used. For example, temperature in the prediction of serious bacterial infection (SBI) in neonates is a 'U' shape, with both very low and very high temperatures being associated with SBI [62]. Other relationships have different forms, with 'floor and ceiling effects' such as the S-shaped

curve of oxygen dissociation from haemoglobin, or the J-shaped associations of body-mass index and mortality. If these are modelled by assuming only a linear relationship can exist, the variable may be discounted (e.g. U-shaped), over predicted at low values and under predicted at high ones (S-shape) or a complex failure of accurate estimation found in the lower portion of a J-shaped curve.

Related issues include multicollinearity, where variables are highly related, will lead to unstable models and inaccurate predictions. An example of co-linearity may be total white cell count and neutrophil counts in the setting of chemotherapy-induced marrow suppression. When faced with this situation, the most clinically sensible variable of the co-linear group should be chosen. A further assumption in simple models is that the observations are independent of one another. One relevant situation when the observations are *not* independent is when multiple 'cases' actually reflect multiple admissions from the same individual; this 'relatedness' needs to be built into the data analysis method.

Continuous variables (age, blood pressure, absolute neutrophil count) will have their maximum predictive accuracy in a model if used as their actual value, rather than categorised in bands of values. Clinicians seem to find the use of continuous variables in this setting unhelpful, and prefer to use categories. To build the most clinically effective CDR, a sensible approach would be to combine these approaches, exploring the association with a continuous value and making a clinically usable CDR with a categorical one. Undertaking this adds further challenges. Repeated studies examining prognostic model building have shown that the collapsing of continuous variables into ordinal (ordered) categories or dichotomies is often undertaken using methods which are highly likely to give spurious results [78, 80]. The problem arises from analyses where a particular set of data is examined to find the cut-point at which the greatest differentiation between the diagnostic or prognostic categories is achieved. In doing so, effectively multiple tests are being undertaken and the reported p-value associated with the final choice is likely to be a gross exaggeration of the true 'significance' of the value. Approaches using clinically or pathophysiologically meaningful values, or ones previously described in the literature, reduce these problems.

Selection of explanatory variables for a short and usable CDR is a further area of potential problems. The best approaches are to think carefully about what relationships are expected to exist in advance of data analysis. This approach will reduce the chance of spurious, data-driven associations slipping into a CDR. More apparently 'rigorous' and 'statistical' techniques can be undertaken. Such selections can be performed by taking all possible

explanatory variables, and excluding those which are not statistically significant (backwards elimination), or by taking in the most statistically significant individual factors (forwards selection) or a combination of the two, adding and removing to build the statistically best fitting model (stepwise). These techniques, when driven by a 'p-value' are seriously at risk of choosing variable with chance relationships and making unstable models [81]. These techniques will also exclude variables that confound each other from entering a multivariable analysis [82]. In essence, selecting variables only by looking at how significantly they are associated with outcomes in the dataset being examined produces highly effective descriptions of that dataset; it doesn't improve the ability of the model to describe the real population from which those data were drawn [83].

Building stable predictive models requires a minimum of 10 to 20 events per variable considered [84], with more being better. Small numbers of events increase the possibility of finding spurious associations that existing only in that dataset, and not in the general population [83]. There are a range of techniques that have been developed to try to reduce the chance of such problems occurring, described as ways of 'shrinking' the overinflated estimates from the model, or 'penalising' the model as it tries to build in too many overly optimistic variables [85]. Such modifications may lead to models where the predictive values are retained in future studies [86].

To make things usable in a clinical environment, it is often far more sensible to present clinicians with a simple table of signs and symptoms with a numerical score than a complex equation. Remarkably, the use of very simple versions of the weights from regression equations often work in practice as well as the mathematically precise numbers [87], and it may even be worth ignoring weights all together and just calling each 'one point' [88].

CDR and other predictive models - Validation

Given the issues off over optimism and generalisability described above, a newly developed CDR should be validated before use. In this context, validation means that the CDR has been shown to accurately discriminate between those with and without disease, or accurately estimate the proportion of patients with the disease. This is analogous to the clinical trial testing of a drug which has shown positive results in cell cultures or mice; inaccurate predictive information both as false positive [89] or false negative test results [90] can be as harmful as an untested therapy.

As previously noted, various levels of validation are described, including temporal validation (the rule is tested again by the same clinical team), geographical validation (testing the rule in a different location, but similar clinical setting) and domain validation (testing the rule in an alternative clinical setting, such as secondary rather than tertiary care). This last stage may be irrelevant if it is to be applied in the same setting as rule development occurred. The use of these various steps is important to demonstrate there is a practical ability of the rule to be used widely and effectively.

The final step should be to demonstrate the efficacy of a CDR in routine practice with multi-site randomised controlled trials [66]. The trial doesn't seek to examine if the CDR is accurate, but instead to randomise between application of the CDR and no application measuring key patient-important, or health-system- important outcome such as length of hospital stay, invasive testing, or improved quality of life. This stage is rarely undertaken, but can be a very powerful way of demonstrating improved care through the use of a diagnostic intervention [91].

CDR and other predictive models - Implementation

Understanding how clinicians use a clinical decision rule, or any diagnostic information, involves understanding how medical professionals 'think'. A number of researchers from a range of backgrounds have examined the diagnostic practices of physicians. There is a wealth of research that demonstrates the common clinical myth of diagnosis following the doctors actions of 'take a history and do a physical examination' is inaccurate [92-93], and that instead health professionals apply an array of mental shortcuts (heuristics) [93-94] which both speed up working practice and at the same time can lead to dangerously wrong conclusions. In fitting a CDR into the practice of managing a clinical problem, it can be useful to take a straightforward model as to how doctors make a diagnosis and move on to treat the disease.

The diagnostic process can be thought of in three stages: initiation, refinement and conclusion [95]. The initiation stage is where a differential diagnosis begins to be considered. The refinement is a working through of these differentials, and the conclusion is a point where a decision to act has been made.

In some situations, such as seeing a toddler with Down's syndrome, initiation is the only step. While this is clear to many lay people, there is research to suggest that the more expert a physician is in a particular area, the more rapidly they come to a diagnosis, and that this in part is because of pattern recognition or fitting new cases to a mental 'categorisation' [94].

For most situations, initiation is truly the first idea that undergoes a process of refinement. It is in this process that CDR can help guide clinicians, and lead to a diagnostic conclusion. Other approaches include a formal Bayesian analysis, pattern fitting, or a stepwise rule-out of significant serious diagnoses.

The conclusion part of the process may be an actual pathological diagnosis (e.g. pneumonia), or a rule-out (e.g. no evidence of bacterial meningitis) or an admission of remaining diagnostic uncertainty.

The CDR should help in refinement by providing good quality guidance to avoid diagnostic errors and minimise unnecessary tests. Commonly diagnostic errors occur because of both systems and cognitive errors [96-97]. Such errors can include: a failure to syntheses diagnostic information correctly and come to a premature diagnostic conclusion; a lack of appreciation of the value of a sign or symptom in making a diagnosis; or exaggeration of the accuracy of a test finding. Other reasons for misdiagnosis would not be helped by the use of CDRs, such as the true diagnosis being rare, or failure in the technical skill of the individual doctor, for example in reading an x-ray or eliciting a physical sign.

However, there remains an almost emotional difficulty in turning to a CDR when instead clinicians should be like House [98], Holmes [99] or Thorndyke [100] in making diagnoses from skill and knowledge. This is despite the widely publicised data which suggest that in many cases a CDR performs better than 'expert opinions' [101-104]. Why healthcare professionals do not follow where the best evidence should steer is a matter of ongoing debate and research.

The most effective uses of CDR seem to have been where the rule has a clear clinical utility, has been championed by well-respected local clinicians, and clearly improves outcomes for patients and clinicians [105-106]. Implementation of any well-derived and validated rule will require skilled advice and a multi-factorial approach to maximise real clinical gains. This should highlight that CDR are potentially a way of making predictions more accurate, and/or quicker and less unpleasant to achieve. They don't necessarily do this [107] and so each rule requires clear sighted critical appraisal before implementation.

Systematic reviews of CDRs

In exactly the same way that therapeutic studies should be viewed and reviewed within the context of all the unbiased information, preferably in the form of a systematic review with meta-analysis, CDR and prognostic studies should be seen within the same context. The rationale behind these studies is identical to that which drives therapeutic reviews: by pooling information chance associations can be minimised and a more precise and accurate estimate of effect can be obtained. The nature of the numerical outputs from these studies is explored in the next sections.

It is notable that reviews of predictive studies, in keeping with reviews of therapeutic interventions, can sometimes produce useful new results [60]. They can also confirm the inadequacy of the current studies to derive a clinically applicable result [108], or highlight the poor quality of underlying studies and the need for higher quality primary research [109-110]. Even more prevalent in predictive studies than in therapeutics are the difficulties produced by the poor quality of reporting of studies [111], and marked publication bias [78]. In many areas, these lead to the need to undertake research which pools the raw individual patient data (IPD) from high quality studies. Undertaking an IPD meta-analysis would increase the number of events studied, which allows more confidence to be placed in the estimates of association between predictive variables and outcome and allow for more consistent handling continuous outcome data, which may well have been categorized differently in differing datasets. It would also permit the independent assessment of episodes (e.g. using only the first episode for each patient) and then analyze the degree to which episodes, patients and outcomes are interdependent.

Simple numerical descriptions of test accuracy

In order to describe accuracy of a test result is it is usually necessary to produce a numerical estimation. Many tests give results as continuous values (for example biomarkers), yet for simplicity these are reported as being positive (e.g. above a certain value) or negative (below the cutoff value). The group under study, in this case children presenting with febrile neutropenia, can be thought of as coming from two populations: those with the 'disease', for example pneumonia, and those without it. The values of "lung injury protein" (LIP) are distributed differently in the two groups. (See Figure 1, over)

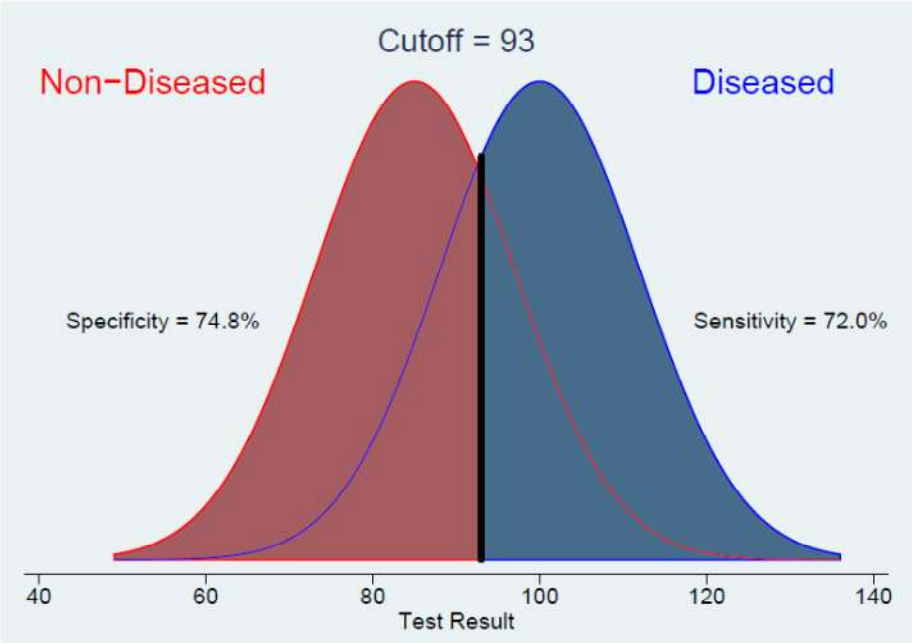
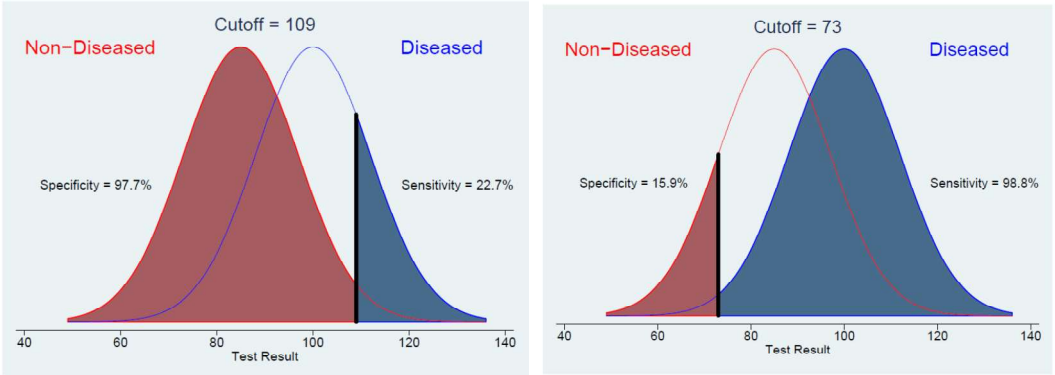


Figure 1: "Lung injury protein" diagnostic test distribution

The threshold then has part of both populations on each side; the proportion of people with pneumonia who have a positive test is the *sensitivity*. The proportion of people without pneumonia who have a negative test is the *specificity*. Shifting where the threshold is drawn alters both these proportions. As Figure 2 shows, pushing the line for 'positivity' upwards makes the test more specific (captures fewer people without the disease in the definition) but becomes less sensitive (fails to diagnose a greater number with the disease). The reverse is true when the level for positive results is reduced.

Figure 2: "Lung injury protein" diagnostic test: different cutoffs



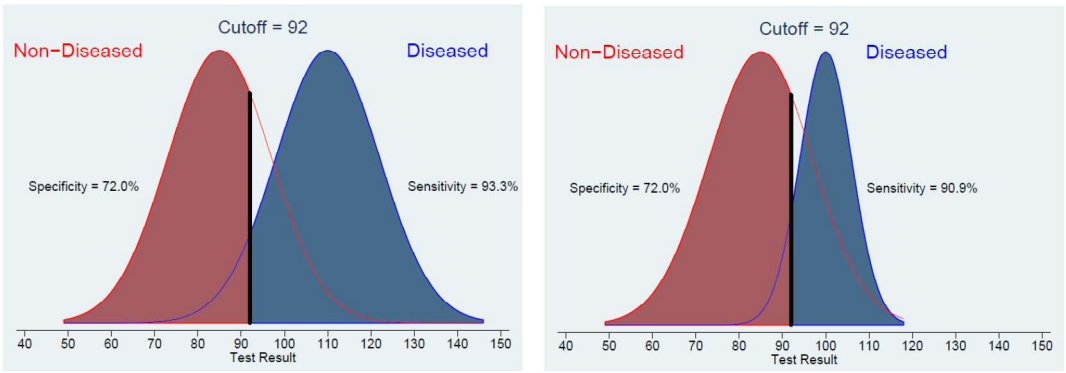
Meta-analysis of diagnostic test accuracy

Meta-analysis is the statistical combination of results from more than one study. In the setting of a systematic review, this process should produce the most precise values which represent the sensitivity and specificity for a diagnostic test and explore the limits of our certainty.

When such a meta-analysis is undertaken, there are three elements in combining the test results that may vary. The first is that the group under study are a (random) selection of the 'true' population of children with and without pneumonia. Any single estimate of test effectiveness is only an estimate of the 'true' test accuracy, and each study reports this chance uncertainty be providing estimates of the variance of sensitivity and specificity.

The next aspect of variation is that the 'true' population from which the sample was drawn may actually be a mixture of slightly different populations. For example, it may be that slightly different LIP values are present in Scottish children who have higher normal values than children in London, through genetic polymorphisms, or less variation between them (see Figure 3). This aspect of variation can be evaluated by the use of a random effects meta-analysis procedure. This examines variability between study populations and provides an estimate of the 'average' sensitivity and specificity in an 'average' population, and also provides a numerical range in which the truly different values may lie in different populations.

Figure 3: "Lung injury protein" diagnostic test: different mean and SD



The third aspect of variability may be that a different threshold is used between studies. In studies that report 'hard' laboratory findings, this should be easily assessed (although different assay techniques mean this is not necessarily the case). In those using clinical criteria, for example "looked clinically unwell", this is much more difficult to judge. One way

to assess this is to examine the differences in sensitivity and specificity, which should be paired and move in opposite directions, as the above example shows, and try to estimate this variation by taking into account the correlation between these two aspects of the singular line of threshold. This is the core principle of a bivariate analysis. The situation has an added level of complexity with two cutoffs (three levels of test result: low, medium and high) requiring multinomial approaches, but the core concept remains the same.

Alternative descriptions of test accuracy

The relationships between sensitivity and specificity can also be visually demonstrated by plotting them against each other on a graph. A common standard for such displays is in receiver operator curve (ROC) space. As Figure 4 demonstrates, this shows the results of a series of studies, plotting their sensitivity and (in reverse axis) specificity.

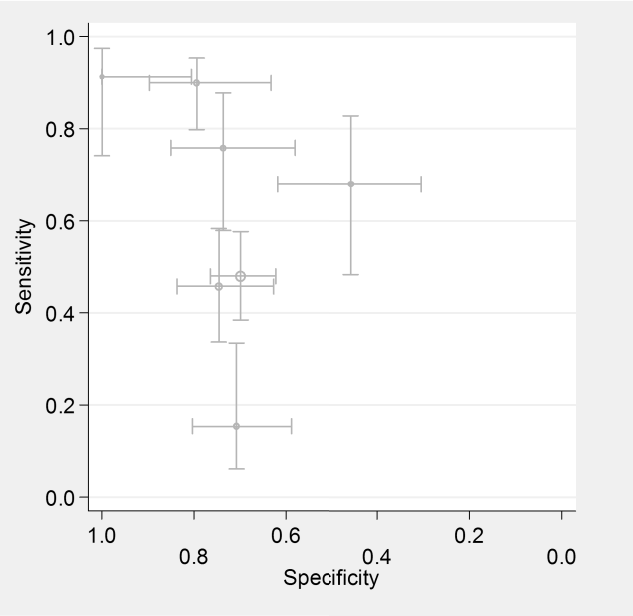


Figure 4: ROC 'cross-hairs' plot showing the results of seven different diagnostic test accuracy studies.

In reading such graphs, tests with a better discriminatory ability fall in the top left corner of the plot (with very high sensitivity and specificity) and non-discriminatory tests fall on a 45° line between the bottom left and top right corners (where the result describes an identical proportion of those with and those without the diseased and so is uninformative). The

capped lines display the 95% confidence intervals associated with the sensitivity or specificity values.

This same graphical display can be used to show the relationship between the varying threshold of a continuous or multiple-layer diagnostic test as a curved line (see Figure 5, example of a single study).

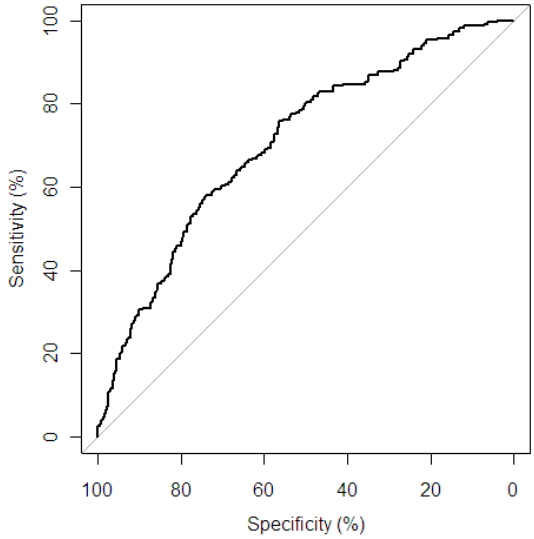


Figure 5: ROC plot showing the results of one continuous scale diagnostic test to determine the presence or absence of disease.

Using test accuracy estimates in clinical practice

For clinical application, the important values are generally not the proportion of people with a disease (e.g. pneumonia) who have a positive test (*sensitivity)* and proportion of people without pneumonia who have a negative test (*specificity)* but rather the proportion of people with a negative test result who truly didn't have disease (*negative predictive value, NPV)* and the converse, the proportion of people with a positive test who did have pneumonia (*positive predictive value, PPV)*. These values are the compilation of both the diagnostic accuracy and prevalence of the disease in the population.

A clear illustration of this comes from an analysis of the different NPV and PPV of testing for *Clostridium difficile* infection with commercially available stool toxin test kits [112]. As Figure 6 demonstrates, the 'truth' of a positive result varies from 50% correct to 92% correct, depending on how prevalent *C. difficile* is in the population under study. (This Figure shows

a theoretical Clostridium difficile toxin assay with a sensitivity of 92% and a specificity of 97%. NPV=negative predictive value. PPV=positive predictive value. [112])

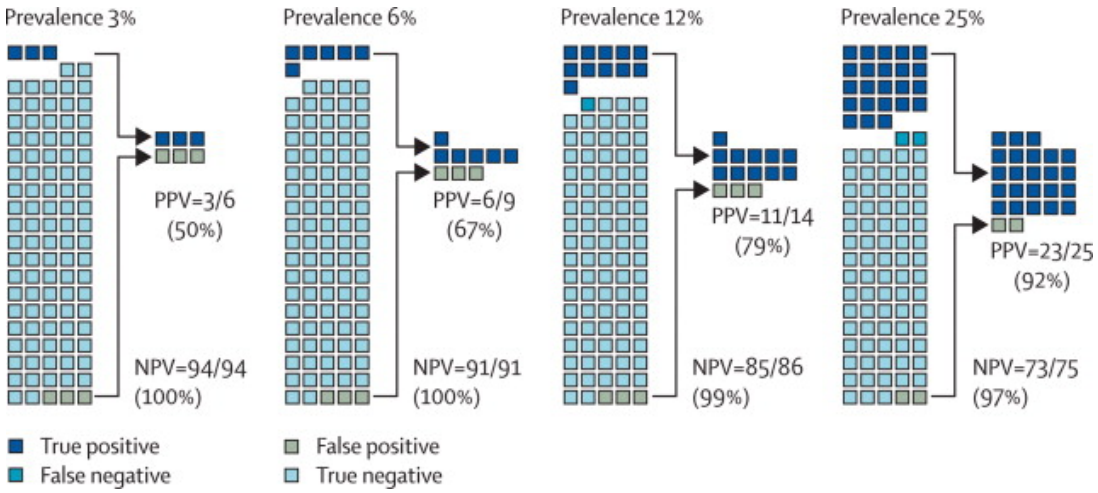


Figure 6: Effect of varying prevalence on the PPV and NPV

The practical implication of this is that the direct translation of sensitivity and specificity values from a single study or meta-analysis into practically meaningful PPV and NPV can only be undertaken if the prevalence of the condition is known, or can reasonably be estimated.

An alternative expression of the sensitivity and specificity of a test is the use of likelihood ratios (LR). These values compare the proportion of patients with the disease and without the disease for a given test result. In the LIP example (Figure 1, specificity 74.8% and sensitivity 72.0%) these values would be LR+ (likelihood ratio for a positive test) = sensitivity / (100% - specificity) = 74.8/28.0=2.67 and the LR- = (100% - sensitivity) / specificity = 0.35. Such values can be calculated for each level of a test result, and so are useful for multi-level as well as dichotomous tests (compare Tables 1 and 2 for LIP).

Table 1: Sensitivity & specificity for Lung injury protein values to detect pneumonia

Test cutoff	Sensitivity	Specificity
73	98.8	15.9
93	72.0	74.8
109	22.7	97.7

Table 2: Likelihood ratios for Lung injury protein values to detect pneumonia

Test cutoff	LR+	LR-
73	1.17	0.11
93	2.99	0.38
109	9.87	0.79

These likelihood ratios can then be used in formal Bayesian analysis converting pre-test odds of disease into post test-odds of disease, or informally interpreted as how far the diagnostic pendulum is pushed towards a disease (e.g. LIP value of 110 makes it about 10-times more likely the child has pneumonia, and LIP of 70 makes it about one-tenth as likely).

Conclusion

Defining a subset of patients as low risk of infection reduces the intensity of treatment in that group, and facilitates rational intensification of anti-infective treatment in the high-risk group. To make this decision rational and repeatable, a logical approach is to use a clinical decision rule. This rule needs to be developed in a robust manner, to reduce the effects of chance, confounding and bias obscuring the true relationships between proposed predictor variables and the outcome of each episode of febrile neutropenia. Furthermore, a rule needs to be tested, to make sure it works effectively and is practically useful.

Chapter 3: Methods of Systematic reviews of Clinical Decision Rules and Serum Biomarkers

Two linked systematic reviews [113-114] were undertaken of existing rules and the value of serum biomarkers, in order to determine whether an IPD meta-analysis was necessary; to identify suitable data sets; and to guide such a study in collecting appropriate variables between 2008 and 2010. These reviews were later updated [115-116], and as a by-product a further review related to a specific aspect of early assessment was also produced: do all children presenting with fever and neutropenia require a chest radiograph to exclude pneumonia [117]? These reviews provided the evidence on which to propose the key predictor variables to be used in the following IPD analysis.

This chapter described the methods generic to both groups of reviews; the clinical decision rule studies [113, 115] and the biomarkers papers [114, 116]. The results of the reviews are then described in Chapter 4, and a discussion and conclusions to the extensive background work for this thesis are presented in Chapter 5.

Methods

The reviews were conducted in accordance with “Systematic reviews: CRD's guidance for undertaking reviews in health care” [118] Protocols were written for each review and in advance of starting the review were registered with the HTA Registry of systematic reviews, CRD32009100453 and CRD42011001684.

Search & retrieval strategy

Electronic search strategies (See Appendix 1 and Appendix 2) were developed which examined the following databases: MEDLINE, MEDLINE(R) In-Process & Other Non-Indexed Citations, EMBASE, CINAHL, Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessment Database (HTA), Cochrane Central Register of Controlled Trials (CENTRAL), Conference Proceedings Citation Index - Science (CPCI-S), Literatura Latinoamericana y del Caribe en Ciencias de la Salud (LILACS)

Reference lists of relevant systematic reviews and included articles were reviewed for further relevant articles. Published and unpublished studies were sought and no language

restrictions applied. Non-English language studies were translated. Searches were conducted on date for the initial review published in 2008 and on for the update published in 2012.

Inclusion and exclusion criteria

Studies were included in the review if they met the following criteria.

Methodology Studies which aimed to derive or validate a CDR in either all or a defined subset of patients were included. Both prospective and retrospective cohorts were included, but those using a case-control ("two-gate") approach were excluded as these have been previously shown to exaggerate diagnostic accuracy estimates [119].

Population Children or young people (aged 0 – 18y) who were receiving treatment for cancer or leukaemia (including extra-cranial and intra-cranial tumours) presenting with febrile neutropenia. Studies which examined children and adults were included if the paediatric data were available separately.

Predictor variables for CDR reviews Clinical decision rules (CDR) using clinical and haematological or biochemical variables used to predict outcome for the particular episode of febrile neutropenia

Predictor variables for Biomarkers reviews Serum inflammatory/infectious markers (for example including, but not limited to, C-reactive protein (CRP), procalcitonin or interleukin levels) measured within the first 12 hours where timing of samples was reported.

Outcomes (At least one of) Survival, need for intensive care, need for high-dependency care, single organ impairment (oxygen requirement, renal impairment, hepatic impairment), invasive bacterial or fungal infection, any documented infection (including radiologically confirmed pneumonia), duration of admission.

Study selection

Two reviewers independently screened the title and abstract of studies for inclusion, and then the full text of retrieved articles. Disagreements were resolved by consensus.

Validity assessment

The validity of each study was assessed using 11 of the 14 questions from the QUADAS assessment tool for diagnostic accuracy studies [120]. The QUADAS tool was adapted specifically for the review, as suggested by "Systematic reviews: CRD's guidance for undertaking reviews in health care" [121], omitting questions on "time between index and reference test", "intermediate results" and "explanation of withdrawals". (See Appendix 3.) The CDR and reference tests are necessarily related, and the design of a CDR and the reporting of the biomarkers studies meant that "intermediate" results are included in any analysis. The issue of incomplete data was addressed in the analysis of the method of derivation or validation, and as such was not included as a quality criterion.

Data extraction

Data were extracted by one researcher using a standardised data extraction form and checked by a second. The data extracted included participant demographic details such as age and sex, geographical location of the study, the participant inclusion/exclusion criteria, antibiotics used, and the performance of the CDR as a 2*k table (where k refers to the number of strata described). Information was extracted on the methods used to derive the CDR (where applicable), including the variables considered, methods of statistical analysis, and methods of dealing with multiple episodes in individual patients and missing data. An example of the form used for the biomarkers review is given in Appendix 4. Authors were not contacted for clarification in the event of 'unclear' risk of bias assessments or to seek additional information.

Methods of data synthesis

The studies were reviewed using both narrative and quantitative synthesis.

Quantitative synthesis was undertaken for studies which tested the same CDR or biomarker, and investigated for sources for heterogeneity.

For dichotomous test data in this review, where possible analyses used a bivariate model (using the 'metandi' command for STATA10 [122]). The bivariate approach, when possible, accounts for the paired nature of dichotomous test characteristics as described in chapter 2. For tests that included very small numbers of studies ($n \leq 4$) fitting a bivariate model is

problematic as the procedure frequently fails to converge. In these cases, a univariate approach was used (pooling sensitivity and specificity separately).

For tests where three-level (low, mid- and high-risk) results were produced, an innovative approach adapted from a previous method used to pool three-level results for the diagnosis of deep venous thrombosis was developed in [60]. The initial method used random-effects meta-analysis was undertaken using WinBUGS 1.4.3 [123] to estimate the proportions of individuals classified as low, medium or high risk in the 'diseased' (e.g. bacteraemic) and non-diseased groups. The extension developed to this method allowed multivariate random effects were applied to the calculation of each cutoff value. Data from studies which used a similar rule but provided only two of the risk categories were also included in this analysis [124]. These proportions were used to calculate likelihood ratios for each risk category and corresponding 95% credible intervals. In such cases, where cutoff thresholds are fixed between studies, not using a multinomial approach which accounts for variability of threshold is less likely to introduce biases.

Two main types of analysis were used for the biomarkers meta-analysis, one using classical statistical methods and one based upon Bayesian analysis. For the maximum likelihood estimate approach, the data were pooled from studies reporting the same marker and similar outcomes using a single cutoff from each study using the STATA routines metandi and midas for analyses of HSROC curves and bivariate analyses with >3 studies, for those with <4 studies a random effects linear regression was fitted using xmelogit. Where possible, the most common cutoff value was chosen for greatest precision of estimate.

Analysis using out innovative Bayesian multinomial random effects method was undertaken to derive proportions of the population with/without the outcome at each cut-off level of the serum markers using monte-carlo markov chain modelling via WinBUGS 1.4.3 [123] with non-informative priors. These results were then used to derive likelihood ratios for each level with corresponding 95% credible intervals.

Where data for continuous variables were presented as mean and standard deviation, rather than 2*2 tables, conversion was undertaken using the assumption of Normality (or log-Normal in the case of serum proteins) and deriving the assumed 2*2 table for cutoffs reported by other studies [Anzures, Cochrane Colloquium Freiburg 2008].

Heterogeneity between study results was explored through consideration of study populations, study design, predictor variables assessed and outcomes chosen. Sensitivity

analysis was undertaken by comparing results with the original (derivation) data set included and excluded. The sparse nature of the data rendered statistical approaches such as consideration of the I^2 statistic inappropriate.

Where quantitative synthesis was not possible, a hypothesis generating narrative approach was used. The narrative synthesis was undertaken according to the framework described in "Systematic reviews: CRD's guidance for undertaking reviews in health care" [125]. This proposes an iterative approach to developing a theory underpinning the data, using this to structure a preliminary synthesis of the findings of the included studies, and exploring how relationships within and between studies support or refute the hypothesis, with an assessment of the robustness of the synthesis.

Methods of data display

Results of dichotomous meta-analyses are displayed using 'cross-hairs' plots. We developed this innovative graphical augmentation of ROC space plotting to assist disseminating the concepts of diagnostic meta-analysis to clinicians, and published this in a descriptive methodological article [126].

Traditional approaches to displaying information from diagnostic meta-analysis have been to use side-by-side forest plots of sensitivity and specificity, which allow the reader to view the univariate heterogeneity of sensitivity and specificity, but are difficult to appreciate a biaviate relationship. An alternative has been to open circles of the point estimate of each study in ROC space, with a larger marker indicating a larger study. While this shows the bivariate relationship, physically this plotting inverts the relationship most clinical readers are familiar with from forest plots.

The cross-hairs plot combines the clinically familiar idea of a forest plot, with box of point estimate and whiskers of individual study 95% confidence intervals, with the enabling the bivariate relationship of sensitivity and specificity to be assessed easily and maintaining the relationship between size-of-arms and uncertainty.

The basic plot uses a single colour or tone to identify the individual studies, with an identified marker icon, and a distinct icon and colour/tone to display the meta-analytic summary. This can be displayed as paired univariate meta-analyses (see Figure 7), or as a

confidence ellipse (see Figure 8). A prediction interval can also easily be plotted onto this type of graph.

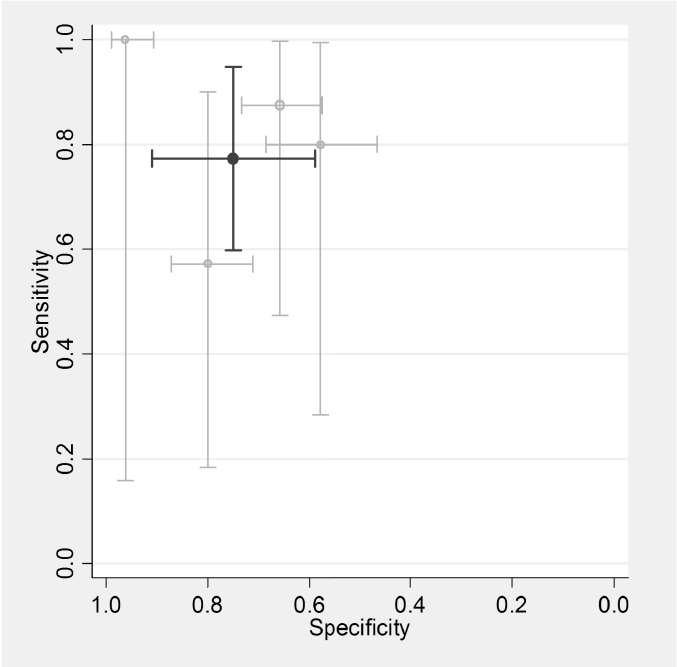


Figure 7: Monochrome cross hairs plot with two univariate meta-analyis results

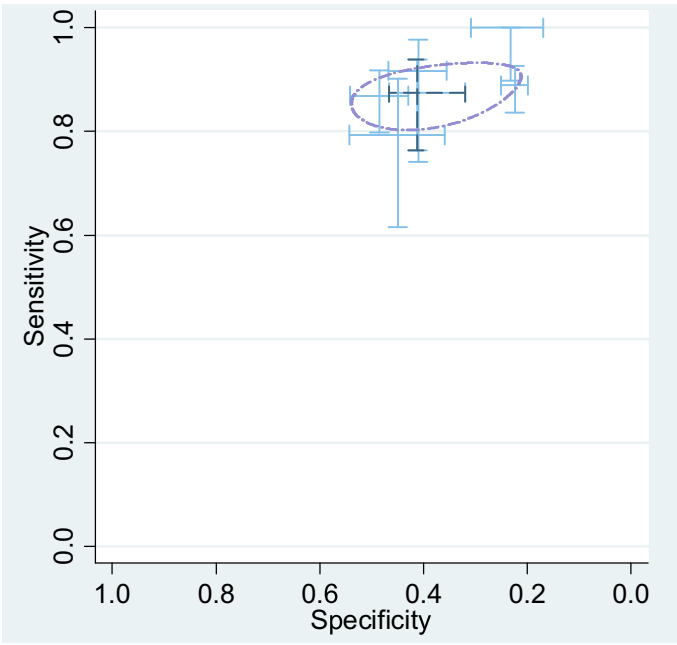


Figure 8: Colour cross hair plot with bivariate confidence ellipse

Modifications of this plot enable further information, such as the threshold level of a serum marker, to be used. This further aid interpreting the results using logical colour selection (for

example, the rainbow sequence) by demonstrating how different thresholds affect the sensitivity and specificity of a test (see Figure 9).

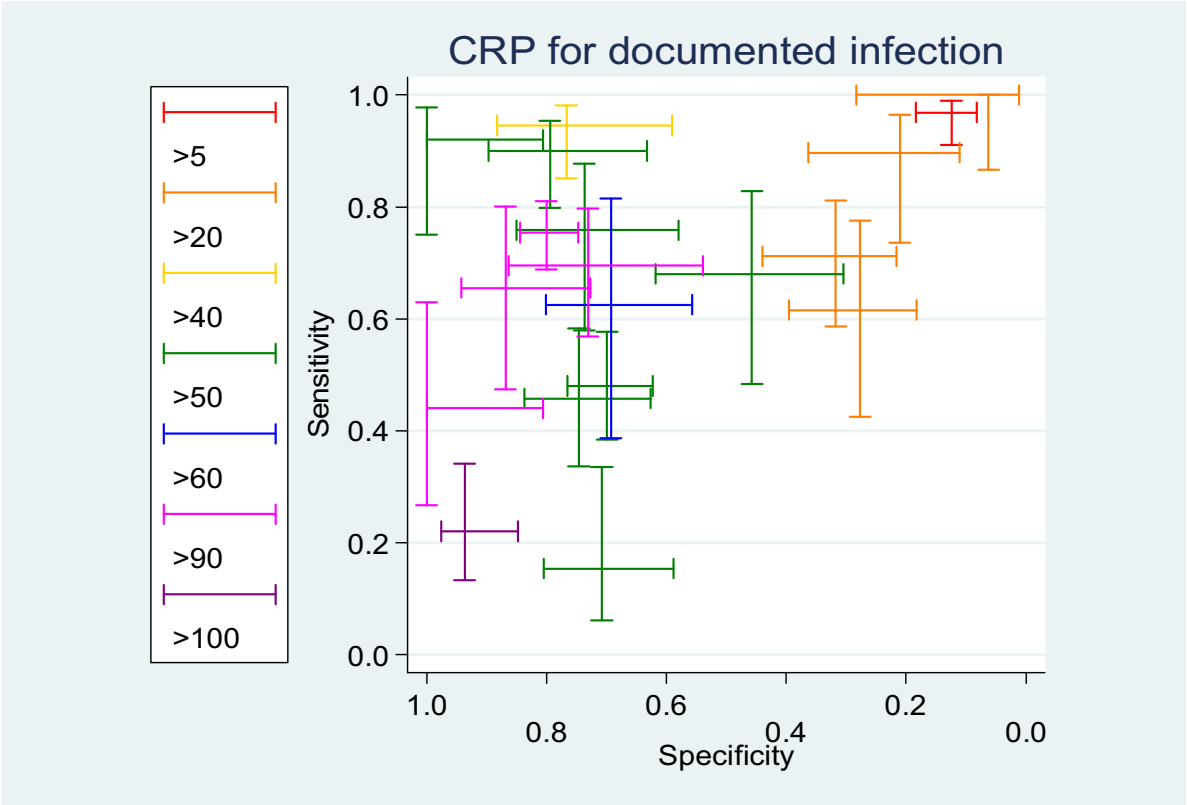


Figure 9: Multicolour cross-hair plot with explicit different test thresholds

Code is available for STATA in Appendix 5, and for users of the R statistical environment, the package mada [127] has been adapted to include this display format after discussion with the author.

Chapter 4: Results of Systematic reviews of Clinical Decision Rules and Serum Biomarkers

This chapter presents the results of the systematic review of clinical decision rules and serum biomarkers in the prediction of adverse outcomes from episodes of febrile neutropenia. They were undertaken in 2008-2010, prior to the IPD analysis explored in the later chapters, and updated during the collection phase of the IPD study; 2010-2012. The updates were, in part, triggered by undertaking the role of Clinical Lead on the National Institute of Health and Clinical Effectiveness commissioned guideline on the Prevention and Management of Neutropenic Sepsis [128] and the international Guideline on the Management of Paediatric Febrile Neutropenia [129].

Clinical decision rules

This section of the Chapter deals with rules based primarily upon clinical assessments at the point of admission or recognition of an episode of FNP, and addresses how they have been derived and validated in predicting risks of infectious complications.

Study inclusion and exclusion

Figure 10 and Figure 11 describe the flow of candidate and eligible articles though the review process. In the initial review, 2057 articles were identified from electronic searches undertaken in Febuary 2009 and 3 further articles were identified from examining the bibliographies of systematic reviews and included studies. From this, 89 articles were identified for detailed examination, of which 25 articles reporting on 24 studies were eligible for inclusion in the review. The update of the review (searches undertaken in September 2011) added a further 9 articles reporting on 8 studies. A total of 10,431 patients were included in these reviews.

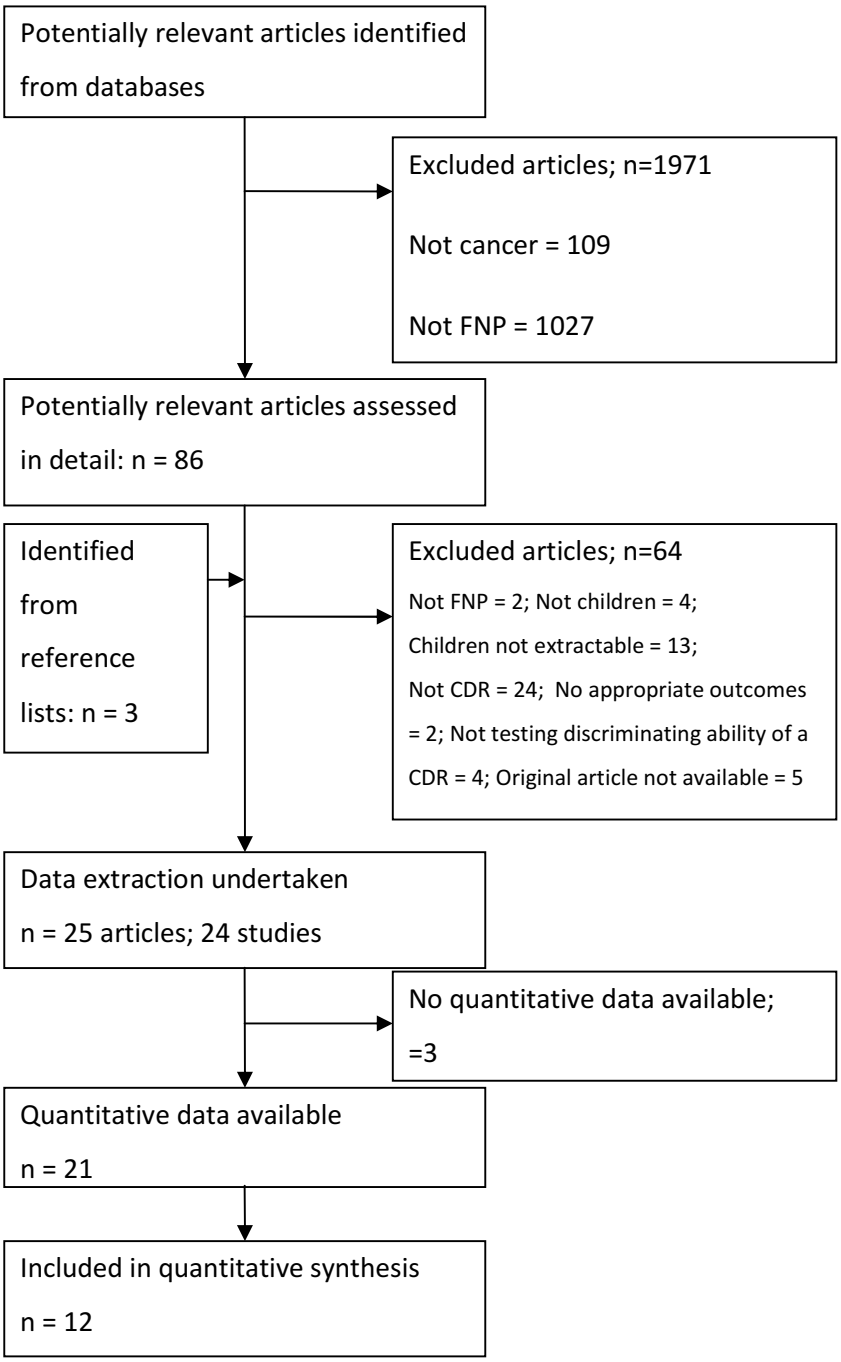


Figure 10: Flow diagram of study selection process; original CDR review

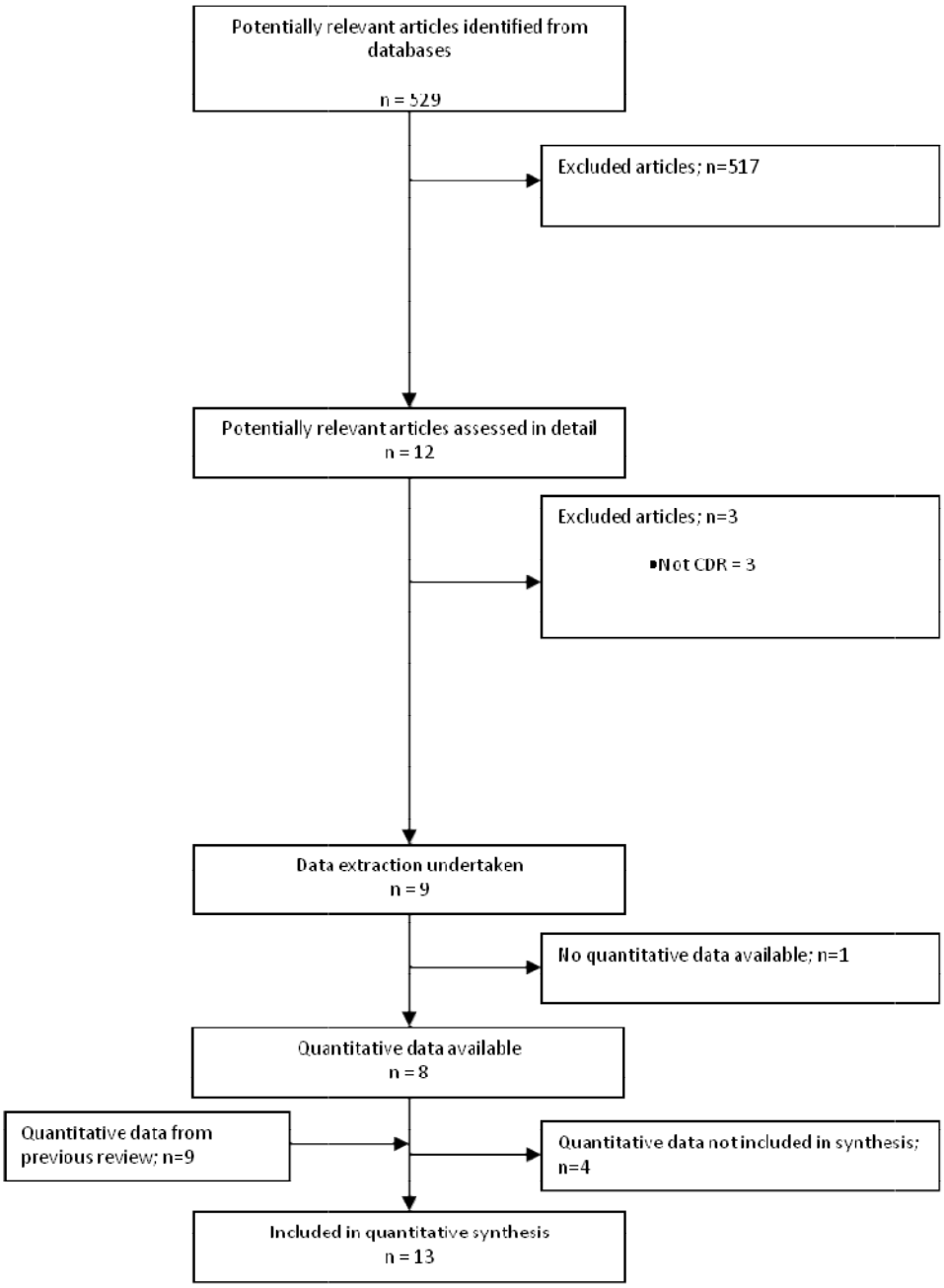


Figure 11: Flow diagram of study selection process; update CDR review

All studies included patients with a wide range of malignancies and patients from 1 month to 23 years old. They included between 29 and 759 patients (median 132) with between 47 and 1117 episodes of febrile neutropenia per study (median 240), where stated. See Table 3 for more detail.

Of the studies 29 deal with general infectious complications and 'routine CDRs' [5, 11-12, 22, 24-26, 69, 79, 130-148]and four address the specific issue of the detection of pneumonia [149-152]. As described in the previous chapter, these studies address a distinct and separate question than the use of a model of assessment of risk of complications of febrile neutropenia. The review has been published [117] and its recommendation (to only undertake chest radiography in the setting of signs or symptoms of lower respiratory tract disease, or in patients with other comorbities which increase their risk of pneumonia) has been incorporated into FNP guidelines [128-129]

Evaluation of Clinical decision rules

The 29 studies which examined general infectious complications included 15 which aimed to derive a CDR [5, 11-12, 22, 24-26, 69, 79, 131-147]. Five studies did not describe a CDR [5, 69, 130, 142-143] as the data collected did not produce statistically significant predictors. A total of 21 CDR were described. 6 studies sought primarily to validate a model's discriminatory ability [11-12, 22, 24, 130, 136], three also recalibrated a rule [12, 136, 144] (see Table 3).

Four studies used a split sample to validate their rule [12, 26, 137, 141], and one study provided data to test an alternative rule [69]. Six CDR have been subject to validation in separate data sets from the derivation set [22, 134, 137, 141, 144, 153]. Bootstrap analysis has been used in four cases [79, 140, 144, 147]. The remaining CDR were only explored in a single dataset. Thirteen individual outcomes (Table 3) were predicted, these can be summarised in five clusters: death, critical care requirement, serious medical complication, significant bacterial infection, and bacteraemia.

Table 3: Clinical decision rules and outcomes under study

Citation	Clinical prediction rule*	Outcome #1	Outcome #2	Outcome #3
Derivation Studies				
Adcock 1999	High risk = hypotension/septic shock, inflamed central line site, recent high dose Ara-C	Gram positive bacteremia		

Alexander 2002	Low risk = Not AML/Burkitts/Induction ALL/Progressive-relapsed with marrow involvement ("Anticipated neutropenia <7days") and no significant comorbidity (defined as hypotension, tachypnea/hypoxia <94%, new CXR changes, altered mental status, severe mucositis, vomiting or abdominal pain, focal infection, other clinical reason for in-patient treatment).	Bacteraemia	Serious medical complicatio n	Death
Ammann, 2003	Final decision tree model: 4 covariates were used to classify low risk; bone marrow involvement, leukocyte count >0.5 x 10^9/L, with clinical signs of a viral infection, and aged up to 6 years at presentation. For those with a leukocyte count ≤0.5 x 10^9/L, they were further classified according to CRP level (≤ or >50mg/L).	Severe bacterial infection, (death from bacterial infection, a positive culture of normally sterile body fluids, radiologically proven pneumonia, clinically unequivocal diagnosis of a bacterial infection, or CRP>150 mg/L)		
(model #2)	Low risk ≤3 factors. Risk factors = bone marrow involvement, absence of clinical signs of viral infection, high serum CRP level, low leukocyte count, presence of a central venous catheter, high haemoglobin level, and Pre-B-cell leukaemia.	Severe bacterial infection (death from bacterial infection, a positive culture of normally sterile body fluids, radiologically proven pneumonia, clinically unequivocal diagnosis of a bacterial infection, or CRP>150 mg/L)		
(model #3)	Low risk ≤4 factors. Risk factors = bone marrow involvement, absence of clinical signs of viral infection, high serum CRP level, low leukocyte count, presence of a central venous catheter, high haemoglobin level, and Pre-B-cell leukaemia.	Severe bacterial infection (death from bacterial infection, a positive culture of normally sterile body fluids, radiologically proven pneumonia, clinically unequivocal diagnosis of a bacterial infection, or CRP>150 mg/L)		
Ammann, 2004 (same population as Ammann 2003)	Low risk = all of: maximum temp ≤39.7C, no comorbidity requiring hospitalisation, leukocyte count >0.5 x 10^9/L, and in partial or complete remission	Bacteraemia		
Amman 2010	Applied after 24 hours: 4 points for chemotherapy more	Significant adverse outcome; Severe bacterial		

	intensive than ALL maintenance,5 points for hemoglobin > 90 g/L, 3 points each for white blood cell count <0.3 x10^9/L, platelet < 50 x10^9/L, any adverse event occurred in preceding 24h. Scores ≤9 are low risk.	infection, admission to HDU/ICU for organ support, severe sepsis or septic shock, potentially life-threatening event. death		
Ageyman 2011 (Same population as Amman 2010)	Applied after 24 hours: shaking chills ever observed, haemoglobin > 90 g/L, platelet < 50 x10^9/L, any other need for IP treatment. No risk factors = low risk	Bacteraemia after 24h		
Badiei 2011	Platelets <20 x10^9/L, temperature ≥39°C, ANC <100/mm^3, mucositis, abnormal CXR on presentation. Risk of infection greater with more risk factors: no single threshold applied	Life threatening infection		
Delebarre 2010 [abstract only]	1 point for hematological malignancy, chemotherapy at high-risk of prolonged neutropenia, 1 point for clinical signs of local infection, fever >39°C, white cell count <500/mm^3 or monocytes <100/mm^3 and procalcitonin >0.3ng/ml. TWO points for severe sepsis. High risk >1 point.	Severe infection		
Hakim 2008	Score from cancer diagnosis: AML = 20, ALL/lymphoma = 7, Solids = 0 Clinical presentation of serious unwell or toxic = 14, fever at presentation: ≥39°C = 11, ANC <100/mm^3 = 10 points, Total score <24 = low risk of serious infection or sepsis	Serious infection or sepsis		
(complications rule)	Score from cancer diagnosis: AML = 11, others = 0. Relapsed disease = 11. Non-white patient = 8, Clinical presentation of serious unwell or toxic = 20. Total score <20 = low risk of any medical complication	Any medical complication		
Hann 1997	No rule described. Individual features = disease type, IV line, shock, duration of granulocytopenia and admission	Bacteraemia		

	temperature.			
Jones 1996	Low risk = ANC ≥200/mm^3, outpatient at onset, in remission	Bacteraemia	Clinical infection	
Klaassen, 2000	Low risk = AMC >100/mm^3; Mid-risk= AMC <100/mm^3, and temp ≤39°C; High-risk = AMC<100/mm^3, but temp >39°C	Bacteraemia	Significant bacterial infection (defined as any blood or urine culture positive for bacteria, interstitial or lobar consolidation on CXR, or unexpected death from infection before ANC recovery (>0.5 x 10^9/L))	
(validation set)	As original			
Lucas, 1996	Low risk = no chills, hypotension, or a requirement for fluid resuscitation at admission	Positive blood culture	ICU	Septic death
Mian 2010 [abstract only]	No clear rule – includes blood culture and CRP results	Admission to critical care		
Paganini 2007	Low risk <4. Mid-risk = 4. High risk = >4. Advanced stage of disease = 3 points, Comorbidity = 2 points, Bacteraemia = 1 point	Death		
(validation set)				
Rackoff, 1996	Low risk = AMC >100/mm^3; Mid-risk= AMC <100/mm^3, and temp ≤39°C; High-risk = AMC<100/mm^3, but temp >39°C	Bacteraemia	Clinical reason for admission	
(validation set)	Low risk = AMC >100/mm^3.			
Riikonen 1993	No rule described. No variables emerged as	Bacteraemia	Suspected sepsis/Fever of Unknown	Focal infection

	significant.		Origin	
Rojo, 2008	No rule described. No variables emerged as significant.	'Unfavourable outcome' - Compound of: haemodynamic instability, new focus if bacterial infection, 72h persistent fever, unresponsive CRP, or continuing +ve blood cultures 72 hours after treatment		
Rondinelli, 2006	Low risk = 2.5 to 5 points: Intermediate risk = 5.5 to 9 points: High risk = Greater than 9 points. 4.5 points for: clinical site of infection; 2.5 points for: **no** URTI; 2 points for: CVC; 1 point for: aged ≤5y, fever >38.5°C, Hb ≤7g/dL	'Serious infectious complication' – sepsis, shock, +ve blood cultures, infection-related death		
Santolaya, 2001	Low risk = 0 factors or isolated low plts or <7 days from chemotherapy. High risk = >1 risk factor, or isolated high CRP, hypotension or relapsed leukaemia. Risk factors: CRP ≥90mg/L, hypotension, relapsed leukaemia, plts ≤50 x10^9/L, chemotherapy within 7 days	Invasive bacterial infection (positive blood culture – 2 for CoNS, positive bacterial culture from usually sterile site, or sepsis syndrome and/or focal organ involvement and haemodynamic instability and severe malaise)	Death	
Tezcan 2006	No rule described. Significant association between hypotension, uncontrolled cancer and mortality. Duration of fever only independent risk factor for microbiologically documented infection.	Death	Clinically suspected infection	Microbiologically documented infection
West, 2004 (internally validated using bootstrap)	Very high risk = temp >39.5°C and CRT >3s; High risk = temp >39.5°C or CRT >3s; Low risk = neither	Requirement for critical care within 24 hours of presentation (fluid boluses ≥60ml/kg, inotropes or ventilation)		
Validation Studies				
Amman 2010	Klassen, Amman 2003, Santolaya, Alexander and Rondellini rules	Bacteraemia	Serious medical complication	Invasive bacterial infection
Baorto, 2001	Low risk = AMC >100/mm^3.	Bacteraemia	ICU/Death related to bacteraemia within 72 hours of	

			admission for FN	
Recalibration	Low risk = AMC >155/mm^3.			
Dommett, 2009	Low risk = Not AML/Burkitts/Induction ALL/Progressive-relapsed with marrow involvement ("Anticipated neutropenia <7days") and no significant comorbidity (defined as hypotension, tachypnea/hypoxia <94%, new CXR changes, altered mental status, severe mucositis, vomiting or abdominal pain, focal infection, other clinical reason for in-patient treatment) and fever responding at 48h	Bacteraemia		
Gala-Peralta, 2005	Low risk ≤2 of: <1yr, poor bone marrow response (plt <75, ANC <100/mm^3.),uncontrolled solid tumour or relapsed leukaemia, chemotherapy <10d earlier, rapid neutropenia, cardiac & renal dysfunction	Positive blood culture		
Macher, 2009	Klassen, Amman 2003, Santolaya, and Rondellini rules	Bacteraemia	Serious medical complicatio n	Invasive bacterial infection
Madsen, 2002	Low risk = AMC >100/mm^3; Mid-risk= AMC <100/mm^3, and temp ≤39°C; High-risk = AMC<100/mm^3, but temp >39°C	Positive blood culture		
Recalibration	Low risk = AMC >10/mm^3; Mid-risk= AMC <10/mm^3, and temp ≤39.5°C; High-risk = AMC<10/mm^3, but temp >39.5°C	Positive blood culture		
Meidema, 2010	Amman 2010 (SPOG rule)	Serious medical complication		
Petrelli, 1991	Low Risk: patients with solid tumors and lymphomas stage I-II.High Risk: patients with leukemias and lymphomas stage III-IV	Positive blood culture		
Santolaya, 2002	Low risk = 0 factors or isolated low plts or <7 days from	Invasive bacterial infection (positive blood culture – 2		

	chemotherapy. High risk = >1 risk factor, or isolated high CRP, hypotension or relapsed leukaemia. Risk factors: CRP ≥90mg/L, hypotension, relapsed leukaemia, plts ≤50x10^9, chemotherapy within 7 days	for CoNS, positive bacterial culture from usually sterile site, or sepsis syndrome and/or focal organ involvement and haemodynamic instability and severe malaise)		
Tezcan, 2006 validation	Low risk = AMC >100/mm^3	Death	Clinically suspected infection	Microbiol ogically document ed infection

* unless stated, the rule dichotomises into low and high risk groups

ALL = acute lymphoblastic leukaemia. AML = acute myeloid leukaemia. AMC = absolute monocyte count. CoNS = Coagulase-negative *Staphylococcus* CRP = C-reactive protein. CRT = capillary refill time. CXR = chest X-ray. Hb = haemoglobin. Plt = platelets.

QUADAS criteria

There was variation in the quality and applicability of the studies with respect to population under study (QUADAS questions "Was the spectrum of patients representative of the patients who will receive the test in practice?" and "Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?"). Thirteen definitions of febrile neutropenia were used, with twelve definitions of fever and four of neutropenia. However, all definitions are clinically similar, with any variation at the 'lowest risk' part of the spectrum of classification. In brief, most of the studies allowed patients who presented with febrile neutropenia following standard chemotherapy to be included. Some variations were found: eight studies excluded any inpatients, and examined only new episodes in outpatients. Ten studies excluded patients following stem cell transplants, and a further study stated no bone marrow transplant patients were included. One study examined only 'lower risk' patients, to further discriminate in this group [143]. The inclusion and exclusion criteria of the studies can be seen in detailed form in

Table 4.

Table 4: Participant characteristics by study, assessing applicability

Citation	Study location	Study years	Inclusion criteria	Exclusion criteria	Total number of patients	Total number of episodes	Age of patients
Adcock 1999	North Carolina, USA	1995 - 1996	ANC <1000cells/ mm^3, temperature ≥38°C, HIV-ve		33	88	Median 5y (range 1y to 18y)
Alexander 2002	Boston, USA	1994 - 1995	ANC ≤500/ mm^3, temperature >38.5°C. Outpatient status.	Post stem cell transplant	104	188	Mean 8.9y (SD 5.7y)
Ammann 2004	Berne, Switzerland	1993 - 2001	ANC ≤500cells/ mm^3 or ≤1000cells/ mm^3 and falling, axilliary temperature ≥38.5°C for ≥2h, or once ≥39°C	Patients with FN due to malignant bone marrow suppression, or following myeloablative therapy.	132	364	Not reported
(Amman 2004, subset used)	As above		As above	Patients with established severe bacterial infection.	111	285	Median 6.3y (interquartile range 3.2y to 12.1y)
Amman 2010	Multiple Swiss and German centres	2004-2007	ANC ≤500cells/ mm^3, temperature ≥38.5°C for ≥2h	Post stem cell transplant, aged <1y or >18y	206	423	Median 6.9y (interquartile range 3.8y to 11.6y)
Agyman 2011	Same as Amman 2010						
Badiei 2011	Iran, Asia	2008-2009	ANC <500cells/ mm^3, oral temperature ≥38.0°C ≥11h, or once >38.5°C	Inpatients, post stem cell transplant, newly diagnosed with malignancy	68	120	Mean 6.5 in "life-threatening infection" group, 5.6y in non~ group
Baorto 2001	St Louis, Dallas & Houston, USA	1990 - 1996	ANC <500cells/ mm^3, temperature ≥38°C, 12m or older	History of BMT	558	1171	Mean 8.0y (range 1y to 23y)
Delebarre 2010	France, Europe	2007-2009	Unclear	Unclear; abstract only	146	316	Mean 8y, (range:0.5y – 17.5y)
Dommett 2009	South-East England, UK	2004-2005	ANC ≤1000cells/ mm^3, temperature ≥38.0°C twice in <12h, or once	None	368	762	Median age 5 years 7 months (range 1 month to 17

			≥38.5°C				years 6 months).
Gala-Peralta 2005	Barcelona, Spain	2002	ANC ≤500/ mm^3, 'fever' (temperature not defined)		30	62	Mean 8.7y (range 1.2y to 14.7y)
Hakim 2008	Boston, USA	2004-2005	Outpatient, ANC <500cells/ mm^3, oral temperature ≥38.0°C ≥11h, or once >38.3°C	Inpatients, stem cell transplant recipients	332	332	Median 6y (range 2.4 months – 21.6 years)
Hann 1997	Multiple centres across Western Europe	1986 - 1994	ANC ≤1000cells/ mm^3, temperature ≥38.0°C twice in <12h, or once ≥38.5°C, in an EORTC trial		759	759	Median 8y
Jones 1996	North Carolina, USA	1987 - 1993	ANC <500cells/ mm^3, oral temperature ≥38.0°C ≥12h, or once >38.5°C	None reported, but 'none of the children were undergoing BMT'	127	276	Mean 8y (range 2m to 21y)
Klaassen 2000	Toronto, Canada	1996 - 1998	ANC <500cells/ mm^3 or ≤1000cells/ mm^3 and falling. Temperature >38.0°C ≥2 occasions in ≥12h, or once >38.5°C, or localised infection	New malignant diagnosis; bone marrow or stem-cell transplantation in preceding 6 months. Another medical condition that independently required inpatient observation. Interstitial infiltrate or lobar consolidation on chest x-ray	140	227	Median 6.8y (range 6m to 17y: derivation set)
(validation set)				Unclear	Unclear	136	Median 7.6y (range 1y to 18y: validation set)
Lucas 1996	New York, USA	1990 - 1992	ANC <500cells/ mm^3 or <1000cells/ mm^3 and falling, temperature ≥38.0°C ≥2 occasions in ≥12h, or once ≥38.5°C. Outpatient status	Received blood product transfusions within 6 hours or cytosine arabinoside within 2 days of presentation	161	509	Mean 9.2y (range 1y to 18y)

Madsen 2002	Indianapolis, USA	1997	New admissions 'coded' as 'fever of unknown origin' and ANC <500cells/ mm^3	History of BMT. AML. In-patient status	76	157	Mean 8y (range 2m to 18y)
Mian 2010	Arkansas, USA	Unclear	Unclear	Unclear	29	51	Rrange:1y – 21y)
Paganini 2007	Multiple centres across Argentina (derived 1 institution, validated in 7 further ones)	2000 - 2004	ANC <500cells/ mm^3 or <1000cells/ mm^3 and falling, temperature >38.0°C ≥2 occasions in 24h, or once >38.5°C	History of BMT	458	714	Mean 7y (range 1m to 17.9y: derivation set)
(Paganini 2007 validation set)					523	806	Mean 7.1y (range 1m to 17.5y: validation set)
Petrelli 1991	Camargo, Brazil	1988 - 1989	ANC ≤500cells/ mm^3, temperature ≥37.5°C ≥3 occasions in ≥24h, or once ≥38.0°C. Outpatient status	Fever associated with blood product transfusions or drugs	146	240	Mean 7.3y
Rackoff 1996	Indianapolis, USA	1994 - 1995	ANC <500cells/ mm^3, temperature >38.0°C ≥3 occasions in ≥24h, or once >38.5°C. Outpatient status		72	115	Range 9m to 18y: derivation set
(validation set)		1993				57	Validation set not reported
Riikonen 1993	Helsinki, Finland	1989 - 1990	ANC <200cells/ mm^3, temperature >38.0°C ≥2 occasions in ≥4h, or once >39.0°C	Antibiotics (excluding Septrin) in the preceding 3 weeks	46	91	Range 1y to 16y
Rojo 2008	Santiago, Chile	2003 - 2006	Episode of febrile neutropenia which was 'low risk' according to the PINDA criteria		33	47	Median 5.8y (1.1y to 15.7y)
Rondinelli 2006	San Paulo, Brazil	2000 - 2003	ANC <500cells/ mm^3 or ≤1000cells/ mm^3 and falling, temperature ≥37.8°C ≥3	Second or subsequent episode. Episodes in progressive disease (<6m from between completing therapy	283	283	Mean 5.2y

			occasions in ≥24h, or once >38.0°C. First episode per patient (new or relapsed disease)	and relapse). History of BMT			
Santolaya 2001	5 centres in Santiago, Chile	1996 - 1997	ANC ≤500cells/ mm^3, axilliary temperature ≥38.0°C ≥2 occasions 1h apart, or once ≥38.5°C	Not reported	257	447	Mean 7y (range 6m to 18y)
Santolaya 2002	6 centres in Santiago, Chile	1999 - 2000	ANC ≤500cells/ mm^3, axilliary temperature ≥38.0°C ≥2 occasions 1h apart, or once ≥38.5°C	Not reported	170	263	Mean 7y (range 7m to 17y)
Tezcan 2006	Antalya, Turkey	1996 - 2004	ANC <500cells/ mm^3 or <1000cells/ mm^3 and falling, axilliary temperature ≥38.0°C ≥2 occasions at 4h intervals, or once >38.3°C	Fever that occurred following transfusion of blood and blood products or administration of G-CSF	240	621	Median 6y (range 1m to 17y)
West 2004	California, USA	1994 - 1998	ANC <500cells/ mm^3 or <1000cells/ mm^3 and falling, axilliary temperature ≥38.0°C ≥3 occasions in 24h, or once ≥38.5°C, within 21d of chemotherapy	Induction, relapse and refractory disease. Collapse within 1h of admission	143	303	Mean 7.6y (SD 4.6y)

Age: y = years, m = months. ANC = absolute neutrophil count. HIV = human immunodeficiency virus. BMT = bone marrow transplant.

Other QUADAS criteria

Biases due to threats to independent outcome assessment were present in some studies (see Table 5). Note the three studies which used aspects of the outcome assessment in the decision rule [19, 23, 132, 153]. In Alexander [23] the outcome of 'significant medical

complication' included 'hypotension and severe mucositis', as did the rule describing high-risk, making these features tautolgous and artificially inflating the sensitivity of the 'predictive' rule. Hypotension was found in 55% (5/9) patients and severe mucositis in 12% (1/9) patients with a 'significant medical complication'. In Ammann's studies [19, 153], the outcome of severe bacterial infections included episodes where CRP > 150 mg/dL without other microbiological confirmation, and the rule included CRP ≤50 mg/dL. CRP > 150 mg/dL was found in 50% (53/106) of episodes, but is unclear how many of these individuals had a further reason to be classified as suffering 'severe bacterial infection'. The study of Delebarre [132] is only an abstract and the degree of incorporation bias cannot be accurately assessed.

Table 5: Further informative QUADAS measures; CDR reveiw

Citation	**Study design**	**Verification procedure biases**			**Interpretation biases**	
		Partial verification	*Differential verification*	*Incorporation bias*	*Review bias*	*Review bias*
	Prospective or retrospective?	Did the whole sample or a random selection of the sample, receive adequate outcome assessment?	Did patients receive the same outcome assessment regardless of the CDR result?	Was the outcome assessment independent of the CDR?	Were the CDR results interpreted without knowledge of the results of the outcome assessment?	Were the outcome assessment results interpreted without knowledge of the results of the CDR?
Adcock 1999	Retrospective	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Unclear – not blinded	Unclear - not stated
Alexander 2002	Retrospective	Not stated	Not stated	No – serious medical complication included hypotension & mucositis (which are part of the CDR)	Yes	Yes
Ammann, 2003 & 2004	Retrospective	Yes	Yes, although some tests were undertaken if	No – one variable from CDR was in outcome	Unclear - not blinded	Unclear - not blinded

			clinically indicated	assessment (C-reactive protein level, although cutpoint differed; 50mg/L in CDR vs 150mg/L in outcome assessment).		
Amman 2010 and Agyman 2011	Prospective	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Yes	Yes
Badei 2011	Prospective	Yes	Yes	Yes	Yes	Yes
Baorto, 2001	Retrospective	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Yes	Unclear - not blinded
Delebarre 2010	Prospective	No. Signs of local infection or severe sepsis were in rule & outcome.	Unclear	Yes	Unclear - not stated	Unclear - not stated
Dommett 2009	Prospective	No. Signs of local infection or severe sepsis were in rule & outcome.	Yes, although some tests were undertaken if clinically indicated	Yes	Yes	Unclear – not blinded
Gala-Peralta, 2005	Retrospective	Yes	Yes	Yes	Unclear – not blinded	Unclear - not blinded
Hann 1997	Retrospective (RCT trial data)	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Yes	Unclear - not blinded
Jones 1996	Prospective	Yes	Yes, although some tests were undertaken if clinically	Yes	Yes	Unclear - not stated

			indicated			
Klaassen, 2000	Prospective	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Yes	Yes - blinded
Lucas, 1996	Retrospective	Yes	Yes	Yes	Yes	Unclear - not blinded
Madsen, 2002	Retrospective electronic record	Yes	Yes	Yes	Yes	Unclear. Review was blinded, but of unblinded case notes
Macher 2011	Retrospective	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Unclear - not stated	Unclear - not stated
Mian 2010	Prospective	Unclear	Unclear	Unclear	Unclear - not stated	Unclear - not stated
Paganini 2007	Prospective	Yes	Yes	Yes	Yes	No - but mortality
Petrelli, 1991	Prospective	Unclear	Yes	Yes	Yes	No
Rackoff, 1996	Prospective (Derive) and Retrospective (Validate)	Yes	Yes	Yes	Yes (Derive) and Unclear – not blinded (Validate)	Unclear - not blinded
Riikonen 1993	Prospective	Yes	Yes	Yes	Yes	No
Rojo, 2008	Retrospective	Yes	Yes	Yes	Unclear - not blinded	Unclear - not blinded
Rondinelli, 2006	Retrospective	Unclear	Yes	Yes	Yes	Yes
Santolaya, 2001	Prospective	Yes	Yes, although some tests were undertaken if clinically indicated	Yes	Yes	Yes - blinded
Santolaya,	Prospective	Yes	Yes, although some tests	Yes	Yes	Yes - blinded

2002			were undertaken if clinically indicated			
Tezcan 2006	Retrospective	Yes	Yes	Yes	Yes	Unclear - not stated
West, 2004	Retrospective	Yes	Yes	Yes	Unclear - not blinded	Unclear - not blinded

Quality of CDR derivation

The 22 reports of attempts to derive a CDR varied in population, outcomes chosen and the number of those outcomes as discussed above. They also varied in the variables assessed, model-building technique, the way that missing data were reported and handled, the way that multiple-episode data were used and in the use and categorisation of continuous and categorical variables. All of these features may have influenced the CDRs produced and provide some explanation of the differences between them.

The number of events per variable considered is generally important in producing replicable studies. Most studies building a CDR used a large number of variables (median 16, range 2 to 39) and had a small number of events (median 41, range 4 to 179) with 76% (16/21) studies having less than ten events per variable under consideration. No study had more than 14 events per variable (see Appendix 6).

The technique used to build the model also varied. Almost all were built using multivariable regression (see Appendix 7 for details). Five models from four publications used alternative approaches. Two models did not use multivariable analysis [23, 133], two used CART (classification and regression tree) techniques [131]. One model [79] was offered alongside a logistic regression, and came to different conclusions from the same dataset. No model building study clearly assessed if relationships between the outcome and the explanatory variables could hold non-linear functional forms, and with the exception of one study [144, 147], nor did they clearly examine co-linearity (the issue of multiple variables being highly correlated as described in chapter 3).

Continuous variables, such as age, blood pressure and absolute monocyte count, were used in the model derivation by seven studies as continuous data and they went on separately to create categories (three by recursive partitioning [141, 144, 147], two by ROC analysis[139, 146],and unstated in the remaining two which were reported as abstracts only [132, 145]). Five studies used some variables in continuous form and making others ordinal [5, 69, 137, 140, 142] In these studies, four did not state clearly how the cutpoints had been determined, one stated 'clinical judgment' was used [5]. A further eight used only categorised variables. Only one of these described clearly a literature-based choice of cutpoints [26]. Of the derivation studies, three did not use purely categorical variables [137, 141-142], 15 did not clearly state the reasons for defining their cutpoints, 3 used 'clinical judgement' [5, 26, 146] and one study stated they had used 'trend to significance' from bivariable analysis [133] (see Appendix 7).

Multiple episodes in individual patients were treated primarily as if they came from unconnected individuals in twelve studies (see Appendix 7). A further three studies performed a secondary analysis which looked at only the 'first included case' and found 'no significant differences' [79, 139, 153]. The first case approach is not necessarily the patient's first-ever presentation with a febrile neutropenic episode, but is the first recorded during the study in question. Other approaches to address the issue of multiple episodes included the use of only first episode data (four studies) [5, 134, 138, 146] and extended modelling techniques that try to account for the clustering; a generalised estimating equation (four studies) [25, 141, 144, 147] or generalised linear mixed models (two studies) [26, 140].

The issue of missing data was described in only eight of the derivation studies [26, 79, 134, 141, 144, 146-147, 153], six of them used a form of complete case analysis (after exclusion of potential variables where <90% of cases had collected the information in two linked studies) and two linked studies used imputation [144, 147]. No study details an assessment of the type of missingness of the data, although seven of the eight who commented on missing data analysis also described the extent of the problem. The remaining studies neither clearly defined the quantity of missing data nor how it was addressed (details in Appendix 7).

Clinical Decision Rule performance

The CDRs designed to predict general infectious complications have diverse test performance. This heterogeneity has largely been explored using a narrative structure, as pooling across all the studies was not possible due to the varied rules, outcomes and populations studied. Initial hypotheses to explain the differences included: the design of the study (derivative better than validating), the population (both geographical, where developing-world studies would be different than developed-world, and case-mix, where less success was predicted from populations where higher-risk cases had been systematically removed), the complexity of the rule (more complex rules would be better) and outcomes chosen (the rules differing between outcomes, without a clear *a priori* hypothesis of which outcomes may be easier to detect). These were examined by analysis of the tabulated CDR performance data (Appendix 8) and graphically with plots of sensitivity and specificity (Figure 12, Figure 13, Figure 14, Figure 15, Figure 16)

Examining potential reasons for the differences found that the derivation studies as expected, had better accuracy than validation studies. For example, for the outcome "serious/invasive infection" the median LR- = 0.06 (range 0 to 0.33) in derivation studies compared with median LR- = 0.35 (range 0.11 to 2.28) for validation studies. This was less marked but similar for "bacteraemia", the median LR- = 0.21 (range 0 to 0.72) in derivation studies compared with median LR- = 0.33 (range 0 to 0.74). The results of a pooled analysis of the 'Rackoff' and 'Alexander' rules (see later for details) supported this previously noted overestimation of the rule performance. The choice of outcome also appears to alter the rule performance, but the different number of studies and the heterogeneity of rules and populations make this difficult to examine clearly.

Those CDRs developed in a population where the highest risk patients were excluded (e.g. bone marrow transplant recipients) did not seem to be particularly better or worse than the rules developed without these exclusions. The few rules derived in South America appeared to be from higher quality studies [22] but not significantly different in terms of performance in their original setting than other rules, but did not show geographical transportability (see Figure 14 and Figure 16, explored below). The issue of geographical and temporal replication has been infrequently examined in these studies.

Examination of the detailed content of all the proposed rules shows they address four major domains (Appendix 9. Individual factors used in clinical prediction rules). The first can be considered stable patient-related factors, including age and the underlying disease. The second group reflects treatment; the presence of a central venous catheter and the type of or duration since last chemotherapy. The third group reflects episode specific clinical features, such as maximum temperature, the patient's blood pressure or clinical features of infection. The final group contains episode specific laboratory test values. These are various markers of bone marrow function where, excepting [24], each rule uses a single item which reflects one of the three major cellular components: haemoglobin, platelets, total white cells or subset of neutrophils or monocytes, and serum inflammatory markers (C-reactive protein). The results of the detailed systematic review of the predictive value of such serum inflammatory markers follows in the later part of this chapter.

The complexity of the rule e.g the Rackoff rule of AMC > 100 cells/mm^3 compared with the five items of the PINDA rule does not seem to have importantly improved their predictive value, though this is difficult to judge effectively as the rules have not been subject to the most extensive validation to enable such comparisons to be undertaken.

When addressing the nature of individual factors found to be significantly associated with adverse outcomes there are many similarities (Appendix 10, tables a-c, subdivided by outcome class). In predicting bacteramia, the disease state (induction/remission or bone marrow involvement) appears important. Age does not appear a strong linear predictive factor. The presence of a central line and use of higher-intensity chemotherapies may be. Episode related factors of importance include outpatient status, other co-morbidities, the presence of respiratory distress (including proven pneumonia), hypotension or shock, mucositis and maximum temperature. A clinical site of infection is probably not a predictive factor. Blood tests with importance include platelet and absolute monocyte count, and potentially higher levels of haemoglobin; neutrophil count appears unimportant.

To predict significant infection (rather than bacteramia alone), various other factors are added. The age of the patient achieves a significant linear prediction ability and a clinical site of infection is important. Outpatient status has not been assessed in this setting, so comment is not possible. Low haemoglobin is also associated with adverse outcomes in some studies [137], and the opposite in others, and once again neutrophil counts are not important.

The few studies that address ICU and death find that the age and disease state remain important, as do clinical assessments of circulatory and respiratory compromise. Higher temperatures remain highly predictive, and neutrophil counts appear unrelated to these outcomes, where monocyte and platelet counts retain some value.

Quantitative meta-analysis

The results of combining studies which used identical clinical decision rules was undertaken in three cases in the original review and supplemented with a further four in the update review. The three-level Rackoff rule [141] to examine bacteraemia (a total of 7 data sets: [12, 26, 69, 136, 141] was not updated, nor was Paganini's rule to predict mortality [137] (with one derivation and one validation set). There was sufficient data to update a meta-analysis of the Santolaya (PINDA) rule for serious infectious complications [139] and additional data to undertake a meta-analysis of the validity of the Alexander, Amman 2004, SPOG and Klaassen rules.

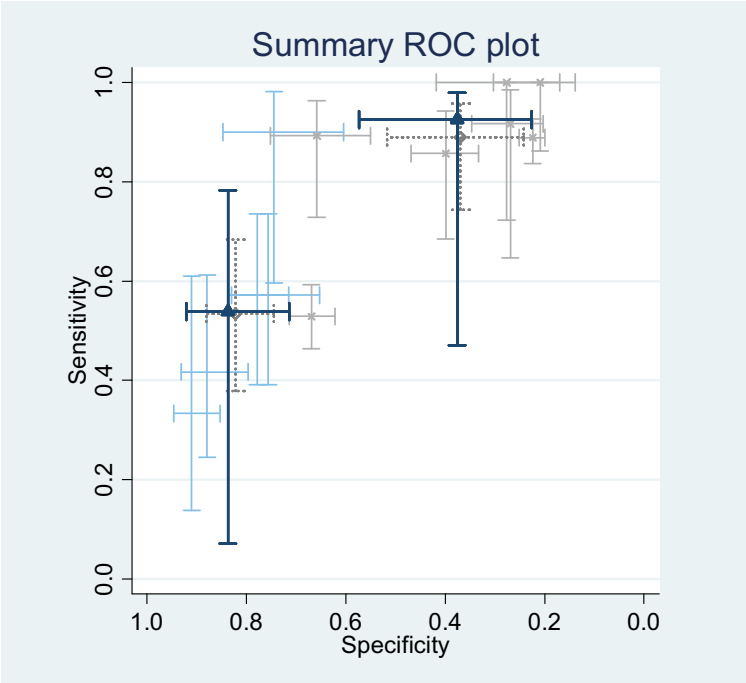
The results of the Rackoff rule combined analysis show a moderate ability to discriminate between three groups of individuals at low, moderate and high risk of bacteraemia. Exclusion of an outlier ([69]; see Figure 2) led to a more Normal distribution of the posterior probability plots, in keeping with it being qualitatively different than the other studies. A further sensitivity analysis excluding the initial rule derivation study demonstrated reduced discriminatory ability (see Table 6.)

The most accurate estimate of predictive accuracy is likely to come from the analysis of 5 data sets (excluding the derivation and outlier); LR [low] = 0.26 (95% CrI 0.08 to 0.72) , LR

[medium] = 0.72 (95% CrI 0.14 to 2.15), LR [high] = 3.11 (95% CrI 1.25 to 8.01). (See Figure 3.)

Table 6: Combined analysis of Rackoff rule data with alternative exclusions

	LR[High]	LR[Middle]	LR[Low]
All Included	3.03 (95% CrI 1.28 to 6.5)	0.81 (95% CrI 0.14 to 2.15)	0.27 (95% CrI 0.06 to 0.79)
Excluding Tezcan	3.2 (95% CrI 1.49 to 6.88)	0.76 (95% CrI 0.3 to 1.73)	0.22 (95% CrI 0.06 to 0.6)
Excluding Derivation	2.9 (95% CrI 1 to 7.2)	0.77 (95% CrI -0.08 to 2.82)	0.32 (95% CrI 0.08 to 0.93)
Excluding Derivation and Tezcan	3.11 (95% CrI 1.25 to 8.01)	0.72 (95% CrI 0.14 to 2.15)	0.26 (95% CrI 0.08 to 0.72)



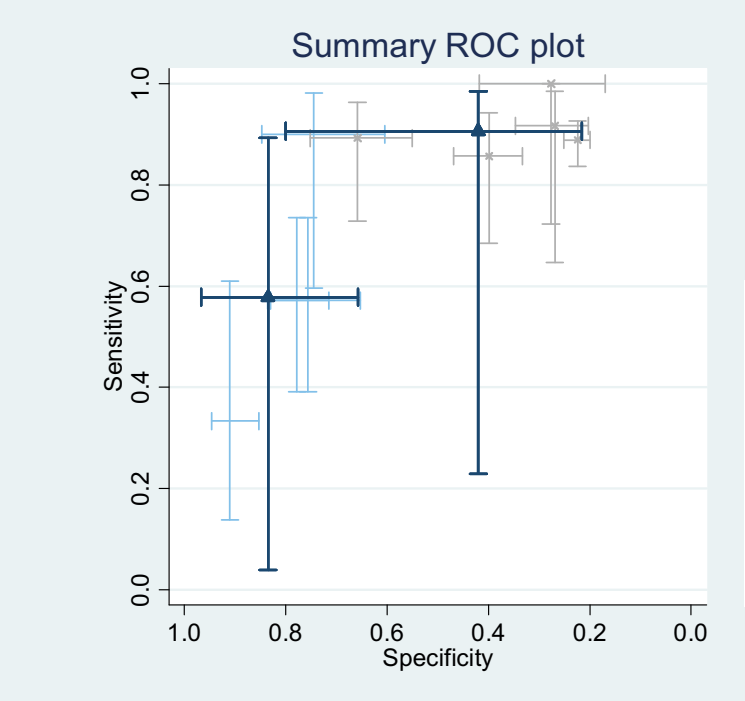
Key for Figure 12

O & blue = low vs. medium-high studies
X & grey = low/medium vs. high studies

Dark blue = MCMC summary estimates
LR [low] = 0.20 (95% CrI 0.052 to 1.54)
LR [medium] = 0.83 (95% CrI 0.31 to 1.29)
LR [high] = 3.28 (95% CrI 0.40 to 7.32)

Grey dotted = Bivariate summary estimates
LR [low] = 0.30 (95% CI 0.14 to 0.63)
LR [medium] = *not estimable*
LR [high] = 3.01 (95% CI 2.26 to 4.00)

Figure 12: Pooled and individual results of the 'Rackoff' model studies

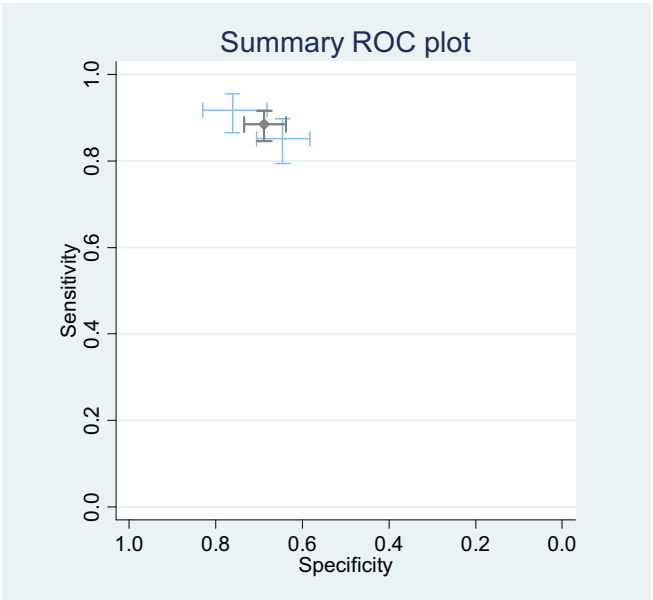


Key for Figure 13

Dark blue = MCMC summary estimates
LR [low] = 0.26 (95% CrI 0.08 to 0.72)
LR [medium] = 0.72 (95% CrI 0.14 to 2.15),
LR [high] = 3.11 (95% CrI 1.25 to 8.01)

Figure 13: Pooled and individual results of the 'Rackoff' model studies excluding derivation and Tezcan

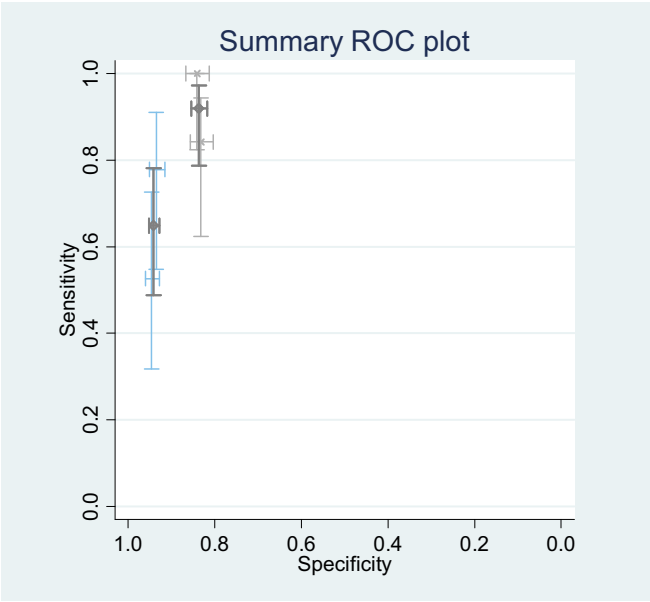
If the meta-analysis is restricted to studies in Chile (as with the original review), the PINDA group (Santolaya) model shows a similar ability to differentiate between low- and high- risks groups when considering a wider definition of 'serious infection', (LR [low] = 0.17 (95% CI 0.12 to 0.23) LR [high] = 2.87 (95% CI 2.43 to 3.38)). The Paganini model demonstrates an ability to quite accurately predict mortality LR [low] = 0.11 (95% CI 0.04 to 0.30), LR [medium] = *not estimable,* LR [high] = 11.0 (95% CI 8.08 to 15.0). However, in undertaking this geographically restricted analysis, the results may be falsely reassuring.



Key for Figure 14

Random effects model summary estimates
LR [low] = 0.17 (95% CI 0.12 to 0.23)
LR [high] = 2.87 (95% CI 2.43 to 3.38)

Figure 14: Pooled and individual results of the 'PINDA' model studies from South America



Key for Figure 15

O & blue = low vs. medium-high studies:
X & grey = low/medium vs high studies

Grey = Random effects model summary estimates
LR [low] = 0.11 (95% CI 0.04 to 0.30)
LR [medium] = *not estimable*
LR [high] = 11.0 (95% CI 8.08 to 15.0)

Figure 15: Pooled and individual results of the 'Paganini' model studies

When the PINDA rule is used in datasets from Europe, there is a marked inconsistency and lack of repeatability in the results (see Figure 16) . This apparent lack of geographical transportability is highly important when deciding to practically use a decision rule.

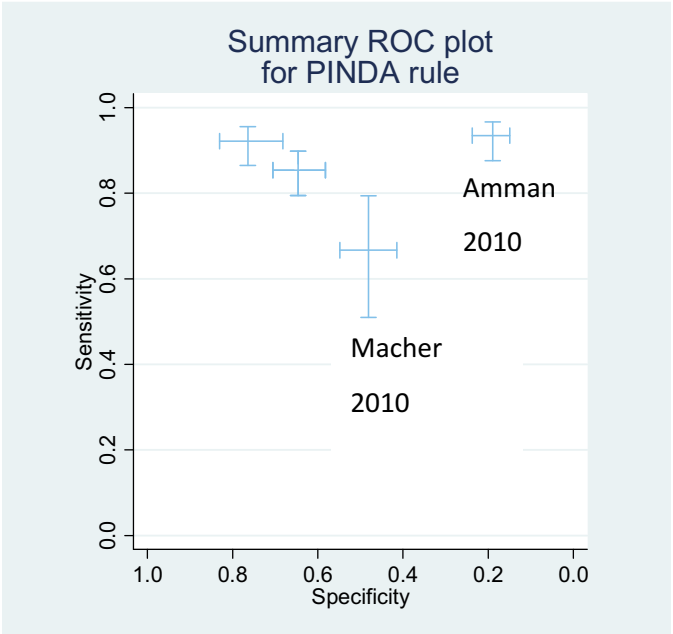


Figure 16: Pooled and individual results of the 'PINDA' model studies from Europe and South America

As no further studies allow evaluation of the Paganini rule, no further comment can be made.

The update review allowed for the analysis of five further rules: Klaassen, Amman 2003 (in which meta-analysis was undertaken), and the Rondellini, SPOG and Alexander rules (where data were not suitable for meta-analysis).

The "Klaassen" rule is based on a single feature: an absolute monocyte count of greater than 100/mm^3 to predict patients less likely to have significant infection. Data were pooled from four studies from the original review [12, 26, 136, 141] and two new sources [130, 144]. The results of this analysis give a pooled average sensitivity of 88% (95% CI 84 to 91%) and specificity of 36% (95% CI 27 to 45%), see Figure 17.

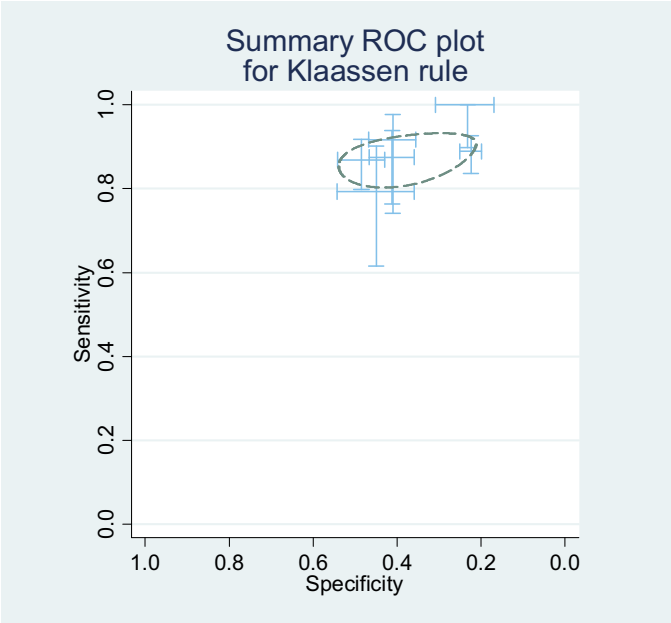
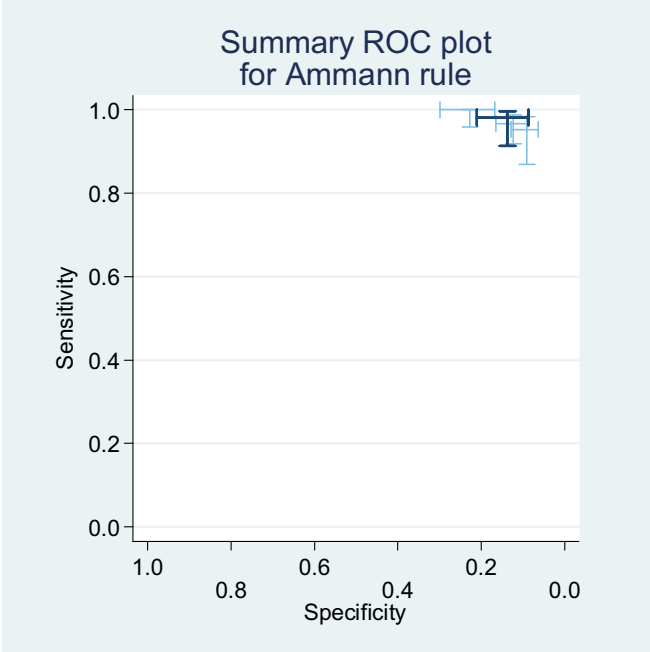


Figure 17: Pooled and individual results of the 'Klaassen' model studies

Key for Figure 17

Random effects model summary estimates
LR [low] = 0.33 (95% CI 0.20 to 0.59)
LR [high] = 1.38 (95% CI 1.15 to 1.65)

The "Ammann" rule was assessed in the three studies providing data to test this rule to detect serious consequences of FNP [79, 130, 144]. The combined average sensitivity was 98% (95%CI 91 to 99%) but pooled average specificity only 13% (95% CI 8% to 21%), see Figure 18.



Key for Figure 18

Random effects model summary estimates
LR [low] = 0.15 (95% CI 0.04 to 1.12)
LR [high] = 1.12 (95% CI 0.98 to 1.25)

Figure 18: Pooled and individual results of the 'Amman' model studies

The “Alexander” rule again examined adverse clinical consequences. This rule was assessed by three studies [134, 144, 154]. There was marked heterogeneity in the results of these three studies (see Figure 19). When used at reassessment after 48hrs of hospitalisation, there was marked improvement in the discriminatory ability of the rule [154] (sensitivity = 100%, specificity = 39%). The derivation was undertaken in North America, the evaluations in Europe.

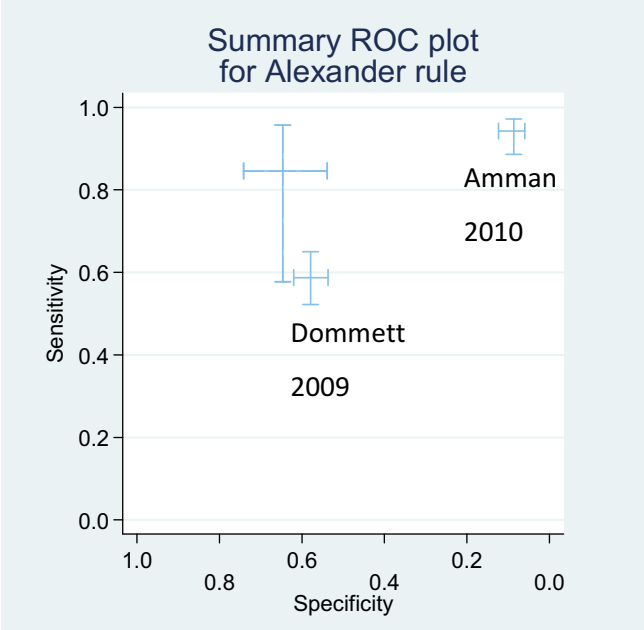


Figure 19: Individual results of the 'Alexander' model studies

The rule of Rondellini [138] describes a low-risk group for adverse clinical consequences, and was assessed in two validation datasets. These demonstrated a sensitivity of 84%[144] and 62%[130] and both estimated specificity at 43%.

The SPOG2003 was only evaluated in one study and varies from the other systems in that it is applied after 8-24 hours of hospitalisation. This model was shown to have a sensitivity 92% and specificity of 45%[144]. A validation of this model demonstrated poorer sensitivity (82%) and slightly better specificity (57%) [155], using data from a similar region (both European) but in countries using slightly different primary treatment regimes.

Biomarkers studies

This section of the Chapter addresses studies which have examined the role of serum biomarkers in the prediction of infectious complications. The assessment of such biomarkers would be ideally undertaken as additional to previously gathered clinical data, but the studies undertaken did not commonly appear to have this design. Accordingly, a decision was made to review them separately, and update that review independently.

Study inclusion and exclusion

Figure 20 describes the flow of candidate and eligible articles though the original review process. 368 articles were identified from electronic searches, of which 72 articles were identified for detailed examination. Seven further articles were identified from examining systematic reviews and the bibliographies of included studies. From this, 27 articles reporting on 25 studies were eligible for inclusion in the review, with 26 articles providing outcome data. Of these, 13 could be included in the quantitative synthesis.

The update review found 13 further studies, of which quantitative data were included from 12 studies (Figure 21).

In total 38 studies were included in the review, of which 37 provided quantitative data and 22 studies were included in the meta-analyses.

Figure 20: Flow diagram of study selection process; original biomarkers review

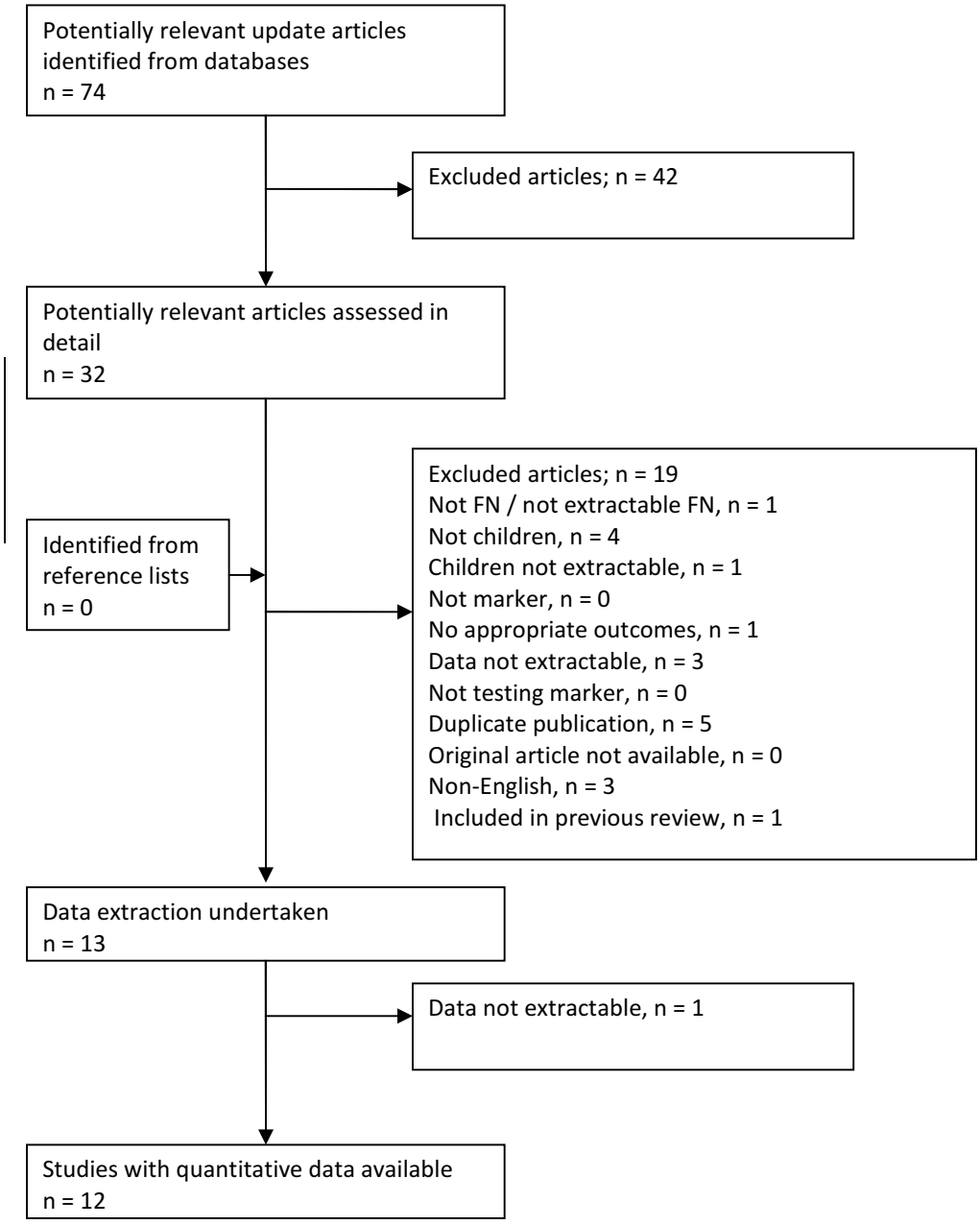


Figure 21: Flow diagram of study selection process; update biomarkers review

Evaluation of Biomarkers studies

The studies included a total of 3071 patients (each included between 19 and 278 individuals, median 56) and over 5169 episodes (between26 and 566, median 94, where stated). Twenty four different markers of inflammation or infection were assessed (see Table 7). The mean age of children ranged from 5.7 to 10 years (where age stated). The studies were undertaken between 1989 and 2009 (where stated) in Europe, North America, South America, North Africa and Asia and Australasia.

Table 7: Summary of biomarkers studied.

Marker	Total studies
CRP	29
IL6	16
IL8	15
PCT	14
IL10	2
TNF-alpha	2
IL1	2
IL5	2
IL2-R	2
IL12	1
IL10	1
MCP	1
ESR	1
tADA	1
SAA	1
IFN-gamma	1
T-reg	1
sTNFRII	1
sTREM	1
Derivative of rO2 metabolites	1
Biological antioxidant potential	1
LDH	1
Glucose	1
Blood urea nitrogen	1

Outcomes examined

The studies reported diverse outcomes including bacteraemia, fungal infection, gram negative and gram positive bacteraemia, significant/documented bacterial infection, systemic inflammatory response syndrome (SIRS), sepsis, intensive care unit (ICU) admission, death, and prolonged (>5days) hospital stay (Table 8).

Table 8: Markers and endpoints in each included biomarkers study.

Citation	Markers assessed	Endpoint #1	Endpoint #2	Endpoint #3	Comments on endpoints
Asturias 2010	CRP	Fever	Bacteraemia		
Ammann 2003	CRP	Significant bacterial infection			Defined as death from bacterial infection, a positive culture of normally sterile body fluids, radiologically proven pneumonia, clinically unequivocal diagnosis of a bacterial infection, or a serum C-reactive protein level (CRP) above 150 mg/L as an indirect sign suggesting severe bacterial infection
Avbratha 2009	CRP	Bacteraemia or clinically documented infection			
Barnes 2002	PCT	Length of stay			Stay of <5d or ≥5d
de Bont 1999	CRP, IL6, IL8	Bacteraemia			
Cost 2011	IL8, IL5	Bacteremia & clinical sepsis			
Delebarre 2011	PCT	Severe infection			Defined by bacteraemia, severe bacterial infection, invasive fungal infection or probable infection
Diepold 2008	IL6, IL8, CRP	Bacteraemia	Fever lasting ≥5d but culture -ve		
Dylewska 2005 a & b	PCT, CRP	Bacteraemia	Clinically defined infections (UTI, neurological, GI or respiratory)	Microbiologically defined other infection	FUO was the default category

El-Maghraby 2007	CRP, IL8, MCP	Bacteraemia or clinically documented infection			
Hatzistiliano u 2007	PCT, CRP	Microbiologic al or clinically documented infection (excludes viral)			
Hatzistiliano u 2010	PCT, CRP, TNF-alpha, IL8, IL1b, sTNFRII	Bacterial infection	Viral infection or PUO		
Heney 1992	CRP, IL6	Bacteraemia			
Hitoglou-Hatzi 2005	CRP, PCT, tADA	Significant bacterial infection			
Hodge 2006	IL5, IL8, IL12, CRP	Positive blood culture			
Hodge 2011	IL2, TNF-alpha, TNF-gamma, T-reg cells	Positive blood culture			
Katz 1992	CRP	Clinically or bacteriologica lly documented infection	Septicemia (+ve blood cultures & unwell clinical appearance)		
Kharaya 2010	IL6, PCT	Bacteraemia			
Kitanovski 2006	CRP, PCT, IL6	Bacteremia & clinical sepsis	Clinically or microbiologica lly documented local infection		
Lehrnbecher 1999	CRP, IL8, IL6	Clinically documented infection	Fungal infection	Bacteraemia (gram-type)	FUO was the default category
Lehrnbecher 2004	IL6, IL8	Significant bacterial infection			Defined as bacteraemia, localised infection or pneumonia

Lodhal 2011	PCT, CRP	Bacteraemia			Split into gram-positive and gram-negative infections
Mian 2009	CRP, IL6, IL2, IL10, IL8, TNF-alpha	Intensive care admission			
Miedema 2011	CRP, IL8, PCT, sTREM	Proven or suspected bacterial infection			Defined as documented bacteria from sterile site or clinically documented infection or clinical sepsis
Nishikwa 2010	dROM, Biological antioxidant potential, CRP	Systemic inflammatory response syndrome			
Reitman 2010	PCT	Bacteraemia			
Richardson 2009	CRP	Bacteraemia			
Riikonen 1992	IL1, IL6, TNF-alpha, SAA	Bacteraemia	suspected sepsis	Focal infection	"No infection" was the default category
Riikonen 1993	CRP	Bacteraemia	suspected sepsis	Focal infection	"No infection" was the default category
	CRP	Documented bacterial infection	Probable bacterial infection	Viral infection	Documented bacterial infection defined as bacteraemia (two sets positive for commensals) or sterile site infection; Probable bacterial infection defined as cultures negative but severe medical course e.g. purulent gingivostomatitis, CXR+; FUO was the default category
Santolaya 2001	CRP	Invasive bacterial infection			Defined as positive blood cultures – 2 for CoNS, positive bacterial culture from usually sterile site, or sepsis syndrome and/or focal organ involvement and haemodynamic instability and severe malaise
Santolaya 2002	CRP	Invasive bacterial infection			Defined as positive blood cultures – 2 for CoNS, positive bacterial culture from usually sterile site, or sepsis syndrome and/or focal organ involvement and haemodynamic instability and severe malaise

Santolaya 2007	CRP	Death			
Santolaya 2008	CRP, IL8, PCT, Glucose, Blood Urea Nitrogen	Severe sepsis			Defined as sepsis + respiratory or cardiac compromise, or + 2 other-organ compromise) not apparent during the first 24h of admission
Secmeer 2007	CRP, PCT, ESR	Bacteraemia	Documented bacterial infection (microbiologically or clinically)	Duration of fever	
Soker 2001	IL2-R, IL6, IL8, TNF-alpha, IL1	Bacteraemia			
Spasova 2005	CRP, IL6, IL8, IL10	Bacteraemia	Microbiologically or clinically proven local infections without bacteraemia		
Stryjewski 2005	PCT, IL6, IL8	Sepsis	Septic shock		Sepsis (positive culture - two consecutive +ve if CoNS, fever, tachycardia, or tachypnoea); septic shock defined as sepsis plus need for inotropes/vasopressors

QUADAS criteria

Analysis of the study quality according to modified QUADAS criteria revealed few informative items. The quality was on the whole good (Appendix 11). The worst identified flaw was in Amman study where there was potential contamination of the reference standard with the diagnostic test (the outcome included CRP >150 mg/dl while the predictive test included CRP). One short report did not detail the exact outcome used [156]. The major deficiencies in most studies were failure to report whether the marker test and outcomes were interpreted blind to each other; only three of the studies by Santolaya clearly documented that this was the case. Detailed analysis of the criteria presented in the published abstracts from conferences in which results were reported was very difficult. Applicability to a general clinical population was fair for those studies that presented information about the included population (Table 9,) although most studies failed to clearly

describe their selection criteria. The Santolaya 2008 [157] study was specifically designed to examine only the high-risk (by CDR) group, and their data may be considered as belonging to an importantly clinically distinct population.

Table 9: Participant characteristics by biomarkers study

Citation	Study location	Study years	Inclusion criteria	Exclusion criteria	Number of patients	Number of episodes	Average age of patients
Ammann 2003	Bern, Switzerland	1993 to 2001	ANC ≤500cells/ mm^3 or ≤1000cells/ mm^3 and falling, axillary temperature ≥38.5°C for ≥2h, or once ≥39°C	Established severe bacterial infection at presentation, or episodes of FN due to bone marrow involvement by the disease itself, i.e. at the time of diagnosis or following a myeloablative therapy.	111	285	Median age at first episode = 6.3y (interquartile range 3.2y to 12.1y)
Asturias 2010	Guatemal City, Guatemala	April 8 to October 15 2008	Children with cancer, age <18ys and hospitalised with fever and neutropenia	Hospitalisation <48hrs, antibiotic therapy 7 days before admission, prior bone marrow transplant	88	102	Mean age: 6.5 yrs; SD: +/- 4.4 yrs; range: 8 months to 18 yrs
Avabratha 2009	Mangalore, India	Not stated	Children aged 1-15yrs with malignancy and febrile neutropenia	Liver disease	33	50	Mean age 6.9 y
Barnes 2002	Melbourne, Australia	Not stated	"Febrile neutropenia" not further specified	Not stated	37	39	Not stated
Cost 2011	Dallas, Texas, United States	March 2010 to Dec 2010	Paediatric oncology patients hospitalised with	Not stated	120	120	Not stated

			febrile neutropenia				
de Bont 1999	Groningen, The Netherlands	1998	Chemotherapy related neutropenia (granulocytes <500cells/ mm^3 or leucocytes <1000cells/ mm^3) and ≥38.0°C for 6h, or once ≥38.5°C	On antibiotics or post BMT/stem cell transplant	19	72	Not stated , all <16y
Diepold 2008	Freiburg, Germany	Not stated	ANC ≤500cells/ mm^3, temperature ≥38.0°C for >1h, or once >38.5°C	Not stated	69	123	Median 7y 8m (range 1m to 20y)
Dylewska 2005 a & b	Bydgoszcz, Poland	Not stated	ANC ≤1000cells/ mm^3, temperature ≥38.0°C (once, axilliary)	Not stated	66	108	Mean 9.6y (Range 2y to 20y)
El-Maghraby 2007	Cairo, Egypt	2004-2005	ANC ≤500cells/ mm^3, temperature ≥38.0°C twice in <6h, or once ≥38.5°C	Systemic antibiotics (except Septrin) wihin previous week	76	85	Mean 7.1y (range 1.5y to 18y)
Hatzistilianou 2007	Not stated	Not stated	ALL patients only. ANC ≤500cells/ mm^3 or leukocytes ≤1000cells/ mm^3, temperature ≥38.0°C for 6h, or once ≥38.5°C	Not stated	29	94	Mean 5.8y (SD 2.9y) range 1y to 14y
Hatzistilianou 2010	Greece/Europe	Not stated	Children with acute leukemia and febrile neutropenia Study also looked at febrile non-neutropenic and afebrile non-neutropenic children	Not stated	0	0	Group A - mean 5.8; range 1-14y; Group B - not stated
Heney 1992	Leeds, UK	Not stated	Temperature ≥38.0°C twice in <24h, or once	Not stated	33	47	Mean 7y (range 7m to 15y)

			≥38.5°C				
Hitoglou-Hatzi 2005	Thessaloniki, Greece	Not stated	ANC ≤500cells/ mm^3 or leukocytes ≤1000cells/ mm^3, temperature ≥38.0°C for 6h, or once ≥38.5°C	Not stated	67	Not stated	Mean 6.4y (range 1y to 14y)
Hodge 2006	North Adelaide, Australia	Unclear	ANC ≤1000cells/ mm^3, temperature ≥38.0°C (sustained) or once ≥38.5°C	Not stated	31	31	Not stated
Hodge 2011	Australia	Not stated	Paediatric oncology patients with febrile neutropenia	Not stated	27	26	Not stated
Kharya 2010	Dehli/India	Not stated	Children with febrile neutropenia	Not stated	*	129	Not stated
Katz 1992	Dallas, USA	November 1989 to June 1990	Outpatients only. ANC ≤500cells/ mm^3, temperature ≥38.0°C for 6h, or once ≥38.5°C	Already on antibiotics (except Septrin)	74	122	Mean 6.3y (range 2m to 17y)
Kitanovski 2006	Ljubljana, Slovenia	Not stated	ANC ≤500cells/mm3 or ≤1000cells/mm3 and falling,tympanic temperature ≥38.0°C for 6h, or once ≥38.5°C	Not stated	32	68	Median 7.6y (range 1y to 18y)
Lehrnbecher 1999	Wurzberg, Germany	Unclear	ANC ≤500cells/ mm^3 or within 72hrs of chemotherapy and falling, temperature ≥38.0°C twice in <4h, or once ≥38.5°C	Febrile >24h before admission and antibiotics (except Septrin) within previous 72h	56	121	mean 8y (range 3m to 20y)
Lehrnbecher 2004	Bonn, Frankfurt &	Not stated	ANC ≤500cells/ mm^3, temperature	Fever >24h before	146	311	Mean 9y (range 0.5y

	Wurzburg, Germany		≥38.5°C	presentation			to 28y)
Lodahl 2011	Denmark/ Europe	25 September 2000 to 28 June 2001	Children (<16 y) admitted to hospital with febrile neutropenia	Not stated	85	230	Median 5.7 y; rang 4 month to 15 years
Mian 2009	USA/North America	Not stated	Febrile neutropenia	Not stated	29	51	Range 1-29
Miedema 2011	The Netherlands	April 1999 to August 2002	Children with malignancy and febrile neutropenia	None	29	43	Median age 8y in both groups and range 6-13y (bacterial infection) and 6-12y (no bacterial infection)
Nishikawa 2010	Japan	Feb 2008 to June 2009	Patients with lymphoma or leukaemia in remission	Not stated	27	36	Mean 10y (range 1-19)
Reitman 2010	United States of America	11 month period (year not stated)	Fever (>38 degrees celcius) and severe neutropenia (not defined)	Not stated	89	89	Not stated
Richardson 2009	United States of America	Jan 2006 to April 2008	Children with cancer or aplastic anaemia, fever and neutropenia	Not stated	48	142	Mean ages in 2 groups were 8.4y (SD 5.6) and 8.5y (SD 5.1)
Riikonen 1992	Helsinki, Finland	Not stated	ANC ≤1000cells/ mm^3, temperature ≥38.0°C twice in 4h, or once ≥39°C, or clinically 'poor condition'	Antibiotics in previous 3 weeks (except Septrin)	46	105	Not stated
Riikonen 1993	Helsinki, Finland	1989-1990	ANC ≤200cells/ mm^3, temperature ≥38.0°C twice in 4h, or once ≥39°C	Antibiotics in previous 3 weeks (except Septrin)	46	91	Not stated
Santolaya 1994	Santiago, Chile	1991-1992	ANC ≤500cells/ mm^3, temperature	Antibiotics in previous 72h or surgery	75	85	Not stated

			≥38.0°C twice in <24h	within 5d			
Santolaya 2007	6 hospitals in Santiago, Chile	2004 to 2005	PINDA 'High Risk' only. ANC ≤500cells/ mm^3, temperature ≥38.0°C ≥2 occasions ≥1h apart, or once >38.5°C	Low risk episodes and BMT patients	219	373	Not stated
Santolaya 2008	6 centres in Santiago, Chile	2004-2006	ANC ≤500cells/ mm^3 and 'fever'	Low risk episodes. Early onset (<24h) severe sepsis	278	566	Mean 7.75y
Secmeer 2007	Ankara, Turkey	January 2004 to January 2005	'Neutropenia' with temperature ≥38.0°C for ≥1h, or once ≥38.3°C	Not stated	49	60	Median age 7.7y (range 2y to 16y) in patients without documented infection and 7.2y (range 2.5y-18y) in patients with documented infection.
Soker 2001	Diyarbakir, Turkey	Not stated	ANC ≤500cells/ mm^3, temperature ≥38.0°C twice in <4h, or once ≥38.5°C	Antibiotics within 72h (except Septrin)	23	48	Mean 7y (range 2y to 14y)
Spasova 2005	Plovdiv, Bulgaria	January 2003 to June 2004	ANC ≤500cells/ mm^3, axilliary temperature ≥38.0°C ≥2 occasions ≥1h apart, or once >38.3°C	Not stated	24	41	Average not stated. Range 2m to 19y
Stryjewski 2005	Washington, DC, USA	Not stated	ANC ≤500cells/ mm^3, axilliary temperature ≥37.5°C or oral/rectal temperature ≥38.0°C	Fever >24h before admission or antibiotics (except Septrin) within previous 72h	56	Not stated	Mean 6.7y (range 5m to 17y)
Santolaya 2001	5 centres in Santiago, Chile	1996-1997	ANC ≤500cells/ mm^3, temperature ≥38.0°C ≥2 occasions ≥1h apart, or once	None stated, but no BMT patients	257	447	Mean 7y (range 6m to 18y)

			>38.5°C				
Santolaya 2002	6 centres in Santiago, Chile	1999-2000	ANC ≤500cells/ mm^3, temperature ≥38.0°C ≥2 occasions ≥1h apart, or once >38.5°C	None stated, but no BMT patients	170	263	Mean 7y (range 7m to 17y)

Data handling

Nine studies used a statistical technique to investigate if the predictive value of the serum marker was affected by other measured factors [30, 79, 139, 145, 158-162]. In five of these [79, 139, 145, 158, 161] an adjusted estimate was produced, using linear multivariable approaches. In these studies, the number of primary adverse events per predictive variable assessed ranged from 2.4 [79], to 11.6 [161] with one study having too little information to be able to assess this [145]. The other four studies [30, 159-160, 162] concluded the other measured variables did not affect the marker's diagnostic value. There was a lack of clarity in reporting the statistical approaches they used.

Assessment of the effect of multiple episodes per patient was undertaken in four studies; de Bont [158] used patients as a random-effect in their regression analysis, the other three studies [79, 160-161] undertook 'first-vs. last' episode comparison and found 'no significant difference'. Three papers describe no adjustment [139, 163-164], and the other studies make no mention of adjustments for clustered episodes. In three papers, only one episode per patient is used [165-167]

Thirty three of the 38 studies did not comment upon missing data. The five studies that did consider this issue used a complete-case analysis (with one study excluding potential variables with >10% missing values [79]). No studies clearly examined the nature of the missing data [79, 168-169].

Twenty five of the 38 included studies used a cut-point for marker test results, in 12 of the studies this was determined by the dataset being examined (eg by 'ROC analysis' or by maximising the sensitivity of the test)[30, 139, 158, 161, 163, 168, 170-175]. In six studies

the cut-point chosen was based on previous literature or alternative datasets [27, 156, 176-178], and in eight studies the choice was not explained[79, 142, 159-160, 169, 179-180]. Four studies did not go on to undertake an analysis with a dichotomised result [164, 166-167, 174, 181-187].

Six studies examined categorical variables as potential modifiers of the predictive ability of serum markers. Five of these used a grouping schema which was not explicitly justified [79, 139, 159, 172, 179] and one study [158] appeared to use ungrouped categorical data (malignancy). (See Appendix 12 for more detail.) As no study concluded these had any effect, no positive bias can have been introduced. However, given the lack of justification of the groupings, important interactions may have been missed.

Predictive performance of biomarkers

For the original review, quantitative data were pooled by three meta-analysis techniques, to explore the strengths and weaknesses of a variety of approaches. The review publication focussed on exploring both the methodological and clinical findings from the review [114]. The studies used in this were eleven studies providing data on CRP [27, 79, 159-160, 169, 172-173, 176, 179, 184], four studies also provided data on PCT [163, 173, 179, 184] and four provided data on the use of IL6 [163, 168-169, 171].

Analyses were originally possible for CRP (microbiologically or clinically documented infection), PCT (microbiologically or clinically documented infection) and IL6 (microbiologically or clinically documented infection, and gram –ve bacteraemia). Individual results for the four most frequently reported markers (CRP, PCT, IL6 and IL8) and outcomes are given in Table 10. These data have been used below to illustrate the advantages and challenges of conventional approaches to meta-analysis of diagnostic test data.

Table 10: Individual biomarkers study results used in pooled analyses presented by marker, outcome and cutoff.

Citation	Cutpoint	Sensitivity (95% CI)	Specificity (95% CI)	Method of derivation
CRP: Bacteraemia				
Spasova 2005	20	1 (95% CI 0.78 to 1)	0.04 (95% CI 0.01 to 0.18)	mean/sd
Spasova 2005	50	0.21 (95% CI 0.08 to 0.48)	1 (95% CI 0.88 to 1)	mean/sd
Spasova 2005	90	0 (95% CI 0 to 0.22)	1 (95% CI 0.88 to 1)	mean/sd
Riikonen 1993	20	0.65 (95% CI 0.41 to 0.83)	0.3 (95% CI 0.21 to 0.41)	2*2 extracted from text/graph
Riikonen 1993	50	0.18 (95% CI 0.06 to 0.41)	0.73 (95% CI 0.62 to 0.82)	2*2 extracted from text/graph
CRP: Documented & Clinical Infection				
Spasova 2005	20	1 (95% CI 0.87 to 1)	0.06 (95% CI 0.01 to 0.28)	mean/sd
Spasova 2005	50	0.92 (95% CI 0.75 to 0.98)	1 (95% CI 0.81 to 1)	mean/sd
Spasova 2005	90	0.44 (95% CI 0.27 to 0.63)	1 (95% CI 0.81 to 1)	mean/sd
Secmeer 2007	50	0.68 (95% CI 0.48 to 0.83)	0.46 (95% CI 0.3 to 0.62)	sensitivity/specificity reported
Katz 1992	20	0.71 (95% CI 0.59 to 0.81)	0.32 (95% CI 0.22 to 0.44)	sensitivity/specificity reported
Katz 1992	50	0.46 (95% CI 0.34 to 0.58)	0.75 (95% CI 0.63 to 0.84)	sensitivity/specificity reported
Katz 1992	100	0.22 (95% CI 0.13 to 0.34)	0.94 (95% CI 0.85 to 0.98)	sensitivity/specificity reported
Riikonen 1993	20	0.62 (95% CI 0.43 to 0.78)	0.28 (95% CI 0.18 to 0.4)	2*2 extracted from text/graph
Riikonen 1993	50	0.15 (95% CI 0.06 to 0.34)	0.71 (95% CI 0.59 to 0.8)	2*2 extracted from text/graph
El-Maghraby 2007	90	0.69 (95% CI 0.57 to 0.8)	0.73 (95% CI 0.54 to 0.86)	2*2 extracted from text/graph
Ammann 2003	5	0.97 (95% CI 0.91 to 0.99)	0.12 (95% CI 0.08 to 0.18)	sensitivity/specificity reported
Ammann 2003	50	0.48 (95% CI 0.38 to 0.58)	0.7 (95% CI 0.62 to 0.77)	sensitivity/specificity reported

Santolaya 1994	40	0.95 (95% CI 0.85 to 0.98)	0.77 (95% CI 0.59 to 0.88)	2*2 extracted from text/graph
Kitanovski 2006	60	0.63 (95% CI 0.39 to 0.82)	0.69 (95% CI 0.56 to 0.8)	sensitivity/specificity reported
Hitoglou-Hatzi 2005	20	0.9 (95% CI 0.74 to 0.96)	0.21 (95% CI 0.11 to 0.36)	mean/sd
Hitoglou-Hatzi 2005	50	0.76 (95% CI 0.58 to 0.88)	0.74 (95% CI 0.58 to 0.85)	2*2 extracted from text/graph
Hitoglou-Hatzi 2005	90	0.66 (95% CI 0.47 to 0.8)	0.87 (95% CI 0.73 to 0.94)	mean/sd
Santolaya 2001	90	0.75 (95% CI 0.69 to 0.81)	0.8 (95% CI 0.75 to 0.84)	sensitivity/specificity reported
Hatzistilianou 2007	50	0.9 (95% CI 0.8 to 0.95)	0.79 (95% CI 0.63 to 0.9)	sensitivity/specificity reported
CRP: Gram-ve Bacteramia				
Lehrnbecher 1999	20	0.88 (95% CI 0.53 to 0.98)	0.34 (95% CI 0.26 to 0.43)	sensitivity/specificity reported
Lehrnbecher 1999	50	0.88 (95% CI 0.53 to 0.98)	0.5 (95% CI 0.41 to 0.59)	sensitivity/specificity reported
Lehrnbecher 1999	100	0.88 (95% CI 0.53 to 0.98)	0.78 (95% CI 0.69 to 0.85)	sensitivity/specificity reported
CRP: Death				
Santolaya 2007	90	0.79 (95% CI 0.52 to 0.92)	0.61 (95% CI 0.56 to 0.66)	2*2 extracted from text/graph
CRP: Sepsis				
Katz 1992	20	1 (95% CI 0.65 to 1)	0.32 (95% CI 0.24 to 0.41)	sensitivity/specificity reported
Katz 1992	50	0.71 (95% CI 0.36 to 0.92)	0.67 (95% CI 0.58 to 0.75)	sensitivity/specificity reported
Katz 1992	100	0.71 (95% CI 0.36 to 0.92)	0.71 (95% CI 0.62 to 0.79)	sensitivity/specificity reported
Santolaya 2008	90	0.54 (95% CI 0.45 to 0.63)	0.63 (95% CI 0.59 to 0.67)	2*2 extracted from text/graph
PCT: Documented & Clinical Infection				
Secmeer 2007	0.1	0.2 (95% CI 0.09 to 0.39)	0.74 (95% CI 0.58 to 0.86)	sensitivity/specificity reported
Secmeer 2007	0.2	0.12 (95% CI 0.04 to 0.29)	0.89 (95% CI 0.74 to 0.95)	sensitivity/specificity reported

Secmeer 2007	0.3	0.11 (95% CI 0.04 to 0.28)	0.94 (95% CI 0.81 to 0.98)	sensitivity/specificity reported
Secmeer 2007	0.4	0.07 (95% CI 0.02 to 0.23)	0.94 (95% CI 0.81 to 0.98)	sensitivity/specificity reported
Hitoglou-Hatzi 2005	0.55	0.97 (95% CI 0.83 to 0.99)	0.58 (95% CI 0.42 to 0.72)	mean/sd
Hitoglou-Hatzi 2005	0.1	0.97 (95% CI 0.83 to 0.99)	0.47 (95% CI 0.32 to 0.63)	mean/sd
Hitoglou-Hatzi 2005	0.2	0.97 (95% CI 0.83 to 0.99)	0.5 (95% CI 0.35 to 0.65)	mean/sd
Hitoglou-Hatzi 2005	0.3	0.97 (95% CI 0.83 to 0.99)	0.5 (95% CI 0.35 to 0.65)	mean/sd
Hitoglou-Hatzi 2005	0.4	0.97 (95% CI 0.83 to 0.99)	0.53 (95% CI 0.37 to 0.68)	mean/sd
Hatzistilianou 2007	0.2	0.97 (95% CI 0.89 to 0.99)	0.97 (95% CI 0.85 to 0.99)	sensitivity/specificity reported
Kitanovski 2006	0.55	0.94 (95% CI 0.72 to 0.99)	0.71 (95% CI 0.58 to 0.82)	sensitivity/specificity reported
PCT: Sepsis				
Santolaya 2008	2	0.57 (95% CI 0.48 to 0.66)	0.46 (95% CI 0.41 to 0.5)	mean/sd
PCT: Bacteraemia				
Kitanovski 2006	0.1	0.33 (95% CI 0.1 to 0.7)	0.78 (95% CI 0.65 to 0.87)	sensitivity/specificity reported
Kitanovski 2006	0.2	0.33 (95% CI 0.1 to 0.7)	0.89 (95% CI 0.79 to 0.95)	sensitivity/specificity reported
Kitanovski 2006	0.3	0.33 (95% CI 0.1 to 0.7)	0.93 (95% CI 0.83 to 0.97)	sensitivity/specificity reported
Kitanovski 2006	0.4	0.33 (95% CI 0.1 to 0.7)	0.95 (95% CI 0.86 to 0.98)	sensitivity/specificity reported
IL6: Gram-ve bacteraemia				
Lehrnbecher 1999	235	1 (95% CI 0.82 to 1)	0.63 (95% CI 0.53 to 0.72)	sensitivity/specificity reported
Lehrnbecher 2000	1000	0.74 (95% CI 0.51 to 0.88)	0.96 (95% CI 0.9 to 0.98)	sensitivity/specificity reported
IL6: Documented & Clinical Infection				
Kitanovski 2006	235	0.88 (95% CI 0.64 to 0.97)	0.87 (95% CI 0.75 to 0.93)	sensitivity/specificity reported

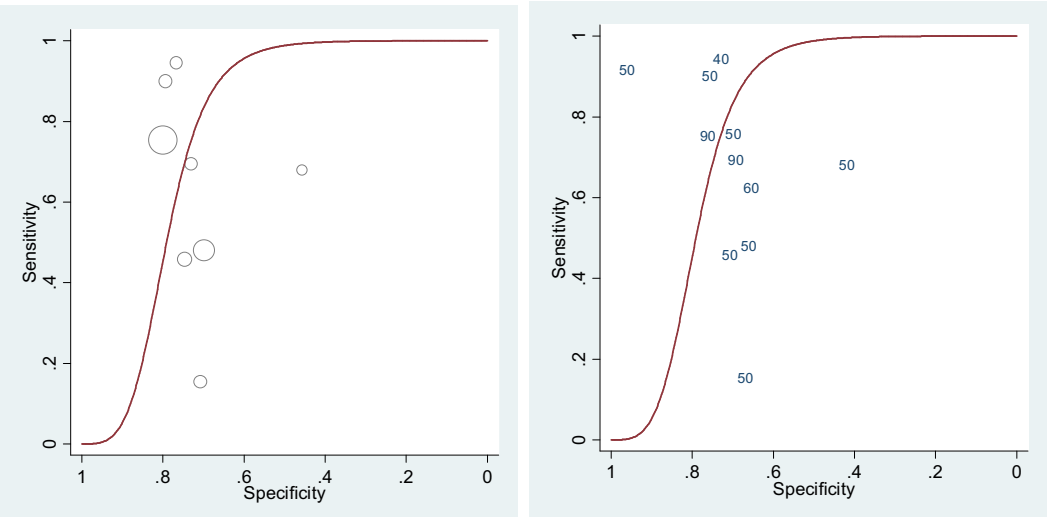
Riikonen 1992	235	0.1 (95% CI 0.03 to 0.3)	0.98 (95% CI 0.91 to 1)	2*2 extracted from text/graph
Riikonen 1992	1000	0 (95% CI 0 to 0.16)	1 (95% CI 0.94 to 1)	2*2 extracted from text/graph
Lehrnbecher 2004	235	0.89 (95% CI 0.82 to 0.94)	0.91 (95% CI 0.86 to 0.94)	sensitivity/specificity reported
Lehrnbecher 2004	1000	0.11 (95% CI 0.06 to 0.18)	0.99 (95% CI 0.97 to 1)	sensitivity/specificity reported
Diepold 2008	42	0.9 (95% CI 0.81 to 0.94)	0.86 (95% CI 0.69 to 0.94)	sensitivity/specificity reported
IL6: Bacteraemia				
Diepold 2008	240	0.64 (95% CI 0.39 to 0.84)	0.75 (95% CI 0.65 to 0.82)	sensitivity/specificity reported
IL8: Sepsis / Prolonged illness				
Santolaya 2008	200	0.49 (95% CI 0.4 to 0.58)	0.71 (95% CI 0.67 to 0.75)	2*2 extracted from text/graph
Diepold 2008	30	0.87 (95% CI 0.78 to 0.93)	0.61 (95% CI 0.42 to 0.76)	sensitivity/specificity reported
IL8: Bacterial infection				
Diepold 2008	90	0.64 (95% CI 0.39 to 0.84)	0.62 (95% CI 0.52 to 0.71)	sensitivity/specificity reported
IL8: Documented & Clinical Infection				
El-Maghraby 2007	62	0.71 (95% CI 0.59 to 0.81)	0.77 (95% CI 0.58 to 0.89)	2*2 extracted from text/graph
Lehrnbecher 2004	320	0.56 (95% CI 0.46 to 0.65)	0.79 (95% CI 0.73 to 0.84)	sensitivity/specificity reported
Lehrnbecher 2004	500	0.44 (95% CI 0.35 to 0.54)	0.89 (95% CI 0.84 to 0.93)	sensitivity/specificity reported

To illustrate the methods and associated challenges, syntheses were undertaken using both classical and Bayesian approaches.

***Method 1* Classical statistical analyses**

Data were combined using a single cut-off from each study using the STATA routines `metandi` and `midas` for analyses of 3 studies and over. For analyses of four or more studies, a random effects linear regression using `xmelogit` was fitted for bivariate estimates.

The HSROC curve (Figure 22a) derived from 11 studies [27, 79, 139, 159-160, 163, 169, 173, 176, 179, 184] demonstrates moderate diagnostic ability for CRP to detect 'documented infection' (Area under the ROC curve 0.78 (95% CI 0.74 to 0.81)). This assumes that a higher cut-off produces a lower sensitivity and higher specificity. However, the plot demonstrating each study's cut-off (in mg/dl) shows that the assumption of threshold variation is not adhered to (Figure 22b): rather than the threshold value steadily falling from high cutoff values in the bottom left through middle values in the mid-point of the curve, to low values in the upper right, we see values of 50 and 60 preceding 90. This should raise doubts about the validity of the summary ROC curve produced. As demonstrably different thresholds are used in creating this pooled analysis, the production of a single bivariate estimate of the 'test effect' is clearly meaningless.

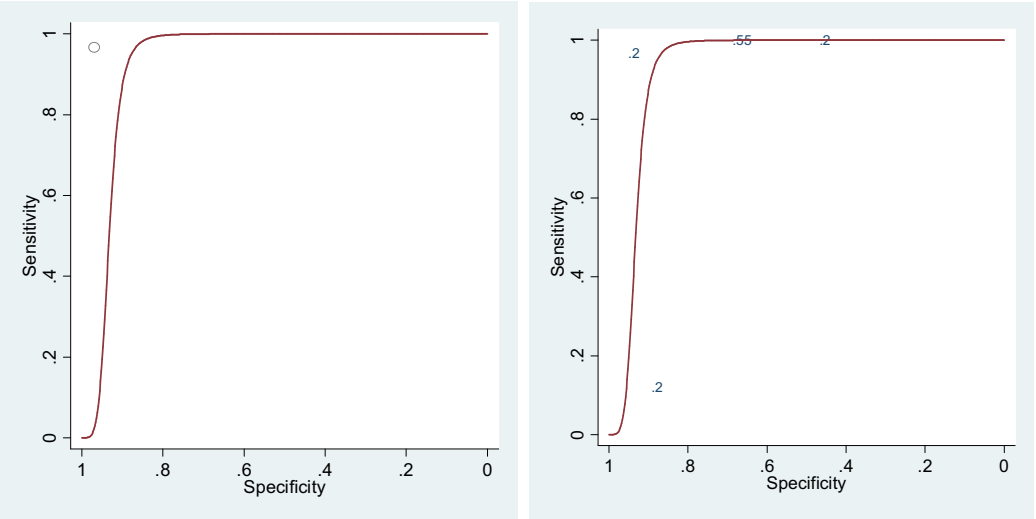


a) Circles weighted according to study precision b) Marker points showing threshold (mg/dl)

Figure 22:HSROC curve plots of CRP for the diagnosis of 'documented infection'

Analysis of PCT[163, 173, 179, 184] suggest a better discriminatory ability (Area under the ROC curve 0.93 (95% CI 0.90 to 0.95)). Though based on only two different cut-offs, the threshold findings are replicated in the PCT data (Figure 23).

While this finding represents only a pair of markers across one outcome, this should raise doubts about the validity of the technique of HSROC determination, which assumes threshold values will follow an expected path, when data about the actual threshold are available and could be used more effectively by an alternative meta-analysis technique.



a) Circles weighted according to study precision b) Marker points showing threshold (mg/ml)

Figure 23: HSROC curve plots of PCT for the diagnosis of 'documented infection'

For studies with similar outcomes and cut-off values, meta-analysis was undertaken using a random effects bivariate approach. Data were sufficient to undertake this in two outcome groups over three markers (see Table 11: Bivariate estimates of diagnostic precision of various markers and outcomes and Figure 24).

Table 11: Bivariate estimates of diagnostic precision of various markers and outcomes

Marker	Outcome	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)
CRP (7 studies)	Documented infection	>50 mg/dl	0.65 (0.41 to 0.84)	0.73 (0.63 to 0.82)
PCT (3 studies)	Documented infection	>0.2 mg/ml	0.96 (0.05 to 0.99)	0.85 (0.53 to 0.97)
IL6 (3 studies)	Documented infection	>235 pg/ml	0.68 (0.15 to 0.96)	0.94 (0.84 to 0.98)
IL6 (2 studies)	Gram –ve bacteraemia	>1000 pg/ml	0.78 (0.57 to 0.91)	0.96 (0.92 to 0.99)

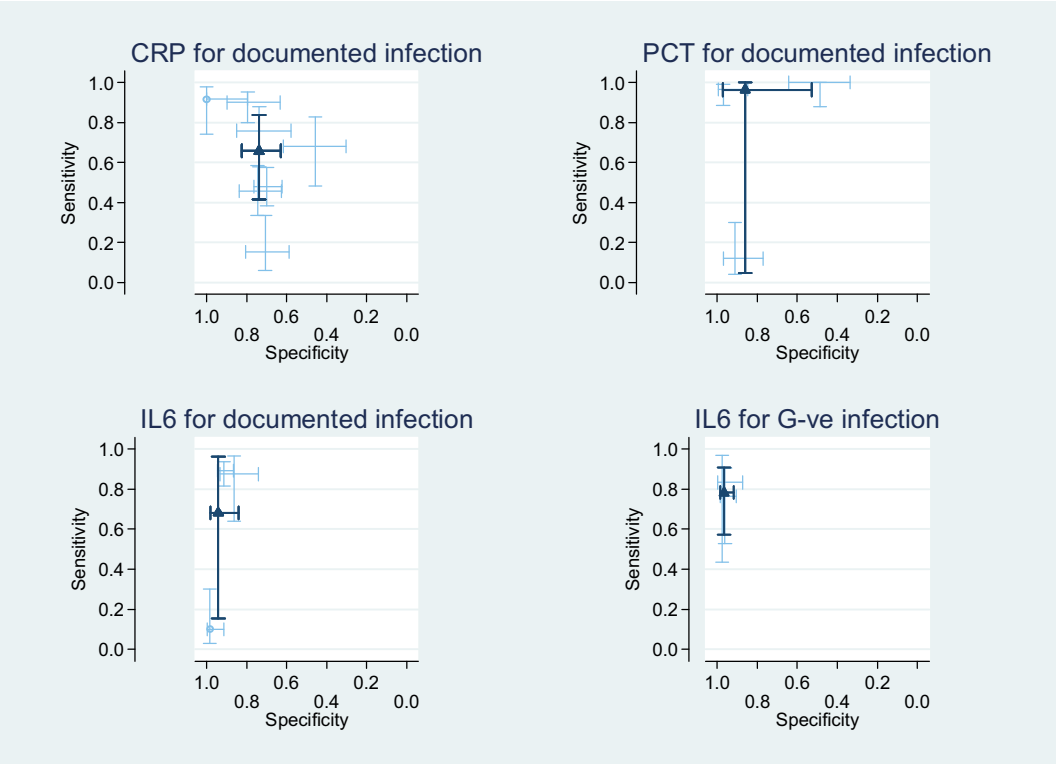


Figure 24: Bivariate pooled estimates of sensitivity and specificity for CRP, PCT & IL6

There is considerable heterogeneity in these results; sensitivity being the most heterogeneous in all markers, and specificity being most heterogeneous in PCT and CRP.

***Method 2.* Prediction within a Bayesian framework**

Meta-analysis of data for IL6, PCT and CRP was attempted for documented infections using a similar approach to that used in the clinical decision rules review.

The analysis of data from the IL6 studies to predict documented infection demonstrated a very wide range of average estimates of diagnostic accuracy. In particular, the uncertainty around the proportion of individuals with disease/non-disease in the groups 235-1000pg/ml and >1000pg/ml led to the median estimates reversing the 'sensible' order of results, implying that higher levels of IL6 were less likely to be associated with disease. (LR <235pg/ml 0.35 (95% CrI 0.02 to 0.96), LR 235-1000pg/ml 9.54 (95% CrI 0.02 to infinite) and LR >1000pg/ml 8.0 (95% CrI 0.05 to 9.8). The heterogeneity between individual study estimates is extreme, particularly for sensitivity, where one of the three available data points for 235pg/ml overlies the cut-offs at 1000pg/ml. Meta-analysis is therefore inappropriate (see Figure 25).

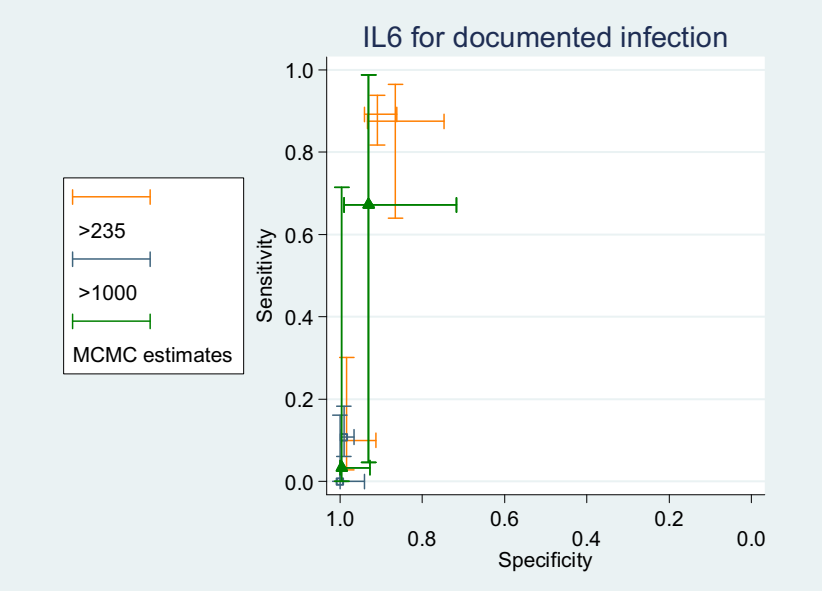


Figure 25: IL6 for documented infection

Analysis of the four PCT studies and eleven CRP studies proved impossible using this analytical technique.

The data from the four PCT studies show extreme heterogeneity in sensitivity and specificity (see Figure 26). In order to reduce the variables under consideration, the cut-offs analysed were limited to 0.2pg/ml and 0.55pg/ml, but this still require six data points (vs. seven in the IL6 example) to provide information on seven independent variables (a total of three proportions in diseased and undiseased populations, and four variance-covariance estimates). Assessments based on such limited data tend to be very unstable.

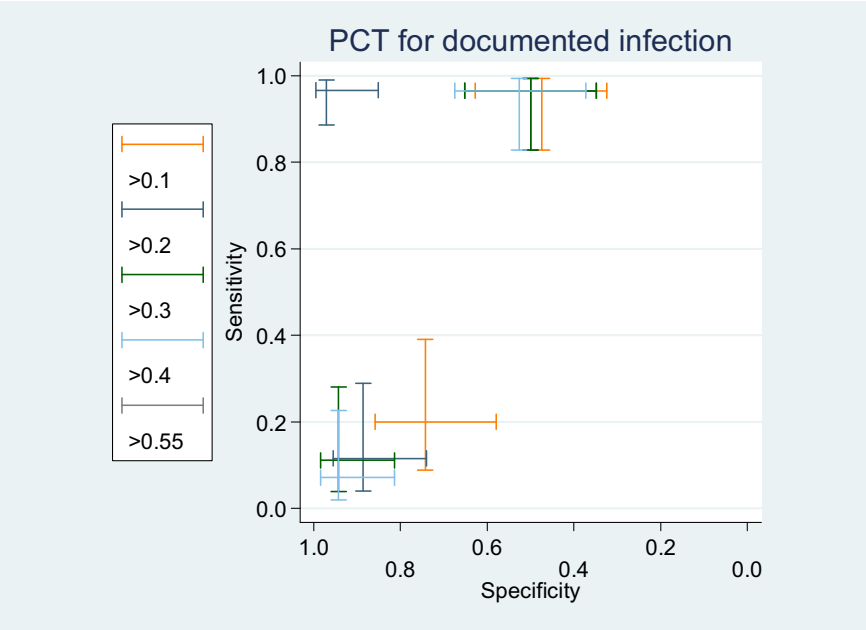


Figure 26: PCT for documented infection

Attempts to reduce the variables to be fitted further, by simplification to an assumption of a single variance in the diseased/undiseased populations was also unsuccessful. This is likely to be due to the extreme heterogeneity and sparse data. Extreme simplification to a fixed

effects model did not succeed, probably as some simulation instances require the (nonsensical) reversal of the arrangement of proportions of individuals in the ordered categories and occasionally 'negative' proportions.

Similar problems were encountered when attempting to fit the model to the eleven studies with CRP values for documented infection (see Figure 27). Here reduction to three cut-offs was undertaken (20, 40-60, 90-100) along with univariate and fixed effects approaches. The model produced extremely uncertain results, particularly estimating the proportion of individuals with disease/nondisease whose CRP ranged from 40 – 100 (between cutoff 2 and 3) where the 95% "credible" interval ranged from -3% to +39%.

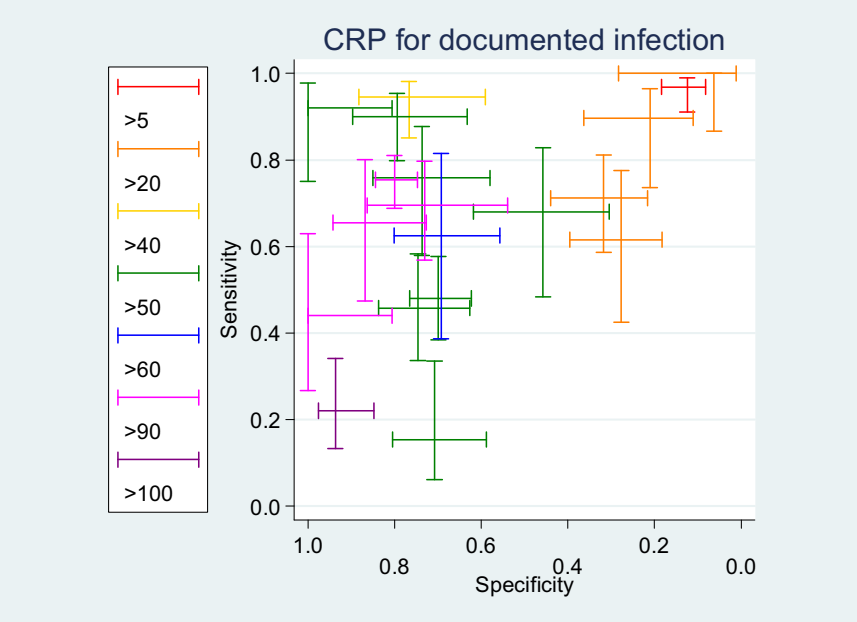


Figure 27: CRP for documented infection

Performance

The update of the systematic review [115-116] provided further data to undertake meta-analysis on a range of biomarkers, with sufficient studies reporting on admission CRP(sixteen studies [27, 79, 139, 142, 159-160, 162-163, 166, 173, 175-176, 179-180, 184, 188]), PCT (nine studies [148, 163, 173-175, 179, 184-185, 189]), interleukin-6 (IL-6; five studies [142,

156, 163, 169, 174]) and additionally five interleukin-8 (IL-8) studies [156, 166, 168, 176, 185] in their ability to detect significant infection.

The pooled estimates remain clinically and statistically heterogenous, with the most appropriate and advanced synthesis technique (multiple threshold approaches using a Bayesian multinomial framework) producing clinically uninterpretable results (see Table 12 and Figure 28).

Table 12: Multivariate meta-analysis of biomarkers to detect significant infection (clinically or microbiologically documented infection)

Threshold	Likelihood ratio	95% credible interval*
CRP		
CRP <20 mg/dL	0.25	0.07 to 1.14
CRP 20-50 mg/dL	-0.44	-8.81 to 8.27
CRP 50-90 mg/dL	0.39	-1.04 to 2.77
CRP >90 mg/dL	2.41	0.87 to 16.74
PCT		
PCT <0.2 ng/mL	0.42	0.009 to 2.1
PCT 0.2-0.5 ng/mL	-0.11	-22 to 23
PCT >0.5 ng/mL	3.1	0.9 to 8.8
IL-6		
IL-6 <235 pg/ml	0.353	0.005 to 1.052
IL-6 235-1000 pg/ml	7.981	-1.669 to 65.45
IL-6 >1000 pg/ml	7.05	0 to 1699
IL-8		
IL-8 <60 pg/ml	0.3	0.12 to 0.59
IL-8 60-320 pg/ml	-0.95	-14.55 to 7.34
IL-8 320-500 pg/ml	0.31	0.06 to 3.89
IL-8 >500 pg/ml	9423	0.02 to 1.19E+10

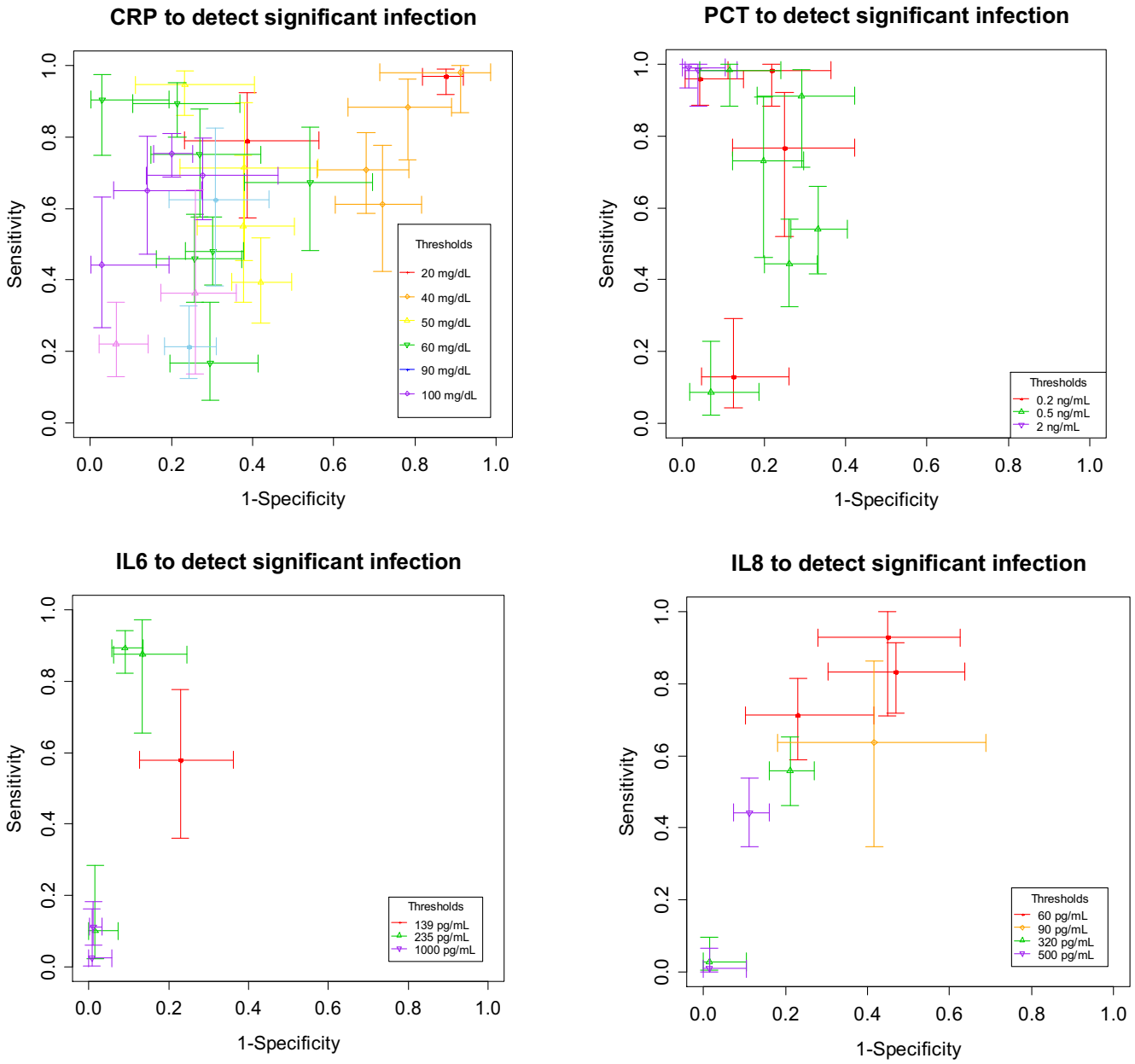


Figure 28: ROC plots of biomarkers detection of infection

Additional studies provided information that could not be incorporated into the meta-analysis, the details of which are given in Appendix 13. PCT was further examined and shown to have moderate sensitivity (66%) and specificity (85%) with a very high study-defined cut-off of 3.3 ng/ml in patients with bacteraemia [174], had different median values between septic and non-septic patients (0.5 vs 0.24 ng/ml)[30], or baldly stated to be associated with significant infectious complications in a multivariate analysis [190] (but all had insufficient data to produce variance estimates). An early study examined the values of PCT levels in patients with short (<5 day) and prolonged admissions and demonstrated a difference in means[170]. The other study to report PCT values [161] did so in only the high-risk group of patients, and failed to find a difference in mean values at admission or 24 hours.

CRP was also reported in studies that could not be added to the meta-analysis because of they provided insufficient data to calculate variances [22, 158, 172, 177, 183, 186-187, 191]. These studies also produced a range of point estimates, generally indicating a small increase in average values in those patients with adverse infectious outcomes, compared to those without. Where entered into multivariable models [22] CRP did not have any independent predictive value.

IL6 and IL8 were examined in a small number of additional studies. An analysis of all-age patients demonstrated an independent predictive value for IL8 and IL6 [158] against a limited range of clinical variables, and both were shown to be significantly higher in patients requiring ICU admission [145]. A study examining multiple cytokines found a higher median IL8 value in bacteraemic patients (0.3 vs 0.02 ng/ml) and showed this varied by the type of organism isolated (gram negative bacteria

0.91 vs gram positive 0.13 ng/ml) [164]. A similar finding was reported by Strewjeski [30] showing 0.45 ng/ml vs. 0.15 ng/ml IL8 median values when using a broader definition of bacteraemia and culture negative sepsis. Santolaya's group also examined IL8 values [161] in the high-risk group of patients, and failed to find a difference in mean values at admission or 24 hours. IL6 was examined in distinguishing bacteraemia and was reported as showing around 65% sensitivity and 70% specificity with a data derived cut-off of 137 pg/ml. [174]

One study (reported as an abstract) showed a combined result for IL5 and IL8 of "sensitivity 0.88, specificity 0.48" but no cut-offs were given [165], and when compared with an earlier report of 100% sensitivity and 88% specificity [177], this in keeping with the general trend of decreasing accuracy with repeat assessments of tests.

There are a range of other novel biomarkers under limited investigation that have been revealed by these reviews. These data are reported in detail in Appendix 13. These are very sparse, based on one or two studies, and clinical conclusions cannot be reasonably drawn from them.

Trajectory of biomarkers

Six studies explored the role of serial biomarkers to detect documented infection or sepsis.[173, 180] [157, 189, 192-194] There were insufficient data available for meta-analysis. In one study, the difference between mean CPR, PCT and IL-8 at 24 hours in children with and without sepsis was more pronounced than at presentation.[157] Similarly, the sensitivity of PCT in predicting bacterial infection was higher at 24 to 48 hours compared with presentation in another study.[194] These finding are in keeping with the results of the clinical decision rules where a 24h+ assessment has been performed.[144, 154] In the study by Hatzistilianou et al the seven-day trend of PCT and CRP was depicted graphically and PCT showed a more rapid decline in patients treated for bacterial infection as compared to CRP.[192]

Four studies provided direct comparisons of the discriminatory power of admission values of PCT and CRP.[157, 192-193, 195] Three of these studies reported area under the receiver operator curve (ROC) estimations.[192-193, 195] In these three, the data showed PCT consistently had a better discriminatory estimate than CRP with AUC range of 0.66 to 0.869 compared with 0.43 to 0.728. The fourth study reported no significant benefit of PCT over CRP.[157] Procalcitonin also had higher discriminatory power than IL-6 in one study[195] and Il-8 in another,[192] which was not confirmed in a further study.[157] Meta-analysis of these

direct comparisons of diagnostic accuracy assessments was not possible from the data available.

Conclusions

This series of systematic reviews of clinical decision rules studied patients with a wide range of malignancies and including between 29 and 759 patients (median 132) per study and with between 47 and 1117 episodes of febrile neutropenia per study (median 240). The biomarkers reviews included 4689 episodes of FN, investigating 24 different makers of inflammation or infection (14 biomarkers from original review; additional 10 updated review). The following Chapter explores the implications of these data and identifies where they point to the development of an IPD meta-analysis.

Chapter 5: Discussion of the results of systematic reviews of Clinical Decision Rules and Serum Biomarkers

A robust risk stratification model that reliably predicts which children are at very low or very high risk of having a significant infection could have important implications for clinical care. Those at very low risk could be treated with reduced intensity antibiotic therapy and spend a shorter period in hospital. Those at high risk of complications could be targeted for more aggressive management. While the systematic reviews were being conducted for this thesis, there was clear evidence of many differing policies for the management of FNP in practice [51, 53] with lack of agreement about how risk stratification, if any, was used.

The previous chapters presented the methods and results of a series of systematic reviews with updates assessing and summarising existing research evidence. These studied clinical examination to identify infectious complications; and the value and added value of specific serum biomarkers in this regard.

Studies were reviewed for their ability, as diagnostic tests, to accurately differentiate groups of patients who did and did not have the condition of interest. Appraisal was undertaken using the QUADAS criteria for quality. Synthesis of these data were undertaken, where possible, to provide the most accurate estimates of predictive accuracy available. Poor quality of execution and design of studies may produce problems which introduce bias (systematic difference) or significant variation that limits the generalisability of a study's findings. In studies of a diagnostic test, these may be categorised as: those which arise from the population studied; the technology used in undertaking the testing; the outcome assessments made; and the nature of the test interpretation [74]. Although clinical examination in the rules described here did not raise issues around the technologies used, it does raise a potential issue with studies of multiple individual investigators undertaking a 'physical skill'. Some physicians are likely to have better auditory or tactile discrimination, and there will be differences between them in the accuracy of their measurements (for example, the reproducibility of precise auditory findings in chest examination is poor [196]). Technical issues about the reproducibility of measurements of biomarkers were raised in the third group of reviews, though the

use of similar quality-assured techniques for the common and commercially available tests constrains these problems.

Studies using clinical features to predict infectious complications

The development and verification of clinical decision rules to predict significant infectious complications were studied in a systematic review that was updated during the development of national[128] and international guidelines [129]. The review was updated specifically in response to the international guideline, as part of my role as Chair of the risk stratification section of the guideline.

The studies examining clinical decision rules to predict infectious complications produced 21 models, and contained eleven datasets used to validate previously derived models. They studied a variety of outcomes, with individual differences in definitions, but covered five main categories: death, critical care requirement, serious medical complication, significant bacterial infection, and bacteraemia.

The validity assessment undertaken suggests that the biases in study design were relatively minor. The most common was a potential for clinical review bias, where clinical information may lead to one final outcome being favoured unfairly over another, which occurred in 19 of the 26 studies. However, the potential effect of this is mitigated by the largely objective nature of the outcomes, for example: microbiologically positive blood culture results, severe sepsis and death. Evidence for the theoretical reduced influence of study design on objective outcomes is present for therapeutic studies [75], but not for those of diagnostic accuracy [74]. Partial verification bias may technically be a threat to studies in which certain 'outcome assessment tests', for example swabs of lesions were undertaken only when clinically indicated, but these are clinically reasonable variations and unlikely to lead to a strong bias. If a lesion is not present to be swabbed, it cannot be the unidentified source of an infection.

Two potential problems are in the use of aspects of the outcome assessment in the decision rule [19, 23, 153]. This was most marked in the Alexander study [23] where the outcome of 'significant medical complication' included 'hypotension and severe mucositis', as did the rule describing high-

risk. This incorporation bias leads to diagnostic tautology, theoretically improving the accuracy of a 'test' [74]. The results from this study are likely to be overly optimistic.

The studies which set out to derive a CDR varied in their populations, the number and type of outcomes studied, variables assessed, model-building technique, reporting and handling of missing and multiple-episode data and in the use and categorisation of continuous and categorical variables. All of these features may have influenced the CDRs produced and provide some explanation of the differences between them.

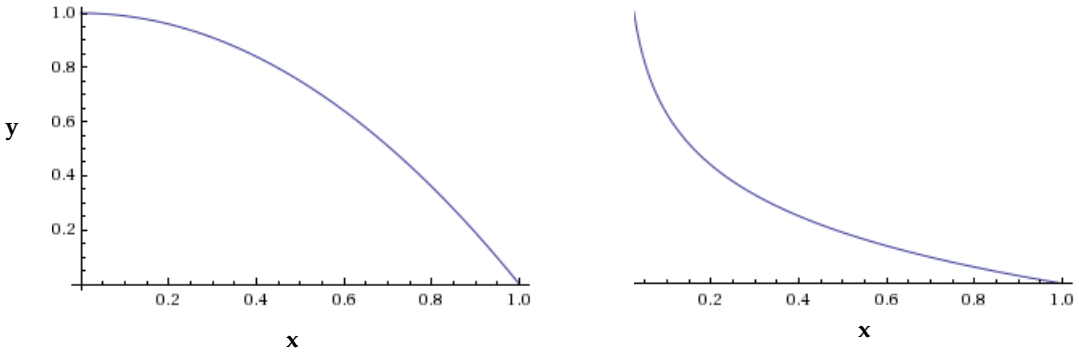
Building stable predictive models requires between 10 – 20 events per variable considered [84] and while some argument has been made to relax this value [197] these simulation studies have only examined single predictors rather than the multiple predictors used here. We found that 76% of the CDR derivation studies had less than ten events per variable under consideration, and no study had more than 14 events per variable. The small sample size makes models more likely to be overfitted to their original dataset and disappointing in clinical practice [83]. Only one study [140] modified model parameters to account for the small sample size and low number of events per variable.

The technique used to build the model is also extremely important. There are a number of families of techniques, including multivariable regression, neural networks and classification and regression trees (CART models). No clear superiority for one technique has been demonstrated [78]. Of the studies that derived CDRs in this review, almost all were built using multivariable regression. One model used CART techniques [79] alongside a logistic regression and came to different conclusions from the same dataset. This highlights an acknowledged difficulty with model building (that differing techniques may reach different conclusions from the same information without it being clear which is the 'most correct'.

The models all assumed linear relationships between the outcome and the explanatory variables, but for some variables relationships have different forms. A classic example of a non-linear relationship is the S-shaped curve of oxygen dissociation from haemoglobin, or the J-shaped associations of body-mass index and mortality. It is not clear in these studies if this assumption of linearity of the variables was assessed, but failing to do so may misjudge a predictor as unimportant

[198]. There are plausible reasons to assume that patient age may have a non-linear 'U'-shaped relationship with infection and outcome [199], as should time-from-chemotherapy. Bone marrow suppression may have a more complex influence upon likelihood of infection than a simple linear relationship (for example of non-linear relationships, see Figure 29a & b).

Figure 29: Examples of non-linear relationships



a) $y = 1-x^2$ b) $y = -\ln(x)$

The relationships shown in Figure 29 are inventions, and extreme, in that they proposes that the risk of infection (y axis) is always present with 'zero' neutrophils (x axis). Figure 29a supposes that the risk is at a very high level, a near plateau, when the count is below around 0.175. This could be justified by a hypothesis which requires a certain number of circulating cells to be present for adequate infection surveillance. Figure 29b supposes that there is a log-linear inverse relationship, with the risk relating not to a straight line of the neutrophil count but to the natural log of the count. This could well be possible if the neutrophil count was log-normally distributed.

The selection of variables for the final model is crucial. Selections can be performed by taking all possible explanatory variables, and excluding those which are not statistically significant (backwards elimination), or by adding, one-by-one, the most statistically significant individual factors (forwards selection), or a combination of the two, adding and removing variables to build the statistically best fitting model (stepwise). These techniques, when selection is driven by a 'p-value' are seriously at risk of choosing variables with chance relationships and resulting in unstable models [81]. These techniques will also exclude variables that confound each other from entering a multivariable analysis [82]. This happens if two variables which measure very highly related parameters (for example, the concentration of haemoglobin in the blood – which is carried within red blood cells,

and the red cell count) are both related to the risk of infection, any model which uses one of these will show the variable to be significant and important. If the model uses both, the 'strength' of each will be mopped up as the model attempts to account for the other variable, making both appear 'insignificant'. In essence, although selecting variables only by looking at how significantly they are associated with outcomes in the dataset being examined produces highly effective descriptions of that data, it doesn't improve the ability of the model to describe the real population from which those data were drawn [83]. Only one of the models clearly examined co-linearity; the issue of multiple variables being highly correlated which may account for the differing 'marrow suppression' markers being used in different CDRs.

When moving beyond the very first stages of exploration in a new area, variables should be selected on the basis of clinical evidence or physiological reasoning [85]. In these studies, this was stated to be the case in seven of the studies, and although unstated in the others, the selection of similar variables implies congruent thinking. This could ameliorate the potential inflation of results.

Continuous variables, such as age, blood pressure and absolute monocyte count, will have their most accurate predictive value in a model if used as their actual value. Clinicians seem to find the use of continuous variables in this setting uncomfortable, and prefer to use categorised values. Repeated studies examining prognostic model building have shown that the collapsing of continuous variables into ordinal categories or dichotomies is often undertaken using methods which are highly likely to give spurious results. [78, 80] The problem comes from analyses where a particular set of data is examined, by looking at the ROC curve or recursive partitioning analysis, to find the cut-point that achieves greatest differentiation between the diagnostic or prognostic categories. In doing this, effectively multiple tests are being undertaken and the reported p-value associated with the final choice is likely to be a gross exaggeration of the true 'significance' of the value. Approaches using clinically or pathophysiologically meaningful values, or ones previously described, avoid these problems. The choice of how to group categorical variables may give rise to similar problems. In the studies where an explanation is given are evaluated, the decisions seem to have been made with a combination of data-driven 'optimal' cut-points then modulated to give clinically sensible numbers (e.g. Rackoff [141] and the use of 100 cells/mm^3 as a cut-off for absolute monocyte count).

A further issue is the assumption of independence that underlies most of the techniques used. In 12 studies, multiple episodes in individual patients are treated as if they come from unconnected individuals. Underlying this needs to be the chance of the first, second, third etc. episodes having the same outcome: this clearly cannot be the case if one of the outcomes under consideration is death. In studies that undertook a further analysis which looked at only the 'first case' and found 'no significant differences' between the approaches, they used this as justification for assuming independence, but this is likely to be underpowered for the rarer outcomes of death and severe infectious complications. Other approaches to address this problem include the use of only first episode data. This has the disadvantage of decreased numbers of episodes analysed and consequent decreased power and efficiency. Four studies reported the use of extended modelling techniques that assessed and accounted for clustering to try to avoid such problems. No study reported the degree of interpersonal and intrapersonal variability to quantitatively estimate the degree of bias introduced by undertaking the simplistic approaches, and so it becomes difficult to assess the potential error introduced.

The way that missing data are handled can also introduce bias and reduce efficacy. Data can be lost or go missing in ways that introduce bias, or in ways that do not introduce bias but reduce the efficacy of the study. Non-biased data loss, for example by the bad luck of a power failure in the lab meaning a blood test can't be analysed, is described as "missing completely at random" (MCAR). The data are missing for no reason but random chance. Potentially biased missing data comes in two sub-categories: the first is where the missing element is intimately linked to something known and recorded, for example the patient's condition meaning arterial blood gas measurements are available on only the sickest children in a cohort. These missing values, which are related to other known and measured factors, are confusingly called "missing at random" (MAR). These missing data are potentially imputable from the data that exists. The second sub-category of bias-inducing data loss is where there is no possibility of linking the missing data to known items. For example, it may be that patients presenting during the first few weeks of a new physician joining a hospital team are less likely to have all the correct blood tests done as the admitting doctor is not familiar with the study protocol. There will be a systematic difference between those with missing data and those without (new doctor versus experienced doctor) but the reason why the data were missing is not linked to anything the researchers can know (assuming no-one tells them the doctors changed jobs). This type of bias-inducing missingness is called "missing not at random" (MNAR).

The CDR derivation studies described how missing data were handled in only eight of the 21 studies. In six of these a form of complete case analysis was used. No study details an assessment of the type of missingness of the data. While MCAR cases can be ignored, using a 'complete case' or 'available case' analysis, it reduces the number of episodes, but doesn't introduce bias. However, undertaking this type of 'available case' analysis when there is a MNAR or MAR problem introduces a form of selection bias. The development of imputation techniques, where the missing elements are replaced by one of a number of reasoned methods, provides a way of increasing the efficiency of a study without introducing bias when data are MAR.[200] No study used such techniques, though it should be acknowledged that these approaches are not without problems. [201]

The CDR to predict infectious complications had diverse test performance across diverse outcomes. Initial hypotheses to explain the differences included: the design of the study, the population (both geographical and case-mix), the complexity of the rule and outcomes chosen. Tabular and graphical analysis, supplemented by minimal quantitative data, supported the following assertions: validation studies produced estimates of lower test accuracy and rule complexity, case-mix did not clearly explain differences between test performance, and geography appears very important. Differences related to the outcome of interest may be present, with rules to predict infectious complications being more sensitive and less specific, rules to predict death/ICU admission being more specific but less sensitive, and rules predicting bacteraemia spanning a range of results, although this was difficult to separate from the other proposed factors.

Where the aim was to define a group of patients who would not develop adverse outcomes from their episode, high sensitivity (capturing all the diseased individuals within the high-risk category) was of primary importance. This would enable those in the low-risk group to be treated with reduced intensity, without concern of 'missing' patients who would develop problems. There remains a need to trade off sensitivity against specificity (as discussed earlier): the most sensitive rule would be to call all patients 'high risk'. This would result in no missed adverse outcomes, but would over-treat a large proportion of patients.

The performance of the AMC/Temperature criteria proposed by Rackoff [141] to exclude bacteraemia was assessed across multiple datasets. This model, being tested by different groups across time and in different centres, has the greatest strength of evidence. The most appropriate

pooled estimate of the rule's effectiveness comes from a random effects model assuming no threshold variability, and excluding both the derivation sample and an outlying study using a different outcome definition. This led to estimates of moderate discriminatory ability LR [low] = 0.26 (95% CrI 0.08 to 0.72) , LR [medium] =0.72 (95% CrI 0.14 to 2.15), and LR [high] = 3.11 (95% CrI 1.25 to 8.01); a low-risk result led to the odds of infection being roughly one quarter of the overall prevalence, a medium risk result was associated with a marginally reduced chance, and a high-risk result approximately three times the odds of an infection being diagnosed.

The exclusion of the derivation sample is justified as this data produced the rule, and would always improve accuracy (this was also demonstrated in reviewing the test performance of individual studies). The single non-US study excluded showed a strikingly lower utility for the rule; this differs in geographical area and reports the wider outcome of 'documented microbiological infection' rather than a narrow bacteraemia diagnosis. As explored above, resolving the reasons for this heterogeneity is very difficult within this group of studies.

The technique used to summarise the data from the data sets used a Markov-chain Monte Carlo approach to estimate the proportions of bacteraemic and non-bacteraemic patients in each risk group. Data from studies which used a similar rule but provided only low versus medium/high risk categories [12, 69] were also included in this analysis. These proportions were used to calculate likelihood ratios for each risk category and corresponding 95% credible intervals were derived from the posterior probability distributions. This analysis technique accounts for heterogeneity due to the sampling variation within populations, and variation of sampling from different populations.

Multivariate models were investigated to assess how well they 'fit' the data under investigation. Compared to the simple assumption of a random effect variation between studies independently in affected (outcome positive) and non-affected individuals, two layers of multivariate model were tested. The first proposes a bivariate relationship between the cut-offs within each study: that is, that the population of 'low', 'middle' and 'high' risk individuals may vary differently in each study and is best estimated by two random-effect variables. The second attempts to model a further layer of heterogeneity: connections between the differences in the affected and unaffected populations across the studies. This is usually explained as different cut-off thresholds for the tests actually applied in different studies. In the analysis undertaken here, the test cut-offs are explicit and

objective (AMC >100, or AMC <100 with maximum temperature measured as either under or over 39°C), with minimal room for intra-study variability in how the rule is applied, and so there is minimal or no threshold variation. The data produced by these three models showed that there was a benefit from multivariate modelling within the affected/unaffected populations, as measured by the Deviance Information Criterion (DIC). The DIC is a value representing how poorly the data fit the statistical model, with lower numbers indicating a better fit. It has no direct, absolute, interpretation; rather should be interpreted to inform the choice of models that produce the lowest DIC. In this case, the multivariate modelling technique reduced the DIC from 180 to 105. Adding further complexity to the procedure did not reduce the DIC any further, and a combination of the statistical and theoretical advantages of the second technique led to this being the favoured approach.

The study of one other model is worth noting particularly. The Santolaya model showed a good ability to differentiate between low- and high- risk groups when considering a wider definition of 'serious infection' where it was developed and tested in Chile; LR [low] = 0.17 (95% CI 0.12 to 0.23), LR [high] =2.87 (95% CI 2.43 to 3.38). However, when the rule was applied to data collected in Europe it showed very poor discriminatory ability, well outside of that expected by chance variation. This highlights the need for models to be evaluated within different geographical settings, as undetermined factors may vary the diagnostic utility.

Unlike the first review, which focussed fully upon CDR at the point of presentation with FN, the update review also examined CDR with applied criteria to information applied beyond this. These showed that re-evaluation at eight to 16 hours [144] or 48 hours [154] was more efficient that initial examination, probably explained by the declaration of initially occult infections within the first few hours of admission.

Studies using biomarkers to predict general infectious complications

The predictive value of serum markers of inflammation and infection in children presenting with febrile neutropenia was studied in an updated systematic review that included a total of 38 studies, examining 24 biomarkers. Of these, 37 provided quantitative data and 22 studies could be included in the meta-analyses.

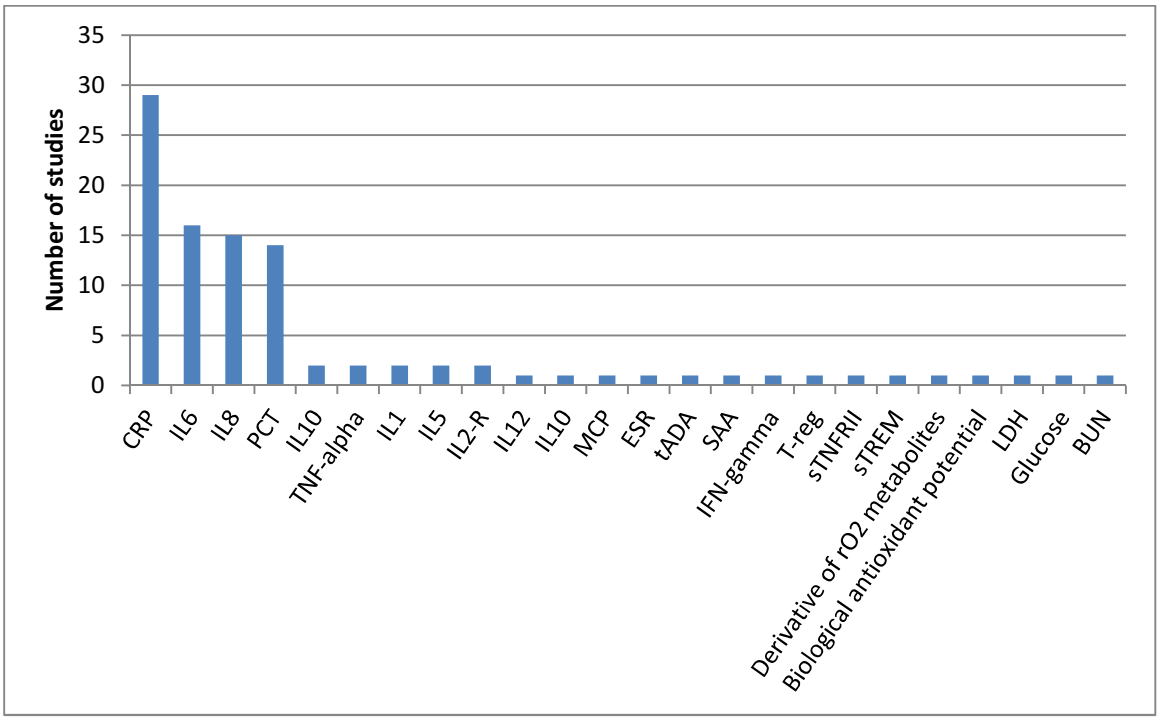


Figure 30: 'Inch -deep, mile-wide' approach to biomarker investigation

The studies presented similar methodological challenges to the decision rules review and had problems of reporting and analysis.

It was seldom reported if the test was interpreted 'blind' to the results of the outcome analysis, and *vice-versa*. Most studies failed to assess if the marker had any supplementary value over and above the simple admission data or clinical decision rules. In itself this does not undermine the interpretation of the predictive value of the marker; it merely reduces the ability of the healthcare practitioner to understand how to value this information when combined with the clinical knowledge they already possess.

As with the clinical decision rules, analysis of the data was frequently undertaken at the level of independent episodes, taking no account of the potential of multiple admissions for the same patient being present. When undertaken, the comparisons used appeared underpowered to detect

small but meaningful differences. Missing data were not examined for the nature of their absence, and no attempt at imputation was reported.

The studies frequently used different test cut-off values to report their findings, and these were largely driven by the dataset from which they were then applied. In these cases, the estimates produced are likely to be significant overestimations of accuracy, as data driven choices best describe the dataset they are derived from rather than estimate the data structure of the wider population. The use of previously defined cut-off values (in six studies) probably provides more trustworthy estimates. Unlike the CDR reviews, there were moderate event-per-variable ratios and few assessments of multiple outcomes within these studies as generally a limited number of potential predictors were under investigation.

Quantitative pooling of the results of the studies presented challenges of sparse data in specific and different subgroups, producing great uncertainty in pooled estimates.

In the included studies, a series of cut-off levels are reported to predict selected outcomes with the marker in question. Pooling these different levels into a single estimate of 'test effect' is meaningless: the estimated sensitivity and specificity do not have a clear relationship to a measurable cut-off value. One approach would be to only use a single cut-off value, but with so few data points this 'wasteful' approach is extremely unhelpful. A more useful approach is to create a hierarchical summary receiver operator curve (HSROC) which describes the average ROC curve derived from the individual curves produced from each study. In this way, it describes the 'average' relationship between a continuous cut-off value and discriminatory ability in the 'average' population. This is unlike the setting of artificial scores generated in a clinical decision rule, where the ordinal cut-offs do not reflect a continuous variable. A reasonable alternative to the HSROC approach would be to undertake a series of summaries at the variously reported cut-offs, making sure the data are only used once for each study by creating a series of *2*k* tables, where (*k*-1) is the number of cut-offs.

The functions used to create the HSROC take the data points from different studies as reflecting a series of individual ROC curves that vary between studies because of sampling, population and

threshold variation: the key elements of multivariate meta-analysis. The summary drawn from this maximises the fit of a curve combining the individual curves: a 'hierarchical' summary ROC. The function does not take into account the actual value of the thresholds. This is frequently reasonable, as it is impossible to quantify the thresholds used by different operators to call an X-ray 'positive' for pneumonia or a vessel 'compressible' on ultrasound examination and so demonstrating blood flow and ruling out thrombosis. In cases where the values are known though, an ordered relationship should be possible to determine.

A technique to undertake this ordered pooling was undertaken with the same meta-analysis technique developed on the systematic review of clinical decision rules for risk prediction in febrile neutropenia. It estimated the true proportion of diseased or non-diseased individuals in each category, constraining each cut-off to be generated from data specific to the reported value of the serum marker, and linking each cut-off with a multivariate normal distribution to reflect different population samples.

This approach failed to produce meaningful results for the ability of IL6, IL8, PCT or CRP to distinguish patients who developed a documented infection from those who did not. This is likely to be due to the massive heterogeneity of the data and the small number of data points available to estimate a large number of model parameters.

In two studies [79, 139] where adjustments were undertaken for other elements of clinical information, CRP added to the predictive ability of simple decision rules. Given this, and the unconfirmed impression of better predictive ability of the other serum markers, it is reasonable to hypothesise that these will add even greater benefit to clinical decision rules.

Direct comparisons of the different biomarkers were very limited, and unsuitable for meta-analytic pooling. They suggested that PCT or IL8 may be better than CRP, and that CRP may have a small additional value above clinical examination. Data for the other markers were too sparse to reasonably interpret. These conclusions should be read with the understanding that these are uncertain and unstable, and only small amounts of new data may substantially alter the findings.

Studies to detect radiographic pneumonia

As a by-product of the initial systematic review of clinical decision rules [113], four studies [149-152] were identified that examined the role of clinical examination in excluding pneumonia. They were undertaken in similar clinical populations which allowed meta-analysis of results and pooling of the higher quality studies using a classical binomial random effects model produced imprecise estimates of sensitivity 75% (95% CI 56.4% to 93.6%) and specificity 67.9% (95% CI 55.9% to 79.9%).

The implications of these results are that for populations with a similar prevalence of pneumonia (~5%), the absence of signs or symptoms of infection on clinical examination produces a post-test probability of pneumonia of about 1.5%. Given low level of risk, this can justify the routine withholding of chest radiographs to children who do not have signs or symptoms of lower respiratory tract infection. This will reduce the cost, resource demand and exposure of the child to radiation. However, the clinician must remember that a number of children will have an occult pneumonia and chest X-rays undertaken in a patient with an unresolving fever may be fruitful despite an absence of signs.

The conclusions we reached in a published review [117] were incorporated into national [202] and international [129] guidelines for the management of FN, which recommend only undertaking chest radiography in the setting of specific clinical indications.

Conclusions

The reviews undertaken and updated for this project demonstrated that a wide range of CDR for the prediction of poor outcomes during episodes of febrile neutropenia in children had been derived, and that there was potential for additional value to be gained from the incorporation of serum biomarkers. None of the rules identified had been subject to the extensive geographical and temporal discriminatory validity assessments that mark the highest quality CDR, and many potential difficulties with different outcomes, variable selection and model building were identified. Many of these issues arose from the challenges of combining the aggregate information presented in printed reports of the studies undertaken.

To maximise the value of the information already collected by these and other cohorts of children with FN, an individual-patient-data (IPD) meta-analysis was justified. This was required to develop and test new and existing prediction models; enable the construction of 'true' ROC curves based on the original data; allow comparison and alignment of different clinical outcomes; and accurately assess the effect of within-patient clustering of episodes. The effective added-value of markers to clinical rules could also be measured more comprehensively. The next sections of this thesis explore in detail the theoretical and practical methods used in forming the collaborative and undertaking the IPD analysis, and report the results of the main analyses in detail.

This intention of this endeavour is to provide a firmer basis for stratified treatment, either in the context of randomised trials of reducing intensity and duration of therapy for those at low risk of severe infectious complications, or of novel methods of early support for those at highest risk. Only in the collation of large quantities of data can we seek to address such questions in this common and occasionally fatal complication of therapy.

Chapter 6: Methods for the Individual Participant Data meta-analysis

Previous chapters identified a wide range of rules that have that been developed to predict poor outcomes during episodes of febrile neutropenia in children who have been treated for cancer. None of these has been subject to the extensive geographical and temporal discriminatory validity assessments that mark the highest quality CDR. The systematic review of these existing CDRs identified many potential difficulties with different outcomes, variable selection and model building and consequently was unable to reach any firm conclusions. A complementary systematic review of studies of serum markers used similarly to predict outcome found similar problems of extremely heterogeneous data and only tentative conclusions could be drawn.

The problems identified are inherent to meta-analysis of aggregate data. Limitations of reporting in published studies mean that we do not have access to the exact distributions of data, or the full range of univariable estimates of predictive power. These issues could have been partially addressed by collecting more detailed summary data from the authors of the original studies. However, this would not allow cross-study validation of different rules or alternative rule building. To meet these challenges, and to maximise the value of the information already collected by these groups and in other cohorts of children with febrile neutropenia, we initiated an international collaborative systematic review and individual participant data meta-analysis. This was intended to enable us to develop and test new prediction models in order to provide a firmer basis for risk stratification, including deriving a simple clinical decision rule, and to test existing rules. Subsequent to this formulation, treatment trials in this common and occasionally fatal complication of therapy could be undertaken.

Rationale for individual patient data meta-analysis in risk stratification in febrile neutropenia

Individual patient data meta-analysis in therapeutic studies has been developed over two decades to improve the precision and reliability of answers to questions of treatment.[203-204] More recently, the approach has been promoted for the synthesis of diagnostic[205] and prognostic[206] studies to

improve the quality of answers to important prognostic questions [111] and matters of diagnostic accuracy.[207] These techniques have been applied to real world clinical datasets [208-209] where they have clarified existing understanding of particular prognostic variables and enhanced an understanding of how different diagnostic tests can be used.[210]

Failure to approach meta-analysis and prediction model building in a coherent and technically sound way does not just lead to mathematical or statistical problems. Failure to address the problems of statistical interpretation has clear and real clinical implications.[211] Systematic review and use of summary prognostic data may be unreliable as the published data may be incomplete (missing vital information for meaningful meta-analysis)[78], and appear very susceptible to significant publication bias (with prognostic markers showing 'highly significant' responses being more likely to be published).[212] It has been suggested that the use of IPD in predictive settings may be even more valuable than in therapeutic reviews.[78]

It has been shown that smaller published studies are much more likely to demonstrate powerful relationships [71, 109] and nearly all studies of prognostic markers in cancer are 'positive'.[76] These problems are compounded by widespread over-citing of articles with high and unrepresentative predictive values[213] and the selective reporting of specific outcomes with 'significant' associations[76].

These problems suggest that the classical systematic review approach will have the potential to introduce greater problems that it solves, and any approach to such analysis should clearly account for these potential difficulties. One method is to use a clearly defined and 'complete' population of studies (e.g. the EORTC breast cancer marker studies[214]) another is the use of only large published studies (e.g. the Fibrinogen Studies Collaboration[215]), both aiming to avoid publication/selective reporting biases.

In the realm of therapeutic assessments, there is clear empirical research demonstrating that study design affects outcome.[216] The issues of study design and the introduction of bias have also been assessed by empirical research in predictive studies, but with less conclusive results. Kyzas has examined a series of 20 meta-analyses and evaluated how they assessed potential sources of bias,

and the effect they had on the overall conclusion.[76] This failed to show a significant effect of any of the study design measures they examined (blinding, prospective/retrospective, outcomes, time period, assay description or reference). This finding leads to the conclusion that there is no clear reason to exclude studies purely on the basis of their study design.

Equally important are the harmonisation of study data sets. This will allow the standardisation of endpoints, where the reports in papers show inconsistent reporting of ostensibly similar outcomes, or differing assay methods for proposed markers.[78] In a similar way, these benefits apply to issues of diagnostic accuracy.[210]

A further challenge avoided in the use of IPD is the un-categorisation of continuous outcome variables, the categorisations of which themselves may have be biased[80], driven by 'significance' based testing.

This is allied to the frequent use in primary studies of multiple data-driven analyses. The method used in the IPD analysis is based on firmly pre-specified potential predictor variables, built upon the clinical experience of the collaborative group and the systematic reviews explored in the preceding chapters. This guards against purely data-driven analyses which have a tendency to over-estimate any predictive value.[111]

In the reviews, we found the studies building a CDR used a large number of variables (median 13, range 2 to 39) and had a small number of events (median 36, range 4 to 178) with 76% studies having fewer than ten events per variable under consideration, and no study having more than 14 events per variable. These low event-per-variable ratios make predictive conclusions drawn from them to be unstable, and estimates of predictive power to be over-optimistic.[83] A collaborative IPD approach allowed us to consolidate the information and greatly increase the number of events studied from the same number of predictive variables.

The raw data also allowed a detailed analysis of the clustering of events (multiple episodes per patient) and variation at the level of the individual patient. This issue is significant when assessing

the problems identified in the aggregate data reviews. Multiple episodes in individual patients were treated primarily as if they came from unconnected individuals in most of the CDR and serum marker studies, which may have been inappropriate.

The functional form of the data, examining *a priori* non-linear/fractional polynomial relationships, can be assessed in detail in a large IPD analysis. No study assessed in the systematic reviews attempted to fit non-linear forms to the data. This was unsurprising, as the development of practical techniques to undertake this was very recent.[198]

Finally, IPD allowed us to not only test existing rules and combine data which have attempted to examine the rules, but potentially develop a more robust rule for future use worldwide.

In summary, the key benefits of prognostic IPD analysis generally are that:

- Analyses are not restricted to those of the published results or subgroups
- Analysis techniques, inclusion criteria and outcomes definitions can be standardised across studies
- Larger numbers of data points allow more powerful statistical conclusions to be drawn, including checking modelling assumptions
- The detailed data allows assessments to be made to account for missing data at the individual-level
- IPD can model data more appropriately, for example analysing continuous variables on their continuous scale (unlike in many prognostic studies, where such variables are reported categorised)
- Analysis can account for clustering (e.g. of patients within studies) and correlated information (e.g. multiple events per individual)
- Multivariate models can be created across differing health care settings
- Individual data sets can be reviewed for completeness and accuracy
- The analysis can provide extensive internal cross-validation to guard against data-driven exaggerations of predictive power

Forming the Collaborative Group

The “Predicting Infectious ComplicatioNs In Children with Cancer” (PICNICC) collaboration was formed around a nucleus of an international group of clinical experts who I had met and discussed potential collaboration as part of the development of the MRC fellowship proposal with. The systematic reviews described in the preceding chapters identified further key studies and researchers who were then invited to join the collaborative group.

Oral presentations on the problem of risk stratification in febrile neutropenia at separate conferences for the Royal College of Paediatrics and Child Health (2009) and the International Society of Paediatric Oncology (SIOP, 2008) also identified further studies and partners. Parallel to these approaches to clinicians and researchers, there was an integrated move to include parents/carers in the Collaboration (see next section). These presentations led to further interested groups contacting the Secretariat.

Following the SIOP presentation, the slides were placed on the international, though developing-world focussed, paediatric oncology website "Oncopedia". From this, I was approached by three more groups and located another group working on FNP stratification and through a mutual colleague, approached the main author, who also agreed to join the collaborative. The Centre for Reviews and Dissemination (CRD) website hosted the project page, which drew in one further group.

Ethical approval was obtained from University of York Health Services Research Ethics and Research Governance Committee, and from York NHS Research Ethics Committee after considerable input into assessing the ethical implications of IPD projects like PICNICC (see subsequent section for detail).

A full draft protocol was presented at SIOP in 2010. Following the presentation and distribution of copies of the IPD protocol, letters of invitation were sent by email and paper to principle authors (of the studies identified in the systematic reviews) and those not already engaged. This generated further contacts from follow-up emails to this group and their contacts and then included. A complex

series of approaches and telephone conferences also led to the inclusion of data from 4 EORTC trials. The flowchart of how study groups were contacted and their involvement requested is detailed in Appendix 14, and examples of the nature of the documentation in Appendix 15.

An important element in confirming the nature of the relationships between the data and the collaborative group was to set a clear publication policy. It was agreed that the main results of the meta-analysis would be published and presented in the PICNICC name, comprising groups supplying data for analysis and the Advisory group. Any subsequent technical papers which describe innovations in the methodologies used in the meta-analysis would acknowledge the Collaborative as the source of the data.

Rationale for parent/carer involvement

The development of shared research initiatives between patient/clinician/researchers has been a notable change in the practice of clinical research over the last decade.[217] It remains shocking to many researchers, clinicians and patients to learn that their views are often strikingly different than each other.[218] A systematic review of studies which describe the process of research planning and priority setting undertaken by the James Lind Alliance [219] demonstrated that the involvement of patients and parent/carers was extremely infrequent.

The PICNICC group has sought to involve parent representation from early in the process. Experiences of other researchers who had engaged patients in IPD collaborative were sought. Clare Vale, MRC Trials Unit, had worked with women in a cervical cancer IPD collaboration [220] and found the patient experiences redirected the focus of the IPD group onto many patient important elements. The benefit of lay involvement in improving the clarity of presentation of information and structure of investigation has also been suggested.[221]

Initial approaches were made to the Chair of the Patient Advocacy Committee (PAC) at the Children's Cancer and Leukaemia Group (CCLG: a charity networking parents, clinicians and allied health professionals in the UK and Ireland who treat childhood cancer) and to the Manager of

Candlelighters (the Yorkshire Children's Cancer charity) to seek their advice and suggestions for volunteers. From these meetings, a lay summary of the project evolved, written in a 'journalistic' style and commented upon by the CCLG PAC and Play Leaders at the Leeds children's cancer unit (see Appendix 16).

The project request for volunteers was highlighted by the CCLG PAC team and a short article printed in the Candlelighters monthly magazine, and a web page made available from the CRD site (http://www.york.ac.uk/inst/crd/projects/risk_stratification_febrile_neutropenia.htm). This led to the involvement of two parents, one of whom had experienced the death of her child, one whose child had been free from disease for over four years.

The involvement of these individuals led to a discussion about the nature and parent/family view on the understanding of risk in the setting of febrile neutropenia. After involvement in the initial refinement of the protocol, one volunteer withdrew, but the second continued and inputted through the process including attending and taking part in the Collaborators meeting. It was clear that the representatives involved had not wished to be actively involved in the process of systematic reviewing, data extraction or analysis, but added opinions to discussions about the nature of the adverse effects of FNP and provided their own professional (non-medical) expertise in advancing the project, particularly in respect of ethical issues and dissemination of data.

Ethical and regulatory considerations

It has been suggested that the re-use of individual participant data from randomised trials within meta-analyses that address the same clinical questions should be exempt from further ethical review requirements. This is because the data are from studies which have already obtained individual consent and ethical approval.[215, 222] The use of data that had been obtained outside specific research studies, or where the meta-analysis has different aims, remains unclear. A consultation exercise undertaken in 2008 by the National Cancer Research Institute (NCRI) found that the belief of most respondents was that material and data collected from cancer patients should be used, without identifiable information, as broadly as possible and that retrospectively seeking consent was inappropriate.[223] The European Treaty on Biomedical Ethics permits the use of data without specific consent [224] (15.2.i/ii) where there is minimal risk and potential benefit to similar persons

Within the UK, legislation controls the use of patient data for the purposes of research, most recently the National Health Service (NHS) Act 2006. This has been interpreted by the UK Medical Research Council (MRC) and summarised in a guidance document. These guidelines state that where possible, data should be released under specific consent. Where this is impractical, anonymised data should be used, and if this is impossible then an application to the Ethics and Confidentiality Committee of the National Information Governance Board for Health and Social Care is required to obtain access.[225] Wherever data is used that has not had specific consent, consent should be sought from an appropriate Research Ethics Committee (REC).[226]

The data sought for the "Predicting Infectious ComplicatioNs In Children with Cancer" (PICNICC) Collaborative IPD review was anonymous (i.e. the Collaboration could not identify the patient from their data) and unlinked (i.e. their data could not be mapped onto a subsequent dataset, with the potential for breaking anonymity). The project Advisory group could not conceive of any harm that may have been occasioned by the use of such anonymous, unlinked data, and that there was a considerable benefit of an improved risk stratification system for episodes of febrile neutropenia for children and young people with cancer. This view was also supported by the parent representatives in the collaboration. Data were sought from formal randomised controlled trials and prospective observational studies, and also informal studies of data routinely collected in clinical practice or as part of quality improvement projects. The transfer of the information from the original researchers to the Collaborative was requested by secure, encrypted electronic methods.

Within the UK, it was considered the project would require NHS REC approval for the use of patient data that had been recorded without specific research consent. Similar processes were discovered to apply in Australia [227], New Zealand[228], and Canada.[229] In other locations (such as Germany or the United States of America [230]) such data are exempt from the need for formal REC approval, but researchers are advised to have such protocols reviews by ethics boards to assure quality and ease publication.[224] In respect of this, we applied for and received approval for the PICNICC IPD protocol from the University of York Health Services Research Ethics and Research Governance Committee, and from the York NHS REC, both of whom determined that a full application was not required and gave consent from the Chair.

Methods

The full protocol of the PICNICC IPD analysis is provided in Appendix 17. It was developed, registered and published prior to commencement of the analysis.[231] The remainder of this chapter outlines the key methods and most important aspects of approach and analyses, paying particular relevance to the part of the PICNICC project undertaken for this PhD submission.

Aims

The primary clinical aim of this IPD analysis was to quantify the risk of adverse clinical outcomes according to clinical variables in children and young people undergoing treatment for malignant disease who present with an episode of febrile neutropenia; i.e. to identify which variables are prognostic, and which have the most independent prognostic importance. This was planned to lead to the development of a new risk prediction model containing multiple prognostic factors in combination, and permit this to be validated.

A further aim was to develop and explore practical and methodological issues around the use of pooled IPD analysis in the development of prediction models, and in the graphical display and communication of such information.

Inclusion and Exclusion Criteria

Studies were considered for inclusion in the IPD meta-analysis if they were:

- cohort studies of children and young people
- presenting with febrile neutropenia
- with either prospective or retrospective data collection, including randomised trial data
- provided data for all essential predictive variables in >50% of included episodes
- provided two or more study-defined-outcomes in >90% of each individual episodes of FNP

These criteria were selected to efficiently gather information which would inform the better understanding of the predictive ability of a range of pre-specified factors, chosen from the systematic reviews conducted to underpin this investigation.

Studies were excluded if they:

- Were case-series (for example, of only 'gram negative bacteraemia')
- Did not record data on all 'essential' predictive variables or could not provide sufficient outcome data

These exclusions were intentionally minimal, and produced to remove datasets which could not be informative about the outcomes of their patients or contained so few of the predictive variables that they would not be able to be used in developing a prediction model.

Studies were included which focus on collection of data from children and young people (between 0 and 24 years old). The inclusion of young people up to the age of 24 years is to address a paucity of research on individuals in the 'young adult' age range.[232] Data from individual patients aged 25 years and older were excluded from this analysis. The median age of inclusion in the children's cohorts examined in the systematic reviews reported in the previous chapters was around seven years old (ranging from one month to 23 years), and the adult study from the MASCC group [13] has a median age of 52 years (ranging from 16 to 91 years old).

Mapping Procedures

In order to harmonise outcomes and maximise the usefulness of the data to be collected, a series of *a priori* mapping procedures were planned, in consultation between two clinical experts (RSP and Julia Chisholm). These procedures were verified by the collaborators' meeting.

The procedures undertaken included microbial infection types/sources and classification in to severe/non-severe, summarising the "intensity of chemotherapy", addressing issues with different approaches to reporting "vital signs" and clarifying the value of continuous variables when they fall below the limits of assay detection.

Mapping of microbial sources to outcomes

The principle of this process was to create an *a priori* list of microbiological infections/sources which can be used to classify infections as 'severe' or 'non-severe', as detailed in Appendix 18. While this will never be a perfect system – for example some patients with *Pseudomonas* pneumonia may not

be significantly unwell, where some children with rhinovirus infection may be severely unwell – examples which cross these boundaries will be rare. In some cases the information on the microbiological outcome was supplemented by clinical site information in the dataset, in others it was not.

The rationale for this mapping was to create a homogenous and unconfounded outcome; one unaffected by any therapeutic manoeuvre, for use in model building and verification.

Mapping of chemotherapy Intensity

A range of treatment intensity approaches have been previously described, including variations of the "intensity of treatment rating scale".[233] The data delivered to the PICNICC collaboration contained a range of information, from the highly specific sub-elements of treatment courses (e.g. BFM acute lymphoblastic leukaemia (ALL) Induction Phase Ib) to the general (e.g. "More intense than ALL maintenance treatment").

In view of this, a three-intensity plan was undertaken to homogenise the information and maximise the quality which was included:

- Equal to or less intensive than ALL maintenance
- Standard chemotherapy more intensive than ALL maintenance
- Stem cell transplant procedures

Mapping of Respiratory & Circulatory results

A number of studies provided continuous variables (heart rate, respiratory rate, blood pressure) where a number of other data sets provided statements of respiratory or circulatory compromise. Where a description of respiratory or circulatory compromise has been given – for example, by explicit statement of use of supplemental oxygen – this has been used. For those where continuous variables alone were given, a mapping exercise was undertaken.

Normal children have been extensively studied for the variation and distribution of respiratory and circulatory parameters, with the development of centile charts for such variables. In view of the extreme nature of compromise in respiratory rate (where tachypnoea alone is not always associated with a failure of gas exchange or the need for other support) those greater than the 99th percentile [234] have been mapped to "compromised". For blood pressure, the lower 5th percentile has been

used (in keeping with the definition of systemic inflammatory response syndrome[235]), calculated as age(yrs)*2 + 65 mmHg.[236] The latter approach assumes a 50th percentile height of patient.

Mapping of biomarkers and age

A number of studies provided age as months; other studies provided the information in days. To convert to a common metric, the months data was multiplied by (m/12 * 365.25). Rounding is assumed to have happened both up and down, so that a 10.6m old would have been recorded as 11m, as would a 11.4m old, making the 'round' month the mid-point.

Inflammatory marker continuous variables (e.g. CRP, interleukins, PCT) have a log-normal distribution.[156, 237] For values below assay detection limits, the mean of the log-normal of the distribution of the 'counted' values was taken as the true mean, and the proportion of patients below the limit of detection calculated, with the median 'unmeasured' value imputed for all those below the cutoff.

Core dataset and variables

The predictor variables and adverse outcomes sought from studies were based on our systematic reviews of aggregate data and clinical experience.

Predictor variables requested were divided into '**essential'** and '*desirable'* items and categorised as patient-related, episode-related-clinical and episode-related-laboratory variables.

Patient-related variables

- **Age**
- **Underlying tumour type**
- *Marrow involvement/remission status*
- *Chemotherapy type and time elapsed since last cycle*
- *Presence of central venous line*

Episode-related clinical variables

- **In-patient or out-patient at onset of episode**
- **Maximum temperature**
- **Antibiotic therapy used**
- *Respiratory rate (or compromise)*
- *Circulatory (or compromise)*
- *Severe mucositis*
- *Global assessment of illness severity*

Episode-related-laboratory variables

- *Haemoglobin*
- *Platelet count*
- **White cell count**
- **Neutrophil count**
- *Monocyte count*
- *C-reactive protein*
- *Procalcitonin*
- *Interleukin-6*
- *Interleukin-8*

The outcomes of primary interest from each episode were:

- Death
- Intensive care admission
- Need for moderate organ support (fluid bolus, oxygen)
- Clinically documented infections
- Microbiologically documented infections

To be eligible for inclusion, studies had to be able to provide two or more of these outcome measures for at least 90% of episodes.

If available, data were also requested on:

- Duration of fever
- Duration of admission

The adequacy of data sources was assessed with an initial survey of data available from collaborators is provided in Appendix 19.

Providing Data

Anonymised de-identified data

Datasets were requested in anonymised format with all directly identifiable material such as name, address, postcode, medical number removed. A patient identification number was requested to facilitate communication and data queries. For the purposes of the analyses planned, the age of the patient (an indirect identifier) was considered essential, and requested to be provided [238] despite some concerns that in small population this could provide a potential patient-level identifier.

Data format

The data were accepted in any electronic formats, but the ideal was a 'flat' spreadsheet format (such as Excel), with one episode per row and variables in columns. To make the cleaning and checking of the data as straightforward as possible guidance on data provision was provided, such as "Each patient should have an in-cohort unique identifier (such as a simple number 1,2...n) to highlight repeated episodes in the same patient". Suggested coding was also provided (Appendix 20) along with an example flat file. Data were re-coded on receipt to ensure consistency.

Transfer of data

Data were transferred using a secure password-protected web server (Dropbox.com) or via PGP-encrypted email. This permitted a secure and identifiable connection to the University of York servers and minimised the possibility of data loss.

The raw files were named according to a specified convention and archived as an unmodified record of the original provided data. Copies of the received files were made and used in the subsequent cleaning and analysis work (see Appendix 21; Data manipulation SOPs).

Data checking

Simple checks of data integrity were undertaken prior to analysis:

The first review was to confirm that the supplied coding sheet and data file corresponded, and to log any initial uncertainties. Data columns were reordered in line with the PICNICC master data file structure, and the presence of the essential variables and outcomes verified at the 'column' level. This was followed by recoding and examination for missing data.

Further data checks were undertaken in Excel:

- age checking (not negative or zero and not older than 9,125 days (25 years); consistency of patient DOBs, and sensible diagnosis & age relationships)
- episodic checking ordering by age and DOB and then by admission date and looking for odd/inconsistent elements (>6m in between FNP episodes)
- time-since-chemotherapy (not negative or >42 days), and looking for consistency with other episodes
- white cell indices (not 'zero', and components e.g. ANC and AMC are not greater than the total WCC)

Subsequent queries and their resolutions were recorded, and when finalised, the data source was locked. Any problems or inconsistencies flagged during these procedures were discussed with the individual responsible for each study and amended as appropriate by consensus.

Plan of investigation

Method of analysis

The key elements in the development of a good clinical decision rule are: high quality rule-building with unbiased data, sensible validation and assessment of generalisability, and finally implementation in a real-world clinical setting. The building of a rule requires the data to be collected without systematic errors, subsequent construction of an accurate prediction model, and the development, using clinical criteria, of an appropriately usable decision rule.

The construction of a prediction model could be accomplished using one of a series of different models. These different analysis techniques include: multivariable regression analysis, classification and regression tree (CART) models, and neural nets. There is no clear evidence that one method is superior to any other.[109] The chosen primary method of analysis for the PICNICC study was logistic regression as this has the widest clinical understanding and applicability.

Logistic regression modelling

This technique seeks to quantify the relationships between predictor variables and the chance of a specified outcome by estimating the relative likelihood of the outcome occurring with increasing values of the predictor.

The equation produced at the simplest level takes the form:

[1] $\text{Logit}(p_{ik}) = \beta_0$

$\text{Logit}(p_{ik})$ is the log-odds (natural logarithm of $p_{ik} / (1- p_{ik})$) of an outcome, for example, bacteraemia in the k'th patient of the i'th study.

β_0 is the intercept of the slope described by the equation – the expected log-odds of bacteraemia.

This equation is clinically meaningful only if each patient in each study has the same risk of bacteraemia, which cannot be predicted by any other factor. This is obviously not the case.

Adding another level of complexity is to try to explain some of the variation:

[2] $\text{Logit}(p_{ik}) = \beta_0 + \beta_1 t_{ik}$

t_{ik} is the value of a covariate, a candidate of a predictor variable (potential prognostic factor, for example maximum temperature) in the k'th patient of the i'th study.

β_1 is the co-efficient, alternatively understood as the 'weight', 'slope' or 'multiplication factor' for that explanatory variable. The value $\exp(\beta_1)$ gives the odds ratio of the candidate predictor; it compares the odds of the outcome for two episodes that vary by one unit of t_{ik}. This value, along with a 95% confidence interval, can be used to summarise the predictive effect of t_{ik}.

Categorical variables can also be used in these equations by the use of 'dummy' or coding variables.

This formula assumes that the risk of bacteraemia can, to some extent, be predicted from the maximum temperature prior to admission. The equation also presumes that the risk is the same in each study, and that the change in risk (per degree of temperature) is also the same – essentially treating the whole dataset as if it were one large study. This may well be untrue, with altering rates of bacteraemia across studies. To incorporate this, we need to add:

[3] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_1 t_{ik}$

β_{0i} is a parameter for each study, which can be thought of as allowing different intercepts , which are the baseline risks of bacteraemia, to account for differences in populations. This keeps the same slope of the temperature-bacteraemia line ('fixed effect' covariate).

If we suggest that there may be a real difference in the relationship between temperature and bacteraemia in different studies which is beyond that expect by chance sampling, then we must also allow the β_1 to vary in from study to study ('random effect' covariate):

[4] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_{1i} t_k$

$\beta_{1i} \sim N(\beta_1, \tau^2_{\beta 1})$

This introduces a second assessment of between-study variability related this time to the temperature covariate. τ is the between-study standard deviation in the β_1 values. If τ is zero, then β_1 is the same (fixed) in each study, and this equation becomes the same as [3]. In either case, β_1 is the effect of a degree-change in temperature on the risk of bacteramia, on average, across all the various studies.

In straightforward terms, estimating a fixed effect covariate assumes the same effect is present across each study, and any differences are due to chance sampling. A random effect covariate assumes that the estimates are drawn themselves from a normal distribution of true effects; that the estimates are both different by sampling, and that real differences may also be present between studies. The clinical interpretation of this can be difficult, if the heterogeneity in this estimate is large, as it means it becomes difficult to predict what the value of a one-degree temperature change is in any given setting. An exception to this is in settings where a study has contributed to the analysis; in these areas a reasonable estimate of the specific value of the covariate can be made more accurately using 'shrunken' estimates.

When multiple predictors are considered the equation stays very much the same, but adds in further covariates. In adding further covariates we move from assessing univariate models to multivariable ones:

[5] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_1 t_{ik} + \beta_2 m_{ik} + \beta_3 y_k$

Each further covariate (e.g. m = monocyte count) has a different corresponding co-efficient. Again, these may be fixed effects (as above) or allowed to have a random-effects estimate. It also allows some covariates to be entered at study-level only, for example the study 'year' may be entered as a potential explanation: $\beta_3 y_i$ (note the lack of 'k' in the subscript).

Combining these approaches allows random effects for some aspects of the model, for example, differences in the intercept and some covariates, and for other covariates assume the same slope across studies (fixed effects).

In clinical terms the multi-level model can describe how baselines rates of an outcome (e.g. proportion of bacteraemia) vary between studies – this is the intercept component. It can also explore which predictive factors have different strengths of influence in different studies – the 'slope' or coefficient. If a predictive factor is found to have strikingly different (heterogeneous) coefficients across different studies the potential explanation for this needs to be explored. If no consistent pattern emerges then the practical implication is that it is impossible to use this in a model which will be generalisable in future as the power of the predictor will be impossible to judge. The alternative situation, where slopes are similar between studies, strengthens the confidence in it being predictive in future practice.

Advanced issues [optional]

Further layers of clinically reasonable complexity can be added to this situation. The first is that there may be situations where the explanatory factor has qualitatively different effects in different settings. For example, it may be that platelet counts have little predictive value in areas of the world where transfusions are simple, cheap and safe, and have strong predictive value in areas where platelet transfusions are difficult to give. In this setting, the basic equation is similar, but the values of β_1 vary between studies where transfusions are common ($_c$) and uncommon ($_u$):

[6] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_{1Ci}p_k + \beta_{1Ui}p_k$

$\beta_{1Ci} \sim N(B_{1C}, \tau^2_{B1})$

$\beta_{1Ui} \sim N(B_{1U}, \tau^2_{B1})$

The logical extension to this variant is that $\tau^2_{\beta 1}$ may also differ between groups, allowing more or less variability in the platelet-bacteraemia relationship in high versus low transfusions settings. These are 'group random effect' models.

A further factor which may need consideration is the how predictions may differ between studies because of unidentified biases. For example, temperature may seem valuable, yet this is only a reflection of different studies having different mean values of maximum temperature (ecological bias).[239]

This can be assessed by looking at how the covariates of the 'study mean' temperature (β_{1A} for 'across study' and 't́ for mean-temperature) differ from those of the individual patient temperature (β_{1W} for within-study). Technically, the individual element component is 'centred' to make it more comparable by subtracting it from the mean study temperature. The equation becomes:

[7] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_{1A}\, t\acute{}_i + \beta_{1W}\, t_{ik} - t\acute{}_i$

This then allows a statistical test of the difference in β_{1A} and β_{1W} to be undertaken, with the null hypothesis that there will be no difference (i.e. there is no 'ecological bias').

There may also be situation where a candidate predictor variable (e.g study year 'y') may not by itself have any predictive value, but it alters how effectively monocyte count 'm' predicts bacteraemia. This can be estimated by looking at the 'interaction' term of $y_k * m_{ik}$

[8] $\text{Logit}(y_k) = \beta_{0i} + \beta_1\, t_{ik} + \beta_2\, m_{ik} + \beta_3\, y_k + \beta_4\, y_k * m_{ik}$

Assessment of study and data quality

At the time the PICNICC protocol was produced there was very little advice in the literature for assessing the quality of prognostic studies. Altman and Lyman have presented suitable criteria that those initiating a primary prognostic study should consider [240], and they suggest that every effort should be made to limit potential biases and to emulate the design standards of a clinical trial. Ideally the data should be collected prospectively, with little missing data for predictors or outcomes, and with pre-defined hypothesizes. We chose to use the format of QUADAS, as used in the systematic review, to help inform the assessment of the quality of the IPD obtained. The influence of any studies considered problematic (e.g. those with large missing data, or lots of incomplete follow-up) in the prediction model was also considered in the later analyses. Since the protocol was developed, there have been further publications exploring the assessment of bias in prognostic studies [241-242], and these issues have also been considered.

Model development

The protocol [231] for model development set out that after data checking for consistency, model building would initially incorporate the simplest predictor variables (malignant diagnosis, age, time since chemotherapy, and maximum recorded temperature) before standard additional variables (such as clinical assessments of compromise, in/out-patient status, white cell counts or other haematological parameters) were added. Further specialist tests (e.g. CRP and IL6 levels) were finally to be added. The type of antibiotic therapy used was always incorporated into model as a categorical variable in a sensitivity analysis.

Potential sources of heterogeneity (e.g. in effects of particular variables across studies, or by individual-level variation) were incorporated as random-effects when appropriate and the effect assessed. The models were assessed for improvement in fit using an Akaike's Information Criterion, with a p-value of < 0.15 used for inclusion; we use a 15% level rather than a 5% as we felt this was more conservative and would avoid missing important covariates. However, at the stage of determining our final model, we checked that the model's predictive accuracy (discriminatory ability) would be improved by the inclusion of variables whose significance was between 5% and 15%. If predictive accuracy was not improved then these variables would be removed.

This approach (of adding specialist tests only after considering the simpler tests) maximizes the utility of a model by ensuring that if extra tests with additional costs are required, they are shown to add considerable predictive power to existing simpler variables.[80] We used bootstrapping and shrinkage to adjust for potential over-optimism (bias) in parameter estimates.

The bootstrap procedure creates a series of 'new' datasets which are compiled from rows re-sampled from the original dataset at random, with replacement, i.e. allowing any individual patient-episode to enter the new bootstrapped dataset multiple times.[85] This is based on the principle of random sampling reflecting the true value of a studied item within a population, and simulates the expected random variations that will appear in when a prediction model would be used in clinical practice. These new datasets are then subject to the analyses which are under consideration. The results of these bootstrap analyses are examined and an average value, along with observed or calculated confidence intervals for each of the chosen parameters, can be drawn. For the analyses in this thesis, the bootstrap procedures were undertaken using R.

Shrinkage [243] is process of producing a reduction in the predictive estimates of a regression equation because there is an empirically proven expectation that prediction models generally perform less well in validation datasets than derivation ones. By applying a 'calculated pessimism' to the estimates this may be avoided. The approach used here follows the shrinkage after estimation approach using a heuristic uniform shrinkage factor *s*, calculated as

[1] $s = (\text{model } \chi^2 - p) / \text{model } \chi^2$

$\text{model } \chi^2$ = likelihood ratio of the fitted model

$\text{AIC} = \text{model } \chi^2 - 2p$

where p = number of fitted predictors in the final model, taken from the formulae of [243] Chapter 13 (p233 and p235).

Continuous candidate variables were assessed using the best fitting functional form considering appropriate transformations or fractional polynomials (also assessed using an Information Criterion) as suggested by previous evidence.

An analysis comparing the new model that we develop with other validated models, for example that of Santolaya [22] was also planned *a priori* to exclude data sets used to derive any of the models. This provided an opportunity to test these rules against data from other geographies and eras, particularly in light of the demonstration of lack of geographical transportability.[115]

The protocol acknowledged there would be unforeseen challenges caused by the variations in the data formats and completeness of studies, and acknowledged establishing the definitive analysis plan will be an iterative process and could even require novel methodological developments.

Assessing model performance

An important use of a prediction model is to classify patients into risk groups. The developed model will produce a risk score for each individual, based on their own predictor values. The calibration of the prediction model was assessed by placing children into deciles ordered by predicted risk and considering the agreement between the mean predicted risk and the observed events in each decile; the slope of this line should be one if the model and reality agree.

To produce a clinical decision rule (CDR) a cut-off value was required. In order to do this, the collaborators, including patient representation, discussed the value below which it would be considered acceptable to be termed “low risk” of bacteraemia at the congress. Through expert opinion, and in keeping with the previous publications of the SPOG group, a “5% risk of bacteraemia” was agreed. Further issues with this decision and approach are explored in Chapter 10.

The decision rule derived from this was simply to classify the output of the prediction model as being “low risk” if it was less than 5%, and “high risk” if 5% or higher. This was cross-validated by comparing the classification of each patient with their actual outcome, allowing an estimate of the sensitivity and specificity of the prediction model. Then, by varying the chosen cut-off level, a

receiver operating characteristic (ROC) curve summarising the sensitivity and specificity of the prediction rule across the range of cut-offs was produced by the R package pROC, along with an evaluation of the overall discriminatory ability summarised as the Area Under Receiver Operating Characteristic curve (AUC ROC) with 95% confidence interval.

The prediction model was tested by checking performance against the data from all, bar one, of the studies in turn (cross validation of intrinsic prognostic performance)[209] and using the bootstrap procedure.[81] This approach is intended to adjust for over-optimism in the estimation of model performance due to validation in the same dataset that was used to develop the model itself. The cross-validation approach (leave one study out at a time) has been referred to as internal-external validation, and is a way of maximising the data toward the prediction model development whilst also externally examining model performance. It tests the systematic biasing of the data in order to assess robustness to variations in study-level variation, for example population, geography and era. Both methods are limited in the reliance in re-using the dataset which derived the rule, but are as robust a method as possible in internally testing the rule.

The improvement in model performance by adding prognostic factors when deciding between more complex model sets was assessed by net reclassification improvement (NRI) [244]. This is a measure of the overall 'benefit' of a new classification model. It is calculated by taking patients with, and without, the outcome separately. Patients who are correctly classified with the new score, but were incorrect in the old one, are given a score of +1, and those who are reclassified incorrectly are scored -1. The unchanged are scored zero. The totals are summed, and divided by the number of patients in that outcome group. For patients with the outcome, this value is the improvement in sensitivity, for those without it is the improvement in specificity. These two values are then added together to give the net reclassification improvement. A larger value indicates a greater improvement.

Validation and future implications

A comparison of the predicted and observed event rates to assess calibration (as described above) and the area under the ROC curve to assess discriminatory ability in new data was proposed as a test bed for the newly generated model. However, such an analysis was outside the initial scope of this project.

These steps should produce the most precise and accurate prediction model that can be created from the IPD data set. The next step is to take these estimates to derive a clinical prediction rule from which management decisions may be made. This requires clinically informed decisions to be about where alternative strategies should be undertaken. For example, what risk of an emerging clinically documented infection would be acceptable before patients could be considered suitable for out-patient therapy? The setting of these thresholds can inform the rational derivation of a rule, along the lines suggested by Vickers.[245] Such decision will require more involved engagement from a wider panel of parents and young people, and should be subject to further detailed study.

Chapter 7: Description of the Individual Participant Data

Introduction

Preceding chapters have described the issues in managing febrile neutropenia, focussing on the possibility a risk stratified approach to initial management, to improve quality of life and not increase any infectious complications of anti-cancer treatment. An extensive analysis of the existing research evidence has been undertaken, including systematic reviews undertaken in 2008/9 and updates done in 2011/12. These led to the clear decision to progress to an IPD analysis to develop these ideas further and attempt to make best use and maximise the utility of the existing data sets.

This thesis reports the primary results of the IPD analysis, describing the development of the collaborative group, a description of the datasets demographics, and the results of analyses for the main outcome of "microbiologically documented infection".

The Collaboration

For this project we established the Predicting Infectious ComplicatioNs In Children and young people with Cancer" (PICNICC) Collaborative, the formation of which is described in Chapter 6. This consists of 22 different study groups from fifteen countries.

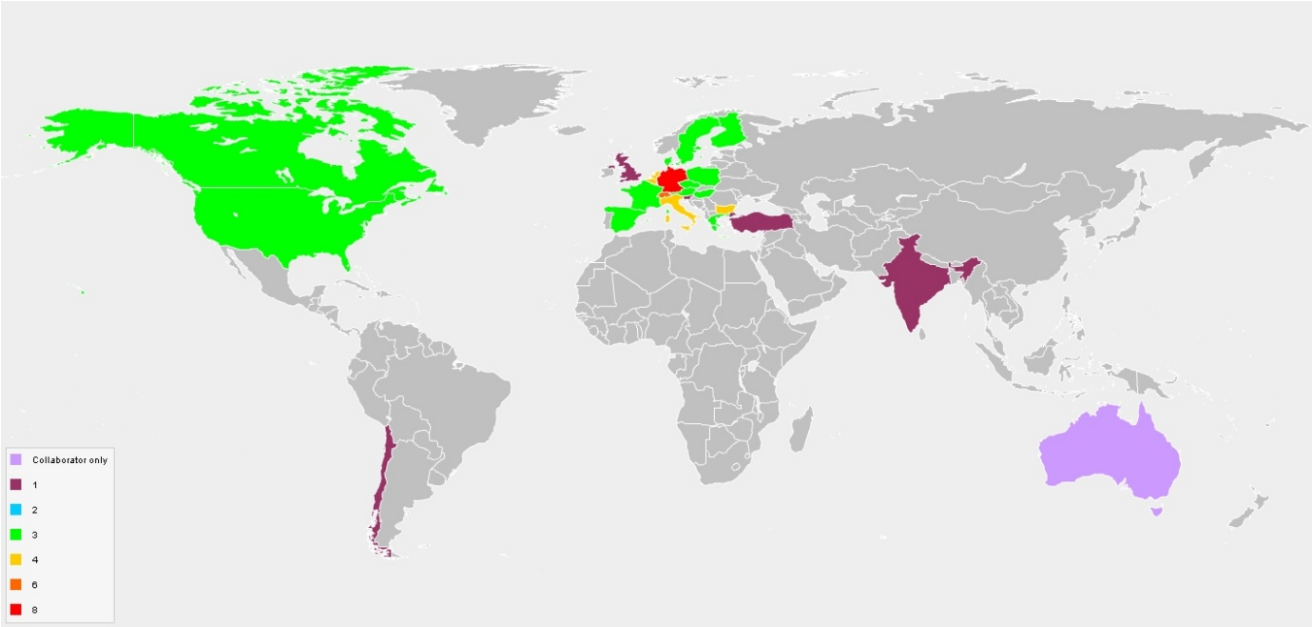


Figure 31: Map of the World indicating the location of Collaborators.

The PICNICC collaboration comprises those who have contributed data and/or for significantly developed the project. It includes paediatric oncologists & haematologists, infectious disease

specialists, statisticians and methodologists and parent/carer representatives. Each brings a different and important element to the discussions and direction of the Collaboration.

Current members are: Roland A Ammann, Thomas Kuehne , Felix Niggli, David Nadal (Switzerland), Ian Hann (Ireland), Lillian Sung, Robert Klaassen, Sarah Alexander (Canada), Thomas Lehrnbecher, Arne Simon (Germany), Karin Meidema,Wim JE Tissing (Netherlands), Neil Ranasinghe, Sally Amos, Susan Hay, Lesley Stewart, Bob Phillips, Daniel Yeomanson, Alex J Sutton, Richard Riley, Julia Chisholm, Rachel Dommett (GB), Elio Castagnola (Italy), Pamela Silva, Juan Tordecilla (Chile), Maria Spassova (Bulgaria), Hana Hakim, Glen Stryjewski (USA), Gulsun Tezcan (Turkey), Lidija Kitanovski (Slovenia), Ajay Gupta (India), Gabrielle Haeusler (Australia), Tiene Bauters, Geneviève Laureys (Belgium), Marianne Paesmann, Peter Donnelly (EORTC).

Ethical and Regulatory Barriers

We undertook an auxiliary investigation into the ethical and regulatory considerations involved in sharing IPD for risk stratification work, based on the ethical and regulatory principles and information collected and presented in Chapter 6. All 36 groups that were initially approached (including collaborators and those who expressed an interest but did not provide data) were surveyed about their experiences of the process. These results are summarised in the Table 13 and have been published elsewhere [246]. In some European countries and USA, specific applications were made and consent obtained to share the information. Other groups were able to share their data from previous investigations without further formal approval. To our knowledge, no potential collaborative group had their request to share such data declined.

Table 13: Consent sought to collaborate in an IPD analysis of predictive features

Country	Study type(s)	Ethics review board approached/answer
Belgium	Prospective study	Yes, from both University Hospital review board, agreed
Bulgaria	Prospective study	No, prior consent to primary study
Canada	Prospective study	No, prior consent to primary studies
Canada	Prospective study	Yes, from Institutional Review Board, agreed
Chile	Prospective study	No, prior consent to primary studies
Germany	Prospective study	No, prior consent to primary studies
Italy	Prospective study	No, prior consent to primary studies
Netherlands	Prospective study	No, prior consent to primary studies
Slovenia	Prospective study	Yes, from National Medical Ethics Committee, agreed
Switzerland	Prospective and retrospective studies	Yes, from University Hospital review board, agreed
Turkey	Audit	No, not required
UK	Audit	Yes, from NHS Research Ethics Committee, agreed
USA	Retrospective notes review	Yes, from Institutional Review Board, agreed
USA	Prospective studies	No, prior consent to primary studies

Unobtained data

We were unable to obtain data from 30 studies identified in our systematic reviews (see Table 14 for details; this is 58% of all identified studies and 51% of all identified episodes). Explanations were provided by the authors of three studies. In two cases (Riikonen, Heney), data were from studies conducted over 20 years ago and were no longer retrievable. In another case (Hodge), data were not provided by the group despite follow-up emails and a confirmation of interest in the project. The authors of the other 28 studies did not respond to our invitations.

Table 14: Studies where IPD was sought but not obtained

Study	Number of episodes
Adcock 1999	33
Baorto 2001	558
Barnes 2002	39
Diepold 2008	123
Dylewska 2005 a & b	108
El-Maghraby 2007	85
Gala-Peralta 2005	30
Hatzistilianou 2007	94
Heney 1992	47
Hitoglou-Hatzi 2005	67
Hodge 2006	31
Jones 1996	127
Katz 1992	122
Lucas 1996	161
Madsen 2002	76
Paganini 2007	981
Petrelli 1991	146
Rackoff 1996	72
Riikonen 1992	105
Riikonen 1993	91
Rojo 2008	33
Rondinelli 2006	283
Santolaya 1994	85
Santolaya 2001	447
Santolaya 2002	263
Santolaya 2007	373
Santolaya 2008	566
Secmeer 2007	60
Soker 2001	48
West 2004	143

The PICNCC project was focused primarily on the development and evaluation of a new CDR, and the aggregate data from these studies was not sufficient to be included in the analysis, for example, by utilising a two-stage approach to meta-analysis of parameter estimates. Apart from date of publication, where older studies were less likely to be included, there was no clear evidence of systematic variation between the studies included in PICNICC and studies from which data could not be obtained in terms of number of participants (p=0.66), number of episodes (p=0.93), number of events (p=0.67), direction of data collection (p=0.13) or geographical region (p=0.25). An IPD analysis of predictive factors differs importantly from a systematic review of treatments because the issues of within-study bias and publication bias appear much more troublesome in prognostics than in therapeutic trials. Also, it is less clear that all available evidence is required to be collected to produce the most accurate estimate of the chosen effect; rather a comprehensive and unbiased collection of information is preferred so that the IPD studies are a representative sample of the populations to which the CDR is to be applicable.

Quality assessment of the included studies

Quality assessment in prognostic/predictive studies is an area of ongoing methodological refinement. At the time that this protocol was devised there was no published guideline , although Hayden [242] has recently suggested a framework for such assessments. In keeping with the systematic reviews undertaken, the assessment of quality followed the QUADAS approach (see Appendix 22, Table 43). These are very similar to the assessments undertaken for the studies included in the published systematic review, as few extra studies were included (notably the EORTC trials).

There appear to be very few differences between the included datasets in the design features proposed to place studies at increased risk of bias (adequate population sampling, adequate reference standards and unbiased collection of prognostic information; see approach (Appendix 22, Table 43,

Table 5: Further informative QUADAS measures and Appendix 11. Full list of QUADAS criteria for included biomarkers studies). This is reassuring and to be expected in the clear and simple study structure of the collection of a cohort of patients presenting to hospital with a well-recognised cluster of symptoms such as FNP.

Overview of the data collected

Data collected prior to the cut-off in November 2011 for analysis in this thesis (the derivation dataset) included 22 datasets from 16 collaborative groups. These contained information from 5,127 episodes of FNP in 3,504 patients (see Table 15). The PICNICC collaboration aims to collect further datasets to undertake independent analyses of the CDR produced.

Table 15: Location and patient numbers per dataset

Study Group	Origin	Patients	Episodes
Alexander	Boston, USA	103	187
BaselSPOG	Basel, Switzerland	6	9
BernSPOG	Bern, Switzerland	69	171
BonnSPOG	Bonn, Germany	35	44
EORTC-XIV	Pan-European	149	149
EORTC-IX	Pan-European	315	315
EORTC-XI	Pan-European	301	301
EORTC-XII	Pan-European	21	21
Genoa	Genoa, Italy	259	703
Hakim	Memphis, USA	332	332
Kitanovski	Ljubljana, Slovenia	32	68
Klaassen	Ottowa, Canada	226	431
Lehrnbecher	Frankfurt, Germany	146	311
PINE	South-East England, UK	762	812
RetroBern	Bern, Switzerland	132	364
Silva	Santiago, Chile	30	52
Spassova	Plovdiv, Bulgaria	80	199
Styjewski	Washington, USA	56	56
Sung	Toronto, Canada	75	75
Tezcan	Antalya, Turkey	57	145
Tissing	Groningen , The Netherlands	114	258
ZurichSPOG	Zurich, Switzerland	72	154
TOTAL		3504	5127

Missing data

The issue of missing data was more significant than originally envisaged. This is almost entirely at the level of study whereby predictor or outcome variables were not recorded by the studies (see Figure 32). The proportion of partially present, partially missing, data was small.

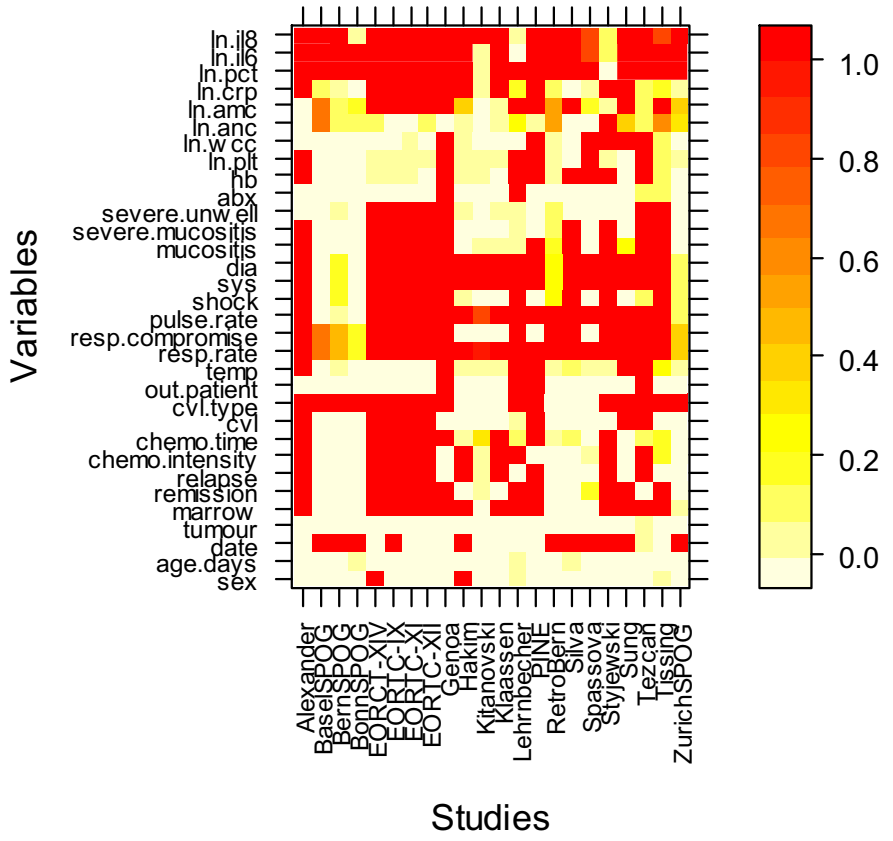


Figure 32: Per study proportion of missing predictors

The graphical representation of multi-dimensional information is a challenge. During this thesis, a variety of approaches have been used, and the most successful seems to be a variant of the "heatmap" approach (see Figure 32 and Figure 33) which have not been used widely in health care research outside of molecular biology, but do have a long tradition in social sciences [247]. This allows the pattern of response in the same variable to be assessed, or the pattern of information delivery by the same study. A colour-coded key displays the information semi-quantitatively, in this case the proportion of missing data, where the 'cool' paler yellows indicate very small quantities of missing data, the orange regions are approximately 50% missing, and the reds indicate high degrees of missingness.

A similar pattern of data absence and missingness is found in the outcome variables (see Figure 33).

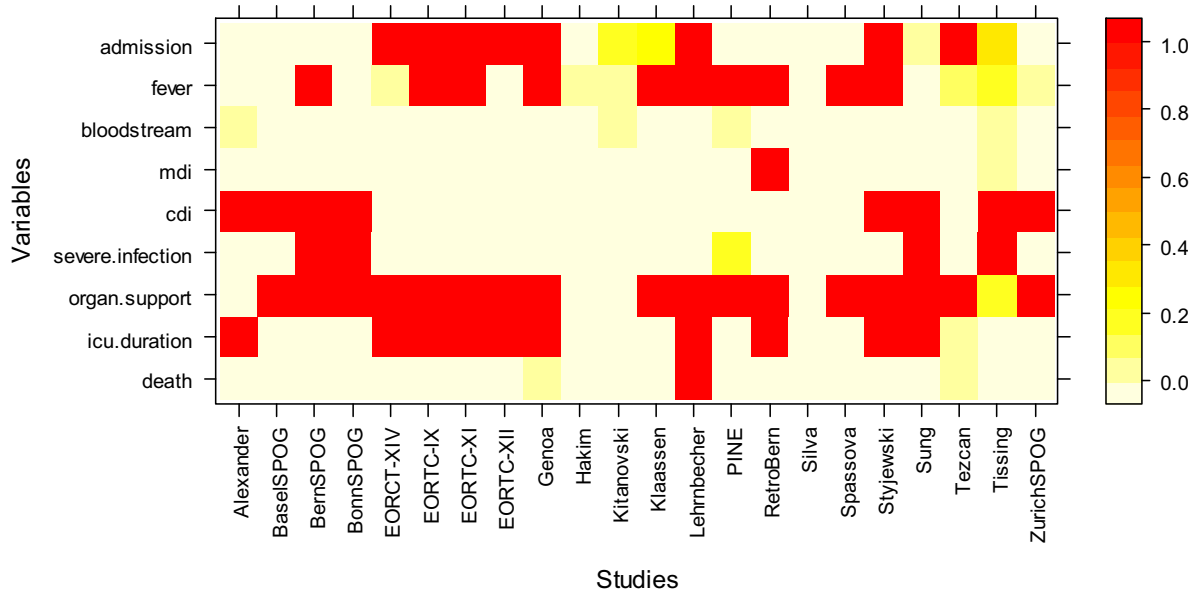


Figure 33: Per study proportion of missing outcomes (heatmap)

The visual impression of the data can be altered by changing the colour gradient, in this case (Figure 34) using a traffic light approach making all combinations of study/variable with more than 50% missing information red and those with less than 10% missing a shade of green.

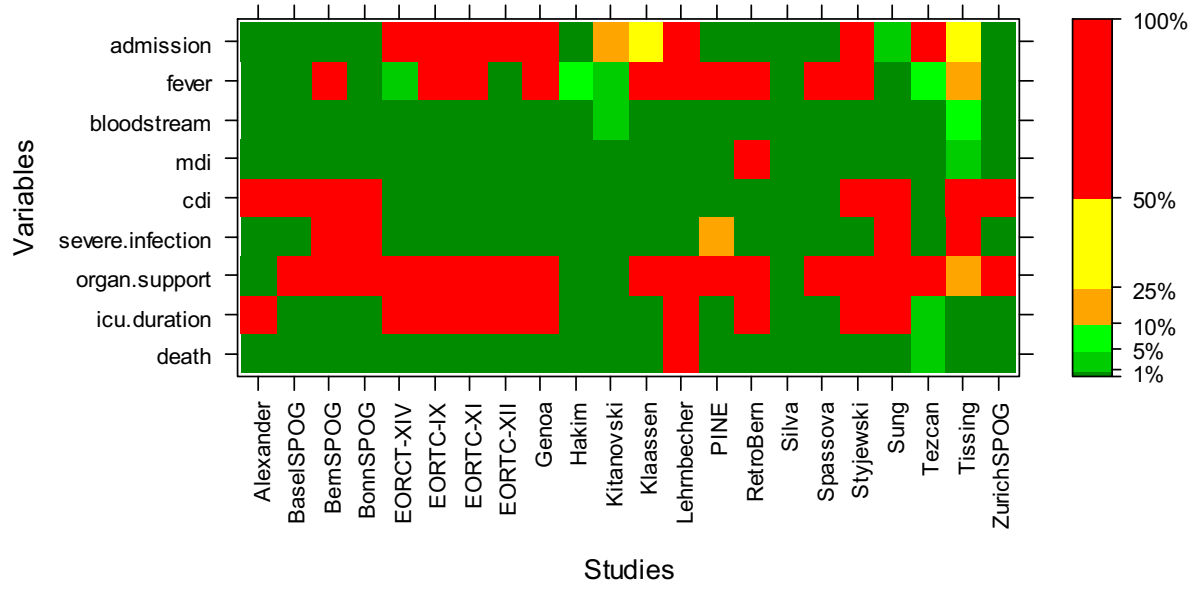


Figure 34: Per study proportion of missing outcomes ('traffic light' colour scheme)

The issue of missingness can also be considered per-episode-per-variable, ignoring the study-level element (Table 16). The missing data ranged from 4 out of 5,127 individual features (tumour type not recorded) to 5,034 out of 5,127 episodes (procalcitonin not reported), that is data being largely absent.

Table 16: Missing data (episodes in whole dataset)

Item	N missing	Proportion missing		Item	N missing	Proportion missing
sex	516	0.1		mucositis	3084	0.6
age in days	5	0.00097		severe mucositis	2228	0.43
date	1772	0.34		haemoglobin	2499	0.48
tumour	4	0.00078		platelets	2411	0.47
marrow involvement	4097	0.79		white cell count	2020	0.39
remission	3225	0.63		absolute neutrophil count	650	0.13
relapse disease	2750	0.53		absolute monocyte count	3468	0.67
chemotherapy intensity	2302	0.45		C-reactive protein	3551	0.69
time since chemotherapy	3111	0.6		procalcitonin	5034	0.98
central venous line (CVL)	2007	0.39		interleukin 6	4701	0.91
CVL type	3711	0.72		interleukin 8	4672	0.91
out-patient at onset	1971	0.38		death	317	0.061
temperature	2336	0.45		ICU duration	2484	0.48
respiratory rate	4933	0.96		need for organ support	4302	0.83
respiratory compromise	3093	0.6		severe infection	685	0.13
pulse rate	4787	0.93		clinically documented infection	955	0.19
shock	2508	0.49		microbiologically documented infection	375	0.073
systolic blood pressure	4555	0.88		bloodstream infection	23	0.0045
diastolic blood pressure	4557	0.88		duration of fever	3748	0.73
severe unwell	1940	0.38		duration of admission	2206	0.43

The patterns of missing data have importantly affected the analyses conducted, as will be explored in detail in subsequent chapters.

Demographics

Age

The overall distribution of ages in the PICNICC dataset is shown in Figure 35;

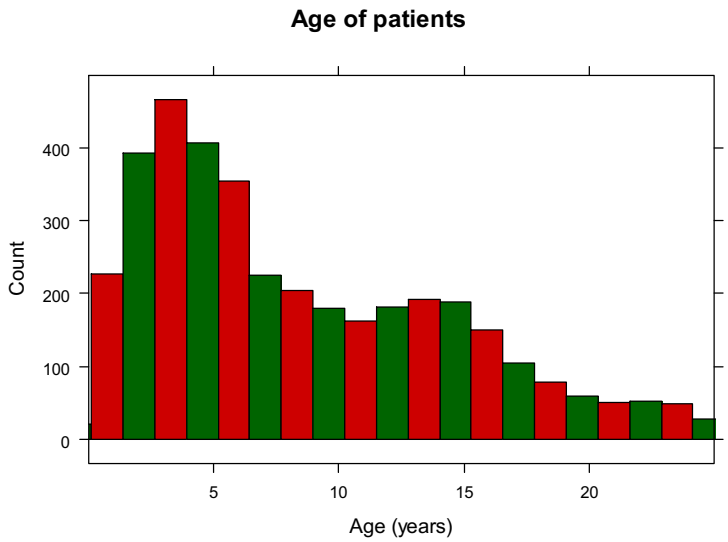


Figure 35: Age distribution of PICNICC dataset

The median age at first episode recorded was 6.5y (mean 8.4y, range 50 days to 25 years, IQR 3.4y to 12.8y). This wide age variation is to be expected from groups undertaking the care of children with cancer, some of whom are born with malignant disease or develop it soon after birth. Twenty-four children aged six months old or younger presented with FN in this dataset; most patients had acute leukaemia; either infant ALL (4) or AML (8).

The age distributions varied in the different datasets (see Figure 36 and more detail in

Appendix 22. Further detailed information on the IPD data, Table 44).

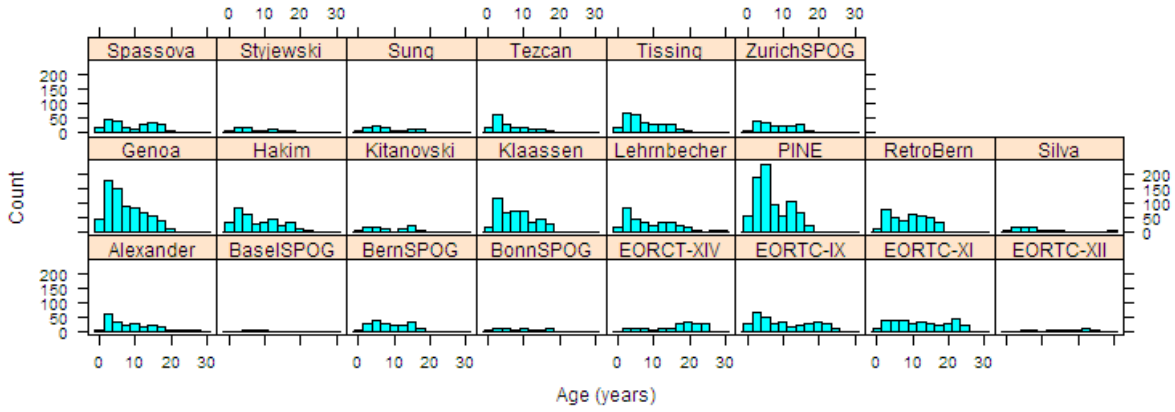


Figure 36: Age distributions of patients per study

The figure demonstrates that the EORTC studies (XIV, XI and XII) have an older population than the other, predominantly paediatric, datasets. This reflects the organisation of health care and the centre-based approach to most of these studies. In most locations (with the exception of the multicentre, all-age EORTC studies), care is delivered to patients who are classified as children (with an upper age limit varying between units; some definitions include <16 years, others <18 years, others <19 years and in full-time education) and these studies were undertaken in paediatric units, rather than across cancer services generally.

Tumour types

A wide variety of malignancies were represented in the included studies, in keeping with the disparate nature of rare and very rare diagnoses treated in paediatric oncology/haematology units (see Figure 37, and per-study data in Figure 57).

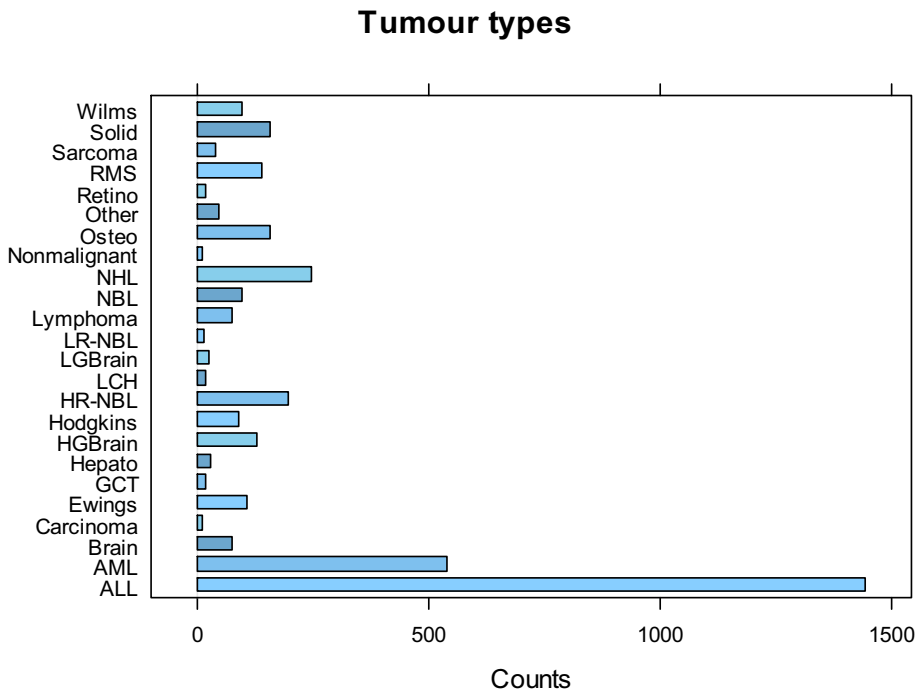


Figure 37: Counts of episodes of FN by tumour type

The data include for two non-cancer groups (the “non-malignant” category and LCH: Langerhans Cell Histiocytosis). In both these categories, the underlying disease types had been treated with chemotherapy; primarily for stem cell transplantation of haemoglobinopathies or immunodeficiencies in the non-malignant group, and relatively low-dose cytotoxic treatments for multi-system or organ-at-risk LCH. The inclusion of this group of non-cancer patients (36 episodes in 19 patients) can be justified on the grounds that they are treated, for all practical purposes, the same way as the patients with malignant disease; the chemotherapy they are exposed to brings similar risks of immunosupression and life threatening infection. Furthermore, the inclusion of such a small number, even if they were to have different predictors of outcome, would be very unlikely to alter any conclusions drawn from the excess of 5,000 episodes in the IPD analysis.

The distribution of age and disease type generally follows the expected pattern of incidence (see Appendix 22, Figure 58). Some examples include how ALL is a disease of younger children (with a further peak in the mid-50s, which is beyond the range of this study), along with retinoblastoma, neuroblastoma (NBL) and Wilm's tumour. This is unlike AML which has a relatively consistent incidence across ages, and osteosarcoma or Ewing's sarcoma which peak in the teenage years.

Gender distribution

There were slightly more males in most of the studies, with 56% male patient-episodes (see Appendix 22, Table 45). Two data sets (EORTC-XIV and Hakim) did not provide gender data on their patients.

This slight male preponderance is in keeping with the male preponderance of cancer in children as recorded by population registries (54.1%) [248]. There is no suggestion in any of the dataset that patient gender prevented access to healthcare.

Multiple episodes

The study designs led to notable differences in the distribution of numbers of episodes of FNP per patient, as illustrated in Figure 38. The design of the EORTC studies, the Hakim and Styjewski and the Sung trial allowed each patient to be included only once.

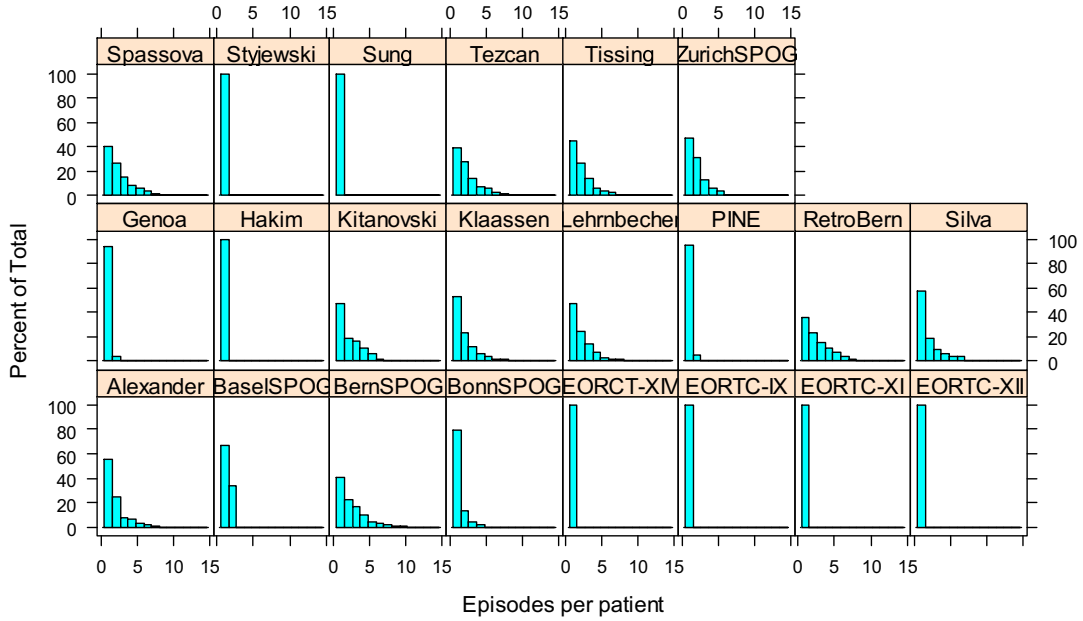


Figure 38: Number of episodes per patient per study

For studies with multiple episodes, there were a median of two (mean 2.4, range 2-14, IQR 1-3) episodes reported per patient (Appendix 22, Table 46). In most of these studies, the counts followed a Poisson distribution, but the Genoa and PINE datasets in particular have an excess of one-episode patients. This may be due to the high number of HSCT patients in the Genoa dataset (who may have a single, prolonged episode but are treated once, usually at the culmination of their therapy). The PINE dataset was gathered across 47 sites over a 12 month period, and there may have been a greater chance of failure to accurately capture linked episodes in this more dispersed study.

A small number of patients (58) had six or more episodes of FN. These did not differ from the less-frequently included patients in gender, date of episode, age, or tumour type (p=1 by χ^2 or Wilcoxon rank-sum tests as appropriate).

Outcomes reported

For each of the IPD studies individual outcome data were requested per episode for: death, intensive care admission (ICU: occurrence or duration), need for moderate organ support (e.g. fluid bolus, oxygen supplementation), any clinically documented infections (CDI) and any microbiologically documented infections (MDI). Where possible, MDI were defined as bloodstream or “other site”. Data were requested on the durations of fever and admission.

Each dataset reported on quite different sets of outcomes. The outcomes most frequently reported by the studies were microbiologically documented infection (MDI), death, bloodstream infection and the *a priori* composite measure "severe infection". The proportion of episodes where data was available for the outcomes is shown in Table 17. The variation is almost entirely at the level of the study, with near-complete outcome assessment for individual elements within the study.

Table 17: Percentage of episodes with known outcomes

	Death	ICU	Organ support	Severe infection	CDI	MDI	Bloodstream infection	Duration of fever	Duration admission
Alexander	100	0	100	100	0	100	99.5	100	100
BaselSPOG	100	100	0	100	0	100	100	100	100
BernSPOG	100	100	0	0	0	100	100	0	100
BonnSPOG	100	100	0	0	0	100	100	100	100
EORCT-XIV	100	0	0	100	100	100	100	96	0
EORTC-IX	100	0	0	100	100	100	100	0	0
EORTC-XI	100	0	0	100	100	100	100	0	0
EORTC-XII	100	0	0	100	100	100	100	100	0
Genoa	99.6	0	0	100	100	100	100	0	0
Hakim	100	100	100	100	100	100	100	93.7	100
Kitanovski	100	100	100	100	100	100	95.6	97.1	85.3
Klaassen	100	100	0	100	100	100	100	0	72.6
Lehrnbecher	0	0	0	100	100	100	100	0	0
PINE	99.9	100	0	83.1	99.9	100	99.3	0	99.9
RetroBern	100	0	0	100	100	0	100	0	100
Silva	100	100	100	100	100	100	100	100	100
Spassova	100	100	0	100	100	100	100	0	100
Styjewski	100	0	0	100	0	100	100	0	0
Sung	100	0	0	0	0	100	100	100	98.7
Tezcan	98.6	98.6	0	100	100	100	100	90.3	0
Tissing	100	100	83.7	0	0	95.7	95	84.1	70.9
ZurichSPOG	100	100	0	100	0	100	100	99.4	100

The proportions of microbiologically documented infection (MDI), death, blood stream infection and calculated “severe infection” differ markedly between studies (see detail in Appendix 22, Table 47).

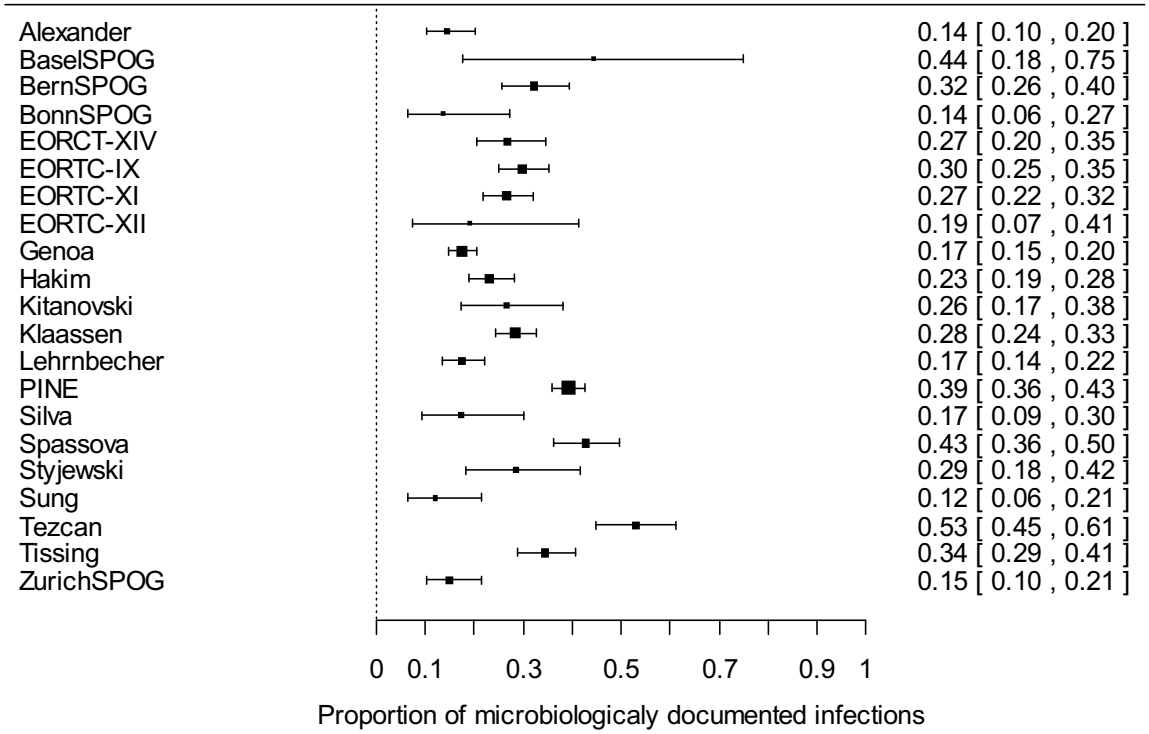


Figure 39: Proportion MDI (per study)

The range of proportions of MDI varied from 12% of episodes in the Sung dataset to 53% in the Tezcan dataset (see Figure 39). The potential reasons for this variation according to the individual’s presenting features are explored in the later part of this chapter (where univariate predictors are examined) and in the next chapter. Study-level differences are difficult to ascertain from study level features: the fever definitions and trial date are not explanatory, nor is retrospective/prospective data collection. Geography, regionalised as Western Europe, Central Europe, North America and South America has some explanatory power; with Central European studies (Kitanovski, Tezcan and Spassova) being significantly associated with greater rates of occurrence of MDI, bloodstream infection and severe infections.

The outcomes of severe infection and bloodstream infection are highly correlated with MDI (Pearson's ρ=0.89 and ρ=0.82 respectively, p<0.001) and show, similar patterns of variability.

Mortality rates in the PICNICC dataset are very small, around 1% in most series. This is compatible with recent experience and reports given the aggressive nature of the antibiotic and resuscitation regimes in use. The datasets with larger point estimates of mortality rates (still 5% or smaller) have confidence intervals in keeping with the overall pattern.

Table 18: Summary statistics for continuous outcomes

	Admission duration (days)				Fever duration (days)			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Alexander	6.8	5.0	1.0	37.0	2.3	2.0	1.0	10.0
BaselSPOG	7.9	7.0	3.0	14.0	2.8	0.0	0.0	6.0
BernSPOG	7.9	5.0	1.0	60.0	NA	NA	NA	NA
BonnSPOG	5.6	5.0	1.0	19.0	2.0	1.0	0.0	10.0
EORCT-XIV	NA	NA	NA	NA	5.1	3.0	0.0	31.0
EORTC-IX	NA	NA	NA	NA	NA	NA	NA	NA
EORTC-XI	NA	NA	NA	NA	NA	NA	NA	NA
EORTC-XII	NA	NA	NA	NA	2.8	2.0	1.0	10.0
Genoa	NA	NA	NA	NA	NA	NA	NA	NA
Hakim	6.2	4.0	0.0	146.0	2.9	1.1	0.0	73.4
Kitanovski	10.0	8.0	4.0	40.0	4.6	4.0	1.0	21.0
Klaassen	5.1	3.0	0.0	93.0	NA	NA	NA	NA
Lehrnbecher	NA	NA	NA	NA	NA	NA	NA	NA
PINE	7.4	5.0	0.0	105.0	NA	NA	NA	NA
RetroBern	6.1	5.0	0.0	58.0	NA	NA	NA	NA
Silva	6.0	4.0	1.0	26.0	12.2	2.0	0.0	24.0
Spassova	4.5	3.0	0.0	23.0	NA	NA	NA	NA
Styjewski	NA	NA	NA	NA	NA	NA	NA	NA
Sung	6.4	4.0	2.0	49.0	1.7	1.0	0.0	14.0
Tezcan	NA	NA	NA	NA	5.3	3.0	1.0	28.0
Tissing	3.1	2.0	1.0	13.0	10.7	8.0	2.0	37.0
ZurichSPOG	7.7	7.0	1.0	28.0	3.4	2.0	0.0	23.0

Data for duration of admission and duration of fever are reported less often, only 12 datasets provided information on duration of fever, and 14 on duration of admission (see Table 18, and Appendix 22, Figure 59 & Figure 60). The duration of admission seems to relate mainly to the policy of patient discharge in each unit, with median durations of 5 days (a "traditional" approach). The duration of fever is more consistent across studies, with a median of around two days. No study level characteristics were found to influence these durations.

Description of the predictors

Predictor variables for microbiologically documented infection have been split into patient-specific background features, the demographics of the patients, as noted in the preceding section; episode-specific background factors, such as the intensity of preceding chemotherapy, and the presence of a central venous line; episode-specific clinical features such as maximum temperature and heart rate; and episode-specific laboratory features including biomarkers of inflammation and elements of the full blood count.

The data are addressed in each section below.

Episode-specific background factors

These features are elements of the treatment that the patient is undergoing at the time of the episode of FN. These are not fixed (like age, and malignant diagnosis), but neither are they clinical impressions of the child or young person's physiological response to FN. They include the intensity of chemotherapy, the time since chemotherapy was delivered, the status of the cancer (in remission, relapsed or not), and the presence of a central venous line.

Remission

The meaning of remission is sometimes interpreted differently in different malignant diseases, and even between different groups examining the same disease. For example, in leukaemia there is a clear and consistent definition of remission (fewer than 5% of the marrow involved by leukaemic cells), but there is a range of alternative interpretations of how a solid malignancy is described as

being in remission. Some definitions may be based on an on-treatment scan showing a good response (defined as >33% reduction in primary tumour volume, or >50% by some groups) or clear scans showing no evidence of disease by physical or metabolic criteria. The definitions used in each dataset were not able to be provided.

Time-since-chemotherapy

Time-since-chemotherapy is also confusing, as the treatment of many malignancies varies, and so the effect of this variable upon immunosuppression may differ between different tumour types, and between the same malignancy at different stages of therapy. The classic approach to treating solid malignancies is to give cycles of chemotherapy, waiting for clinical and bone marrow recovery before commencing a further cycle. The expected nadir in marrow function is between day 10 and 14 of commencing chemotherapy. In contrast, in some protocols for acute leukaemia there is a maintenance phase of treatment where chemotherapy is given as an oral medication on a daily basis with dose titrated against toxicity. In other parts of acute leukaemia treatment, chemotherapy is given intermittently over 10-21 days. Details of which phase of therapy was undertaken prior to each episode were not requested, so the uncertainty of interpreting this variable has meant that the IPD analysis has not used this information.

Central venous lines

Central venous lines (CVL) are used to deliver chemotherapy directly into major veins (usually the superior vena cava in the upper part of the chest) and allow blood to be taken for regular tests while minimising trauma to the child. They vary by number of lumens (separate tubes within the line), usually having between one and three. They also differ in that they may be tunnelled (with the tube exiting the major vein and passing under the skin for a distance before the access point) or untunnelled. There are two major types of tunnelled lines which have different types of access for administration of products or taking of blood tests. If the end of the line emerges out of the skin they are known as Hickman-style lines, and if it ends beneath the skin with a palpable metal/plastic port which can be easily accessed with a short needle, known as a Port-a-cath. The different types of lines are used in different ages of patients and sometimes for different types of chemotherapy.

Table 19: Presence of CVL (per study)

	Percentage with CVL	Number of episodes with CVL	Number of episodes without CVL	Line type if given			NA
				Port	Hickman	Untunnelled	
BaselSPOG	100%	9	0				0
BernSPOG	76%	130	41				0
BonnSPOG	93%	41	3				0
Genoa	94%	664	39				0
Hakim	95%	317	15	27%	69%	4%	0
Kitanovski	63%	43	25	13%	0%	0%	0
Klaassen	84%	364	67	81%	33%	0%	0
Lehrnbecher	97%	301	8				2
RetroBern	50%	183	181	51%	6%	0%	0
Silva	98%	51	1	9%	6%	1%	0
Spassova	100%	199	0	0%	4%	0%	0
Styjewski	100%	56	0				0
Tissing	100%	258	0				0
ZurichSPOG	83%	128	26				0

Table 19 shows the proportion of patients with a central line, and the type of line in use. For most studies, each patient has a CVL of some type, but with marked variation in the type of line used.

Chemotherapy intensity

Chemotherapy intensity was collapsed into one of three categories for the IPD analysis (as described in detail in Chapter 6) consisting of: low intensity (at or less than the ongoing maintenance treatment used for acute leukaemia); HSCT, haemopoietic stem cell transplantation, very intensive chemotherapy which requires rescue with haemopoietic stem cells, commonly known as a 'bone marrow transplant'; and standard intensity, which covers all of the middle ground between these extremes.

Eleven datasets gave information directly, or indirectly, on the chemotherapy used per patient and so could define the level of intensity (see Table 20). In one very small dataset, only standard-

intensity chemotherapy had been used (BaselSPOG) and only two datasets (PINE and Genoa) included patients that had undergone HSCT. In all other studies, patients who had received HSCT were excluded from the data collection.

Table 20: Number of patients receiving each level of chemotherapy intensity (by study)

	Low	Standard	HSCT	NA
BaselSPOG	0	9	0	0
BernSPOG	26	145	0	0
BonnSPOG	4	40	0	0
Genoa	23	504	176	0
Kitanovski	1	65	0	2
PINE	218	577	17	0
RetroBern	32	332	0	0
Silva	1	51	0	0
Spassova	4	194	0	1
Sung	1	74	0	0
Tissing	14	193	0	51
ZurichSPOG	17	137	0	0

Episode-specific clinical features

Features in this grouping relate to the individual as they present with each episode, and are may be more varied between individuals and between different episodes for the same individual than those classified as “background” features. They tend to be assessed by simple clinical examination on presentation by the first-contact healthcare providers.

Maximum temperature

There were generally similar maximum temperatures in the study groups represented; some differences in inclusion were introduced by either strict adherence to a minimum fever, or including patients who were being treated for suspected infection, regardless of temperature or neutrophil count (see Figure 40, giving a box and whisker plot of the distribution per study, and

Appendix 22. Further detailed information on the IPD data, Table 22). For information, a normal body temperature is considered to be between 36.5°C and 37.2°C, depending upon method of measurement.

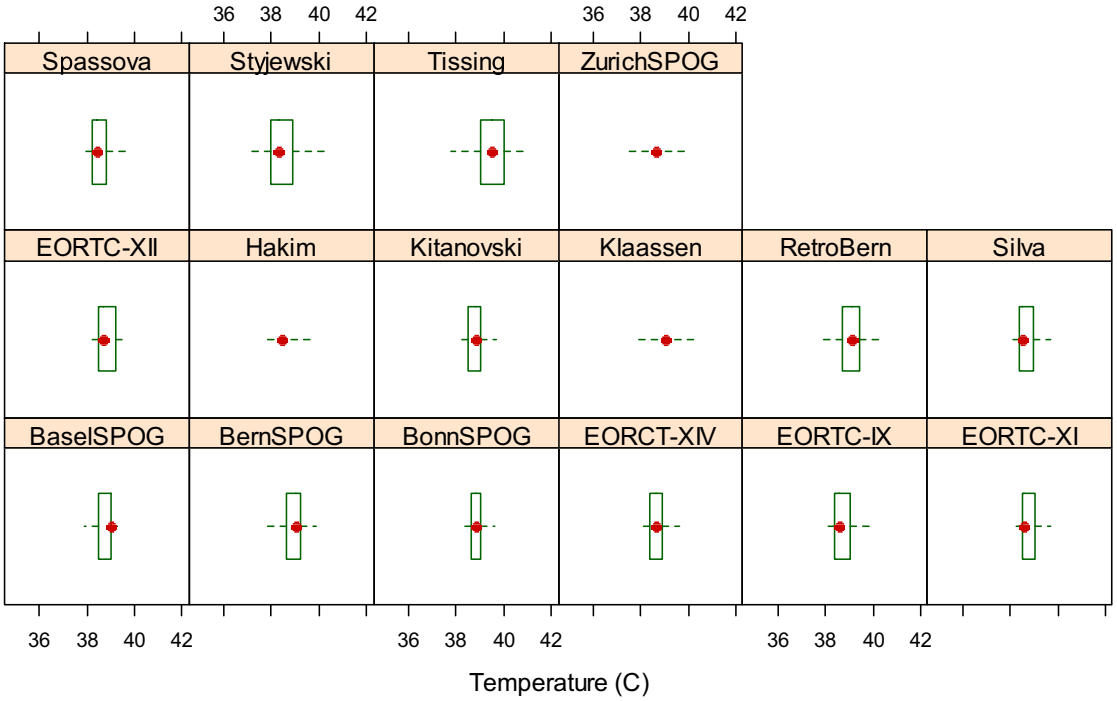


Figure 40: Box-and-whisker plot of distribution of temperature (by study)

Mucositis

Mucositis is the inflammation and ulceration of the gastrointestinal mucosa associated with many chemotherapy treatments and radiotherapy. It can affect all parts of the gastrointestinal tract, from mouth through oesophagus, stomach, small and large bowel to sigmoid colon. Estimates of mucositis in the datasets were provided as measures of severity, or by a dichotomised approach describing the presence/absence of severe mucositis. Sometimes extensive free text comments on the state of patient at admission were present allowing recoding of mucositis into the commonly used 0-IV grading of the CTC (Common Toxicity Criteria) grading structure, or the information was provided as a three-level assessment (where none, mild and severe were coded as zero, I and III respectively). This approach led to a bimodal pattern in some study groups (as shown in Table 21) but the information was generally consistent.

Table 21: Mucositis (graded)

	Grade 0	Grade I	Grade II	Grade III	Grade IV	Severe	Non-severe	NA
Alexander	0	0	0	0	0	0	0	187
BaselSPOG	7	1	1	0	0	1	8	0
BernSPOG	113	22	12	18	6	24	147	0
BonnSPOG	29	7	7	1	0	1	43	0
EORCT-XIV	0	0	0	0	0	0	0	149
EORTC-IX	0	0	0	0	0	0	0	315
EORTC-XI	0	0	0	0	0	0	0	301
EORTC-XII	0	0	0	0	0	0	0	21
Genoa	0	0	0	0	0	0	0	703
Hakim	288	26	0	18	0	18	314	0
Kitanovski	41	13	0	10	0	14	54	4
Klaassen	395	10	5	16	4	21	410	1
Lehrnbecher	229	10	36	23	12	35	275	1
PINE	0	0	0	0	0	53	759	812
RetroBern	278	0	0	26	0	78	246	60
Silva	0	0	0	0	0	0	0	52
Spassova	132	18	0	49	0	49	150	0
Styjewski	0	0	0	0	0	0	0	56
Sung	40	6	5	3	2	19	56	19
Tezcan	0	0	0	0	0	0	0	145
Tissing	0	0	0	0	0	0	0	258
ZurichSPOG	79	20	32	12	11	23	131	0

Out-patient status

The proportion of out-patient episodes varied according to study design, with some studies only examining patients presenting from outside hospital. Other studies did not provide data on the in-patient status of the patient at each episode (see Table 22). Of those studies where all-episodes, regardless of admission status of the patient, were recorded, there was considerable variability in the proportion of episode of FN originating in in-patients and out-patients.

Table 22: Percentage of out-patient, unwell and cardiovascular/respiratory compromise episodes for informative studies

Study ID	% Out-patient	% "Unwell" patients	% Cardiovascular compromise	% Respiratory compromise
Alexander	100%	22%		
BaselSPOG	11%	100%	0%	67%
BernSPOG	82%	14%	6%	42%
BonnSPOG	84%	5%	5%	5%
EORCT-XIV	15%			
EORTC-IX	35%			
EORTC-XI	37%			
EORTC-XII	67%			
Hakim	100%	25%	2%	5%
Kitanovski	65%	4%	1%	0%
Klaassen	100%	27%	4%	2%
Lehrnbecher		5%		
PINE		14%	8%	9%
RetroBern	73%	2%	7%	
Silva	96%	0%		
Spassova	26%	24%	9%	14%
Styjewski	100%	20%		
Sung	97%	41%	7%	
Tezcan			5%	
Tissing	100%			
ZurichSPOG	79%	31%	5%	35%

Clinical impression of significantly unwell patient

The usefulness of the clinical impression of a child/young person presenting being "severely unwell" has been debated[249], but in practice has been held as a firm and important factor. A variable accounting for this gestalt impression was present in 15 of the 22 datasets. The presence of such a feature was again quite variable, but was between 20% and 30% in most

studies (see Table 22). The very small dataset of BaselSPOG is somewhat of an outlier in this group, with all patients having “severe illness”, but this is a very small group of patients.

Vital signs

It is relevant to note again that the expected values for respiratory rate, pulse and blood pressure vary with age, and that the simple values delivered are unhelpful. For example, a one year old breathing at 30 breaths/minute is perfectly normal whereas an adult breathing at the same rate would be panting. Transformations which can address this include z-score (“centile”) modifications or dichotomising values into abnormal/normal, as in the case of shock/respiratory compromise (as noted in Methods chapter 6). As data were provided in raw form by relatively few studies, analysis was limited to the dichotomised versions of cardiac or respiratory compromise.

Respiratory compromise was proposed as a practical way of combining datasets where this dichotomised assessment of respiratory function had been supplied, and those few datasets where respiratory rates had been supplied. The process of mapping from respiratory rate to compromise was undertaken for the SPOG group datasets, and produced markedly higher rates of respiratory compromise (see Appendix 22, Table 22) which makes the interpretation of this “mapped” variable difficult to believe.

A similar mapping exercise was undertaken for blood pressure and cardiovascular compromise. In this instance, a difference between mapped and directly reported data was not noted (see Table 22), and around 1 in 20 episodes presented with cardiovascular compromise.

Episode-specific laboratory factors

The use of laboratory-measured features to predict outcome is considered to be more objective and consequently robust than subjective clinical assessments of mucositis, respiratory compromise or gestalt “unwellness”. The features examined have focussed on the cellular elements of the full blood count (haemoglobin, a reflection of red cell count, platelets, total white cell count and particular sub-types, neutrophils and monocytes) and

inflammatory biomarkers (particularly C-reactive protein, CRP; procalcitonin, PCT; and interleukins 6 and 8, IL-6 and IL-8) measured in serum.

Full blood count

Of these features, all except for haemoglobin showed a log-normal distribution, and were used after natural log transformation (see Figure 41).

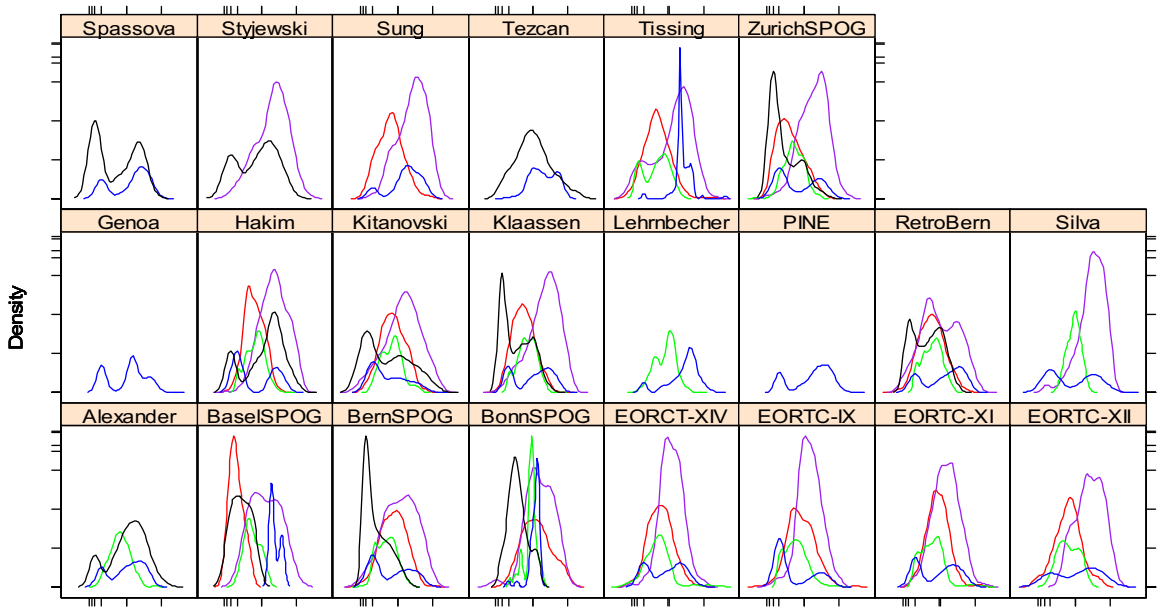


Figure 41: Density of distribution of FBC parameters, by study, transformed where appropriate

Haemoglobin red, Platelets purple, White Cell Count green, Absolute Neutrophil Count blue, Absolute Monocyte Count black

The study groups showed similarity across the range of haemoglobin values, which are to be expected from clinical practice (for greater detail see

Table 50).

Other parameters of the blood count were also as expected (for greater detail see Appenidx 22, Table 51- Table 54). The occasional very high values of white cells recorded are to be sometimes found in patients presenting with a strong marrow response, often in the setting of severe infection or occasionally as a complication of therapy, for example corticosteroids or GCSF.

Different studies provided different elements of the blood count (as noted in earlier in the chapter). The elements were only moderately correlated (calculated where pairs existed; see Table 23).

Table 23: Spearman rank correlation coefficients

	ln(white cell count)	**ln(absolute neutrophil count)**	**ln(absolute monocyte count)**	**ln(platelets)**	**haemoglobin**
ln(white cell count)	1				
ln(absolute neutrophil count)	0.468	1			
ln(absolute monocyte count)	0.43	0.369	1		
ln(platelets)	0.33	0.271	0.408	1	
haemoglobin	0.147 (*)	0.035	0.124	0.22	1

All p<0.001 except * (p=0.0357)

Biomarkers

The serum biomarkers were far less frequently available. The number of studies providing information is grossly variable; eleven with CRP, five with IL-8, four with IL-6 and only two studies reporting on PCT (see Figure 42 and Appendix 22, Table 55 and Table 56).

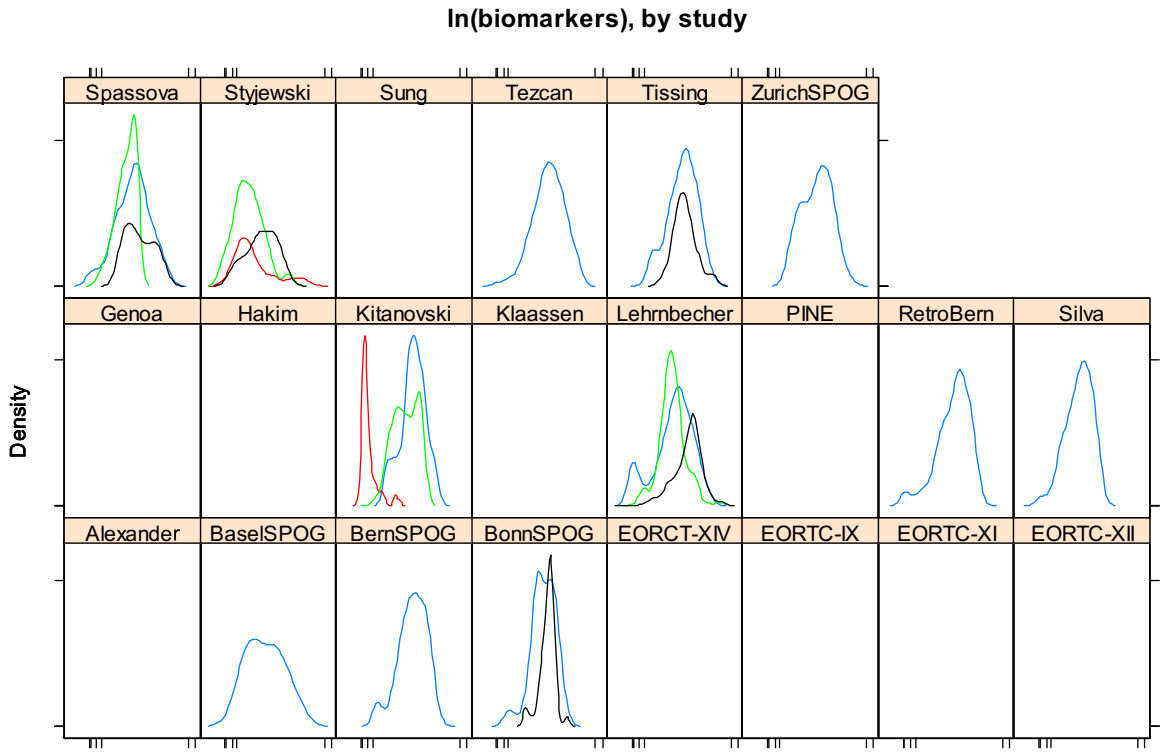


Figure 42: Density plots of ln(biomarkers) per study

Blue – CRP, Red PCT, Green IL-6, Black IL-8

Examining these data for correlations between biomarkers values between studies are suggestive of some inconsistency (see Table 24). For instance, whereas IL-6 and IL-8 are strongly positively correlated, IL-8 is moderately correlated with PCT and IL-6 is (insignificantly) negatively correlated with PCT. One of these three relationships appears incongruent which may be in part due to different datasets providing information (IL-6/8: Spassova, Styjewski, Lehrnbecher, IL-8/PCT: Styjewski, IL-6/PCT: Styjewski and Kitanovski)

Table 24: Correlation of ln(biomarkers)

	ln(CRP)	**ln(PCT)**	**ln(IL-6)**	**ln(IL-8)**
ln(CRP)	1			

ln(PCT)	0.302	1		
ln(IL-6)	0.123	-0.12	1	
ln(IL-8)	0.026	0.429	0.715	1

Normal: p>0.05, **Bold**: p 0.01 to 0.05, ***Bold Italic***: p<0.01

Summary

The PICNICC collaboration collected data prior to the cut-off of November 2011 for derivation in 22 datasets from 16 collaborative groups. These contained information from 5,127 episodes of FNP in 3,504 patients. The median age at first episode recorded was 6.5y (mean 8.4y, range 50 days to 25 years, IQR 3.4y to 12.8y), and has a slight male preponderance (56% male). A wide variety of malignancies were represented, in keeping with the nature of diagnoses treated in paediatric oncology/haematology units.

A wide range of outcomes and potential predictor variables were provided in the PICNCC dataset, all of which show marked differences in completeness and consistency. The issue of missing data was more significant than originally envisaged. This was almost entirely at the level of study, whereby predictor or outcome variables were “not recorded” rather than “missing” for some patients or episodes. Some data items, such as the presence or absence of shock, could be recoded to minimise inconsistencies. Others are inconsistent as reflected in their original study designs (e.g. proportion of episodes commencing as out-patients). Some variables (such as remission and time-from-chemotherapy) have been found to be unusable in the analysis. The distribution of continuous variables followed a Normal or Log-Normal pattern, as expected.

The assessment of these data for consistency and quality is an important first step in preparing to undertake univariate analyses, which are important to clinicians who would ideally wish a single feature to be powerfully predictive of the presence or absence of adverse outcomes, and multivariable analyses from which the decision rule will be built in subsequent chapters.

Chapter 8: Results of the univariate analyses

In the examination of candidate predictors of outcome, a useful first step is a univariate assessment of the association between each predictor and outcome. The term 'univariate' means that each predictor is considered separately and so each association is not adjusted for other variables. Multivariable analyses (which do adjust) are conducted in subsequent chapters.

In this chapter univariate analyses are undertaken to create a list of potential predictors for use in the multivariable analysis, and to inform clinicians of the association between single factors and the risk of infection in their patients. Univariate associations are important to obtain an initial overview of the associations and whether they appear consistent with clinical expectations. Additionally, if a univariate association is extremely strong, it may remove the need for any more complex examination of the data as a clinically effective decision can be made from that one piece of information.

To assess the consistency of these features across the different component studies, the univariate model has been fitted directly to each dataset. This chapter illustrates the general approach using particular examples, both as exemplars and to illustrate specific problems. Additionally, the data on inflammatory biomarkers, which were present in very few datasets and could contribute little to the multivariable analyses, are also explored in depth.

Outcome examined

This thesis considers the outcome of microbiologically documented infection (MDI). This was selected from the wider range of outcomes included in the initial protocol for the PICNICC analysis because it is the most completely reported, is the most frequently concerning to clinicians in paediatric FN, and has an uncomplicated definition.

Models fitted

Three logistic regression models were fitted to explore univariate associations for each of the candidate predictors of microbiologically documented infection. The models were fitted

using maximum likelihood estimation within the package `lme4` in R (see Appendix 23 for code).

These models are presented using the same notation as Chapter 5, briefly :

p – the probability of an outcome (in this case microbiologically defined infection, MDI)

$()_i$ - *i* refers to the 'i-th' dataset

$()_k$ - *k* refers to the 'k-th' patient

$()_{ik}$ - *ik* refers to the 'k-th' patient of the 'i-th' study

$beta_0$ (β_0) – the estimate of the 'intercept', that is, the log-odds of MDI when all the other features have 'zero' value

t - the predictive feature under investigation (e.g. temperature)

beta1 ($\beta_1 t$) – the estimate of the log odds ratio of the probability of MDI in individuals who vary by one unit of 't'

The first models is the "full hierarchical with individual effect" which fitted a hierarchical model estimating independent intercepts by study, with random effect on predictor within study (to allow for a separate predictor effect in each study), and random effect on individuals (to allow for a separate intercept for individuals with multiple records within a study),

[1] $Logit(p_{ik}) = \beta_{0i} + \beta_{1i} t_{ik} + \beta_{2k}$

$\beta_{1i} \sim N(\beta_1, \tau^2_{\beta 1})$

$\beta_{2k} \sim N(\beta_2, \sigma^2_{\beta 2})$

This model, additionally, has

beta2 (β_2) – the estimate of the log odds ratio of the probability of MDI in individuals who vary by study

tau (τ) - an estimate of the variation of the predictive feature under investigation between studies

sigma (σ) - an estimate of the variation of the predictive feature under investigation between individuals

The following model was the "reduced hierarchical with study effect" estimating independent intercepts by study, with random effect on predictor within study alone,

[2] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_{1i} t_{ki}$

$\beta_{1i} \sim N(\beta_1, \tau^2_{\beta 1})$

Finally, a "fixed effect model" allowing intercept to vary by study, with fixed effect on predictor within study, was used

[3] $\text{Logit}(p_{ik}) = \beta_{0i} + \beta_1 t_{ik}$

The previously described difficulties interpreting data on remission, time since chemotherapy, pulse, blood pressure, respiratory rate and respiratory compromise meant they were excluded from meaningful analysis (see Chapter 7). The effect of tumour type could be examined using a fixed effects model only because the fitting of 528 random effect estimates (22 studies * 24 tumour-types) was not feasible.

Results of model comparisons

Full hierarchical model

Full hierarchical analyses with individual effects were undertaken with (a) all patient data and (b) in the subset of patient data where multiple episodes were permitted. The latter approach provides a fairer approximation of the effect of variation attributable to the individual patient

The full comparison is provided in Appendix 24. In brief, estimates of the predictive value of the candidate variables were largely unchanged when approached with either information from every patient-episode, or only those patient-episodes from studies where multiple episodes per patient were permitted, for example in temperature (beta-estimates of 0.81 will all data, 0.77 with multiple-entry-studies only data). Estimates of the individual variation were also largely similar, with the 'temperature' data having individual level standard deviation estimates of 1.1 for all data and 1.4 for multiple-entry-studies . Features where differences in parameter estimates were shown included patient age (which had a very small absolute change in the predictive estimate), chemotherapy intensity (where the estimate of the predictive value of Haemopoetic Stem Cell Transplant, HSCT, varied markedly; see later for discussion) and absolute neutrophil count. Differences in patient-related variability were seen in central line type (where the use of multi-patient data to estimate the individual effect showed reduced variation), out-patient status, and the presenting features of temperature, shock, severe mucositis and clinical "unwellness" which demonstrated greater variability when estimated from multi-episode samples.

Reduced hierarchical model

Comparing the reduced model (see equation [2] above) to the full model led to very little difference in the association estimates for each candidate predictor. The differences introduced by assessing multiple episodes within patients were assessed by comparing the predictor estimates from the dataset with every patient-episode and a subset with just one episode per patient (full data in Appendix 5). This showed only meaningful difference when all episodes (examined as independent events) were compared with one-episode-per-patient; IL-6 was significantly less predictive of MDI when only one episode per patient was included.

Fixed effects model

Given the limited benefits demonstrated from the full multilevel model, a further assessment of the differences between the hierarchical and simple fixed effects model was undertaken. This demonstrated no significant difference in predictive estimates for the univariate predictors of MDI. (The results of this are shown in 26.)

Between-study consistency

As between-study consistency was considered important, an analysis of the predictors was undertaken per-study to assess this.

Univariate predictors of MDI

The fixed effect model was used to examine twenty-six different candidate predictors of MDI. Of these, twelve were significant at p<0.05, and six between p=0.05 to 0.15. The values of the strength of association are considered in order of statistical significance (see Table 25). The assessment of the effect of tumour type relative to acute lymphoblastic leukaemia (the most common diagnostic group) is shown in Figure 43.

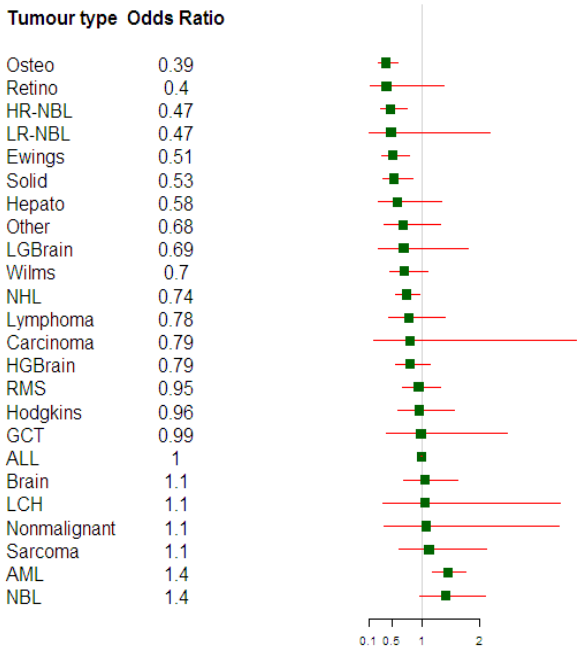


Figure 43: Odds ratio of MDI by tumour type relative to ALL

Table 25: Fixed-effect predictors arranged by order of statistical significance

Predictor name	OR	95% CI	p-value	% of episodes with missing data
Temperature	1.9	1.61 to 2.24	2.60E-14	72%
Log (white cell count)	0.72	0.66 to 0.78	1.40E-13	72%
Severely unwell	2.2	1.78 to 2.73	2.20E-13	72%
Log (IL-6)	2.1	1.71 to 2.63	4.40E-12	72%
Log (absolute monocyte count)	0.8	0.75 to 0.86	4.50E-11	49%
Log (IL-8)	1.8	1.48 to 2.28	1.70E-08	38%
Log (platelets)	0.8	0.74 to 0.87	2.10E-08	45%
Log (absolute neutrophil count)	0.92	0.9 to 0.95	1.20E-07	45%
Chemo.intensity - Low	1			45%
Chemo.intensity - Standard	2.2	1.56 to 3.04	4.00E-06	91%
Chemo.intensity - HSCT	1	0.83 to 1.22	0.96	45%
Shock	2.4	1.69 to 3.43	1.60E-06	98%
Log (PCT)	1.9	1.35 to 2.73	0.00033	91%
Out.patient	0.7	0.53 to 0.92	0.0078	79%
Cvl.type - None	1			53%
Cvl.type - Port	1.1	0.66 to 1.7	0.8	39%
Cvl.type - Hickman	1.4	0.87 to 2.32	0.17	69%
Cvl.type - Untunnelled	3.1	0.95 to 9.55	0.054	0%
Relapse	1.4	1.08 to 1.87	0.012	88%
Mucositis	0.89	0.8 to 1	0.052	48%
Severe.mucositis	0.76	0.55 to 1.03	0.078	10%
Marrow	1.5	0.92 to 2.46	0.095	10%
Haemoglobin	1	0.99 to 1.1	0.11	88%
Log (CRP)	1.1	0.96 to 1.19	0.25	63%
Age.days	1	1 to 1	0.38	13%
Central venous line	1.2	0.82 to 1.65	0.4	60%
Diastolic BP	0.99	0.97 to 1.02	0.47	47%
Sex - F	1	1 to 1		67%
Sex - M	0.96	0.84 to 1.1	0.56	43%
Remission	0.98	0.74 to 1.29	0.87	39%
Systolic BP	1	1 to 1	0.92	38%

These results suggest the following clinically suspected covariates may be associated with MDI: the presence of an untunnelled central line, the clinical appearance of significant unwellness, of documented cardiovascular compromise (shock), a high temperature, raised serum biomarkers, low white cell counts and platelets, a diagnosis of AML, and undergoing treatment for relapsed disease. The following potential associations are more clinically surprising; osteosarcoma/Ewings sarcoma patients, and patients with more severe mucositis, are associated with a decreased risk of MDI.

Patient age has been shown to be associated with the risk of death from FN [250] with teenagers at greater risk of dying. The reasons for this are unclear, and may relate to an increased risk of MDI, delayed presentation to hospital, or a reduced physical reserve than in younger children. Infants too (those less than 12 months old) are felt to be at greater risk, potentially through a natural lack of immunity to disease or the subtlety of clinical signs of severe illness. However, data from the IPD analysis suggest that there is no clear relationship between age and rate of MDI (also see Appendix).

The consistency of these features across studies can be examined graphically and by fitting the model direct to each. The general approach is illustrated using temperature (showing the approach to linearity and consistency), elements of the full blood count (where study-level variation was hypothesised to be important), and the challenges of presenting large amounts of data in the different tumour types. The issue of chemotherapy intensity is also examined in detail, exploring the inconsistencies found. Finally, the data on inflammatory biomarkers, which were present in very few datasets and could contribute little to the multivariable analyses, are explored in depth.

Temperature

Figure 44 shows the association between probability of infection and temperature assessed in each individual study. The dataset of Spassova shows a negative relation between temperature and risk of MDI. This outlier may be partly explained by the inclusion of a hypothermic (rather than febrile) patient with maximum measured temperature of 35^{o}C.

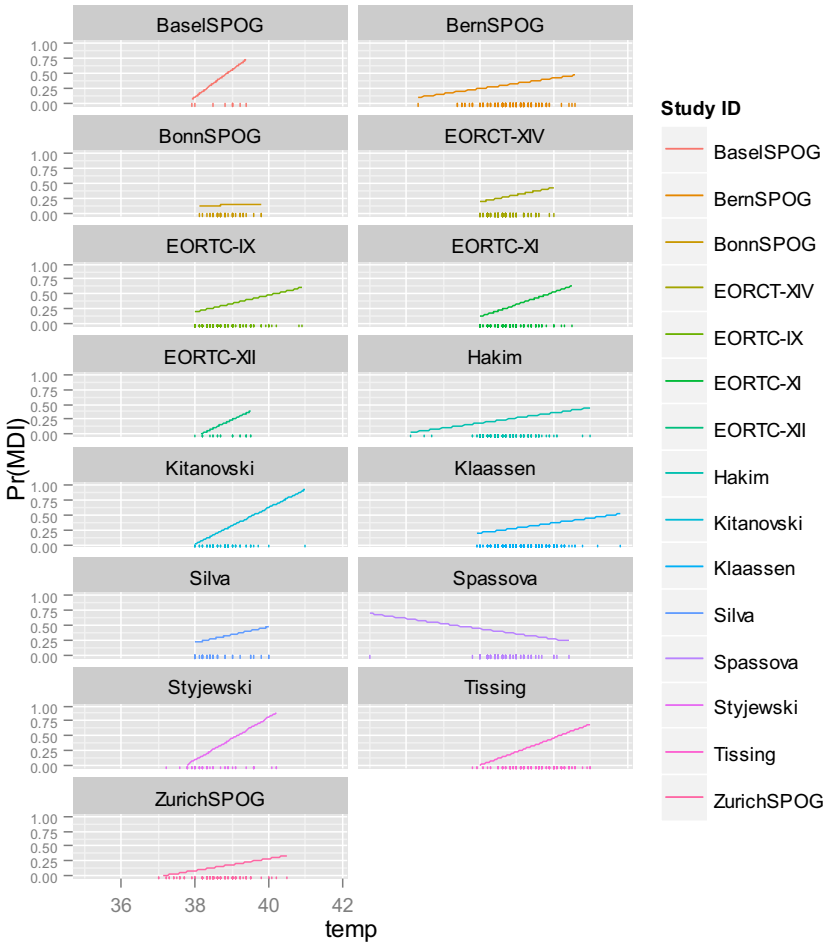


Figure 44: Relation of MDI and temperature by study. Data points indicated by rug plot.

This is demonstrated even more clearly by combining each study estimate onto a single graph Figure 45, and in further analyses, removing the hypothermic outlier reduced the inconsistency importantly.

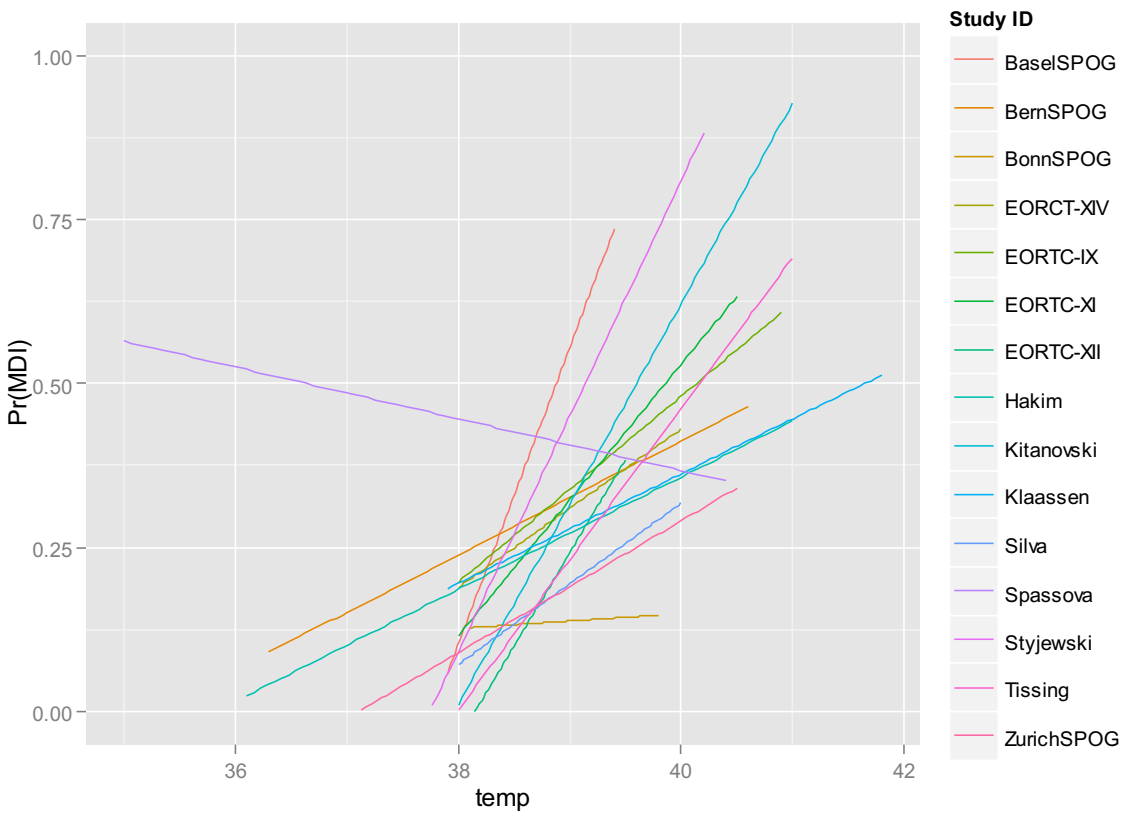


Figure 45: Relation of MDI and temperature by study

An alternative explanation for this finding is that temperature has a non-linear relationship with the probability of infection, with both very low and very high readings being linked to increased risk of MDI. Such a relationship was assessed by using fractional polynomials with common transformations of [-2, -1, -0.5, log, 0.5, 1 or 2]. These did not improve AIC values.

A centralised transformation of the temperature covariate (taking 37^{o}C, normal body temperature, away from each reading) was then used to comparing the linear model fit using quadratics. This led to statistically insignificant decrease in residual deviance and a small decrease in AIC (2832 vs. 2827).

Using splines with df= 2 , 3 or 4 led to one transformation of the formula which showed a statistically significant improvement in fit (placing a single knot at 38.3^{o}C; improved residual deviance p=0.008) but this benefit was of marginal benefit when assessed by AIC (2832 vs. 2828).

A major challenge in assessing the possibility of non-linear associations between the probability of infection and temperature comes from the inclusion criteria of the datasets. Only a very few datasets have selected patients without fever (or with hypothermia) and infection. As the data are sparse, the best fit can really only be based on when fever, measured as >37.5^{o}C in most of these data, is present.

Full blood count

Examining the relationships between the various subcomponents of the common blood test "full blood count" is instructive in assessing the potential utility of this extremely simple and widely available test, and examining for between-study effects. This test assesses aspects of the bone marrow's formation of the cellular components of blood, so may reflect the amount of bone marrow suppression induced by treatment, or in the case of cancer involving the bone marrow, disease.

There is evidence of the expected association (OR 0.8, 95% CI 0.75 to 0.86) between a lower absolute monocyte count (AMC) and risk of MDI is consistent across all studies that report this variable (see Figure 46). This is also seen – though less strongly (OR 0.92, 95% CI 0.9 to 0.95) with the absolute neutrophil count (ANC, Figure 62). This is of clinical importance as the currently "valued" element of the differential white cell count is the ANC, rather than the AMC.

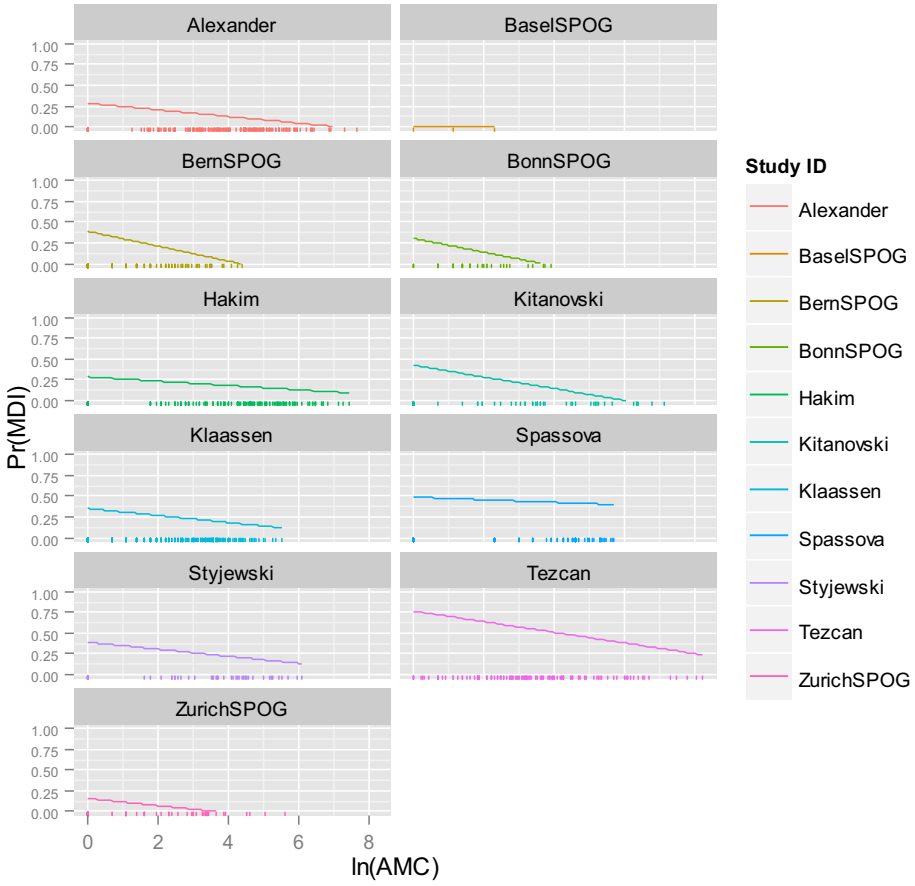


Figure 46: Relation of MDI and ln(AMC) by study.

Haemoglobin (Hb) and platelets are importantly distinct from differential white cell counts, in that low levels may be supplemented by transfusion of blood components. As policies to undertake such transfusions vary between clinical centres and have varied across time, it is important to examine for study-level variation in the predictive power of these variables, before accepting the overall estimate of association to be correct.

The relationship between Hb and risk of MDI appears less consistent than the white cell subsets (see Table 26) with some datasets estimating a positive relationship between increasing Hb values and risk of MDI, and the remainder a small negative relationship. No individual dataset has a “conventionally statistically significant” estimate of association, and no dataset has an estimate incompatible with the overall IPD estimate. The potential explanation of transfusion policies varying does not seem to apply in this group; while there is variation, it is generally in the trigger-level at which to transfuse, and aims to raise Hb to

approximately 12 g/dL. If this was a strong effect we should see both a bimodal distribution of Hb values and clear clustering of MDI cases around this value. The distributions of Hb shown in the previous chapter do not support this, nor do the distribution of MDI cases. As such, the most plausible explanation of these findings is chance variation.

Table 26: Per study estimates of haemoglobin (g/dL) in predicting MDI

Study	OR	95% CI	p-value
BaselSPOG	0.92	0.71 to 1.19	0.49
BernSPOG	3.6	0.53 to 25.53	0.2
BonnSPOG	1.1	0.9 to 1.27	0.48
EORCT-XIV	0.79	0.49 to 1.26	0.32
EORTC-IX	1.2	0.98 to 1.52	0.06
EORTC-XI	0.98	0.87 to 1.12	0.81
EORTC-XII	0.85	0.72 to 1.01	0.06
Hakim	0.88	0.48 to 1.62	0.68
Kitanovski	1.1	0.95 to 1.36	0.16
Klaassen	1.1	0.77 to 1.45	0.72
Sung	1.1	0.99 to 1.28	0.07
Tissing	1.2	0.8 to 1.75	0.39
ZurichSPOG	1.1	0.96 to 1.3	0.16
IPD Model	1	0.99 to 1.11	0.11

When examining the association between MDI and platelet levels, there is cross-study consistency and the expected increased likelihood of MDI with low levels. This may be a response to infection (with platelet consumption being a common finding in bacterial sepsis) or part of the increased susceptibility one expects with marrow suppression.

Tumour type

The strength of this IPD meta-analysis becomes apparent when the variable “tumour type” was examined as an individual-study level predictor (see Figure 43).

Graphically demonstrating the strength of association in such large and complex datasets is challenging. A traditional systematic review approach of forest plots would need to show 23 different plots. An alternative would be to display a heat map of the data points (see Figure 47). This can demonstrate that, for most tumour types, the point estimates (which are very imprecise – see Appendix 27, Tumour type

Table 57) tend in the same direction (shaded greens indicating an association with reduced risk of MDI, or orange/reds indicating an association with an increased risk). Blue represents an estimate of about unity, and white an area with no data.

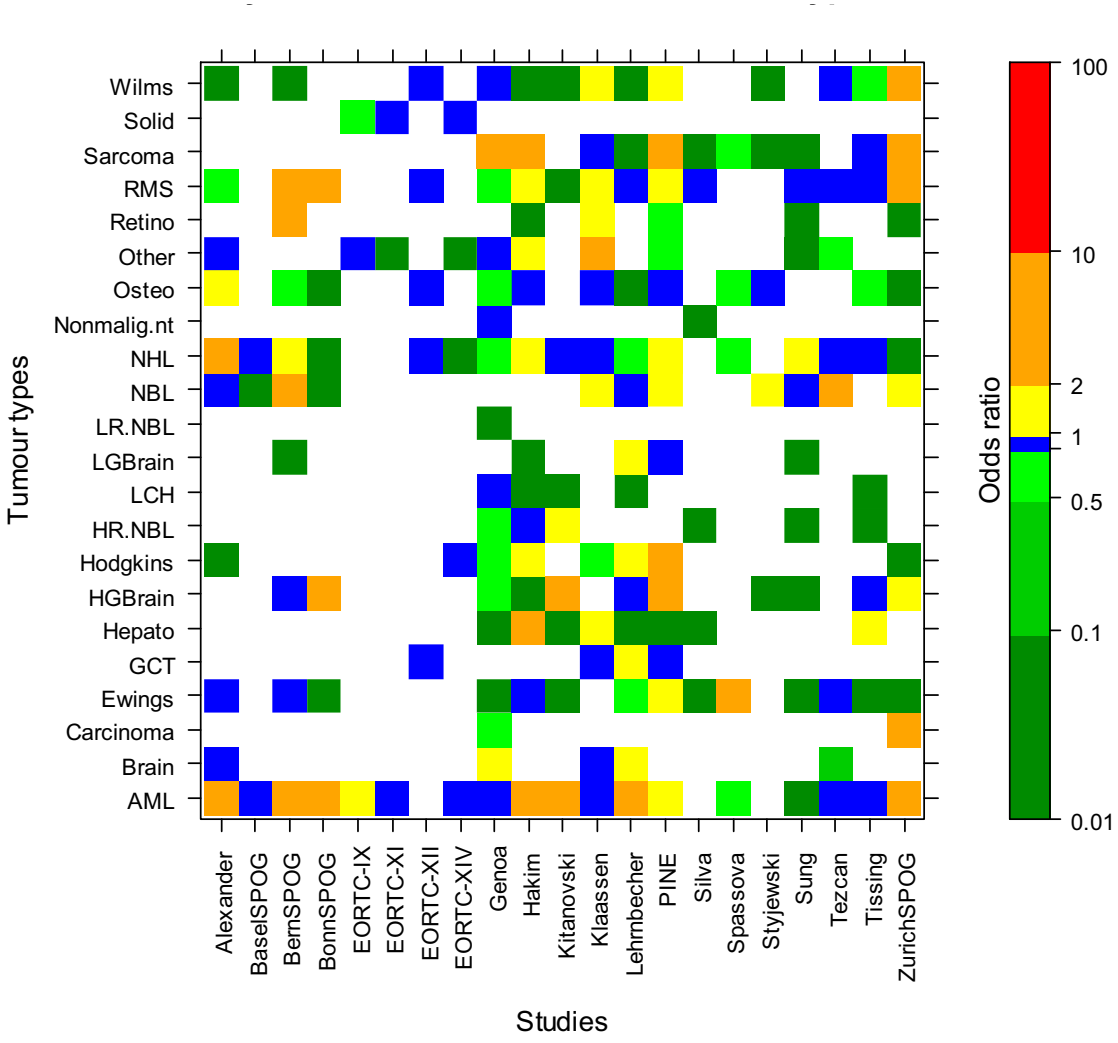


Figure 47: Heatmap of strength of association (OR) between tumour type and MDI

While this graphical display does show concisely a great deal of information from varied sources, it fails to illustrate the uncertainty around each odds ratio displayed, providing only the point estimate, and should be an aid to interpretation alongside the numerical data.

Although the finding of an increased risk of MDI in AML is widely acknowledged, the unexpected finding of an association between a reduced chance of infection in patients with bone sarcomas and febrile neutropenia requires further examination in a multivariable model, as it may affect the views of clinicians on implementing a reduced intensity therapy for low-risk patients.

Chemotherapy intensity

As discussed in Chapter 6, the intensity of chemotherapy refers to a concept encompassing the likely duration and severity of complications attributable to the cytotoxic agents used to treat cancer. For the purposes of the PICNICC analysis, a three-level ordered categorical approach was used: Low, Standard and HSCT; (commonly referred to as "bone marrow transplant"), but only two studies provided data on HSCT.

Table 27: IPD analysis for chemotherapy intensity

Predictor name	OR	95% CI	p-value
chemo.intensity - Low	1		
chemo.intensity - Standard	2.2	1.56 to 3.04	0.000004
chemo.intensity - HSCT	1.0	0.83 to 1.22	0.96

The IPD analysis (Table 27) contradicts the clinical experience that HSCT patients, who have undergone an extremely intense and immunosuppressive treatment, have a higher rate of documented MDI, than those undergoing standard or low-intensity therapy. This "expected" view is supported by the Genoa dataset (

Table 28)

Table 28: Observed association of chemotherapy intensity including HSCT from informative studies

	Chemo intensity	n Patients	OR	95% CI	p-value
Genoa	Low	23	1		
	Standard	504	1.1	0.52 to 2.29	0.82
	HSCT	176	1.5	0.96 to 2.46	0.08
PINE	Low	218	1		
	Standard	577	2.5	1.21 to 4.98	0.01
	HSCT	17	1.0	0.66 to 1.56	0.96

The Genoa dataset included patients currently undergoing HSCT treatment, and in the immediate period after this, unlike the PINE data where the HSCT patients largely consists of those who had recovered from the intensively treated phase and were recuperating with a greatly improved immune response and fewer other toxicities. As such, the HSCT data from the two groups appear to reflect different clinical phenotypes, and the clinical interpretation is challenging.

The relationship between standard intensity and chemotherapy is less contradictory (Figure 48), although quite heterogeneous between groups:

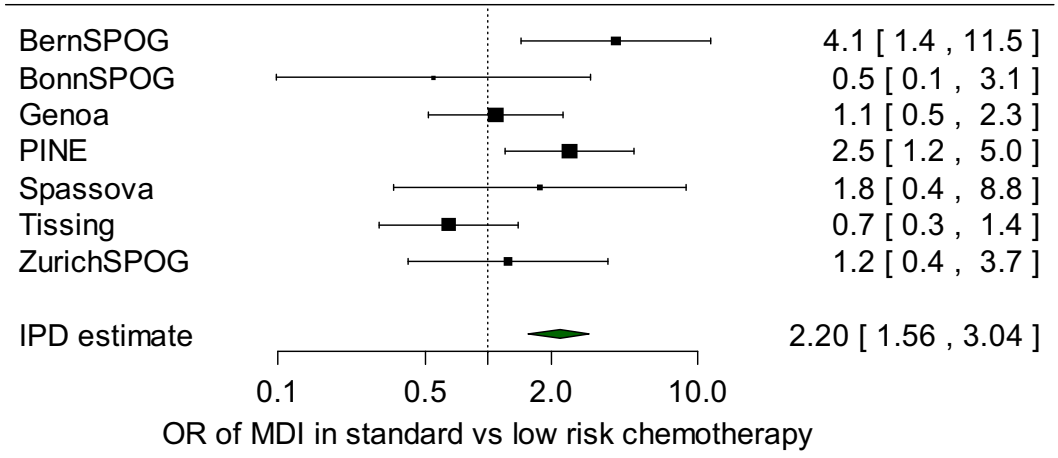


Figure 48: Association of standard intensity vs. low intensity chemotherapy and MDI

The variation in estimates of increased risk of MDI with chemotherapy intensity may be due to chance, to the changes in intensity of therapies across eras, or to different definitions of intensity. There is little to support the latter two explanations in these datasets.

Other clinical predictors

Many of the findings from the univariate analysis are as expected, and show consistency across the datasets (for further details, see Appendix 27). These findings include the associations of relapsed disease, out-patient status, severe cardiovascular compromise, the gestalt appearance of a patient being unwell, and central line type or presence. The absence of an association with patient age, gender and marrow involvement is also consistent.

The relatively surprising association suggested with increasingly severe mucositis decreasing the risk of MDI was consistent across studies, though no single study would have detected this association, and using the grossly categorical approach of "severe" vs. "non-severe" also dilutes this association (see Table 29).

Table 29: Study level associations of mucositis with MDI

	Predictor name	OR	95% CI	p-value
BaselSPOG	Mucositis (per grade)	0	0 to infinite	1
	Severe mucositis	0	0 to infinite	1
BernSPOG	Mucositis (per grade)	0.78	0.57 to 1.07	0.1
	Severe mucositis	0.51	0.18 to 1.45	0.21
BonnSPOG	Mucositis (per grade)	0.66	0.19 to 2.3	0.51
	Severe mucositis	0	0 to infinite	1
Hakim	Mucositis (per grade)	1.1	0.78 to 1.51	0.66
	Severe mucositis	1.3	0.45 to 3.74	0.64
Kitanovski	Mucositis (per grade)	0.99	0.59 to 1.65	0.98
	Severe mucositis	0.71	1.88 to 1.88	0.63
Klaassen	Mucositis (per grade)	0.91	0.67 to 1.25	0.57
	Severe mucositis	0.78	0.28 to 2.16	0.64
Lehrnbecher	Mucositis (per grade)	0.75	0.55 to 1.02	0.07
	Severe mucositis	0.41	0.12 to 1.38	0.15
Spassova	Mucositis (per grade)	1	0.85 to 1.3	0.69
	Severe mucositis	1.1	0.59 to 2.15	0.72
Sung	Mucositis (per grade)	0.73	0.26 to 2.05	0.55
	Severe mucositis	4.6	1.05 to 19.11	0.04
ZurichSPOG	Mucositis (per grade)	0.8	0.54 to 1.19	0.26
	Severe mucositis	0.5	0.11 to 2.29	0.37
PINE	Severe mucositis	0.54	0.29 to 1.02	0.05
IPD estimate	Mucositis (per grade)	0.89	0.80 to 1.00	0.05
	Severe mucositis	0.76	0.55 to 1.03	0.16

Shaded row shows single dataset with only dichotomous mucositis

Biomarkers

The datasets providing information on the four biomarkers studied in the PICNICC collaboration (CRP, IL-8, IL-6 and PCT) were much smaller than for many of the other potential predictor variables (1606 episodes with CRP, 485 with IL-8, 456 with IL-6, and 123 with PCT) and represents a smaller subgroup of the biomarkers studies. As explored extensively in the systematic reviews (with update) the desire to find effective early serum markers of infection is strong and an extension of the PICNICC project with further data collection and synthesis may well be warranted.

CRP was assessed in the greatest number of datasets, and showed a repeatedly statistically non-significant association with risk of MDI.

Table 30: Association of ln(CRP) with risk of MDI

	Predictor name	**OR**	**95% CI**	**p-value**
BaselSPOG	ln(CRP)	2.1	0.55 to 7.87	0.28
BernSPOG	ln(CRP)	0.95	0.7 to 1.3	0.73
BonnSPOG	ln(CRP)	1.2	0.44 to 3.25	0.72
Kitanovski	ln(CRP)	1.9	0.99 to 3.77	0.05
Lehrnbecher	ln(CRP)	0.91	0.74 to 1.13	0.42
Silva	ln(CRP)	0.77	0.38 to 1.56	0.47
Spassova	ln(CRP)	1.2	0.95 to 1.51	0.13
Tezcan	ln(CRP)	1.1	0.8 to 1.44	0.66
Tissing	ln(CRP)	0.94	0.72 to 1.24	0.67
ZurichSPOG	ln(CRP)	1.5	0.99 to 2.25	0.06
IPD estimate		1.1	0.95 to 1.18	0.25

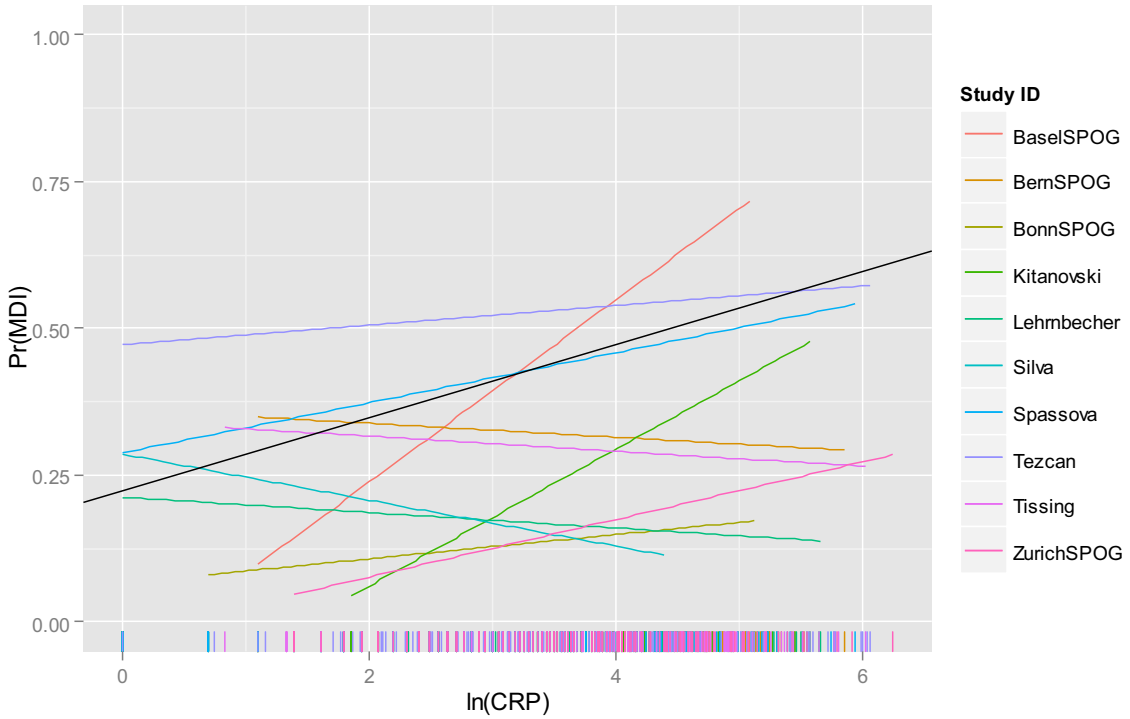


Figure 49: Studies with superimposed IPD estimate (black)

The data for IL-6 are more suggestive of a discriminatory value, with varying but consistent associations. IL-6 values appear to be most strongly associated with gram-negative bacterial infections; if a larger proportion of the 18 cases of MDI in the Kitanovski study were of this type, it may explain the very strong association seen here. Removing this study led to a smaller, but still significant, OR of 1.88 (95% CI 1.52 to 2.33).

Table 31: Association between ln(IL-6) and MDI

	Predictor name	**OR**	**95% CI**	**p-value**
Kitanovski	ln(IL-6)	7.1	2.61 to 20.88	0.00
Lehrnbecher	ln(IL-6)	2.3	1.73 to 2.99	0.00
Spassova	ln(IL-6)	1.4	0.65 to 3.23	0.37
Styjewski	ln(IL-6)	1.1	0.74 to 1.75	0.54
IPD estimate		2.10	1.71 to 2.61	4.4×10^{-12}

IL-8 has been used in the definition of a group of individuals who have received no antibiotic despite being febrile and neutropenic. In this IPD analysis, IL-8 was confirmed as being a potentially very important biomarker. It is the very low estimated intercept values of IL-8

which are of particular interest, giving the estimated proportion of MDI in patients with undetectable IL-8 (see Table 32).

Table 32: Association between ln(IL-8) and risk of MDI

	Predictor name	OR	95% CI	p-value	Estimated intercept (proportion MDI)	95% CI
BonnSPOG	ln(IL-8)	3.6	0.71 to 19.04	0.12	0	0 to 0.51
Lehrnbecher	ln(IL-8)	1.8	1.37 to 2.37	0.00	0.01	0 to 0.04
Spassova	ln(IL-8)	1.7	0.96 to 3.12	0.06	0.07	0 to 0.51
Styjewski	ln(IL-8)	1.9	1.1 to 3.28	0.02	0.04	0.01 to 0.22
Tissing	ln(IL-8)	1.9	0.94 to 4	0.08	0.01	0 to 0.27
IPD estimate		1.8	1.48 to 2.28	1.7E-08		

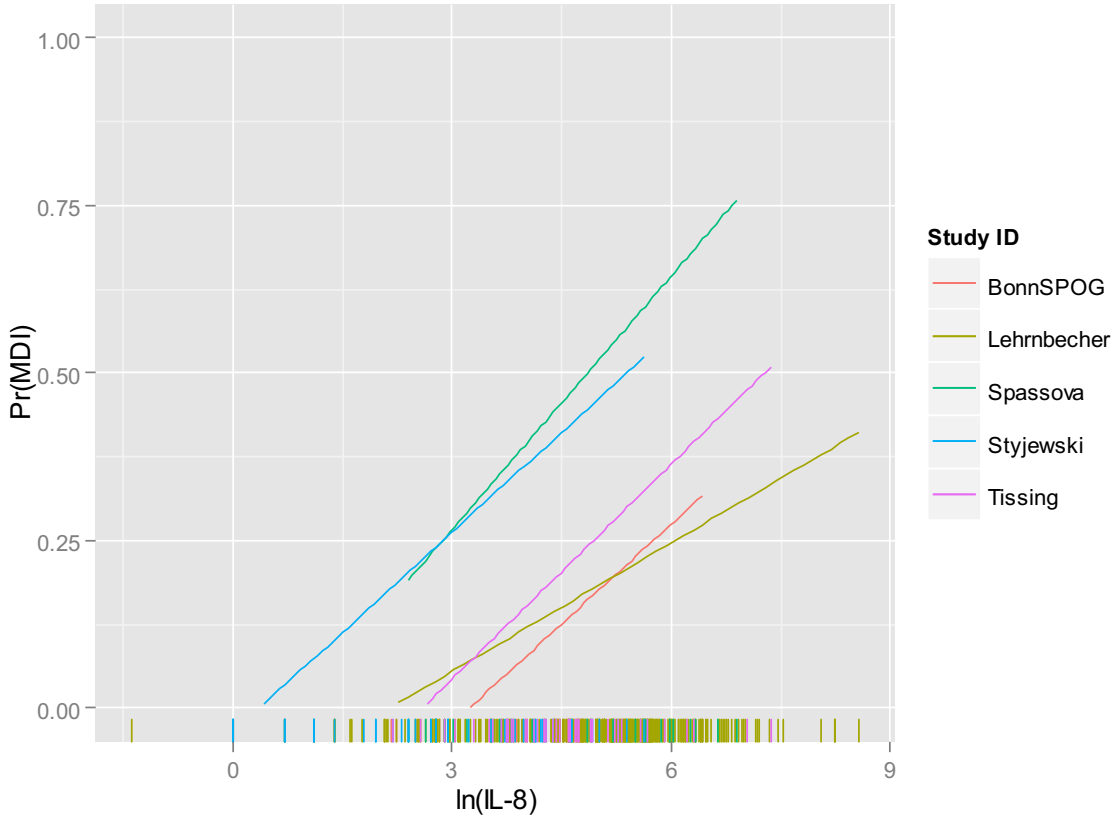


Figure 50: Association of ln(IL-8) with risk of MDI

The paucity of overall data limits the more extensive exploration of this biomarker.

Procalcitonin had the fewest data of the biomarkers studied, despite being used widely in adult respiratory medicine and critical care to help define when to refrain from commencing or discontinuing antibiotics in patients with suspected significant infections. Data were only available from two studies.

	Predictor name	**OR**	**95% CI**	**p-value**
Kitanovski	ln(PCT)	4.7	1.79 to 13.72	0.00
Styjewski	ln(PCT)	1.6	1.12 to 2.28	0.01
IPD Estimate		1.9	1.34 to 2.72	0.00033

While these are both strongly suggestive that PCT is an effective marker of infection, there are far fewer data on which to base this conclusion.

Conclusion

Analysis of the univariate relationships between the patient-specific background features, episode-specific background factors, episode-specific clinical features and episode-specific laboratory features has confirmed many of the previously suspected associations. It also revealed some novel and challenging findings, which may be explained by confounding as these are univariate associations. It has not revealed any single feature as being compellingly associated with either the absence of, or presence of, microbiologically documented infection. The proposed use of models assessing individual clustering through hierarchical random effects assessments, although statistically appealing and theoretically advantageous was not shown to meaningfully affect the association estimates for each predictor, and was therefore not pursued as a technique.

The IPD analysis confirms existing beliefs that gender is unimportant in predicting the risk of MDI, and suggested there is **not** an association between age and risk of MDI. Tumour type is, as expected, related to the risk of MDI, but this very large data set has shown an unexpected but consistent relationship between reduced risk of MDI and a bone tumour (Ewings / osteosarcoma) diagnosis.

Examination of the episode-specific background factors of the type of CVL, and intensity of chemotherapy highlighted challenging aspects of the dataset. The CVL data, providing both dichotomised and type-specific data, has shown that granular categorical data is more informative than lumped data where there are important differences between the

categories. It has also shown that selective inclusion criteria when creating the datasets can lead to paradoxical results, as demonstrated in the chemotherapy intensity variable.

Although patchy in coverage between datasets, there were sufficient numbers and broad enough coverage of out-patient status at onset of the episode to suggest a small decrease in the risk of MDI in patients presenting from home. The clinical impression of a severely unwell child or young person was strongly associated with the risk of MDI, as was an increasing maximum recorded temperature. Again, the power of such large numbers suggests an unexpected negative association between the grade and severity of mucositis, and risk of MDI.

The objective measures of blood tests for marrow function and inflammatory biomarkers showed evidence of the increased risk of MDI with lower platelets and white cell subsets, particularly AMC rather than ANC. Of the serum biomarkers, there were convincing data of the poor discriminatory value of CRP and a suggestion that IL-8 may be a very effective marker. PCT was insufficiently studied to be as convincing in value as the other markers.

The univariate associations described in this chapter are clinically interesting, in showing which features may be key discriminators and which we should not rely too strongly on in practice. If one, or more, predictors had shown very high odds ratios to “rule in” MDI, or alternatively, had extremely low intercept values, where the absence of the feature could “rule out” MDI, and that this feature was consistent and reproducible across data sets, then there would be no need to develop this analysis further.

In the absence of such a feature, the exploration of ideas of how the variables are related to each other, and where one variable provides information above and beyond that provided by others, is needed for an efficient CDR to be produced. Multivariable approaches provide this next step, and are described in the next chapter.

Chapter 9: Results of the multivariable analyses

Introduction

The primary clinical aim of the IPD analysis was to quantify the risk of adverse clinical outcomes; primarily microbiologically documented infection in children and young people undergoing treatment for malignant disease presenting with febrile neutropenia.

To recap briefly, the previous chapter showed how univariable analysis did not indicate any single feature was extremely strongly associated with MDI, or its absence, in this population. A new multivariable risk prediction model was therefore required, and this chapter details how the model was built and evaluated. The methods are fully discussed in Chapter 6.

Data selection

Data selection was very challenging, as there were a high proportion of studies with uncollected variables which considerably reduced the number of patients, episodes and studies available for a complete case analysis. Therefore, a decision was made to work with a dataset containing 1,000 episodes, estimated to contain 200 events of MDI and produce a 20:1 ratio for the analysis of 10 separate predictor variables. The method used to derive this dataset was based on taking the univariate predictors with greatest statistical significance and removing incomplete cases until the limit of fewer than 1,000 episodes was reached. The steps undertaken to reach this dataset are demonstrated in Table 33.

Table 33: Data available for complete case analysis

Constituents	Episodes	Patients	Studies	Proportion of episodes with MDI
Temp	2461	1798	15	28%
Temp + Severe Unwell	1486	925	10	27%
Temp + Severe Unwell + ANC	1348	849	9	27%
Temp + Severe Unwell + ANC + plts	1148	768	8	24%
Temp + Severe Unwell + ANC + plts + wcc	1101	742	7	24%
Temp + Severe Unwell + ANC + plts + wcc + tumour	1101	742	7	24%
Temp + Severe Unwell + ANC + plts + wcc + tumour + OP	1101	742	7	24%
Temp + Severe Unwell + ANC + plts + wcc + tumour + severe mucositis	965	616	7	25%

Temp = temperature, ANC = absolute neutrophil count, plts= platelets, wcc = white cell count, tumour = tumour type, OP = out-patient status

Model selection

The model finally selected to produce the most accurate prediction of the risk of MDI proceeded following the pre-specified method (See Chapter 6). This was by forward selection including variables with p<0.15, commencing with the demographic details and assessing model improvements using Akaike's Information Criterion (AIC; a measure of the goodness of fit of a model to the data). On the basis of this, variables were incorporated or rejected as shown in Table 2.

Table 34: Selection of model prediction terms

Variable	p-value for parameter	Model AIC	NRI (95% CI)	AUC ROC	Included or rejected
Base case		1069		0.646	
Age in days	0.689	1070	NA	0.645	Rejected
Tumour type	0.003	1076	0.012 (0.004 to 0.02)	0.644	Included
CVL	0.625	1076	0.001 (-0.001 to 0.004)	0.645	Rejected
Out patient	0.668	1076	0.008 (0.002 to 0.015)	0.644	Rejected
Temperature	0.00002	1058	0.015 (0.006 to 0.024)	0.671	Included
Shock	0.392	1032	NA	0.670	Rejected
Mucositis	0.295	1031	NA	0.672	Rejected
Severely unwell	0.000005	1039	0.01 (0.001 to 0.018)	0.697	Included
Haemoglobin	0.012	1035	NA	0.701	Included
log(Platelets)	0.028	1029	NA	0.703	Included initially, rejected when WCC included
log(WCC)	0.00003	1021	0.018 (0.007 to 0.028)	0.723	Included, Rejected plt
log(ANC)	0.585	1012	NA	0.723	Rejected
log(AMC)	0.0002	999	NA	0.736	Included
Antibiotics (sensitivity)	0.989	1017	NA		(Essentially unchanged)

CVL – central venous line; WCC – white cell count; ANC – absolute neutrophil count; AMC - absolute monocyte count. NRI – net reclassification improvement on previous model, in classification of low risk (<5% MDI)

The final model

The final model included:

Tumour type + Temperature + Severely unwell + Hb + log(WCC) + log (AMC)

$$\text{Logit}(pMDI_{ik}) = \alpha + \beta_{0k} + \beta_1 \text{ tumour type}_i + \beta_2 \text{ temperature}_i + \beta_3 \text{ unwell}_i + \beta_4 Hb_i + \beta_5 \log_e (\text{white cell count})_i + \beta_6 \log_e(\text{absolute monocyte count})_i$$

The value of the model can be assessed by comparing it to simpler versions to evaluate if the increase complexity can produce meaningful benefits. When the full model was compared with the simplest model possible, using only study-ID to predict the risk of MDI, this produced a net reclassification improvement (NRI; a measure of the clinically relevant improvement of identifying children correctly as infected or non-infected) of 0.079, and AUC ROC improved from 0.646 to 0.736. Using clinical variables only (tumour type, temperature, severely unwell) gave an AUC ROC of 0.697. The addition of the simple full blood count variables (haemoglobin, white count and monocyte count) improved the prediction further with an NRI of 0.042, and AUC ROC improved to 0.736.

The predictive estimates of the variables were then revised using bootstrapping, as described in Chapter 6, re-sampling 5,000 iterations with replacement. For most predictive estimates, there appeared to be little bias (see Table 35). For those data items with sparse data, for example, the tumour type GCT (n=7) the bootstrapped estimates were not normally distributed and the median estimate is markedly different than that initially derived (see Appendix 28. Detailed estimates from final multivariate model).

Table 35: Selected parameter estimates and bootstrap values

	AML	GCT	Osteo-sarcoma	RMS	Temp-erature	Severely unwell	Haem-oglobin	log(WCC)	log(AMC)
Original estimates	0.655	-0.069	-1.193	-0.244	0.566	0.786	0.180	-0.299	-0.209
Bootstrap Median	0.661	-0.148	-1.229	-0.253	0.588	0.809	0.182	-0.309	-0.215
Lower limit 95%ile	0.128	-16.046	-2.889	-0.952	0.296	0.400	0.081	-0.516	-0.332
Upper limit 95%ile	1.158	1.832	-0.288	0.393	0.905	1.203	0.292	-0.103	-0.099
Difference	0.006	-0.078	-0.036	-0.008	0.022	0.023	0.002	-0.010	-0.005
% variation	1%	113%	3%	3%	4%	3%	1%	3%	3%

The shrinkage estimate for the model [243] was calculated to be 0.97, in keeping with the very small differences produced by the bootstrap values.

Validation

The new model was tested by assessing the calibration (comparing the actual proportions of MDI in patients with the predicted proportion of MDI) and by assessing it's discriminatory value in classifying patients at very low risk (<5%) of MDI

Calibration

In order to test the model's calibration, we extracted the predicted risk of MDI using the derivation dataset, but using a generalised intercept based on a meta-analysis of the study-variable intercepts from the model, rather than the study-specific intercept of the individual patient, in an attempt to produce a result more likely to be applicable in future clinical practice (as intercepts for each practice are unlikely to be available in reality). These were then shown graphically by plotting the predicted result against actual outcomes, grouped by either deciles of predicted risk (i.e. the predicted probability from those with <10%, 10% to

<20%, etc.) or by deciles of population (i.e. the predicted probability from the first 10%, 20%, etc.)

Figure 51 shows the actual average % of MDI per predicted MDI decile (0-10%, 10-20%, etc.), comparing the actual rate of MDI in these patients with the value calculated from the model. It also shows, on the categorical axis, the percentage of patients who fall into the risk grouping.

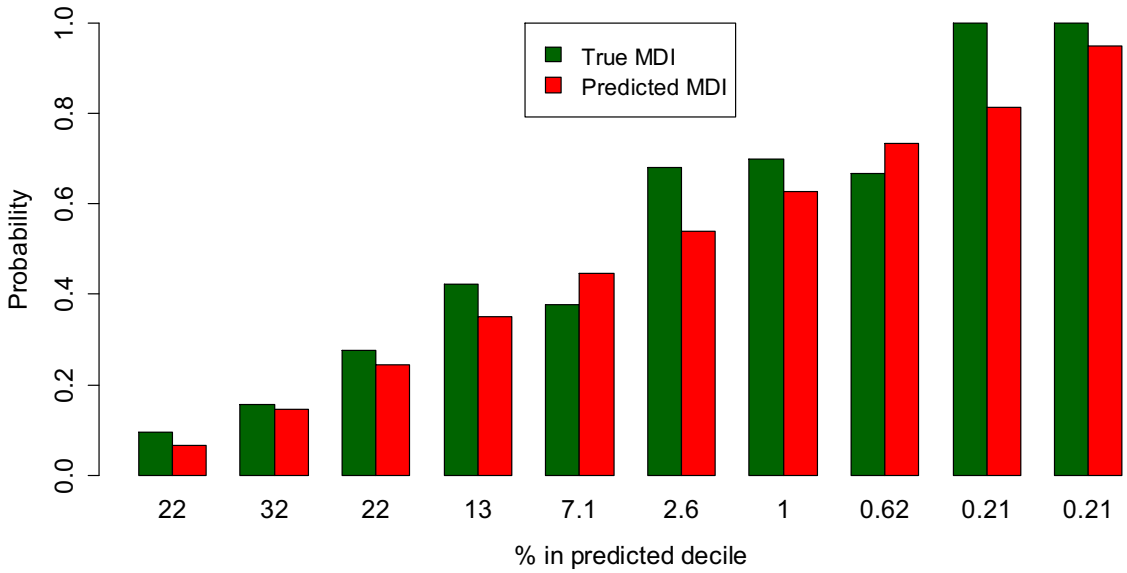


Figure 51: Discriminatory performance of the new model

An alternative visualisation is to plot the grouped observations against each other, see Figure 52.

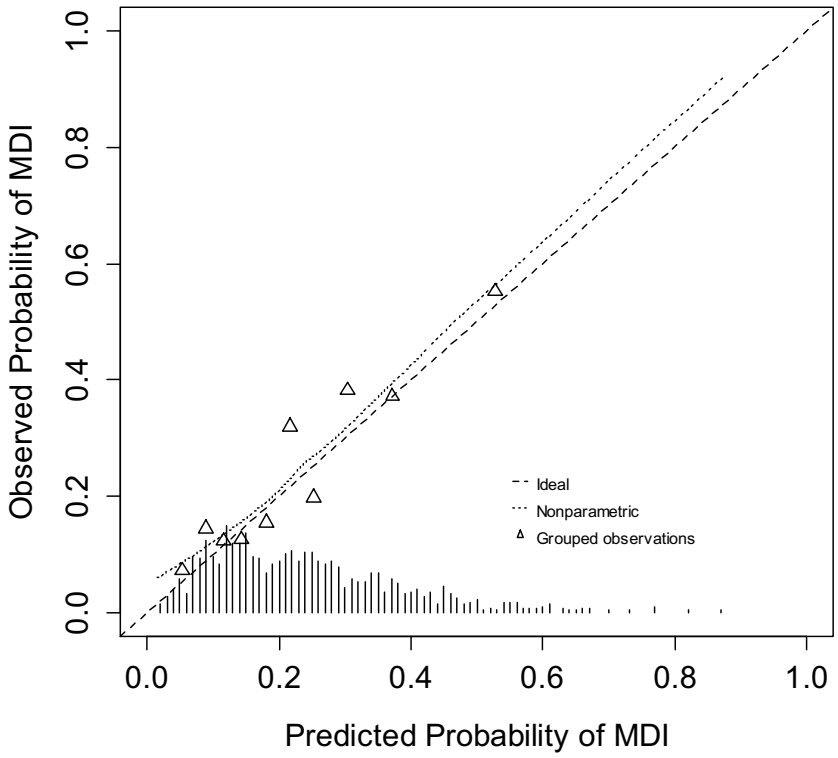


Figure 52: Calibration plot of the new model

This Figure shows the distribution of predicted probabilities as a rug plot, with the height of the line showing the frequency of the predicted probability. The group estimates are shown in deciles of the population, rather than deciles of the predicted values, as triangles. The dashed line of "perfect fit" at 45° shows where predicted and actual probabilities would perfectly intersect; the dotted line is a Lowess smoothed curve of the predictive versus actual values.

These figures show that the predictive value of the model is very close to the actual probability of MDI across the range of predictions, showing no major systematic bias, though there is a slight underprediction of risk between approximately 0.2 and 0.4.

The use of bootstrapped median parameter estimates made no effective difference to the model's predictive ability (see Appendix 28. Detail, Figure 65).

Discrimination

Assessments of discriminatory validity were based on a categorising the results to lower than 5% chance of MDI or greater. This value was chosen as discussed in Chapter 6 by an expert consensus with parent/carer involvement to reflect a very conservative estimate of acceptable risk of MDI. This was the same cut-off as the NRI calculation for model derivation. The results are shown in ROC space in Figure 53, demonstrating an AUC (also known as C-index) of 0.723 (95% CI by bootstrapping 0.685 to 0.785).

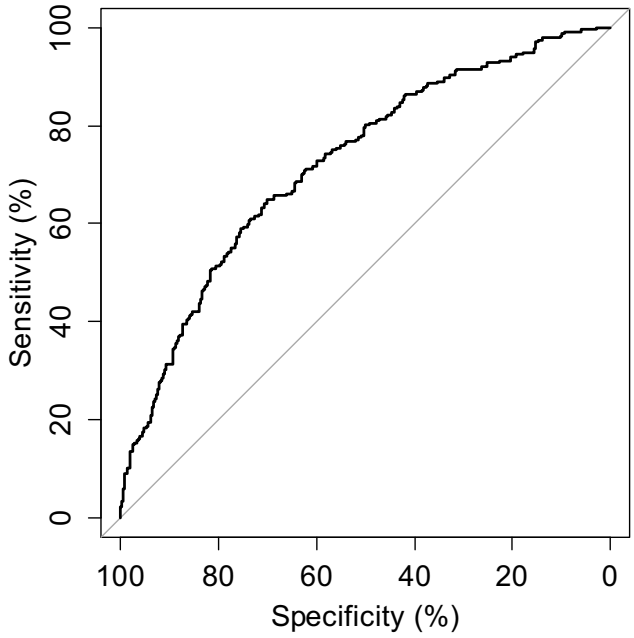


Figure 53: Model discrimination at in ROC space

An alternative approach to this is to examine how many patients fall in the correct category, and to how many this rule would apply. This is shown in Table 36 where the model places 57 patients (6% of the total) into a low risk category, of whom 2 (~4%) were misclassified and have an MDI. This type of dichotomous "rule" use of the model can also be described by its sensitivity (99.2%) and specificity (7.5%).

Table 36: Discrimination matrix (2x2 table) for 5% risk of MDI

	MDI	**No MDI**
High risk	234	676
Low risk	2	55
	Sn= 99.2%	Sp = 7.5%

(LR is 6% total patients and 4% LR patients are misclassified)

Cross validation

Validation was undertaken by comparing the results of the initial analyses against the data from all bar one of the studies in turn (cross validation of intrinsic prognostic performance), without significant differences in calibration (see Appendix 28. Detail) or discrimination (see Table 37).

Table 37: Calibration and discrimination values of all-bar-one analysis

Study removed	AUC ROC	Sensitivity	Specificity	Proportion in LR group	Misclassified
BaselSPOG	0.724	99.2%	7.6%	5.9%	3.5%
BernSPOG	0.718	99.0%	8.1%	6.4%	3.8%
BonnSPOG	0.723	99.1%	7.9%	6.1%	3.5%
Hakim	0.710	100.0%	3.9%	2.9%	0.0%
Kitanovski	0.712	99.1%	7.2%	5.7%	3.9%
Klaassen	0.766	98.2%	11.7%	9.6%	3.8%
ZurichSPOG	0.720	99.1%	7.3%	5.7%	4.0%

The discrimination qualities were also tested by bootstrapping (2,000 samples, with replacement, comparing predicted risk of MDI against actual risk) reporting the sensitivity, specificity, percentage of patients classed as low risk and percentage of patients "misclassified" – with an MDI in the low risk group.

Table 38: Bootstrapped estimates of discrimination

	Median	Lower 95% confidence limit	Upper 95% confidence limit
Sensitivity	99.2%	97.8%	100.0%
Specificity	7.6%	5.6%	9.7%
Proportion in LR group	5.9%	4.3%	7.4%
Misclassified	3.2%	0.0%	8.4%

Sensitivity analyses

The PICNICC model for the prediction of probability of MDI was robust to cross validation of intrinsic prognostic performance and bootstrapping, and the heuristic estimate of uniform shrinkage was also low (0.97), all congruent with a prediction model which was not overfitted or overoptimistic.

A sensitivity analysis adding antibiotic type to the model was planned in the protocol. This did not change the predictive value of the model (see Table 34). This is to be expected as a standard set of antibiotics tends to be given in each hospital, without reference to any variable which may be predictive of MDI.

A series of further sensitivity analyses were undertaken addressing other potential challenges to validity. Examining the model for evidence of study-level variation is important as only seven of the 22 datasets could contribute sufficient variables to provide information for the model. If these studies were importantly different than those which could not be included, the validation procedures undertaken could fail to give an accurate estimate of how useful the model may be in practice.

This potential problem was addressed by undertaking a sensitivity analysis, where an average value for each of the missing variables was used where it had been unrecorded. This is a simple initial step in addressing missing data through imputation, and further research could consider multiple imputation techniques. To fit this "average imputed" model, the mean values of each continuous variable drawn from the entire dataset was placed where an NA was recorded. The "severely unwell" variable was coded at 0.17; the proportion of episodes where severely unwell had been recognised. The tumour type was coded as "ALL"; this was the most common diagnosis, and also the reference diagnosis for tumour type.

Using these values, the model showed much worse discrimination, with an AUC ROC of 0.619 (see Figure 54), as expected with the insertion of large quantities (~50%) of undiscriminating data.

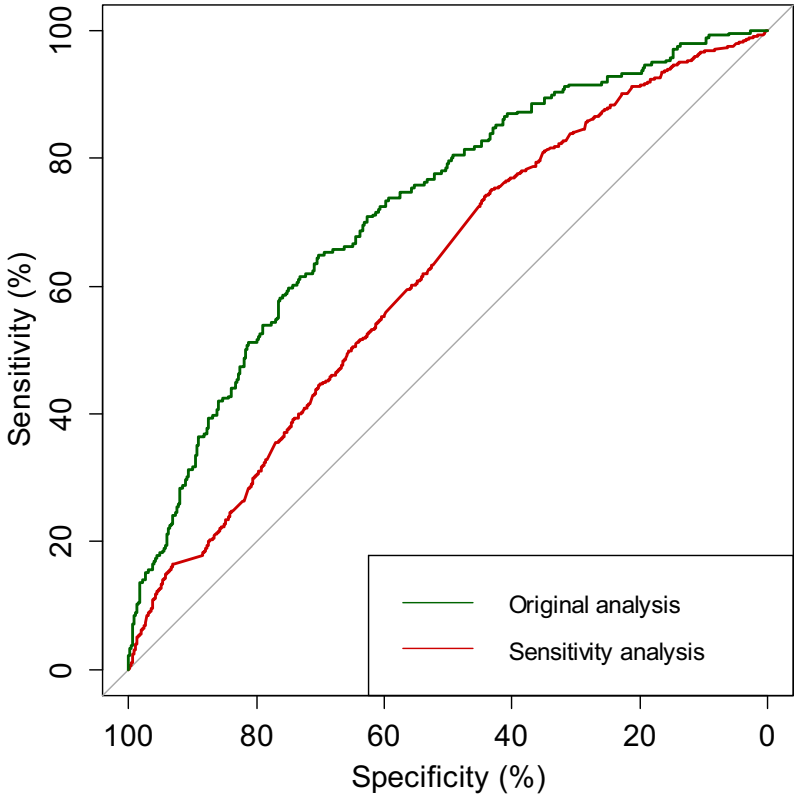


Figure 54: ROC curve for "average imputed" sensitivity model

The test accuracy values were similarly reduced (see Table 39) for a cut-off of 5% chance of MDI for low-risk.

Table 39: Discrimination values for "average imputed" sensitivity model

	Original	**Sensitivity analysis dataset**
Sensitivity	99.2%	98.1%
Specificity	7.6%	4.0%
Proportion in LR group	5.9%	3.4%
Misclassified	3.2%	15.6%

The model also demonstrated worse calibration, as seen by the distribution of actual versus predictive probabilities in Figure 55, particularly in the "bunching" of probabilities in the

~20% range. The plot demonstrates that this sensitivity analysis data typically underestimates probabilities, but follows a reasonably similar slope for most of the curve. Again, this is highly likely to be due to the very large amount of non-discriminatory data.

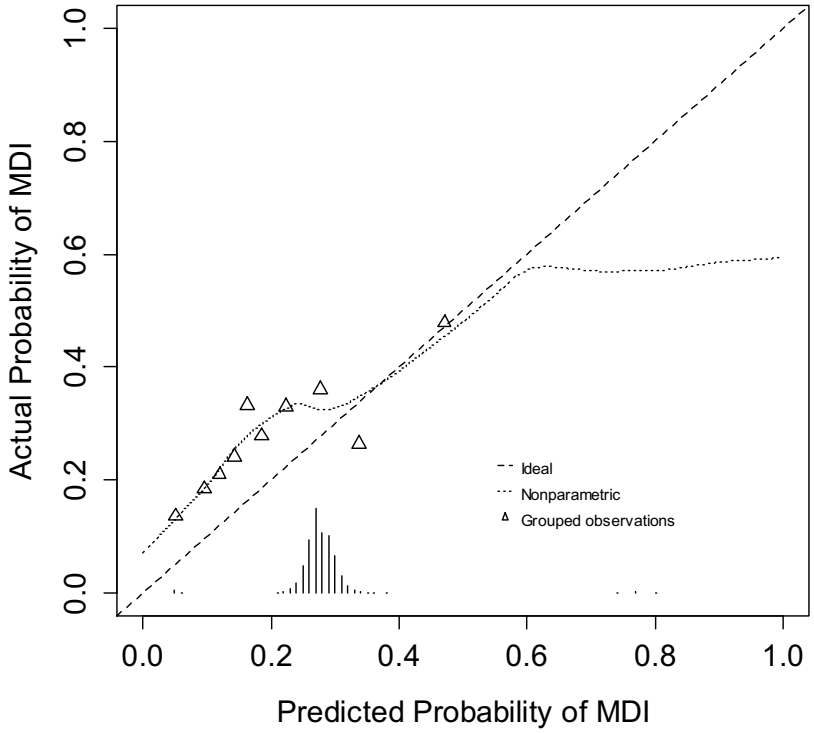


Figure 55: Calibration plot for sensitivity analysis

Examining the parameter estimates obtained by fitting the model to this "average imputed" data shows the predictive values of clinically recognised "unwellness"; temperature, WCC and AMC remain similar. The point estimates of the predictive values of some tumour types changes, and the value of haemoglobin as a predictor is also reduced, although remains predictive (see

Table 40).

Table 40: Parameter estimates in model fitted to original and sensitivity analysis data

	Original dataset OR (95% CI)		Sensitivity analysis OR (95% CI)	
AML	1.92	1.17 to 3.17	1.15	0.91 to 1.45
Brain	0.63	0.3 to 1.35	1.39	0.79 to 2.43
Carcinoma	NA	NA	0.39	0.05 to 3.11
Ewings	0.53	0.14 to 1.93	0.50	0.3 to 0.83
GCT	0.93	0.17 to 5.2	0.91	0.3 to 2.76
Hepatoblastoma	1.61	0.53 to 4.91	0.13	0.02 to 0.98
High grade Brain	0.71	0.29 to 1.75	0.78	0.49 to 1.23
Hodgkins	0.67	0.17 to 2.63	0.97	0.55 to 1.73
HR-NBL	2.51	0.69 to 9.17	0.39	0.23 to 0.67
LCH	NA	NA	1.33	0.41 to 4.37
LGBrain	NA	NA	0.83	0.31 to 2.25
LR-NBL	NA	NA	0.41	0.09 to 1.93
Lymphoma	NA	NA	0.67	0.37 to 1.21
NBL	1.6	0.61 to 4.23	1.52	0.96 to 2.42
NHL	0.62	0.34 to 1.16	0.70	0.5 to 0.96
Nonmalignant	NA	NA	1.02	0.33 to 3.17
Osteosarcoma	0.30	0.1 to 0.92	0.36	0.21 to 0.6
Other	2.22	0.49 to 9.99	0.42	0.18 to 1
Retinoblastoma	1.73	0.32 to 9.26	0.17	0.02 to 1.4
RMS	0.78	0.42 to 1.46	0.76	0.5 to 1.17
Sarcoma	1.21	0.24 to 6.03	0.96	0.48 to 1.92
Solid	NA	NA	0.43	0.27 to 0.69
Wilms	0.61	0.17 to 2.24	0.78	0.47 to 1.29
Temperature	1.76	1.33 to 2.34	2.03	1.61 to 2.56
Severe unwell	2.20	1.5 to 3.21	2.20	1.63 to 2.98
Haemoglobin	1.20	1.09 to 1.32	1.06	0.99 to 1.15
ln(WCC)	0.74	0.61 to 0.9	0.75	0.67 to 0.84
ln(AMC)	0.81	0.72 to 0.91	0.87	0.79 to 0.96

In both models only the osteosarcoma has statistical significance as a predictor, AML losing significance in the sensitivity analysis. Some of the tumour types reverse their direction of association: brain, high-risk neuroblastoma, hepatoblastoma and retinoblastoma.

Modified model

On the basis of the complexity of the tumour type classification and the huge uncertainty associated with many of the subtypes, a pragmatic decision was made to simplify the values with only AML (OR 1.5), Ewings sarcoma (OR 0.5) and Osteosarcoma (OR 0.3) being different to ALL (OR 1, reference value). These were chosen because they were consistent in the data sets (the sarcomas) and clinically recognised as being having a serious concern about repeated severe infection (AML).

This model, when applied to the original derivation dataset, produced a marginally less effective discrimination (ROC 0.709 cf 0.723) mainly through a loss in specificity (see Table 41). When used in the "average imputed" sensitivity analysis dataset it showed smoother discrimination and calibration characteristics (ROC 0.629 cf 0.619), though remained far worse than the original dataset (see Table 41 and Appendix 28. Detail).

Table 41: Discrimination characteristics of simplified model

	Original	Tumour simplified model	Sensitivity analysis dataset	Simplified model in sensitivity analysis
Sensitivity	99.2%	99.2%	98.1%	99.2%
Specificity	7.6%	5.3%	4.0%	2.6%
Proportion in LR group	5.9%	4.2%	3.4%	2.1%
Misclassified	3.2%	4.9%	15.6%	10.5%

Comparison

The new PICNICC model was compared against previously proposed and evaluated risk stratification rules identified in the systematic reviews. The rules were selected because they had previously shown excellent properties (Santolya [22]), had been suggested in national or international guidelines (Alexander [134]), were very recently published (Swiss Paediatric Oncology Group, SPOG [144]) or were extremely simple (Rackoff [141]) (see Table 42). The

rules were all designed to stratify the risk groups into low risk and high risk and so were assessed for their discrimination rather than calibration, and were undertaken in the PICNICC dataset when studies which had been used to derive the rules had been excluded.

Table 42: Previous Rules for comparison

Elements	Rackoff	Alexander	Santolaya	SPOG
Patient and disease related factors	None	AML, Burkitt lymphoma, induction ALL, progressive disease, relapsed with marrow involvement	Relapsed leukemia, chemotherapy within 7 days of episode	4 points for chemotherapy more intensive than ALL maintenance
Episode specific factors	Absolute monocyte count	Hypotension, tachypnea/hypoxia <94%, new CXR changes, altered mental status, severe mucositis, vomiting or abdominal pain, focal infection, other clinical reason for in-patient treatment	CRP ≥90 mg/L, hypotension, platelets ≤50,000/uL	5 points for hemoglobin > 9 g/dL, 3 points each for white blood cell count <300/uL, platelet < 50,000/uL
Rule formulation	Absolute monocyte count > 100/uL= low-risk of bacteremia	Absence of any risk factor = low-risk of serious medical complication	Zero risk factors or only low platelets or only <7 days from chemotherapy = low-risk of invasive bacterial infection	Total score <9 = low-risk of adverse FN outcome

The analyses show that the previous rules have worse discrimination as the rule derived from the PICNICC model, as measured by the AUC ROC, and one rule (Santolaya) included a greater proportion of patients with MDI in the low risk than the high risk group.

This illustrates the dilemma that is faced when choosing a rule to use, and which has been explored initially in Chapter 1 and will be discussed at greater length in Chapter 10. Rules

which have a high sensitivity have a low specificity, and while very “safe” apply to very few patients. If “safety” was the greatest concern then the Rackoff rule may be considered better than the PICNICC rule because the misclassification rate is 0% (but applies only to 3% of episodes). If a greater misclassification rate is acceptable, with lower sensitivity but greater specificity, then the SPOG rule could be considered a better rule as it applies to a much greater proportion of patients (43% cf. 6%) but at the “cost” of an increased misclassification rate (10% vs. 3.5%).

Conclusions

The PICNICC dataset allowed the development of a new clinical prediction rule to determine the risk of microbiologically documented infection consisting of a combination of three clinical features (tumour type, maximum temperature and the clinical appearance of being significantly unwell) and three elements of the full blood count (haemoglobin, white cell count and absolute monocyte count). This model was developed without including biomarkers due to limited quantities of data. This clinical model had moderate discrimination and calibration, with fair agreement between the predicted probabilities and actual rates of MDI. These results were then dichotomised at a predicted value of 5% or less to produce a low risk group. This produced a highly sensitive, through poorly specific, rule which applied to ~6% of the population.

The model and resultant rule appeared to be robust to internal validation techniques, but exploratory sensitivity analyses examining the model performance across the non-included studies offer a suggestion that the inclusion of haemoglobin may not prove to be efficient when the model is used in future populations, and that a simpler approach to tumour type inclusion may be warranted.

Chapter 10: Discussion

Introduction

Children with cancer in Europe now have an 80% chance of cure[2]; this has been possible through meticulous attention to treatments directed at their cancer and supporting them through the side effects of these therapies. The cost of this cure, in terms of intensity of therapy and recurrent admissions with toxic effects, is considerable and a burden upon children, young people and their families [58]. One such toxicity, fever with neutropenia (FN), also known as “febrile neutropenia” or “neutropenic sepsis”, has been the focus of this thesis, which has addressed the issues of balancing risks and personalising care in FN by undertaking risk stratification at each episode to differentiate who was at higher or lower risk of significant infection, and who was potentially eligible for alternative treatment approaches.

What was already known

- Cancer in children is curable in an increasingly large proportion of cases
- Emergency readmission in patients with fever and neutropenia (FN) is the second largest reason for hospitalisation (after chemotherapy delivery)
- Risk stratification had been proposed using a number of different approaches

What this thesis adds

- The range and heterogeneity of studies exploring risk prediction in paediatric FN was captured in five systematic reviews
- The reviews influenced national and international guidelines in FN
- An international collaboration of 22 groups over 15 countries formed to share information from previously undertaken studies and explore clinical and methodological problems
- A robust prediction rule was developed to predict the risk of microbiologically documented infection
- Methodological and graphical refinements of the approaches undertaken were developed and disseminated

Background

At the onset of this programme of study, it was clear that current practice in managing febrile neutropenia in paediatric oncology was variable, both nationally [53] and internationally[51, 54-55]. Some centres used a risk-stratified, reduced intensity approach, directed by clinical decision rules (CDR) whereas others treated all children with aggressive antibiotic therapy.

An ideal system for FN management would predict the risk of adverse outcomes, classically the occurrence of a microbiologically documented infection, using clinical data collected at or soon after presentation. These data can be clinically derived from the simple act of reviewing and examining a patient, performing basic bedside tests, such as measuring pulse rate or blood pressure. They can also be obtained from routine laboratory investigations such as the measurement of full blood count, or from more specialised tests, including the measurement of serum biomarkers of inflammation.

Information from these data sources would then be used in a simple system (a predictive model) to determine the risk of a microbiologically documented infection for each episode. The output of such a predictive model could then be used in two different ways: to decide if the risk was low enough to allow out-patient management; and at the opposite end of the risk scale, to consider the need for increasingly close observation and more aggressive management. Developing predictive models needs to be done in a robust manner to reduce the effects of chance, confounding and bias obscuring the true relationships between proposed predictor variables and the outcome of each episode of febrile neutropenia. Furthermore, a rule needs to be tested, to make sure it works effectively and is practically useful.

Building a robust and reliable model required obtaining large numbers of well-collected data from FN episodes that had occurred in different places and at different times. As with much paediatric research, it was noted that essential problems of research in this area were linked to the condition's rarity and small numbers of cases, and limited collaboration in primary studies.

The first step in addressing this issue was to systematically review existing studies to assess published CDR and the value of serum biomarkers in performing risk stratification, to determine whether an individual patient data (IPD) meta-analysis was necessary; to identify suitable data sets; and to guide the development of such a study. These reviews were

subsequently updated [115-116] to inform the development of national [128] and international [129] guidelines in this area.

The results of the systematic reviews of CDR in the prediction of adverse outcomes from episodes of FN included a total of 10,431 patients from thirty-four studies developing twenty-one different CDR. Four of these rules stood out for their simplicity (using just temperature and absolute monocyte count: Rackoff[141] rule), their predictive ability (Santolaya [22] and SPOG [144] rules) or their current use in some areas of the UK (Alexander [154] rule). The serum biomarkers reviews included over 5,100 episodes assessing twenty-four different markers of inflammation or infection. Those markers with fewer studies appeared to have better characteristics but much greater uncertainties.

Taken together, these reviews suggested a CDR for the prediction of poor outcomes during episodes of febrile neutropenia could be effective, and that there was potential additional value from the incorporation of serum biomarkers. None of the rules had been subject to extensive geographical and temporal discriminatory validity assessments, and many potential difficulties with the studies were identified.

To maximise the value of the information already collected by these and other cohorts of children with febrile neutropenia, an international collaboration was established to facilitate an individual-patient-data (IPD) meta-analysis to develop and test a new prediction model[231].

Collaboration

The “Predicting Infectious ComplicatioNs In Children with Cancer” (PICNICC) collaboration was formed by engaging international clinical and methodological experts, authors of studies identified in the systematic reviews, parent representatives and healthcare researchers. The PICNICC collaboration consists of twenty-two different study groups from fifteen countries. Although PICNICC was created via (and provided the data and drive for) this PhD, the collaboration will continue to develop and progress research into infection in children and young people with cancer beyond this thesis.

Aims and Data collected

A protocol for the PICNICC IPD study was developed, registered and published prior to commencement of the analysis.[231]

The primary aim of the IPD analysis was to quantify the risk of adverse clinical outcomes according to clinical variables in children and young people undergoing treatment for malignant disease who present with an episode of febrile neutropenia; and to develop a new risk prediction model. A further aim was to develop methodological approaches to IPD analysis in the development of predictive models, including the graphical display and communication of such information. This thesis focused on the outcome of microbiologically documented infection (MDI) because it was the most completely reported and most objective and frequently occurring significant complication of paediatric FN.

IPD information from 5,127 episodes of FN in 3,504 patients was provided for analysis. A wide variety of malignancies were represented, in keeping with the disparate nature of diagnoses treated in paediatric oncology/haematology units. The median age of the patients was 6.8 years, with a range of 50 days to 25 years old; 56% of the patients were male.

Assessment of the IPD collected showed a wide range of outcomes and potential predictor variables demonstrating marked differences in completeness, interpretation and consistency. No dataset completely reported every item. Tumour type was the most fully collected data item, with only four episodes having missing data; procalcitonin was only reported in 93 episodes, and missing in 5,034 Data were largely absent because individual studies did not record variables. The nature of the unrecorded data effectively reduced the dataset available to undertake multivariate analysis to around 1,000 episodes in 600 patients over seven studies, still greater than the previous largest study of 447 episodes in 227 patients. [139]

The small quantity of information available on biomarkers in comparison to the clinical data meant that they were assessed only in their univariate relationship with MDI and not as part of a multivariable model.

Results - Associations with microbiologically documented infection

The results of univariate analyses showed expected associations between potentially predictive covariates and MDI including the presence of an untunnelled central line, the

clinical appearance of significant unwellness, of documented cardiovascular compromise (shock), a high temperature, raised serum biomarkers, low white cell counts and platelets, a diagnosis of AML, and undergoing treatment for relapsed disease. There was no clear relationship demonstrated between age and risk of MDI. Two surprising potential associations were found: osteosarcoma/Ewings sarcoma patients and patients with more severe mucositis were associated with a decreased risk of MDI.

The biomarkers studied were C-reactive protein, procalcitonin, interleukins 6 and 8. Of note, CRP was not significantly associated with MDI, and while only studied in five datasets, IL-8 was interesting in both the strength of association (OR 1.8, 95% CI 1.48 to 2.28) and very low intercept implying that very low levels were potentially able to rule out MDI.

Mucositis was reported differently in the different study sets. More data were present on the presence/absence of severe mucositis than a graded response. Contrary to popular belief, the presence of increasingly severe mucositis was associated in univariate analysis with a decreased risk of MDI, and the relationship was consistent across studies. No single study would have detected this association. However, when subject to multivariable analysis, mucositis no longer provided any information above the tumour type and maximum temperature. This may be explained by the strong association between mucositis and type of chemotherapy delivered, and the malignant diagnosis of the patient.

Multivariable model building

The multivariable predictive model derived had six components: Tumour type, temperature, clinical description of being “severely unwell”, and the results of measurements of haemoglobin concentration, total white cell count and absolute monocyte count. The model is:

$$\text{Logit}(p(\text{MDI})_{ik}) = \alpha + \beta_{0k} + \beta_1 \text{ tumour type}_{ik} + \beta_2 \text{ temperature}_{ik} + \beta_3 \text{ unwell}_{ik} + \beta_4 \text{ Hb}_{ik} + \beta_5 \log_e(\text{white cell count})_{ik} + \beta_6 \log_e(\text{absolute monocyte count})_{ik}$$

(where the subscript i refers to the i’th patient, and k the k’th data set)

This predictive model showed moderate discrimination (AUC ROC 0.736) and good calibration between predicted and actual estimates of the risk of MDI when assessed across the range of predictive values. The rule was robust to bootstrap and cross-validation sensitivity analyses.

This model produced predicted probabilities of MDI. A clinically useful dichotomy was then introduced by creating a 'rule' stating episodes with a predicted risk of MDI <5% were 'low risk' episodes and all others 'not-low-risk' episodes.

This clinical decision rule was highly sensitive, which means that those patients classified as low risk are extremely unlikely to have a microbiologically documented infection (~3%). However, the rule was very poorly specific, which means that many of the patients classed as "high risk" do not have a documented infection, and also that the rule only classified a very small proportion (between 4% and 7%) of the episodes of patients presenting with FN as low risk.

Comparison with other low risk rules

The full PICNICC prediction model had better discrimination (based on the AUC ROC) than the other four published models selected for comparison (Rackoff[141] rule, Santolaya [22], SPOG [144] and Alexander [154] rule).

When the PICNICC model was converted into a dichotomous rule by using a threshold of 5% predicted risk of MDI, the PICNICC model had equivalent sensitivity (99.2% vs. 100%) to the Rackoff rule but better specificity (7.5% vs. 3.7%). It was more sensitive than the SPOG rule (99.2% vs. 80.8%) but at the expense of specificity (7.5% vs. 50%), which led to a much smaller proportion of the population being classified as low risk (6% vs. 43%). The Santolaya and Alexander rules performed poorly.

The appropriate choice of a low risk rule to use in clinical practice requires a discussion of: the acceptable threshold value, which influences the proportion of patients who are classified as low risk and the proportion who will have a microbiologically documented infection in this group; the ease of implementation; and the reproducibility and reliability of the rule across different locations.

Threshold choice

The threshold of 5% was based on a consensus of the collaborating members who met at the Congress held to discuss the initial analyses and report the IPD findings. This included a series of clinically active research physicians, a parent whose child had undergone treatment for malignancy and who had experienced FN, and statisticians. If this very strict definition of

low risk is used to decide who may be eligible for out-patient antibiotic therapy, then there will be limited chance to markedly decrease the proportion of patients who are hospitalised for treatment of FN. However, if the threshold value had been set at a 10% risk of MDI, the derived rule would have applied to approximately 20% of patients, which would have an important impact on the application of the rule.

The clinical implications of a choice of threshold are extremely important. “Too low” a threshold will be exceptionally safe. It will not risk sending a child or young person out of hospital that may have a microbiologically documented infection developing. However, it will severely limit the size of the group for whom out-patient therapy will be judged possible, and will result in a large number of children and young adults being hospitalised unnecessarily. A higher threshold may increase the number of patients who will need to be readmitted after an MDI is identified, but will allow a greater number of patients to receive out-patient care. Identifying how this threshold should be set requires a balance of the costs, risks, and benefits of the different thresholds. Such an investigation requires a specific research project to identify the key factors involved and assess the opinions of the various stakeholders: children and young people; parents or carers; and health professionals.

Clinical implementation

The clinical implementation of prediction model or rule will require it to be believed by the clinical teams, based on sound data, and easily usable in practice. The full PICNICC model has complexity (with the series of different predictors for tumour type, and the use of log-transformed data) which makes it likely to be unwieldy unless made easily applicable.

A basic implementation of the predictive model has been made ‘live’ on a shared google-drive spreadsheet: http://tinyurl.com/PICNICC1. This could be easily adapted to work off a standard (or smartphone) web page or ‘app’. An alternative formulation for a dichotomous ‘rule’ based on the model could be created as a nomogram.

Limitations of this study

The lack of commonly agreed clear definitions has restricted what can sensibly be analysed to generate a predictive model for microbiologically defined infection. A further analysis could be undertaken using a definition of “significant adverse outcome” concluded upon by

the Thesis Advisory Panel, or a group of the PICNICC collaborators. This would have some benefits; it would produce a further model, and allow investigations as to the similarities and differences between the models and CDRs produced. The disadvantage of producing a further CDR which would again use an outcome without wide agreement of its utility by clinicians and the families to whom it would be applied, and this may importantly reduce its chance of being taken up and used in practice, and this is felt to outweigh any potential benefits.

The issue of missing data was more significant than originally envisaged, as data were found to be missing almost entirely at the level of study, whereby predictor or outcome variables were "not recorded" rather than "missing". This has led to a much smaller dataset being available for the development of the multivariable analyses, and so reduced precision. In particular, the pattern of data collected has meant that the recipients of bone marrow transplant/autologous stem-cell rescue chemotherapy are not adequately represented in the dataset.

Strengths

- Robust systematic reviews of published studies underpinned the development of the IPD analysis
- Geographically and temporally varied datasets included in the IPD data
- Clinically sensible analyses conducted to derive a meaningful prediction model
- Robust to most sensitivity analyses

Limitations

- Missing data a larger and more comprehensive problem then initially foreseen
- Heterogeneity of definitions for some clinically important outcomes limited analyses to be undertaken
- Lack of collaboration with groups collecting quantities of biomarkers data limited the analyses addressing their predictive ability

The related issue of limited biomarker data also restricted analyses. It is not known exactly why some groups did, and others did not agree to collaborate in this project, despite requesting feedback, and so considerations of why this is the case are tentative. It appears

that fewer infectious disease led groups who have explored biomarkers have taken part compared to the oncology led groups. It may be that those groups, working primarily on the laboratory evaluation of such markers may not have collected sufficient patient-level clinical information to be eligible for inclusion. It may be that an oncology-led group failed to generate sufficient peer-recognition to encourage the clinicians to become involved. The lack of sufficient data provides an opportunity to develop this research further.

The prediction model developed contained five items which were relatively consistent across the different study groups (tumour type, temperature, unwellness, absolute monocyte count and total white cell count) and one item (Hb) which was heterogeneous across studies when assessed in univariate analyses. The limitations in the unreported data meant it was not possible to tell accurately if the heterogeneity found in the univariate analysis would also be present when assessed in a multivariable model. Therefore, it is difficult to tell if the inclusion of haemoglobin would be applicable in alternative datasets and subsequently in clinical practice.

Further research

This thesis has completed an analysis of the PICNICC dataset focusing on the prediction of microbiologically documented infection. There remain a number of research opportunities available from the current PICNICC dataset, and also developments which the Collaboration will drive further.

Further research opportunities

- International collaborative to harmonise endpoint definitions and define a core dataset
- Develop prediction model based on this harmonised endpoint definition, potentially using advanced approaches to missing data handling
- Evaluate model performance on new datasets
- Continue to build the PICNICC group and incorporate more information on biomarkers
- Work with children, young people, their families and clinicians to define a 'low enough' level of risk to make decisions about therapeutic management
- Undertake a RCT of risk adapted management of FN

The multiple definitions of “adverse events” from FN episodes led to complications in synthesising data in the systematic reviews. The problem of inconsistency in trial definitions has become prominent recently with the formation of the international Core Outcome Measures in Effectiveness Trials (COMET) initiative. Through the development of agreed core outcome measures, this initiative aims to promote awareness of the problems of inconsistent data collection and enhance the collection of identical core outcomes for specific clinical questions.[251] In achieving this goal, it will allow greater comparability and synthesis of data to maximise the value of both individual studies and meta-analysis and allow these to influence practice more strongly. Such an approach, though not strictly relevant as these are not “effectiveness trials”, is required in this area of research and it would be sensible to build on the PICNICC Collaboration to achieve a committed consensus.

Following on from such a consensus, further analysis of the PICNICC dataset to produce a CDR addressing the prediction of ‘any adverse event’ will be necessary. This may well be subtly different than the CDR predicting microbiologically documented infection, as has been shown in the systematic reviews of previous CDR.

Developments in the handling of missing data using simulation have produced guidelines using imputation techniques to maximise the value of the IPD data collected[252-254]. The application of such methods in the particular situation of large quantities of unreported data has yet to be fully explored, and provides an opportunity for further study exploiting the PICNICC dataset further.

The model produced as part of this thesis and its consequent CDR, and any future predictive models and CDR addressing a consensus definition of “adverse event”, will require evaluation in clinical practice from alternative datasets collected in different geographical locations. The PICNICC Collaboration will undertake this by continuing the collection of existing datasets in which to evaluate the rules, but the collection of new data from other areas will also be necessary. Such investigations will provide both specifically collected information to analyse, and allow the uptake of a risk stratified approach to treating FN and dissemination of the PICNICC CDR in settings where this has not previously been undertaken.

The small amount of data reporting on the value of serum biomarkers, particularly when taken with the systematic reviews on this subject, suggest that more information and different analyses of the data are required. We need to collect greater quantities of

information on the additional benefit of particular biomarkers and good quality data on their comparative efficacy in initial risk stratification. Moving beyond the initial treatment of FN, and focusing on how we should treat patients with either a defined MDI or those without a clear cause, the patterns of how biomarkers change over time which reflect response to treatment will require evaluation, and also how these patterns may vary both between individuals and within individuals after different elements of their cancer treatment. Undertaking such research will move closer to an ideal of 'personalised medicine', and is part of the continued efforts of the PICNICC team to undertake and combine such studies.

It is clear that the description and decision of what constitutes a 'low enough' risk to draw a threshold at remains a matter of debate. The choice of threshold reflects both a need to predict if a patient has a significant infection requiring specific therapy to avert a poor outcome, and an understanding of the likelihood of a fatal or near-fatal outcome given the particular infection and physiological response to it. The setting of this threshold of 'low enough' risk appears to vary between healthcare professionals and families, and between healthcare professionals themselves [38] , and requires further study. Such a project is under development, with a PhD candidate at the University of York preparing a Thesis including these stakeholders and supported by a local Children's Cancer charity. With this information, combined with solid evaluations of an effective CDR, a randomised non-inferiority trial of discharge of children and young people with low-risk FN within 24 hours of presentation to hospital should be achievable to prove or disprove the utility of this approach to management[255].

Innovations and Impact

The work undertaken during the completion of this thesis has extended methods of undertaking diagnostic meta-analysis (multinomial approaches to multi-level diagnostic tests[113]), and promoted alternative graphical methods of presenting diagnostic test information (cross-hairs plots[126]) which have already been incorporated into computer software for undertaking such analyses (the 'mada' package in R[256]). It has also adapted display techniques from other areas, such as the 'heat maps' of social science research, to display the rich information found in IPD datasets in an accessible way. The practical and ethical framework for sharing such information has been delineated and a better understanding of the different barriers involved has been achieved [246].

The systematic reviews [113-116, 257] forming the basis of the IPD have been cited in national [128] and international guidelines [129] on the management of FN and have been key in forming some of their recommendations. These have influenced changes in clinical practice and as such instituted improvements for people with cancer presenting with FN [246].

Finally, the formation of the PICNICC collaborative has brought together an international core of researchers who remain committed to improving the management of FN and advancing our understanding of how treatments for infection in children and young people with cancer should be trialled and implemented.

Conclusions

The aim of this thesis was to describe the clinical problem of the initial management of febrile neutropenia in children and young people undergoing treatment for malignant disease, to thoroughly examine the existing research, and to seek to synthesise this to quantify the risk of adverse clinical outcomes and develop a new risk prediction model. This was undertaken to inform everyday clinical decisions and future research. A further aim was to develop methodological approaches to IPD analysis in the development of predictive models, including the graphical display and communication of such information. This thesis focused on the outcome of microbiologically documented infection (MDI) because it was the most completely reported and most objective, and frequently occurring, significant complication of paediatric FN.

The work undertaken has formed a global collaboration which has shared thousands of items of data, developed a new predictive model for MDI, and from this derived a CDR which is robust to internal validation techniques. We have demonstrated that such a project is feasible across many different jurisdictions and eras of study, and should provide the impetus for a series of projects which will evaluate and improve the management of FN across the world.

Appendices

Appendix 1. CDR Search Strategy

Example based on OVID-Medline: (this was adapted for other databases)

FNP identification

1 Neutropenia/

2 (neutropenia or neutropenic).ti,ab.

3 1 or 2

4 Fever/

5 (fever$ or febril$).ti,ab.

6 4 or 5

7 3 and 6

Child identification

8 adolescent/ or child/ or child, preschool/ or infant/ or infant, newborn/ or Puberty/

9 schools/ or schools, nursery/

10 (infan$ or newborn$ or new born$ or baby$ or babies or neonat$ or neonat$ or child$ or schoolchild$ or kid or kids or toddler$ or adoles$ or teen$ or boy$ or girl$ or minor$ or underage$ or under age$ or juvenil$ or youth$ or kindergar$ or nursery or puber$ or prepuber$ or pre puber$ or pubescen$ or prepubescen$ or pre pubescen$ or pediatric$ or paediatric$ or peadiatric$ or school or schools or preschool$ or pre school$ or schoolage$).ti,ab.

11 8 or 9 or 10

Cancer identification

12 exp Neoplasms/

13 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or sarcoma$ or leukaemi$ or leukemi$ or chemotherap$).ti,ab.

14 12 or 13

Consolidation

15 11 and 14

16 7 and 15

CDR Hedge

17 (predict$ or clinical$ or outcome$ or risk$).mp.

Final search

18 16 and 17

Appendix 2. Biomarkers Search Strategy

Example based on OVID-Medline: was adapted for other databases

FNP identification

1 Neutropenia/

2 (neutropenia or neutropenic).ti,ab.

3 1 or 2

4 Fever/

5 (fever$ or febril$).ti,ab.

6 4 or 5

7 3 and 6

Child identification

8 adolescent/ or child/ or child, preschool/ or infant/ or infant, newborn/ or Puberty/

9 schools/ or schools, nursery/

10 (infan$ or newborn$ or new born$ or baby$ or babies or neonat$ or neonat$ or child$ or schoolchild$ or kid or kids or toddler$ or adoles$ or teen$ or boy$ or girl$ or minor$ or underage$ or under age$ or juvenil$ or youth$ or kindergar$ or nursery or puber$ or prepuber$ or pre puber$ or pubescen$ or prepubescen$ or pre pubescen$ or pediatric$ or paediatric$ or peadiatric$ or school or schools or preschool$ or pre school$ or schoolage$).ti,ab.

11 8 or 9 or 10

Cancer identification

12 exp Neoplasms/

13 (cancer$ or neoplas$ or oncolog$ or malignan$ or tumo?r$ or sarcoma$ or leukaemi$ or leukemi$ or chemotherap$).ti,ab.

14 12 or 13

15 14 and 11 and 7

Markers identification

16. Biological Markers/

17. (marker$ or serum).ti,ab.

18. (biomarker$ or bio-marker$).ti,ab.

19. or/18-20

20. Cytokines/

21. cytokine$.ti,ab.

22. 22 or 23

23. Interleukin-1/

24. (interleukin-1 or interleukin-i or il-1 or il1).ti,ab.

25. t-helper factor.ti,ab.

26. lymphocyte-activating factor.ti,ab.

27. macrophage cell factor.ti,ab.

28. epidermal cell derived thymocyte-activating factor.ti,ab.

29. or/25-30

30. Interleukin-5/

31. (interleukin-5 or il-5 or il5).ti,ab.

32. eosinophil differentiation factor.ti,ab.

33. t-cell replacing factor.ti,ab.

34. (b-cell growth factor-ii or b-cell growth factor-2).ti,ab.

35. (bcgf-ii or bcgfii or bcgf-2 or bcgf2).ti,ab.

36. or/32-37

37. Interleukin-6/

38. (interleukin-6 or il-6 or il6).ti,ab.

39. plasmacytoma growth factor.ti,ab.

40. b-cell differentiation factor.ti,ab.

41. (b-cell stimulat$ factor-2 or b-cell stimulat$ factor-ii).ti,ab.

42. (bsf-2 or bsf2 or bsf-ii or bsfii).ti,ab.

43. hepatocyte-stimulating factor.ti,ab.

44. hybridoma growth factor.ti,ab.

45. (interferon beta 2 or interferon beta2 or ifn-beta 2 or ifn-beta2).ti,ab.

46. mgi-2.ti,ab.

47. myeloid differentiation-inducing protein.ti,ab.

48. or/39-49

49. Interleukin-8/

50. (interleukin-8 or il-8 or il8).ti,ab.

51. monocyte-derived neutrophil chemotactic factor.ti,ab.

52. neutrophil activation factor.ti,ab.

53. lymphocyte-derived neutrophil-activating peptide.ti,ab.

54. monocyte-derived neutrophil-activating peptide.ti,ab.

55. (alveolar macrophage chemotactic factor-i or amcf-i).ti,ab.

56. anionic neutrophil-activating peptide.ti,ab.

57. cxcl8.ti,ab.

58. macrophage-derived chemotactic factor.ti,ab.

59. neutrophil chemotactic factor.ti,ab.

60. or/51-61

61. Interleukin-10/

62. (interleukin-10 or il-10 or il10).ti,ab.

63. csif-10.ti,ab.

64. or/63-65

65. Interferon-gamma/

66. (interferon-gamma or gamma-interferon or IFN-gamma or IFNgamma).ti,ab.

67. (interferon ii or interferon 2).ti,ab.

68. (type ii interferon or interferon type ii).ti,ab.

69. immune interferon.ti,ab.

70. or/67-71

71. Interferon-beta/

72. (interferon-beta or beta-interferon or IFN-beta or IFNbeta).ti,ab.

73. fibroblast interferon.ti,ab.

74. (interferon-beta1 or beta1 interferon or beta-1 interferon or IFN-beta1 or IFNbeta1).ti,ab.

75. Fiblaferon.ti,ab.

76. or/73-77

77. transforming growth factor beta/

78. (beta transforming growth factor or transforming growth factor beta or tgf-beta or tgfbeta).ti,ab.

79. milk growth factor.ti,ab.

80. platelet transforming growth factor.ti,ab.

81. bone-derived transforming growth factor.ti,ab.

82. or/79-83

83. Antigens, CD70/

84. (CD70 or cd27l or cd27 ligand).ti,ab.

85. 85 or 86

86. Tumor Necrosis Factor-alpha/

87. (tumour necrosis factor or tumor necrosis factor).ti,ab.

88. (tnf or tnfalpha).ti,ab.

89. Cachectin.ti,ab.

90. or/88-91

91. Receptors, Tumor Necrosis Factor, Type II/

92. (tnfrii or tnfr-ii or tnfr2 or tnfr-2).ti,ab.

93. (stnf-ii or stnfrii or stnfr2 or stnfr-2).ti,ab.

94. (tnfr p75 or tnfr p80 or tnf-sr75).ti,ab.

95. (cd-120b or cd120b).ti,ab.

96. tnfrsf1b receptor$.ti,ab.

97. or/93-98

98. C-Reactive Protein/

99. (c-reactive protein or Creactive protein or c-reaction protein or Creaction protein).ti,ab.

100. 100 or 101

101. Receptors, Interleukin-2/

102. (interleukin-2 receptor$ or interleukin-ii receptor$).ti,ab.

103. (il-2 receptor$ or il-ii receptor$ or il2 receptor$).ti,ab.

104. (sil-2 or sil-2r or sil2 or sil-ii or sil-iir).ti,ab.

105. (t-cell growth factor receptor$ or tcgf receptor$).ti,ab.

106. or/103-107

107. (procalcitonin or pro-calcitonin).ti,ab.

108. calcitonin precursor.ti,ab.

109. 109 or 110

110. Receptors, IgG/

111. igg receptor$.ti,ab.

112. (gamma fc receptor$ or fc gamma receptor$).ti,ab.

113. immunoglobulin g receptor.ti,ab.

114. (leu-11 or leu11).ti,ab.

115. (cdw32 or cd-32 or cd32 or cd-64 or cd64 or cd-16 or cd16).ti,ab.

116. (fc gamma ri or fc gammari or fc gamma rii or fc gammarii or fc gamma riii or fc gammariii).ti,ab.

117. (sfc gamma riii or sfc gammariii).ti,ab.

118. or/112-119

119. Adenosine Deaminase/

120. adenosine deaminase.ti,ab.

121. (ada-1 or ada1 or ada-2 or ada2).ti,ab.

122. (adenosine aminohydrolase or adenosine amino hydrolase).ti,ab.

123. or/121-124

124. Blood Sedimentation/

125. ((erythrocyte or blood) adj sedimentation).ti,ab.

126. 126 or 127

127. Serum Amyloid A Protein/

128. (serum amyloid A or serum amyloid protein a).ti,ab.

129. serum a related protein.ti,ab.

130. amyloid serum protein saa.ti,ab.

131. amyloid-related serum protein.ti,ab.

132. (amyloid a adj (precursor or protein)).ti,ab.

133. (amyloid protein adj (saa or aa)).ti,ab.

134. amyloid fibril protein aa.ti,ab.

135. or/129-136

136. Chemokine CCL2/

137. (monocyte chemotactic protein-1 or monocyte chemoattractant protein-1 or mcp-1).ti,ab.

138. ccl2.ti,ab.

139. or/138-140

140. Neopterin/

141. (neopterin or neopterine).ti,ab.

142. (umanopterin or monapterin).ti,ab.

143. or/142-144

144. lipopolysaccharide-binding protein.ti,ab.

145. lps binding protein.ti,ab.

146. 146 or 147

147. 21 or 24 or 31 or 38 or 50 or 62 or 66 or 72 or 78 or 84 or 87 or 92 or 99 or 102 or 108 or 111 or 120 or 125 or 128 or 137 or 141 or 145 or 148

148. 15 and 147

Appendix 3. Modified QUADAS Criteria for Quality Assessment

	Item	Yes	No	Unclear
1.	Was the spectrum of patients representative of the patients who will receive the test in practice?	( )	( )	( )
2.	Were selection criteria clearly described?	( )	( )	( )
3.	Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
~~4.~~	~~Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?~~	~~( )~~	~~( )~~	~~( )~~
5.	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	( )	( )	( )
6.	Did patients receive the same reference standard regardless of the index test result?	( )	( )	( )
7.	Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?	( )	( )	( )
8.	Was the execution of the index test described in sufficient detail to permit replication of the test?	( )	( )	( )
9.	Was the execution of the reference standard described in sufficient detail to permit its replication?	( )	( )	( )
10.	Were the index test results interpreted without	( )	( )	( )

	knowledge of the results of the reference standard?			
11.	Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
12.	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	( )	( )	( )
~~13.~~	~~Were uninterpretable/ intermediate test results reported?~~	~~( )~~	~~( )~~	~~( )~~
~~14.~~	~~Were withdrawals from the study explained?~~	~~( )~~	~~( )~~	~~( )~~

QUADAS criteria which are struck through were not used in the assessment

Appendix 4. Key Data Extraction Fields

Quality of the study

- Was the study retrospective or prospective?
- Were selection criteria clearly described?
    - No
    - Yes – Consecutive
    - Yes – Random
    - Yes – Other
- Was the study population appropriate to clinical practice?
- Did the whole sample or a random selection of the sample, receive adequate outcome assessment?
- Did patients receive the same outcome assessment regardless of the serum marker test result?
- Was the outcome assessment independent of the serum marker test (i.e. the serum marker test did not form part of the outcome assessment)?
- Was the execution of the serum marker test described in sufficient detail to permit replication of the test?
- Was the execution of the outcome assessment described in sufficient detail to permit its replication?
- Were the serum marker test results interpreted without knowledge of the results of the outcome assessment?
- Were the outcome assessment results interpreted without knowledge of the results of the serum marker test?
- Were the same clinical data available when the serum marker test was interpreted as would be available when the serum marker test is used in practice?

Comments

Study & Patient Background

- Where did the study take place (country/continent)?
- What years were the patients studied in?
- What are the inclusion criteria?
- What are the exclusion criteria?
- What number of patients were included?
- What number of episodes were included?

- What was the average age (& range or SD) of the patients?
- What number of the patients/episodes were male?
- What number of the patients/episodes had missing values?

Comments

Results

- Serum marker test assessed:
    - Marker #1
    - Marker #2
    - Marker #3

- List all clinical end points (outcomes) examined.
    - Death
    - Intensive care admission
    - Medical complications (eg need for O2, renal failure)
    - Bacteraemia
    - Significant bacterial infection
    - Absence of adverse sequelae
    - Other

Marker #1	Cutpoint used:	Cutpoint used:	Cutpoint used:
No. true positive			
No. false positive			
No. true negative			
No. false negative			

Marker #2	Cutpoint used:	Cutpoint used:	Cutpoint used:
No. true positive			
No. false positive			
No. true negative			
No. false negative			

Marker #3	Cutpoint used:	Cutpoint used:	Cutpoint used:
No. true positive			
No. false positive			
No. true negative			
No. false negative			

- Other indicator #1 (e.g. correlation or HR) with SE or CI
    - Marker #1
    - Marker #2
    - Marker #3
- Other indicator #2 (e.g. correlation or HR)
    - Marker #1
    - Marker #2
    - Marker #3

- How were continuous variables handled in the analyses?
    - As continuous values
    - Made ordinal (e.g. high/mid/low)

- If relevant, describe methods used for cutpoint determination?

- How were categorical variables handled in the analyses?
    - Grouped (e.g. high/mid/low)
    - Other

- If relevant, describe methods used for cutpoint determination?

- How was missing data handled?
    - Not applicable
    - Not specified
    - Completed data only used
    - Imputation method

- How were multiple episodes in individual patients handled?
    - No discrimination
    - First-last comparison
    - Generalised equivalence equation (GEE)
    - Other

Comments

Appendix 5. Example 'Stata' Code for Cross-Hairs Plots

```
/* Using cxr2.dta for CROSS-HAIRS*/

/* uses llsn = lower limit of sn for individual study (and similar) with other 5 params */

/* uses exbisn = lower limit of sn for summary estimate study ... etc ... */

set scheme s2color

twoway (rcap  llsn ulsn sp, msize(large) lwidth(medthin) lcol(eltblue)) (rcap llsp ulsp sn, hor msize(large) lcol(eltblue) lwidth(medthin)) (sc sn sp [w=d+nd] , msize(*.6) msymbol(oh) mcol(eltblue) yscale(range(1 0)) ylabel(1(0.2)0, format(%03.1f))  ytitle("Sensitivity") xscale(rev range(1 0)) xlabel(1(0.2)0 , format(%03.1f)) xtitle("Specificity") aspectratio(1) legend(off)) (rcap  exbisnll exbisnul exbisp, msize(large) lcol(navy)) (rcap exbispll exbispul exbisn, hor msize(large) lcol(navy)) (sc  exbisn  exbisp, msymbol(O) mcol(navy))



/* USING SM-CRP.DTA: cross-hairs with weighted marker sizes, >50mg only */

twoway (rcap snll snul sp if marker==1 & group==2 & val==50, msize(large) lwidth(medthin) lcol(eltblue)) /*

*/ (rcap spll spul sn if marker==1 & group==2 & val==50, hor msize(large) lcol(eltblue) lwidth(medthin)) /*

*/ (sc sn sp [w=d+nd] if marker==1 & group==2 & val==50, msize(*.4) msymbol(oh) mcol(eltblue) yscale(range(1 0)) ylabel(1(0.2)0, format(%03.1f) angle(360)) /*

*/ ytitle("Sensitivity") xscale(rev range(1 0)) xlabel(1(0.2)0 , format(%03.1f) alt) xtitle("Specificity") aspectratio(1) /*

*/ legend(off))
```

```
/* USING SM-CRP.DTA: cross-hairs with weighted marker sizes, any CRP results */

twoway (rcap snll snul sp if marker==1 & group==2 , msize(large) lwidth(medthin)
lcol(eltblue) lp(solid)) /*

*/ (rcap spll spul sn if marker==1 & group==2 , hor msize(large) lcol(eltblue) lwidth(medthin)
lp(solid)) /*

*/ (sc sn sp [w=d+nd] if marker==1 & group==2 , msize(*.4) msymbol(oh) mcol(eltblue)
yscale(range(1 0)) ylabel(1(0.2)0, format(%03.1f) angle(360)) /*

*/ ytitle("Sensitivity") xscale(rev range(1 0)) xlabel(1(0.2)0 , format(%03.1f) alt)
xtitle("Specificity") aspectratio(1) /*

*/ legend(off))



/* USING SM-CRP.DTA: cross-hairs with weighted marker sizes split across two graphs, any
CRP results */

twoway (rcap snll snul sp if marker==1 & group==2 , msize(large) lwidth(medthin)
lcol(eltblue) lp(solid)) /*

*/ (sc sn sp [w=d+nd] if marker==1 & group==2 , msize(*.4) msymbol(oh) mcol(eltblue)
yscale(range(1 0)) ylabel(1(0.2)0, format(%03.1f) angle(360)) /*

*/ ytitle("Sensitivity") xscale(rev range(1 0)) xlabel(1(0.2)0 , format(%03.1f) alt)
xtitle("Specificity") aspectratio(1) /*

*/ saving(CRPsn, replace) legend(off))

twoway (rcap spll spul sn if marker==1 & group==2 , hor msize(large) lcol(eltblue)
lwidth(medthin) lp(solid)) /*

*/ (sc sn sp [w=d+nd] if marker==1 & group==2 , msize(*.4) msymbol(oh) mcol(eltblue)
yscale(range(1 0)) ylabel(1(0.2)0, format(%03.1f) angle(360)) /*

*/ ytitle("Sensitivity") xscale(rev range(1 0)) xlabel(1(0.2)0 , format(%03.1f) alt)
xtitle("Specificity") aspectratio(1) /*
```

```
*/ saving(CRPsp, replace) legend(off))

graph combine CRPsn.gph CRPsp.gph


/* USING CROSS HAIRS WITH BIVARIATE OVERALL SUMMARY */

/* Cheating, sort of, by running this after a 'metandi' command */

/* code for the dotty elipse of the summary confidence interval taken directly from it */


                tempname covmuAB sAB rconfAB sepredA sepredB rpredAB

                matrix V = e(V)

                scalar `covmuAB' = V[1,2]

                scalar `sAB' = _b[sAB]


                /* derived params */

                scalar `rconfAB' = `covmuAB'/(_se[muA]*_se[muB])

                scalar `sepredA' = sqrt(_b[s2A]+_se[muA]^2)

                scalar `sepredB' = sqrt(_b[s2B]+_se[muB]^2)

                scalar `rpredAB' = (`sAB'+`covmuAB')/(`sepredA'*`sepredB')


                        tempname croot phi confB confA confspec confsens

                        scalar `croot' = sqrt(2*invF(2,e(N)-2,95/100))

                range `phi' 0 `=2*c(pi)' 500

                        gen `confB' = _b[muB] + _se[muB] * `croot' * cos(`phi')

                        gen `confA' = _b[muA] + _se[muA] * `croot' * cos(`phi' +
acos(`rconfAB'))
```

```
gen `confsens' = invlogit(`confA')

gen `confspec' = invlogit(`confB')
```

```
twoway (rcap  llsn ulsn sp, msize(large) lwidth(medthin) lcol(eltblue)) (rcap llsp ulsp sn, hor
msize(large) lcol(eltblue) lwidth(medthin)) (sc sn sp, msymbol(oh) mcol(eltblue)
yscale(range(1 0)) ylabel(1(0.2)0, format(%03.1f))  ytitle("Sensitivity") xscale(rev range(1 0))
xlabel(1(0.2)0 , format(%03.1f)) xtitle("Specificity") aspectratio(1) legend(off) title("Summary
ROC plot" "for all studies")) (rcap  exbisnll exbisnul exbisp, msize(large) lcol(navy)) (rcap
exbispll exbispul exbisn, hor msize(large) lcol(navy)) (sc  exbisn  exbisp, msymbol(O)
mcol(navy)) (line `confsens' `confspec',clpatt(dash))
```

Appendix 6. Numerical aspects of derivation studies

Citation	n patients	n episodes	n events of 1^0 outcome	n variables examined	events per var	All candidate variables initially examined	How were candidate variables selected?
Adcock, 1999	33	88	16	14	1.14	Demographics, primary diagnosis, history of present illness, vital signs, and physical examination. Recent chemotherapy regimen, prophylactic (antibiotic) therapy, leukocyte count and ANC, maximum daily temperature, age, and condition of central line	Not stated
Alexander, 2002	104	104	13	2	6.5	Anticipated neutropenia <7 days, no significant comorbidity at presentation (defined later).	Literature review
Ammann, 2003 (models #1 -	111	285	90	39	2.31	39 variables: age, gender, pre-B-cell leukaemia or other diagnosis, first or later malignancy, relapsed or unrelapsed malignancy, history of episodes of FN without significant	Covariates with possible relevance to severe bacterial

#3)						bacterial infection, history of episodes of FN with significant bacterial infection, history of episodes of FN with bacteraemia, remission status of malignancy, bone marrow involvement, maintenance therapy or more intensive chemotherapy, delay since last chemotherapy, time since diagnosis, year of previous episode(s) of FN, season of previous episode(s) of FN, preventive application of G-CSF, central venous catheter present, hospitalisation history before FN, presence of comorbidity requiring hospitalisation, iatrogenic reason for fever, fever rule (≥38.5°C persisting for at least two hours or once ≥39°C), weight loss since last chemotherapy, BMI, maximal fever at presentation, general appearance, presence of chills at presentation, lowest systolic BP, lowest diastolic BP, presence of oral mucositis, presence of clinical signs of viral infection, haemoglobin level, leukocyte count, neutrophil count, monocyte count, phagocyte count, thrombocyte count, serum CRP level, serum creatinine level, and serum ASAT level.	infections and accessible to the treating physicians within the first two hours after fulfilment of the criteria of FN
Ammann, 2004	132	364	85	39	2.18		
Amman, 2010	206	423	62	33	1.88	Age, gender, past-FN, past-bacteraemia, relapse, AML, any haematological, BM involvement, CR, >1y since diagnosis,	Not stated

						interval from chemo <7d, intensiveness of therapy, GCSF, inpatient status, 'unwell', mucositis (oral & any), CVL, URTI, temp >39.5, raised HR, RR, low BP or sats, Hb, WCC, ANC, AMC, APC, plts, CRP … plus at reassessment: low, Bp or sats, chills, T-max, CXR needed, focal infection, other need for IP care	
Ayegman, 2011	206	423	67	33	2.03	Age, gender, past-FN, past-bacteramia, relapse, AML, any haematological, BM involvement, CR, >1y since diagnosis, interval from chemo <7d, intensiveness of therapy, GCSF, inpatient status, 'unwell', mucositis (oral & any), CVL, URTI, temp >39.5, raised HR, RR, low BP or sats, Hb, WCC, ANC, AMC, APC, plts, CRP … plus at reassessment: low, Bp or sats, chills, T-max, CXR needed, focal infection, other need for IP care	Not stated
Badeiei, 2011	68	120	35	18	1.94	At least Sex, Age, tumour type (solid vs non-solid), relapse, time from chemo (grouped), temp (grouped), duration of fever before admission, URTI symptoms, mucositis, WBC (grouped), ANC(grouped), Hb(grouped), Plt(grouped), CXR finding	Not stated

Delebarre, 2010	146	316	70	N/A	N/A	Not stated	Not stated
Hann, 1997	759	759	165	13	12.69	Gender, underlying disease (AML, ALL, BMT, HD/NHL, CML-aplasia-blast-crisis-other, Solid tumour), disease status (induction, relapse, maintenance), IV line in situ, defined site of infection, shock, granulocyte count, period of granulocytopenia, antifungal prophylaxis, antibacterial prophylaxis, age, temperature, (log) creatinine	Not stated
Hakim, 2010	332	332	41	22	1.86	Age, gender, race, cancer diagnosis, Prior relapse, time-since-relapse, CVL, steroid use, GCSF use, recent antifungal therapy, colonisation (VRE, MRSA, pseudo), tmax>39, clinical appearance, comorbidity at presentation, URI symptoms, Hb, ANC, Plts, anticipated neutropenia <7d, antifungal prophylaxis, time-since-chemo, duration of preceding neutropenia	Literature review and medical expertise
Jones, 1996	127	276	68	5	13.6	Underlying disease and status (i.e. induction therapy, remission or relapse). Age at time of fever episode. ANC at time of onset of fever. Inpatient versus outpatient status	Not stated

Klaassen, 2000	140	227	28	13	2.15	13 variables assessed: age, presence of bone marrow disease, central venous catheter type, general appearance on initial examination, previous granulocyte colony-stimulating factor (G-CSF) therapy, initial ANC, initial lymphocyte count, initial monocyte count, initial platelet count, presence of localized bacterial infection on initial examination, peak temperature, tumour type, and sex.	Systematic review to identify risk factors for significant bacterial infection and expert opinion.
Lucas, 1996	161	509	82	8	10.25	Chills, hypotension, poor perfusion, the need for fluid resuscitation, time from cytotoxic chemotherapy, diagnosis, disease status, and the presence of a focus of infection	Not stated
Mian, 2010	29	51	8	N/A	N/A	Not stated	Not stated
Paganini, 2007	458	714	18	17	1.06	Age, days since chemotherapy, 'advanced stage of disease' (= bone marrow involvement, relapse, second tumour, high-dose therapy, genetic disease), previous antibiotic or CSF use, ANC <100, clinical infection, pneumonia, mucositis, bacteraemia <24h, comorbidity (=incoercible bleeding, refractory hypoglycaemia and hypocalcaemia, hypotension, altered mental status, renal insufficiency, hepatic dysfunction, and respiratory failure). They also state that the	Unclear

						following variables were collected and registered for analysis: facial, anal, oral or catheter-associated cellulitis, sepsis, necrotising gingivitis, sex, underlying disease and staging, predicted period of neutropenia, presence of intravenous device.	
Rackoff, 1996	72	115	24	9	2.67	State of disease (remission vs. not), degree of mucositis, ill appearance, presence of GI symptoms, cellulitis, use of GCSF, admission ANC, admission AMC, maximum admission temperature	Unclear
Rackoff, 1996 revised model	102 (see note)	57	10	7	1.43	AMC, Temperature (39.5C cutoff), ANC, APC, Platelets, age, WBC	By reference to previously published literature
Riikonen 1993	46	91	17	16	0.94	Duration of fever, duration of neutropenia, central line present, prophylaxis with Septrin, general clinical examination, HR, signs of bleeding, BP, temperature, chills, Hb, Plt, prolonged PTT, sodium & potassium ESR, CRP	Unclear
Rojo, 2008	33	47	4	6	0.67	Sex, age, type of malignancy (leukaemia vs. solid), focus of	Unclear

						infection, duration of hospitalisation, microbiologically proven infection	
Rondinelli, 2006	283	283	93	17	5.47	Significantly on univariate: Age, gender, disease type (AML, ALL, Others), disease status (remission/other), CVC, temperature, Hb, WCC, AGC, Plt, AMC, URTI, time from chemotherapy, pneumonia, clinical site of infection, mucositis plus others not reported	Unclear
Santolaya, 2001	257	447	179	17	10.5	(1) demographic variables, i.e., age, sex, and maternal educational level; (2) cancer-related variables, i.e., cancer type, intensity of chemotherapy, use of granulocyte colony-stimulating factors since last administration of chemotherapy, and use of an indwelling catheter; (3) variables related to the febrile episode, i.e., hours of fever before admission, days since last administration of chemotherapy, and use of prophylactic antimicrobial agents; (4) admission clinical and laboratory variables, i.e., axillary temperature, blood pressure, ANC, AMC, quantitative serum CRP level, haemoglobin level, and platelet count	Not stated

Tezcan, 2006	240	621	143	11	13	Age, sex, ANC, AMC, CRP, duration of neutropenia, duration of fever, presence of previous FN, presence of hypotension, uncontrolled malignancy, cancer type.	Unclear
West, 2004	143	303	36	18	2	Age, type of cancer, chills, temperature, HR, RR, SBP, DBP, mucositis, Hb, Plts, WCC, differential WCC, ANC, AMC, monocytes <10%, perirectal abscess, capillary refill time >3s.	Review of literature + medical opinion

Note: 102 minus participants excluded for meeting exclusion criteria

Appendix 7. Variable and missing data handling techniques in derivation studies

Citation	Statistical technique to build the model	Management of multiple episodes.	Management of continuous variables…	… and cutpoint determination	Management of categorical variables.	and cutpoint determinatio n	n patients or episodes with missing values	Management of missing data.
Adcock, 1999	Univariate analysis	No discrimination	Made ordinal (BP described as 'hypotension' or not)	Not stated	Grouped (Ara-C vs. Other chemotherapy)	Based on "trend to significance" from univariate analysis	Not stated	Not stated
Alexander , 2002	Univariate analysis	First episode only used	Made ordinal (hypotension and hypoxia)	Not stated	Grouped ('anticipated neutropenia' group by cancer type – AML/Burkitts/I nduction	Not stated	Two patients excluded due to missing data	Completed data only used

					ALL/Progressive-relapsed with marrow involvement vs not)			
Ammann, 2003	Decision tree, regression type	First-last comparison	Made ordinal, with up to three categories	Not stated	Grouped	Not stated	One patient (two episodes) excluded due to missing data exclusion of 41 episodes where >10% of covariates were missing	Completed data only used. Covariates with more than 10% missing values were discarded from the prediction model
Ammann,	Stepwise	First-last	Made ordinal, with up to three	Not stated	Grouped	Not stated	One patient (two	Completed data only

2004	backward	comparison	categories				episodes) excluded due to missing data. exclusion of 16 episodes where >10% covariates were missing	used. Covariates with more than 10% missing values were discarded from the prediction model
Amman, 2010 (and Ayegman, 2011)	Forwards stepwise logistic regression, with random effect per patient, and with 100-fold cross-validation to protect against overfitting	Generalised equivalence equation (GEE)	Made ordinal (e.g. high/mid/low)	Split using recursive partitioning before introduction as dichotomous variables.	Grouped (e.g. high/mid/low)	Not stated	Unclear	Imputation method

Badeiei, 2011	C4.5 decision tree system in SPSS	No discrimination	As continuous values or as categorical variables	Not stated	Grouped (e.g. high/mid/low)	Not stated	Not stated	Not stated
Delebarre, 2010	Univariate and multivariable analysis – no further detail	Not specified	Some continuous, some as ordinals	Not stated	Not stated	Not stated	Not stated (multivariate analysis was based on 678 children. For one of the included trials data were available for 145/220 children)	Not stated
Hann, 1997	Stepwise backward	First episode only used	As continuous values, or if skewed distribution,	According to clinical	Grouped	According to clinical	Not stated (multivariate analysis	Not stated

			categorised according to clinical judgement	judgement		judgement	was based on 678 children. For one of the included trials data were available for 145/220 children)	
Hakim, 2010	Stepwise multiple logistic regression with bootstrapping	Only one episode per patient selected	As continuous values then Made ordinal (e.g. high/mid/low)	ROC-based efficiency methods: Minimising sn/sp differences, more efficient ROC placement and maximising distance from chance. All	Grouped (e.g. high/mid/low)	According to clinical judgement	43% of episodes had missing data for absolute monocyte and lymphocyte counts	Completed data only used

				methods agreed				
Jones, 1996	Logistic regression - unclear	No discrimination	Made ordinal (e.g. age <2, 2-5, 6-12, 13+, ANC <200, ≥200).	Not stated	Grouped (e.g. solid tumour, leukaemia, other).	Not stated.	Not stated	Not stated
Klaassen, 2000	Forward logistic regression	Generalized linear mixed model	All continuous variables except age were dichotomised.	Using predefined thresholds taken from the literature, or recursive partitioning for monocyte count and peak temperature	Dichotomised (tumour type – AML/NHL versus others).	According to clinical judgement	Derivative set – 1.	When monocyte count was not available the patient was excluded (n=1), for other variables it was unclear (all 13 variables were prospectively collected in 98% of the

								episodes).
Lucas, 1996	Logistic regression - unclear	GEE	Made ordinal (time from chemo: ><10d, ANC ><100/mm)	Not stated.	Grouped	Not stated	Unclear	Not specified
Mian, 2010	Stepwise multivariable logistic regression, cut-off p<0.05	Not specified	As continuous values or as categorical variables	Based on published 'normal range'cut-offs (HR & Sys BP)	Not stated	N/A	Not stated	Not stated
Paganini, 2007	Forward logistic regression	No discrimination	Made ordinal (e.g. ANC <100) or used as continuous (e.g. age & days since chemo)	Not stated.	Not used		Not stated	Not stated
Rackoff, 1996	Backwards logistic regression	GEE	Initially continuous,	then recursive partitioning analysis	Grouped – disease state into remission vs. relapse/progres	Not stated	D: One patient had missing diff WBC for one	Completed data only

					sive		episode – excluded leaving 115 episodes	
Rackoff, 1996 revised model	Logistic regression - unclear	No discrimination	Made ordinal (dichotomous)	The AMC, APC and ANC cut-off values of 250/mm3 and temperature of 39.5C were selected arbitrarily. AMC cut off of 100/mm3 was used due to previous study findings (Rackoff, 1996). Serial NLR determined at intervals of 5 units/mm3	Not applicable		Validation set – 82 (60%) episodes	Not stated

				across the range of AMC, ANC and APC values. Platelet count values >25,000/mm3 were tested at intervals of 25,000/mm3				
Riikonen, 1993	Univariable analysis	No discrimination	Made ordinal: Hb <100g/l, Plts <10, 10-30, 30-100), PT 'prolonged', Na & K: less than normal limits. Kept continuous: ESR & CRP	Not stated	Not applicable		Unclear	Not stated
Rojo, 2008	Univariable analysis	No discrimination	Not applicable		Grouping (solid vs. haematological	Not stated	Not stated	Not stated

					malignancy)			
Rondinelli, 2006	Forward logistic regression	First episode only used	Made ordinal	Not stated, but use previously defined cut-offs	Grouped	Not stated	Not stated	Completed data only
Santolaya, 2001	Forward logistic regression	Secondary analysis undertaken with first episode only	As continuous values initially	Then cutpoints determined with ROC (for CRP and platelets)	Grouped (ALL, AML, lymphoma, sarcoma, relapsed leukaemia, other solid)	Not stated	Not stated	Not stated
Tezcan, 2006	Logistic regression - unclear	No discrimination	Some kept continuous (e.g. age, CRP, duration of fever, duration of neutropenia), some made ordinal (e.g. ANC <100, AMC <100).	Not stated	Grouped (e.g. cancer type: leukaemia and lymphoma vs. solid tumours	Not stated	Not stated – different outcome categories have different total values	Not stated

West, 2004	Stepwise backward & bootstrapping	Multivariate analysis "adjusted for clustering at patient level"	Some continuous (temp, age, heart rate z-score) some categorised (BP dichotomised to -2SD, monocytes <10% and ANC=0)	Not clearly described	Categorised: type of cancer – leukaemia/lymphoma, sarcoma/neuroblastoma, other	Not stated	Not stated	Not stated

Appendix 8. Performance of CDR

Citation	Clinical prediction rule	Number of episodes	Outcome	Number with Outcome	Predictive accuracy		
					% Low	LR Low	LR High
Models with one supporting data set							
Adcock, 1999	High risk = hypotension/septic shock, inflamed central line site, recent high dose Ara-C	88	Gram positive bacteraemia	26	Refers to G+ve	Data not given	

Ammann, 2003	Final decision tree model: four covariates were used to classify low risk; bone marrow involvement, leukocyte count >0.5 x 10^9/L, with clinical signs of a viral infection, and aged up to six years at presentation. For those with a leukocyte count ≤0.5 x 10^9/L, they were further classified according to CRP level (≤ or >50mg/L).	111	Severe bacterial infection, (death from bacterial infection, a positive culture of normally sterile body fluids, radiologically proven pneumonia, clinically unequivocal diagnosis of a bacterial infection, or CRP>150 mg/L)	90	10%	0	1.18

(model #1: bootstrapped)							
(model #3)	Low risk ≤4 factors. Risk factors = bone marrow involvement, absence of clinical signs of viral infection, high serum CRP level, low leukocyte count, presence of a central venous catheter, high haemoglobin level, and Pre-B-cell leukaemia.	-111	As above	-90	20%	0.07	1.39
Agyman 2011 (Same population as Amman 2010)	Applied after 24 hours: shaking chills ever observed, haemoglobin >90 g/L, platelet <50 G/L, any other need for IP treatment. No risk factors = low risk	423	Late bacteraemia	67	36%	0.17	1.58
Badiei [11], Threshold value:	Platelets <20 g/dL, temperature ≥39ºC, ANC <100/mm3, mucositis, abnormal CXR on presentation. Risk of infection greater with more risk factors: for 0 factors:	120	Life threatening infection	35	29.20%	0.07	1.62
	For 1 risk factor				64%	0.36	3.37
	For 2 risk factors				75%	0.41	6.69

	For 3 risk factors				85%	0.63	8.51
	For 4 risk factors				98%	0.94	Infinite
Gala-Peralta, 2005	Low risk ≤2 of: <1yr, poor bone marrow response (plt <75, ANC <100),uncontrolled solid tumour or relapsed leukaemia, chemotherapy <10d earlier, rapid neutropenia, cardiac & renal dysfunction	60	Positive blood culture	16	27%	0.18	1.44

Hakim [8]	Score from cancer diagnosis: AML = 20, ALL/lymphoma = 7, Solids = 0 Clinical presentation of serious unwell or toxic = 14, fever at presentation: ≥39ºC = 11, ANC <100/mm3 = 10 points, Total score <24 = low risk of serious infection or sepsis	332	Serious infection or sepsis	47	69%	0.33	3.16

	Score from cancer diagnosis: AML = 11, others = 0. Relapsed disease = 11. Non-white patient = 8, Clinical presentation of serious unwell or toxic = 20. Total score <20 = low risk of any medical complication	332	Medical complications	40	63.00%	0.32	2.54
Jones 1996	Low risk = ANC ≥200, outpatient at onset, in remission	127	Bacteraemia	68	17%	0.71	1.07
Lucas, 1996	Low risk = no chills, hypotension, or a requirement for fluid resuscitation at admission	509	Positive blood culture	82	87%	0.72	4.05
Petrelli, 1991	Low Risk: patients with solid tumors and lymphomas stage I-II.High Risk: patients with leukemias and lymphomas stage III-IV,	146	Positive blood culture	35	45%	0.58	1.42

Rondinelli, 2006	Low risk = 2.5 to 5 points: Intermediate risk = 5.5 to 9 points: High risk = Greater than 9 points. 4.5 points for: clinical site of infection; 2.5 points for: **no** URTI; 2 points for: CVC; 1 point for: aged ≤5y, fever >38.5°C, Hb ≤7g/dL	283	'Serious infectious complication' – sepsis, shock, +ve blood cultures, infection-related death	93	Odds ratio only:	Low 1.0 Mid 13 High 50	
West, 2004 (bootstrapped)	High risk = temp >39.5C and CRT >3s; Mid risk = temp >39.5C or CRT >3s; Low risk = neither	143	Requirement for critical care within 24 hours of presentation (fluid boluses ≥60ml/kg, inotropes or ventilation)	36	Low 89% Mid 10%	0.73 2.7	Infinite

ALL = acute lymphoblastic leukaemia. AML = acute myeloid leukaemia. AMC = absolute monocyte count. CoNS = Coagulase-negative *Staphylococcus* CRP = C-reactive protein. CRT = capillary refill time. CXR = chest X-ray. Hb = haemoglobin. Plt = platelets.

Performance of CDR with more than one supportive data set

	Clinical prediction rule	Number of episodes	Outcome	Number with Outcome	% Low	LR Low	LR High
Santolaya, 2001	Low risk = 0 factors or isolated low plts or <7 days from chemotherapy. High risk = >1 risk factor, or isolated high CRP, hypotension or relapsed leukaemia. Risk factors: CRP ≥90, hypotension, relapsed leukaemia, plts ≤50, chemotherapy within seven days	407	Invasive bacterial infection = positive blood culture (2 for Coagulase-negative *Staphylococcus* spp), positive bacterial culture from usually sterile site, or sepsis syndrome and/or focal organ involvement and haemodynamic instability and severe malaise	178	42%	0.22	2.41
Santolaya,	As above	263	As above	140	40%	0.11	3.91

2002							
Amman, 2010	As above	423	As above	122	15%	0.35	1.15
Macher, 2010	As above	249	As above		46%	0.8	1.1

	Clinical prediction rule	Number of episodes	Outcome	Number with Outcome	% Low	LR Low	LR High
Rackoff, 1996 (proposed from validation set)	Low risk = AMC >100; High-risk = AMC<100	57	Bacteraemia	10	23%	0	1.45

Baorto, 2001	As above	1171	Bacteraemia	189	21%	0.45	1.45
Madsen [16]	As above	157	Microbiologically documented infection	12	39%	0.21	1.55
Klassen	As above	227	Significant bacterial infection	43	37%	0.30	1.48
Amman [13]	As above	423	Serious adverse medical outcome	67	38%	0.26	1.72
Macher [14]	As above	377	Significant bacterial infection	70	40%	0.46	1.44
Tezcan, 2006	As above	621	Microbiological documented infection	225	60%	0.74	1.59

	Clinical prediction rule	Number in study	Outcome	Number with Outcome	% Low % Mid	LR Low	LR Mid	LR High
Paganini,	Low risk <4.	714	Death	18	82%	0	2.38	12

2007	Mid-risk = 4. High risk = >4. Advanced stage of disease = 3 points, Comorbidity = 2 points, Bacteraemia = 1 point				10%			
(validation set)	As above	806	Death	19	82% 12%	0.12	2.76	9.86

	Clinical prediction rule	Number in study	Outcome	Number with Outcome	% Low % Mid	LR Low	LR Mid	LR High

Rackoff, 1996	Low risk = AMC >100; Mid-risk= AMC <100, and temp <39; High-risk = AMC<100, but temp ≥39	115	Bacteraemia	24	17%	0	0.87	3.44
(derivation set)					65%			
(validation set)	As above	57	Bacteraemia	10	23%	0	0.21	3.52
					40%			
Klaassen, 2000	As above	227	Bacteraemia	28	37%	0.35	0.75	2.57
					37%			
			Significant	43		0.39	0.94	2.29

			bacterial infection					
(validation set)	As above	136	Bacteraemia	19	42% 24%	0.21	3.93	2.8
			Significant bacterial infection	27		0.59	1.59	1.22
Madsen, 2002	As above	157	Positive blood culture	12	25% 64%	0.31	0.91	3.72

	Clinical prediction rule	Number of episodes	Outcome	Number with Outcome	% Low	LR Low	LR High
Amman, 2004	Low risk ≤3 factors. Risk factors = bone marrow involvement, absence of clinical signs of viral infection, high serum CRP level, low leukocyte count, presence of a central venous catheter, high haemoglobin level, and Pre-B-cell leukaemia.	364	Severe bacterial infection (death from bacterial infection, a positive culture of normally sterile body fluids, radiologically proven pneumonia, clinically unequivocal diagnosis of a bacterial infection, or CRP>150 mg/L)	132	14%	0.00	1.29
Amman,	As above	423		122	10%	0.27	1.10

2010							
Macher, 2010	As above	377		122	8%	2.29	0.97

	Clinical prediction rule	Number of episodes	Outcome	Number with Outcome	% Low	LR Low	LR High
Alexander, 2002	Low risk = Not AML/Burkitts/Induction ALL/Progressive-relapsed with marrow involvement and no significant comorbidity	104	Bacteraemia	13	58%	0.24	2.39

Amman, 2010		304	Bacteraemia	122	8%	0.66	1.03
Domment, 2009		762	Bacteraemia	122	53%	0.72	1.38

	Clinical prediction rule	Number of episodes	Outcome	Number with Outcome	% Low	LR Low	LR High
Amman, 2010	Applied after 24 hours: 4 points for chemotherapy more intensive than ALL	423	Serious adverse medical outcome	122	35%	0.18	1.67

	maintenance,5 points for hemoglobin >90 g/L, 3 points each for white blood cell count <0.3 G/L, platelet <50 G/L, any adverse event occurred in preceding 24h. Scores ≤9 are low risk.						
Miedema, 2011		210		57	50%	0.31	1.55

Appendix 9. Individual factors used in clinical prediction rules

Overview

Variable		Citations								
Pt related										
Age										
	<6yr	Amman 2003	(Madsen)	(Tezcan)						
	<5yr	Rondinelli								
	<4, 4-8, >8yr	(Amman 2010)								
Disease										
	AML/Burkitts/Induction ALL/Relapse-progressive/BM involvement	Alexander								

	AML or ALL/Lymph or Other	Hakim	Amman 2010							
	Leuk/Lymph or BMT vs Other	Hann	(Lucas)	(Badei)	Delabarre					
	BM involved	Amman 2003	Amman 2004	(Klassen)	(Amman 2010)					
	Pre-B Leukemia	Amman 2003								
	In PR/CR	Amman 2004	(Rondinelli)	(Tezcan)	(Amman 2010)					
	Induction/relapse vs. remission	Jones	(Amman 2010)							
	Advanced disease	Paganini								
	Relapsed leuk	Santolaya	Hakim							
Rx related										

Type of Rx										
	Ara-C <7d	Adcock								
	Low intensity (anticptaed neutropenia <7d)	Alexander	Amman 2010	(Hakim)	Delabarre					
	Chemo <7d	Santolaya	(Amman 2010)	(Badei)						
Others										
	CVC present	Amman 2003	Hann	Rondinelli	(Amman 2010)					
Episode related										
Airway/Breathing										
	Tachypnoea/hypoxia	Alexander	(Amman 2010)							
Circulation										
	Tachycardia	(Mian)	(Amman 2010)							

	Hypotension	Adcock	Santolaya	(Tezcan)	(Mian)	(Amman 2010)				
	Chills/hypotension/fluid bolus composite	Lucas								
	Shock/Severe sepsis	Hann	Delabarre							
	CRT >3	West								
Neurology										
	Altered mental status	Alexander								
Source										
	Inflammed CVC site	Adcock								
	CXR +ve / pneumonia	Alexander	Badei	(Rondinelli)						
	Mucositis	Alexander	(Rondinelli)	(Amman 2010)	(Badei)					

	Vomiting/diarrhoea	Alexander								
	Viral infection	Amman 2003	Rondinelli (no URTI)							
	Clincally unwell	(Amman 2010)	Hakim							
	Known bactermia	Paganini	Amman 2010	Mian						
	Clincal site	Rondinelli	Delabarre	(Madsen)						
Temperature										
	Cont variable	Hann	Rackoff	Madsen	West					
	<38.5	Rondinelli	(Badei)							
	<39.7 C	Amman 2004								
	<39 (also 39.5 used)	Klassen	Hakim	Delabarre	Badei	(Mian)	(Amman 2010)	(Santollaya)		

Others										
	Other comorbidity	Alexander	Amman 2004	Paganini	(Amman 2010)					
	Abdominal pain	Alexander								
	OP at start	Jones								
Blood work										
FBC										
	WCC >0.5	Amman 2003	Amman 2004	Amman 2010	Delabarre	(Madsen)	(Rondinelli)			
	Hb <7	Rondinelli	Amman 2003	(Badei)						
	Han >10	Amman 2010								
	Granulocytopenia >15d	Hann								
	ANC	Jones	Hakim	(Rackoff)	(Lucas)	(Madsen)	(Tezcan)	(West)	(Amman	

									2010)	
	AMC	Klassen	Rackoff	(Paganini	Madsen	Delabarre	(Santolaya)	(Baorto)	(Tezcan)	(Amman 2010)
	Plts	Santolaya	Amman 2010	Badei	(Madsen)	(Rondinelli)	(Hakim)			
Biochemistry										
	PCT >0.3	Delabarre								
	CRP (continuous)	Mian								
	CRP >50	Amman 2003								
	CRP >90	Santolaya								
	CRP >150	Amman 2010								

Study citations in (brackets) refer to those assessed but not included in models.

Appendix 10. Factors predictive of adverse outcome, by study

Table a – Assessments of individual factors predicting bacteraemia

Variable		Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI
Pt related													
Age													
	<6yr	(Madsen	0.447 (cont	n/s)									
Disease													
	AML/Burkitts/Induction ALL/Relapse-progressive/BM involvement	Alexander											

	Leuk/Lymph or BMT vs Other	Hann	n/a	p<0.0001 (0.23 to 0.77)	(Lucas	0.041	n/a)						
	BM involved	Amman 2004	1.3 to 1.46	n/a	Agyeman	n/a	1.2 to 8.6	(Klassen	0.002,	0.008 on fixed and 0.014 on random)			
	AML	Agyeman	n/a	1.3 to 5.1	Hakim	2.77 to 17.24	2.76 to 20.14						
	ALL/Lymphoma vs. solid	Hakim	0.83 to 4.51	0.84 to 5.02									
	In PR/CR	Amman 2004	n/a	n/a									
	Induction/relapse vs. remission	Jones	1.23 to 10.95	n/a	(Hakim	0.9 to 3.74	n/a)						
Rx related													
Type													

of Rx													
	Low intensity (anticptaed neutropenia <7d)	Alexander			(Hakim	0.23 to 0.83	n/a)						
Others													
	CVC present	Hann	n/a	0.03 (1.06 to 3.31)									
Episode related													
Airway/Breathing													
	Tachypnoea/hypoxia	Alexander											
Circulation													

	Chills/hypotension/fluid bolus composite	Lucas	<0.0001	n/a									
	Shock	Hann	n/a	0.003 (1.80 to 17.3)									
Neurology													
	Altered mental status	Alexander											
Source													
	Inflammed CVC site	Adcock	p>0.05	n/a									
	CXR +ve / pneumonia	Alexander											
	Mucositis	Alexander											

	Vomiting/diarrhoea	Alexander											
	Clincal site	(Madsen	0.044 (cont)	n/s)									
Temperature													
	Cont variable	Hann	n/a	<0.0001	Rackoff	0.002	n/a	Madsen	0.012	0.031			
	<38.5												
	<39.7 C	Ammann 2004	1.5 to 7.1	n/a									
	<39 (also 39.5 used)	Klassen	0.023	0.033 (Fixed) 0.049 (Random)	Agyeman	n/a	1.2 to 7.2	Hakim	1.0 to 2.4	1.3 to 6.5			
Others													

	Other comorbidity	Alexander			Amman 2004	1.3 to 4.2	n/a						
	Abdominal pain	Alexander											
	OP at start	Jones	1.36 to 5.71	n/a									
Blood work													
FBC													
	WCC >0.5	Amman 2004	1.4 to 4.1	n/a	(Madsen	0.321 (cont)	n/s)						
	WCC <0.3	Agyeman	n/a	2.3 to 7.0									
	Hb <7												
	Granulocytopenia >15d	Hann	n/a	0.01 (1.5 to									

				37.9)									
	ANC >200	Jones	0.63 to 2.43	n/a	(Rackoff	>0.5	n/a, as cont var)	(Lucas	>0.05	n/a, for >100)	(Madsen	0.227 (cont)	n/s)
	ANC <100	Agyeman	n/a	2.9 to 17.4	Hakim	1.4 to 5.7	1.2 to 5.7						
	AMC (<100, 10 or 115)	Klassen	0.002	0.031 (Fixed) 0.046 (Random)	Rackoff	0.08	(by partitoning)	(Bartoo	0.28 to 0.72	n/a)	Madsen	0.004 (cont)	0.016 (cont)
	Plts <50	Agyeman	n/a	1.1 to 4.3	Hakim	1.25 to 4.34	n/a	(Madsen	0.036 (cont)	n/s)			

Table b – Assessments of individual factors predicting significant/documented infection and/or complications

Variable		Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI
Pt related													
Age													
	<6yr	Amman 2003	0.67 to 3.41 (to 12) or 0.60 to 3.05 (older than 12y)		(Tezcan	0.01	n/s)						
	<5yr	Rondinelli	1.37 to 3.37, <0.001	1.0 to 3.4, 0.049									
	<4, 4-8, >8yr	(Amman 2010)	0.9 to 2.8 (middle group)	n/a									
Disease													

	AML	Amman 2010	1.7 to 6.1	1.5 to 6.3									
	Leuk/Lymph or BMT vs Other	Delabarre	n/a	n/a									
	BM involved	Amman 2003	1.47 to 8.87	n/a	(Amman 2010	0.7 to 6.2	n/s)						
	Pre-B Leukemia	Amman 2003	0.38 to 1.54	n/a									
	In PR/CR	(Rondinelli	1.3 to 4.0, 0.001	n/s)	(Tezcan docum inf	n/s	n/s)	(Amman 2010	0.9 to 2.5	n/s)			
	Relapsed leuk	Santolaya	n/a	1.7 to 2.3									
Rx related													
Type of Rx													
	Low intensity (anticptaed neutropenia <7d)	Amman 2010	1.1 to 7.1	n/s									
	Chemo <7d	Santolaya	n/a	1.1 to 1.6	Delabarre	n/a	n/a	(Amman 2010	0.4 to 1.6	n/s)			

Others													
	CVC present	Amman 2003	0.79 to 2.98		Rondinelli	2.0 to 6.3, 0.001	1.5 to 5.5, 0.001	(Amman 2010	0.3 to 1.1	n/s)			
Episode related													
Circulation													
	Hypotension	Santolaya	n/a	2.3 to 3.2	(Tezcan	n/s	n/s)						
	Severe sepsis	Delabarre	n/a	n/a									
Source													
	Inflammed CVC site												
	CXR +ve / pneumonia	Badei	n/a	n/a	(Rondinelli	2.3 to 9.9, 0.001	n/s)						
	Mucositis	(Rondinelli	1.8 to 5.4, 0.001	n/s)	(Amman 2010	0.4 to 1.0	n/s)						
	Viral infection	Amman 2003	0.96 to 4.57	n/a	Rondinelli (no URTI)	2.2 to 5.5, p=0.045	1.7 to 15.1, p=0.001	Badei (no URTI)	n/a	n/a			

	Clincal site	Rondinelli	1.2 to 21, 0.001	7.0 to 39.5, 0.001	Delabarre	n/a	n/a						
Temperature													
	Cont variable												
	<38.5	Rondinelli	1.0 to 3.6, p 0.033	1.0 to 3.6, p 0.033									
	<39 (also 39.5 used)	Badei	n/a	n/a	Delabarre	n/a	n/a	(Santollaya	0.19	n/a)			
Blood work													
FBC													
	WCC >0.5	Amman 2003	1.20 to 7.14 (cf >1)		(Rondinelli	0.97 to 2.7, 0.053	n/s)						
	WCC <0.3	Amman 2010	0.4 to 2.1	2.5 to 6.8									
	Hb <7	Rondinelli	1.1 to 3.5, 0.021	1.2 to 3.6, 0.021	Amman 2003	0.26 to 2.04	n/a						
	Hb >9	Amman	1.1 to 1.8	1.5 to 4.3									

		2010											
	ANC <500	Delabarre	n/a	n/a									
	ANC >200	(Tezcan	n/s	n/s)									
		<100											
	AMC (<100, 10 or 115)	(Santolaya	0.04	0.236)	(Tezcan	0.01	n/s)						
	Plts <50	Santolaya	n/a	1.4 to 2.2	Badei	n/a	n/a	Delabarre	n/a	n/a	(Amman 2010	1.2 to 3.2	n/s)
Biochemistry													
	CRP >50	Amman 2003	1.21 to 62.9	n/a									
	CRP >90	Santolaya	n/a	3.6 to 4.8									
	PCT >0.3	Delabarre	n/a	n/a									

Table c – Assessments of individual factors predicting outcomes including ICU & death

Variable		Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI	Citation	Univariate p-value or OR 95% CI	Adjusted p-value or 95% CI
Pt related							
Age							
	<5yr	Rondinelli	1.37 to 3.37, <0.001	1.0 to 3.4, 0.049			
Disease							
	In PR/CR	(Rondinelli	1.3 to 4.0, 0.001	n/s)			
	Advanced disease	Paganini	0.01 (1.02 to 1.06)	0.001			
Rx related							

	CVC present	Rondinelli	2.0 to 6.3, 0.001	1.5 to 5.5, 0.001			
Episode related							
Circulation							
	Tachycardia	(Mian	p=0.27	n/s)			
	Hypotension	(Mian	p=0.34	n/s)			
	CRT >3	West	0.01	1.88 to 16.84, 0.002			
Source							
	CXR +ve / pneumonia	(Rondinelli	2.3 to 9.9, 0.001	n/s)			
	Mucositis	(Rondinelli	1.8 to 5.4,	n/s)			

			0.001				
	Viral infection	Rondinelli (no URTI)	2.2 to 5.5, p=0.045	1.7 to 15.1, p=0.001			
	Known bacteraemia	Paganini	<0.0001 (5.2 to 39)	0.001	Mian	0.031	1.8 to 240, p=0.015
	Clinical site	Rondinelli	1.2 to 21, 0.001	7.0 to 39.5, 0.001			
Temperature							
	Cont variable	West	0.002 (cont)	1.25 to 2.43. p0.001			
	<38.5	Rondinelli	1.0 to 3.6, p 0.033	1.0 to 3.6, p			

			0.033				
	>39	(Mian	p=0.21	n/s)			
Others							
	Other comorbidity	Paganini	0.001 (10.5 to 102)	0.0001			
Blood work							
FBC							
	WCC >0.5	(Rondinelli	0.97 to 2.7, 0.053	n/s)			
	Hb <7	Rondinelli	1.1 to 3.5, 0.021	1.2 to 3.6, 0.021			
	ANC >200	(West	0.69	n/s)			
	AMC (<100, 10 or 115)	(Paganini	0.03 (0.9	n/a)			

			to 12)				
	Plts <50	(Rondinelli	1.2 to 3.9,	n/s)			
		<20	0.007				
	Plts <20	(Mian	p=0.05	n/s)			
Biochem							
	CRP	Mian	0.029	1.0 to 1.03, p=0.021			

Appendix 11. Full list of QUADAS criteria for included biomarkers studies

Author	1	2	3	4	5	6	7	8	9	10	11
Ammann, 2003	Yes	Yes	Yes	Yes	No	Yes	Yes	Unclear	Unclear	Yes	Yes
Asturias, 2010	Yes	Yes	Yes	Yes	Unclear	No	No	Unclear	Unclear	Yes	Yes
Avabratha, 2009	Yes	Yes	Yes	Yes	Unclear	Yes	No	Unclear	Unclear	Yes	Yes
Barnes, 2002	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	No	Yes
Cost, 2011	Yes	Yes	Unclear	Unclear	Unclear	No	No	Unclear	Unclear	Unclear	Yes
de Bont, 1999	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Diepold, 2008	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Dylewska, 2005 a&b	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
El-Maghraby, 2007	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Hatzistilianou, 2007	Unclear	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Hatzistilianou, 2010	Yes	No	Yes	Unclear	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes
Heney, 1992	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes

Hitoglou-Hatzi, 2005	Unclear	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Unclear	Yes
Hodge, 2006	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Hodge, 2011	Yes	No	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Unclear	Yes
Kharya, 2010	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Katz, 1992	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Kitanovski, 2006	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Lehrnbecher, 1999	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Unclear
Lehrnbecher, 2004	Yes	No	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Yes	Yes
Lodahl, 2011	Yes	Yes	Yes	Unclear	Yes	Yes	No	Unclear	Unclear	Yes	Yes
Mian, 2009	Yes	No	Yes	Yes	Unclear	No	No	Unclear	Unclear	Unclear	Yes
Miedema, 2011	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes	Unclear	Unclear	Yes	Yes
Nishikawa, 2010	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Reitman, 2010	Yes	No	Unclear	Unclear	Yes	No	No	Unclear	Unclear	Unclear	Yes
Richardson, 2009	Yes	Yes	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Yes	Yes

Riikonen, 1992	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Riikonen, 1993	Unclear	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Santolaya, 1994	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Santolaya, 2007	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes
Santolaya, 2008	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes
Secmeer, 2007	Unclear	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Soker, 2001	Unclear	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Spasova, 2005	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Stryjewski, 2005	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes
Santolaya, 2001	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Santolaya, 2002	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

1 = representative patients, 2 = clearly described selection criteria, 3 = whole sample, or a random selection of sample, received reference standard, 4 = all patients received same reference standard, 5 = index test not part of reference standard, 6 = index test described adequately, 7 = reference standard described adequately, 8 = blinded interpretation of index test results, 9 = blinded interpretation of reference standard results, 10 = same clinical data available as in clinical practice, 11 = adequate reference standard

Appendix 12. Handling continuous and categorical variables in biomarkers studies.

Citation	How were continuous variables handled in the analyses?	If relevant, describe methods used for cutpoint determination	How were categorical variables handled in the analyses?	If relevant, describe methods used for cutpoint determination
Ammann, 2003	Made ordinal, with up to three categories	Not stated	Grouped	Not stated
Asturias, 2010	Grouped	Not described		
Avabratha, 2009	Grouped	Not described		
Barnes, 2002	Continuous and dichotomised	Not stated. Appears to be maximal diagnostic efficiency (based on study Figure)		
Cost, 2011	Continuous			
de Bont, 1999	Continuous (markers), grouped (age)	Age split into <16, 16-50, and >50. Implied that a level was chosen 'that identified 28% patients as low risk with 100% sensitivity'	Sex and type of malignancy used ungrouped	

Diepold, 2008	Continuous and dichotomised	Based on ROC curve to maximise diagnostic utility		
Dylewska, 2005 a & b	Continuous			
El-Maghraby, 2007	Continuous and dichotomised	Taken from (CRP) literature or cytokines local control upper limit of normal		
Hatzistilianou, 2007	Continuous and dichotomised	Not stated. Appears to be based on ROC analysis.		
Hatzistilianou, 2010	Continuous			
Heney, 1992	Continuous			
Hitoglou-Hatzi, 2005	Continuous			
Hodge, 2006	Continuous and dichotomised	Upper limit of normal defined as Mean + 2SD in control population		

Hodge, 2011	Continuous			
Kharya, 2010	Continuous then grouped	ROC curves		
Katz, 1992	Continuous and dichotomised	Not stated. Appear arbitrary.	Dichotomised	Haematological vs. solid tumours
Kitanovski, 2006	Continuous and dichotomised	'Best predictive value' by ROC analysis		
Lehrnbecher, 1999	Continuous and explored with cutoffs	IL6 & 8 cut at two levels: to maximise sensitivity and median values.		
Lehrnbecher, 2004	Continuous and dichotomised	Based on previous threshold values (not this dataset)		
Lodahl, 2011	Continuous and grouped (alternative models)	Grouped: PCT - PCT cut-off levels from literature, CRP- cutoff levels chosen to give identical sensitivities		
Mian, 2009	Continuous			

Miedema, 2011	Continuous			
Nishikawa, 2010	Continuous			
Reitman, 2010	Continuous			
Richardson, 2009	Grouped	Results from another study suggesting that CRP >/=4 was predictive of bacteraemia		
Riikonen, 1992	Continuous and explored with cutoffs	Not stated	Unclear	
Riikonen, 1993	Continuous and explored with cutoffs	Not stated. Appears to be based on previously defined values		
Santolaya, 1994	Continous and dichotomised	From literature reflecting CRP in non-immunosuppressed patients		
Santolaya, 2001	Continuous and dichotomised	By ROC analysis	Grouped	Not stated
Santolaya, 2002	Continuous and	States cutoffs were determined using ROC		

	dichotomised	analysis		
Santolaya, 2007	Continuous and dichotomised	States cutoffs were determined using ROC analysis	Grouped	Not stated
Santolaya, 2008	Continuous and dichotomised	"best discriminative cutoff value"		
Secmeer, 2007	Continuous and dichotomised	Not stated. Appears to be based on previously defined values	Grouped (e.g. duration of neutropenia was defined as prolonged if more than 72 hours)	Not stated
Soker, 2001	Continuous			
Spasova, 2005	Continuous and dichotomised	Not stated		
Stryjewski, 2005	Continuous and dichotomised	Data derived (maximal efficiency, and to ensure 100% Sn)		

Appendix 13. Further data on infrequently or partially reported markers and outcomes

Citation	Marker and Cutpoint	Outcome	Sensitivity (95% CI) Or measure in 'infected'	Specificity (95% CI) Or measure in 'non-infected'	Method of derivation
De Bont, 1999	CRP	Bacteraemia	Not significantly associated		
De Bont, 1999	IL6	Bacteraemia	Beta value ln(IL6) = 0.658 (se 0.31)		
De Bont, 1999	IL8	Bacteraemia	Beta value ln(IL8) = 0.551 (se 0.28)		
Delebarre, 2010	PCT 0.3ng/ml	Significant infection	'Significantly associated in multivariate analysis'		
El-Maghraby, 2007	MCP 350	Documented infection	0.644 (95% CI 0.517 to 0.754)	0.923 (95% CI 0.759 to 0.979)	2*2 extracted from text/graph
Hitoglou-Hatzi, 2005	tADA 35U/l	Significant bacterial infection	1 (95% CI 0.883 to 1)	1 (95% CI 0.908 to 1)	2*2 extracted from text/graph
Cost, 2011	IL8 and IL5 (no cutoff)	Bacteraemia or clinical sepsis	0.88	0.48	
Kharaya, 2010	PCT 3.3ng/ml	Bacteraemia	Sn 66% (derivation), 80% (validation)	Sp 85% (derivation), 80% (validation	No absolute numbers of bacteraemia to use in calculating appropriate CI
Kharaya, 2010	IL6	Bacteraemia	Sn 67%	Sp 75%	No absolute numbers of

	137pg/ml		(derivation), 60% (validation)	(derivation), 77% (validation	bacteraemia to use in calculating appropriate CI
Riikonen, 1992	TNF 40	Bacteraemia or focal infection	1 (95% CI 0.879 to 1)	0.065 (95% CI 0.025 to 0.154)	2*2 extracted from text/graph
Hodge, 2006	IL5 17	Positive blood culture	0.5 (95% CI 0.215 to 0.785)	Could not calculate	2*2 extracted from text/graph
Hodge, 2006	IL5 & 8 combined >17 and >220	Positive blood culture	1 (95% CI 0.676 to 1)	0.87 (95% CI 0.679 to 0.955)	2*2 extracted from text/graph
Riikonen, 1992	IL1 SAA	Bacteraemia or focal infection	Could not calculate	Could not calculate	
Soker, 2001	IL2R	Bacteraemia	Median (range) 5230U/mL (1120 to 7600)	1190 (724 to 5400)	Medians/range only reported
Soker, 2001	TNF-alpha	Bacteraemia	8.4 (4.0 to 68.2)	7.8 (3.0 to 37.2)	Medians/range only reported
Soker, 2001	IL1	Bacteraemia	Could not calculate	Could not calculate	
Soker, 2001	IL8	Bacteraemia	305pg/ml (16 to 4838)	23pg/ml (5 to 184)	Medians/range only reported
Santolaya, 2008	PCT (ng/ml)	Severe infectious complications	Mean (+/- reported SD) Admission	Mean (+/- reported SD) Admission 6 5.5	Only in patients in the non-low risk group. Clearly non-Normal data reported as

			13.6 +/- 63.6 24h 14.5 +/- 54.5	_+/-30.7 24hrs 8.8 +/- 40.2	mean/SD.
Santolaya, 2008	IL8 (ng/ml)	Severe infectious complications	Mean (+/- reported SD) 444.9 (623)	Mean (+/- reported SD) 232.9 (445.6)	Only in patients in the non-low risk group. Clearly non-Normal data reported as mean/SD.
Secmeer, 2007	ESR	Bacteraemia	"not statistically significantly different between patients with and without documented infection"		Figures not shown
Stryjewski, 2005	PCT	Sepsis or septic shock	502pg/ml (39 to 54774)	201pg/ml (47 to 9862)	Median values (range) for admission comparing non-septic with septic
Stryjewski, 2005	IL6	Sepsis or septic shock	10pg/ml (0 to 597)	11pg/ml (0 to 105)	Median values (range) for admission comparing non-septic with septic
Stryjewski, 2005	IL8	Sepsis or septic shock	45pg/ml (7 to 278)	15.5pg/ml (0 to 219)	Median values (range) for admission comparing non-septic with septic
Mian, 2011	IL6	Admission to PICU	5594(+/-4337)	981 (+/-2252)	Mean +/- SD
Mian, 2011	IL8	Admission to PICU	2465 (+/-2504)	419 (+/-921)	Mean +/- SD
Mian, 2011	IL10	Admission to	3971	322 (+/11819)	Mean +/- SD

		PICU	(+/15251)		
Mian, 2011	TNF-alpha	Admission to PICU	61 (+/-56)	11 (+/-22)	Mean +/- SD
Mian, 2011	CRP	Admission to PICU	243 (+/-112)	130 (+/-91)	Mean +/- SD

IL6 cutoff 137pg/ml - 67% sensitive and 75% specific in predicting bacteraemia in test group IL6 cutoff 137pg/ml - 60% sensitive and 77% specific in predicting bacteraemia in validation group

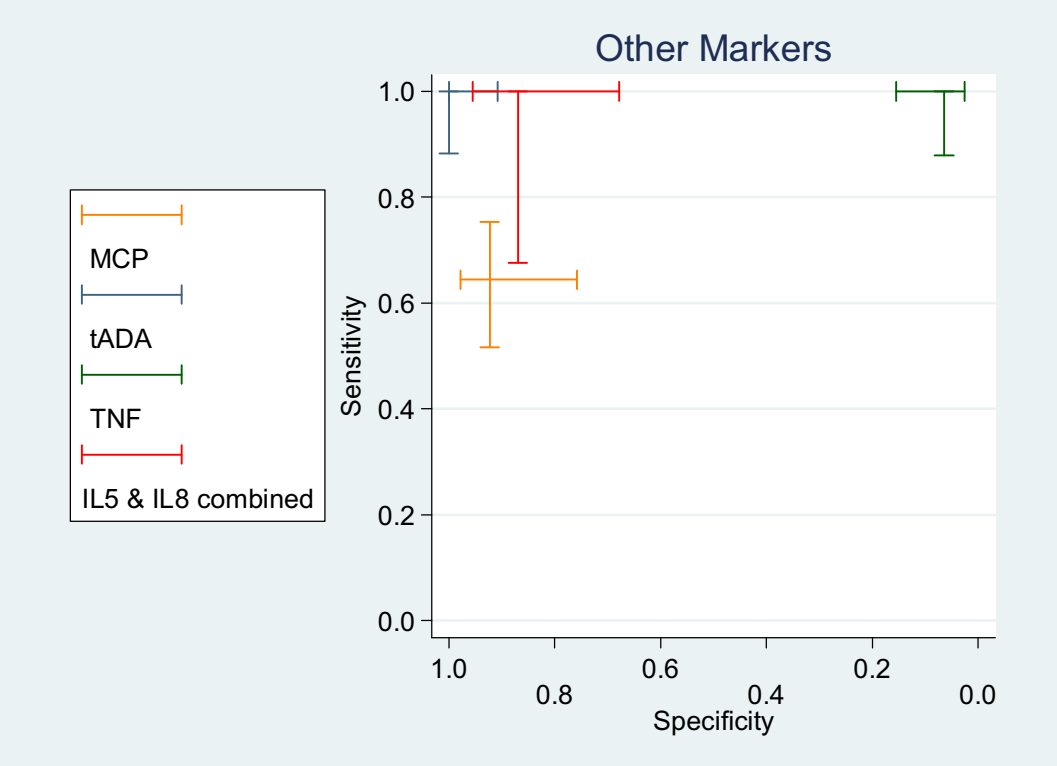
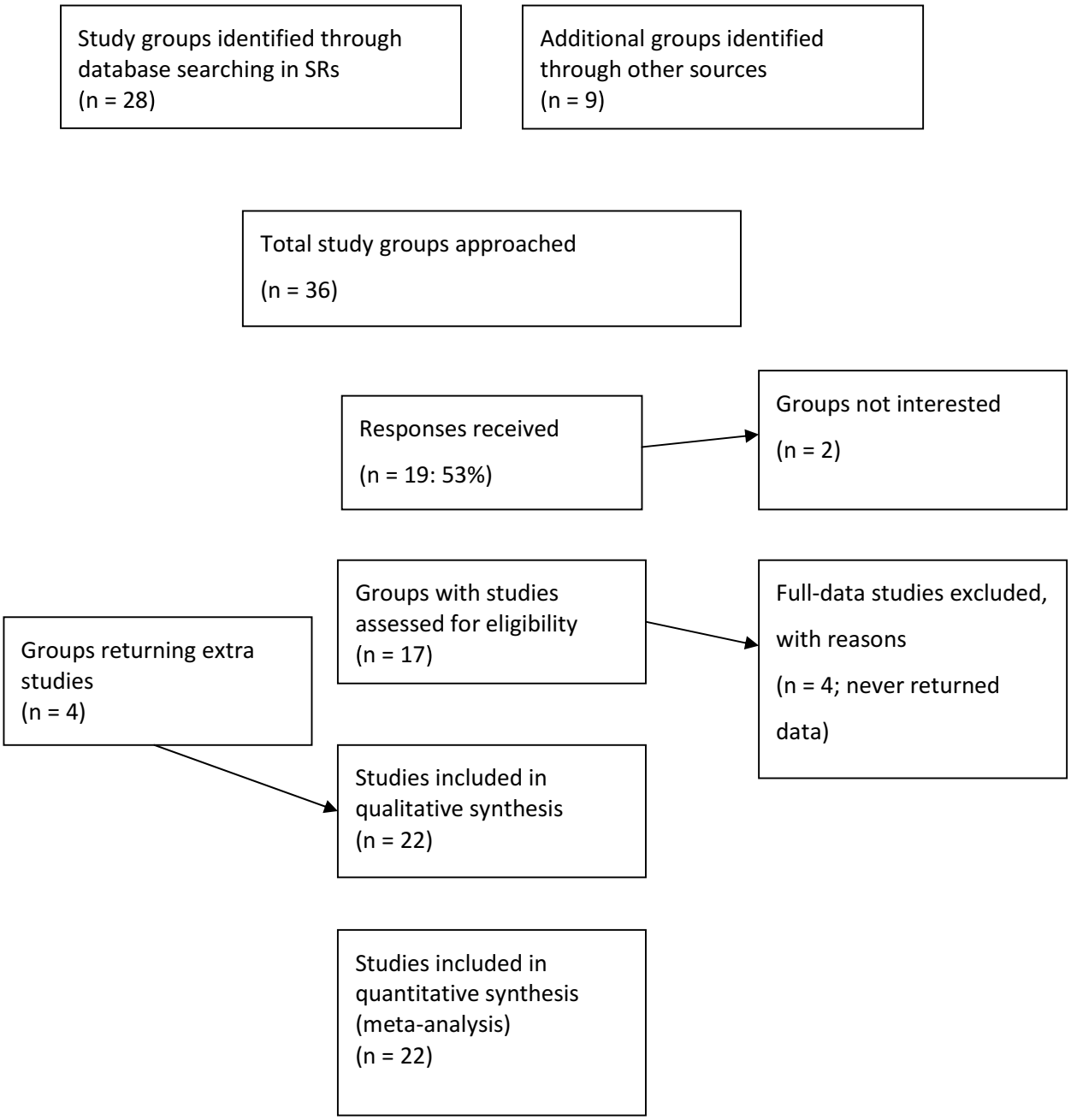


Figure 56: Diagnostic value of exploratory biomarkers for documented infection

Appendix 14. Study inclusion flow diagram



Appendix 15. The PICNICC Invite Letter

The PICNICC (Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer) Collaborative

Optimizing risk predictive strategies in febrile neutropenic episodes in children and young people undergoing treatment for malignant disease

Dr. Michaela Semeraro
Department of Paediatric Oncology
39 Rue Camille Desmoulins
94805 Villejuif
France

3 December 2010

Dr Bob Phillips
PICNICC Collaborative
Centre for Reviews and Dissemination
Alcuin College
University of York
York
UK
YO10 5DD

Dear Dr Semeraro,

We are writing to you on behalf of an emerging collaboration of paediatric oncologists and academics who are undertaking a systematic review and individual patient data meta-analysis of studies evaluating risk stratification of febrile neutropenic episodes in children and young people. Through our systematic searches of the published literature in this field, we have identified you as an author of "Semeraro M, Thomée C, Rolland E, et al. A predictor of unfavourable outcome in neutropenic paediatric patients presenting with fever of unknown origin. *Pediatric Blood & Cancer* 2010;54(2):284-290." and would like to invite you to join the collaboration.

Systematic reviews and meta-analyses provide an opportunity to improve our knowledge in any field of research by drawing together multiple studies to produce a more precise conclusion than any one study could achieve. When published summary results alone are used, reviews can be limited in the conclusions they reach because of the challenges of reporting data in published papers and the varied assessments of differences between important sub-groups of patients. The use of individual patient data techniques preserves the benefits of traditional systematic reviews but overcomes the limitations of reporting within published papers and allows more powerful sub-group analysis.

We would like to invite you to become a collaborator: to be involved in the refinement of the protocol, and provide your dataset for the central collaborative. The collaborative already includes clinicians from the UK, USA, Switzerland, Germany, Mexico, India and Italy, and we would greatly value the input of your group. We intend the results of this review be submitted to a peer-reviewed journal regardless of its findings, under the collaborative group name, of which you would be a member.

To be a member of the PICNICC collaborative, you would need to supply individual patient data from cohort studies of in children and young people, including randomised trial data, who presented with febrile neutropenia. This could be with either prospective or retrospective data collection, but needs to provide data for all 'essential' predictive variables in >50% of included episodes and two or more of the study defined outcomes for >90% of the included episodes of FNP. We will exclude studies which are case-series (for example, of only 'gram negative bacteraemias').

The 'essential' predictive variables are proposed to be age, underlying tumour type, and remission status; chemotherapy type and time elapsed since last cycle; in-patient or out-patient at onset of episode; maximum temperature; antibiotic therapy used; white cell count; neutrophil count and at least one of the following four assessments: respiratory rate (or compromise), circulatory status (or compromise), presence of severe mucositis, or a global assessment of illness severity. The core outcome variables are: death; intensive care admission; need for moderate organ support (fluid bolus, oxygen); clinically documented infections and microbiologically documented infections. These are subject to discussion in the refinement of the study protocol.

We would be delighted hear from you, and happy to answer any questions that you may have about becoming involved in this project.

Dr Bob Phillips

MRC Research Fellow, Centre for Reviews and Dissemination, University of York and Consultant in Paediatric Oncology, Leeds Teaching Hospitals Trust, Leeds

Prof Alex Sutton, Professor of Medical Statistics, University of Leicester
Dr Richard Riley, Senior Lecturer in Medical Statistics, University of Birmingham
Dr Julia Chisholm, Consultant in Paediatric Oncology, Royal Marsden NHS Foundation Trust, Surrey
Dr Susan Picton, Consultant Paediatric Oncologist, Leeds Teaching Hospitals Trust, Leeds
Prof Lesley Stewart, Director, Centre for Reviews and Dissemination, University of York

e: crd-picnicc@york.ac.uk
t: +44 1904 321099
f: +44 1904 321041
w: http://bit.ly/PICNICC

Appendix 16. Parental Advisory Request

The PICNICC (Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer) Study
Optimizing risk predictive strategies in febrile neutropenic episodes in children and young people undergoing treatment for malignant disease
This project will use an individual patient data meta-analysis approach to develop and evaluate a risk stratification model to predict which children and young people have a low risk of adverse outcomes during an episode of febrile neutropenia. This will define a group of individuals who may be treated with reduced intensity or duration of antibiotic therapy, and so reduce the inconvenience and cost of these episodes. The project will also explore specific issues around adapting established techniques of IPD meta-analysis of interventions for use the method of prognostic evidence synthesis and predictive modeling.
Febrile neutropenia (FNP) is a complication of cancer therapy. It is the occurrence of a fever in the context of immunosuppression which may lead to death from overwhelming sepsis. It is the second commonest reason for hospital admission among children & young adults with cancer, with approximately 4000 episodes of FNP occurring annually in the UK. In adopting a policy of aggressive in-patient intravenous antibiotic use in such episodes, the mortality rate related to these episodes has improved dramatically (from 30% in the 1970s to 1% in the late 1990's). However, there remain many episodes of FNP, possibly two-thirds or more, in whom no significant infection is identified, and in whom this aggressive management strategy is likely to be excessive.
The National Institute for Clinical Excellence (NICE) guidance document "Improving outcomes with children and young people with cancer" called for "the development of robust methods of risk stratification in the management of FNP". At present there are many differing policies for the management of FNP in the UK with lack of agreement about how risk stratification, if any, is used. Such models of risk stratification are based on small data sets with relatively few events.
Individual patient data pooled analysis for the synthesis of prognostic information has only recently been begun to be applied to real world clinical data sets where they have clarified existing understanding of particular prognostic variables. It has been suggested that the use of IPD will provide a more accurate and robust assessment of the value of potential risk factors. Systematic review and use of summary prognostic data may be unreliable as the published data may be incomplete (missing vital information for meaningful meta-analysis), often relies on categorization of continuous outcome variables (which themselves may be biased), are susceptible to significant publication bias (with prognostic markers showing 'highly significant' responses being more likely to be published) and may have many simple methodological flaws (such as inconsistent reporting of ostensibly similar outcomes).
Methodological exploration, development and adaptation to the prognostic setting of the techniques of IPD pooled analysis will also be undertaken (the exact analyses will be subject to the data sets obtained). The analyses may address issues regarding the development of clinical decision rules in a meta-analytic setting, the analysis of missing data using different imputation models, the relative merits of prospective and retrospectively collected information, the comparison between episodic and patient-centred analyses and the use of categorical outcome variables.
It is for these reasons that this project was conceived, and awarded funding as part of an MRC Research Training Fellowship. The Fellowship will be undertaken in the Centre for Reviews and Dissemination (University of York) under the supervision of Prof Lesley Stewart, Dr Alex Sutton and Dr Dawn Dowding. The Fellowship has been awarded on a part-time basis, to complement a part-time post as Consultant in Paediatric Oncology in St James's Hospital, Leeds. The two aspects will complement and support both the project and clinical practice, while being separated by being undertaken on different sites on different days in the week.

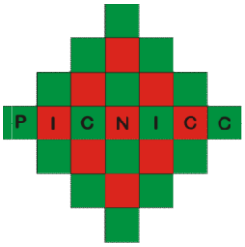
The PICNICC (Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer) Study Clinical Advisory Group

This group will be include clinicians and methodologists and will provide advice throughout the project. This group will help:
ground decisions on data in clinical practice
explore the opportunities for family/patient centered analyses
advise on methodological issues
provide advice as to likely clinical uptake and implementation
encourage networking and data sharing for success of the project
Members of the group will ideally attend the international collaborative group meetings, when this is formed.
The currently proposed structure of this group is:
Lesley Stewart, CRD, York (expert in IPD methodologies & reviews)
Sue Picton, Paediatric Oncology, Leeds (interest in FNP, great experience in field)
Julia Chisholm, Paediatric Oncology, London (interest in FNP, Chair of CCLG supportive care group)
Alex Sutton, Medical Statistics, Leicester (expert in meta-analytic techniques, including Baysian approaches)

This group may be usefully strengthened by the addition of two patient or family representative members, and a statistician with expertise in prognostic models and/or clinical decision rule building.

The group is expected to meet between one and two times per year, with occasional contact by telephone call or email in between the meetings. Such meetings would be unlikely to be more than half-a-day in length. The offer of pre-meeting and post-discussions with one of the PICNICC team would be offered to patient or family representative members in order to discuss any issues that arise during the meetings. Expenses to cover travel and associated expenses have been awarded.

(1st draft)



The PICNICC (Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer) Study

Optimizing risk predictive strategies in febrile neutropenic episodes in children and young people undergoing treatment for malignant disease

Febrile neutropenia (FNP) is a complication of cancer therapy. It is the occurrence of a fever in a patient who is immunosuppressed following cancer therapy, and is potentially an indication of serious infection. It is the second commonest reason for hospital admission among children & young adults with cancer, with approximately 4000 episodes of FNP occurring annually in the UK. Since this was recognised in the early 1960s, patients have been aggressively treated with hospitalisation and intravenous antibiotics. This policy has improved the mortality rate dramatically (from 30% in the 1970s to 1% in the late 1990's). There remain many episodes of FNP, possibly two-thirds or more, in which no significant infection is identified, and in these patients this management approach is unnecessary. Clear risk stratification would enable us to differentiate between these extremes.

In the UK, the National Institute for Clinical Excellence (NICE) guidance document "Improving outcomes with children and young people with cancer (2005)" called for "the development of robust methods of risk stratification in the management of FNP". At present there are many different policies for the management of FNP in the UK with lack of agreement about how risk stratification, if any, is used.

There have been a number of studies in small numbers of patients trying to define which patients need the aggressive treatment, and which could be treated with much reduced intensity of therapy. These risk stratification models vary between the studies, and none have been tested extensively. Small numbers of cases lead to uncertainty about the results of studies: any conclusions drawn may be chance findings, rather than real differences, and this may explain why different models have been developed. Secondly, models which have not been tested may not be usable on a larger scale.

To try to develop our understanding of risk stratification we plan to undertake a "meta analysis" of the previously undertaken studies. Meta-analysis is the technique of collecting and statistically synthesising data from multiple reports. In doing this, it allows larger numbers to be studied, and the data is drawn from a variety of sources. Systematic review and use of only published data may be unreliable as the information may be incomplete (missing vital information for meaningful meta-analysis). Additionally, prognostic and diagnostic reports are susceptible to significant publication bias (where studies showing prognostic markers to have 'highly significant' responses being more likely to be published). There may also be particular statistical problems, such as the categorization of continuous outcome variables. (This refers to the grouping of data that can be measured on a scale, such as age or weight. The groups may vary between reports (e.g. "less than 12kg vs. more than 12kg" in one study, and "less than 10kg vs. more" used in another study), which means they cannot easily be synthesised, and the reasons why a particular cut-point has been chosen may themselves be biased. There may also be simple flaws (such as inconsistent reporting of apparently similar outcomes).

In situations similar to the problem of risk stratification, an approach has been developed which uses the raw data ("individual patient data" or IPD) pooled analysis for the synthesis. There are a range of theoretical reasons why this approach will be better, and overcome some of the problems present in using previously published data alone. This study will use this IPD approach, by creating an international collaborative of researchers in the area to combine and share the data, with the objective of creating and testing an effective risk stratification model.

As the techniques of IPD pooled analysis are still relatively new, part of the project is to test theories that exist about how to undertake the synthesis. The exact issues will depend on the data available, but may address issues regarding the development of clinical decision rules in a meta-analytic setting (rather than just using data from one source), the analysis of missing data using different statistical imputation models, the relative merits of information collected prospectively and information taken by trawling through from medical records, assessing data from people who have few or many episodes of FNP (to determine if there is a difference in how risk should be assessed)

and the use of categorical rather than dichotomous outcome variables (for example "Needed ICU", "Needed oxygen", "Identified infection" and "Well - no complications" compared with "Well" vs. "Not well").
It is to develop a clinically useful decision rule and to advance our understanding of the statistical techniques used that this project was conceived, and awarded funding as part of a Medical Research Council (MRC) Research Training Fellowship. The Fellowship will be undertaken in the Centre for Reviews and Dissemination (University of York) under the supervision of Prof Lesley Stewart, Dr Alex Sutton and Dr Dawn Dowding. The Fellowship has been awarded on a part-time basis, to complement a part-time post as Consultant in Paediatric Oncology in St James's Hospital, Leeds. The two aspects will complement and support both the project and clinical practice.

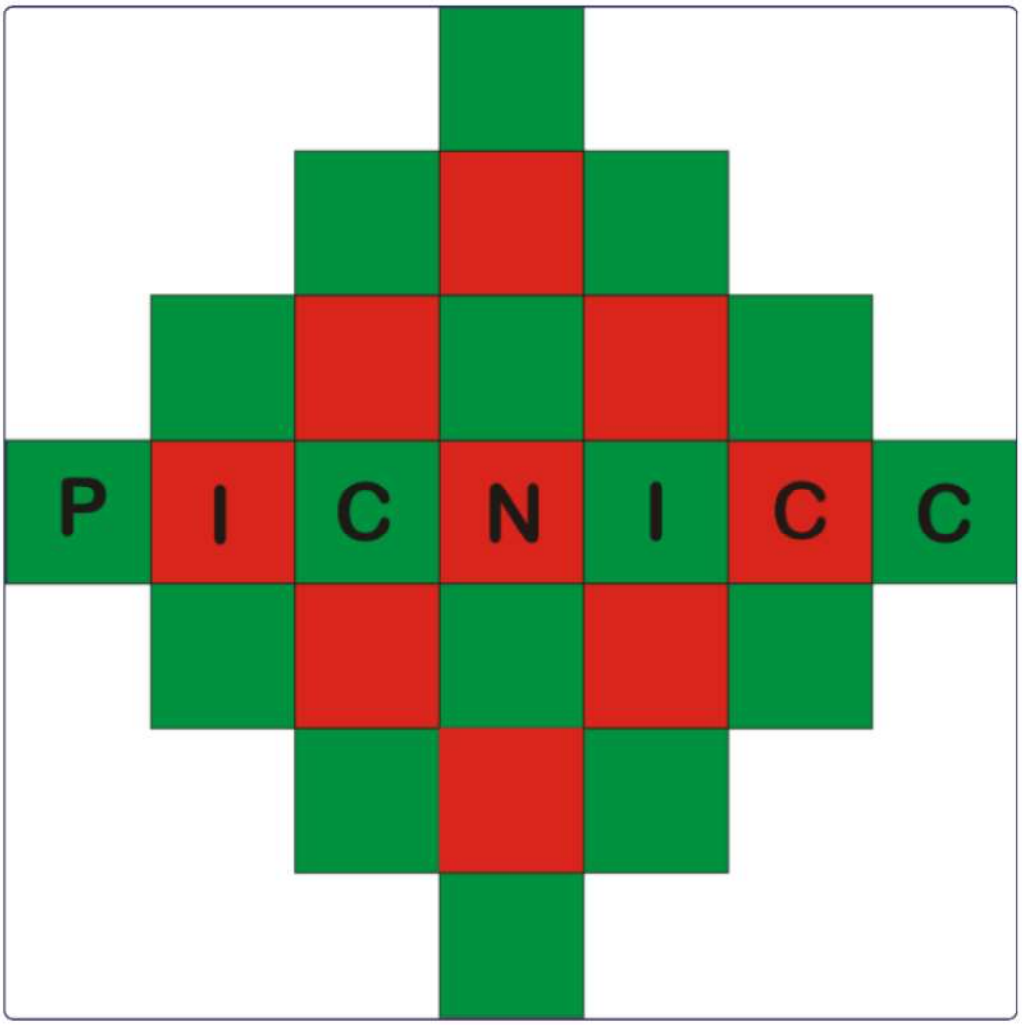
The PICNICC (Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer) Study Clinical Advisory Group

This group will provide advice throughout the project. This group will help:
ground decisions on data in clinical practice
explore the opportunities for family/patient centered analyses
advise on methodological issues
provide advice as to likely clinical uptake and implementation
encourage networking and data sharing for success of the project
This group will benefit from the involvement of two patient or family representative members, and a statistician with expertise in prognostic models and/or clinical decision rule building.

The group is expected to meet between one and two times per year, with occasional contact by telephone call or email in between the meetings. Such meetings would be unlikely to be more than half-a-day in length. Pre-meeting and post- discussions with one of the PICNICC team will be offered to patient or family representative members in order to discuss any issues that arise during the meetings. Expenses to cover travel and associated expenses have been awarded. Members of the group will ideally attend the international collaborative group meetings, when this is formed. These may occur only 2-3 times over 5 years.
The currently proposed structure of this group is:
Prof. Lesley Stewart, CRD, York (Expert in IPD methodologies & reviews)
Dr. Sue Picton, Paediatric Oncology, Leeds (Extensive experience in FNP)
Dr. Julia Chisholm, Paediatric Oncology, London (Chair of CCLG supportive care group)
Dr. Alex Sutton, Medical Statistics, Leicester (Expert in meta-analytic techniques)

Should you wish to find out more to consider becoming involved, please contact Dr Bob Phillips at **picnicc@gmail.com** or phone (10904) 321099.

Appendix 17. Protocol



Predicting infectious complications in neutropenic children and young people with cancer (IPD protocol)

Phillips *et al.*

BioMed Central

Phillips *et al. Systematic Reviews* 2012, **1**:8
http://www.systematicreviewsjournal.com/content/1/1/8 (9 February 2012)

SYSTEMATIC REVIEWS

PROTOCOL Open Access

Predicting infectious complications in neutropenic children and young people with cancer (IPD protocol)

Robert S Phillips[1,5*], Alex J Sutton[2], Richard D Riley[3], Julia C Chisholm[4], Susan V Picton[5] and Lesley A Stewart[1], for the PICNICC Collaboration

Abstract

Background: A common and potentially life-threatening complication of the treatment of childhood cancer is infection, which frequently presents as fever with neutropenia. The standard management of such episodes is the extensive use of intravenous antibiotics, and though it produces excellent survival rates of over 95%, it greatly inconveniences the three-fourths of patients who do not require such aggressive treatment. There have been a number of studies which have aimed to develop risk prediction models to stratify treatment. Individual participant data (IPD) meta-analysis in therapeutic studies has been developed to improve the precision and reliability of answers to questions of treatment effect and recently have been suggested to be used to answer questions regarding prognosis and diagnosis to gain greater power from the frequently small individual studies.

Design: In the IPD protocol, we will collect and synthesise IPD from multiple studies and examine the outcomes of episodes of febrile neutropenia as a consequence of their treatment for malignant disease. We will develop and evaluate a risk stratification model using hierarchical regression models to stratify patients by their risk of experiencing adverse outcomes during an episode. We will also explore specific practical and methodological issues regarding adaptation of established techniques of IPD meta-analysis of interventions for use in synthesising evidence derived from IPD from multiple studies for use in predictive modelling contexts.

Discussion: Our aim in using this model is to define a group of individuals at low risk for febrile neutropenia who might be treated with reduced intensity or duration of antibiotic therapy and so reduce the inconvenience and cost of these episodes, as well as to define a group of patients at very high risk of complications who could be subject to more intensive therapies. The project will also help develop methods of IPD predictive modelling for use in future studies of risk prediction.

Keywords: individual participant data meta-analysis, predictive modelling, paediatric oncology, febrile neutropenia, collaborative studies

Background

Children undergoing treatment for malignancy have an excellent chance of survival, with overall rates approaching 75% [1]. In most cases, children who die following treatment for cancer do so as a result of their disease, but despite huge improvements in supportive care, around 16% of deaths within 5 years of diagnosis are due to the complications of therapy [2,3]. One such life-threatening complication in immunocompromised children remains infection, which frequently manifests as the occurrence of fever with neutropenia [4].

In adopting a policy of aggressive inpatient intravenous antibiotic use in such episodes, the mortality rate related to these episodes has improved dramatically (from 30% in the 1970s to 1% in the late 1990s) [4]. Intensive care management is required in less than 5% of cases [5-7], although a substantial proportion of children have complications which require specialised care [7]. There

* Correspondence: bob.phillips@york.ac.uk
[1]Centre for Reviews and Dissemination, Alcuin College, University of York, York, YO10 5DD, UK
Full list of author information is available at the end of the article

remain many episodes of febrile neutropenia (FNP), possibly two-thirds or more, among patients in whom no significant infection is identified and in whom this aggressive management strategy is likely to be excessive [7].

To better inform the clinical management of children with cancer and FNP, there is increasing interest in using risk prediction models (also known as 'prognostic models') and clinical decision rules (CDRs) [8-10]. Risk prediction models utilise multiple prognostic factors in combination to predict the risk of a future health outcome for an individual on the basis of their set of prognostic factor values. A CDR recommends a particular clinical action (or inaction) for an individual on the basis of the prediction (for example, the predicted probability, or 'risk score') derived from the model.

A robust risk prediction model which identifies those children at very low risk of having a significant infection could result in reduced intensity and/or duration of antibiotic therapy in the hospital. It could also form the basis of a randomised controlled trial (RCT) of alternative management approaches (for example, ambulatory oral antibiotics vs inpatient intravenous antibiotics) and would be the ideal way of informing the sample size required by reliably predicting the proportion of events expected in a low-risk group. This would lead to reduced costs for the healthcare system and the patient and family [11], as well as potentially a better quality of life for all affected. At present, there are many differing policies for the management of FNP in practice [12,13] but a lack of agreement about how and which CDRs, if any, are used.

Assessment of the risk of adverse outcome of each episode of FNP has been undertaken by many different groups, with many of them creating a CDR which aims to allow clinicians to accurately judge risk and treat patients appropriately. However, none of these analyses have resulted in a widely used risk stratification model, and current practice is variable, both in the United Kingdom [12] and internationally [13-15]. Some centres use a risk-stratified, reduced-intensity approach, and others treat all children aggressively. The essential problems with research in this area are common across much of paediatric practice: those of rare conditions with small numbers of cases and limited collaboration in primary studies. The modelling studies that have been done have incorporated different clinical features and outcomes and have used different methodologies, and it is therefore difficult to draw meaningful conclusions from this body of evidence. Calls for collaborative trials [16-18] have led to little progress.

This setting provides an ideal opportunity to undertake a collaborative, pooled analysis of the existing data sets in the form of an individual participant data (IPD) meta-analysis. In this effort, we will collect and reanalyse the original study data, which will permit reanalysis of the same clinical features across studies using a consistent approach and provide sufficient numbers to draw more robust and reliable conclusions. The findings of this work should therefore more robustly inform practice and future therapeutic RCTs. The analysis can be approached from three methodological directions: testing existing CDRs for their ability to 'diagnose' adverse outcomes, assessing the added value of individual prognostic factors and building a more accurate predictive rule containing a parsimonious set of prognostic factors.

Systematic reviews of existing knowledge

In preparation for this project, two systematic reviews were undertaken to assess the prior knowledge of the discriminatory ability of CDRs [19] and inflammatory serum markers (R Phillips, R Wade, T Lehrnbecher, LS Stewart, A Sutton) in children and young people with FNP. The systematic review of CDRs initially identified 2,057 potential studies and finally included 24, of which 21 had data in a usable format. It showed that two groups of studies have been undertaken to risk-stratify children who present with FNP. Researchers in the first group of studies examined the use of clinical examinations to predict radiographic pneumonia (4 studies) [20], and investigators in the second group examined more general infectious complications (20 studies) [19].

Among the studies in which general infectious complications were examined, 16 separate models were produced and contained 9 data sets used to validate previously derived models. The researchers studied a variety of outcomes with individual differences in definitions, but covered five main categories: death, critical care requirement, serious medical complications, significant bacterial infections and bacteraemia.

Only one rule could reasonably be assessed across multiple data sets: that of absolute monocyte count and temperature criteria proposed by Rackoff *et al.* [21] to exclude bacteraemia. The most appropriate meta-analysis of the rule's effectiveness led to estimates of moderate discriminatory ability, with the average probability of bacteraemia in the groups being low risk = 6% (95% CrI = 1% to 34%), middle-level risk = 18% (CrI = 3% to 37%) and high risk = 49% (95% CrI = 6 to 84%).

Of the other rules, the model of Santolaya *et al.* [22] showed a good ability to differentiate between low- and high-risk groups when a wider definition of 'serious infection' was used, with average predictive ability estimated as low risk = 13% (95% CI = 9% to 18%) and high risk = 72% (95% CI = 68% to 75%). The rule has been developed and tested in Chile and may be of limited applicability in Western Europe and North America [23]. Other rules show promise and have clinical physiological similarities, but have not undergone extensive testing.

The systematic review of the predictive value of serum markers of inflammation and infection in children presenting with FNP included 27 studies reporting over 13 different markers derived from an initial screen of 375 studies. The studies included had similar methodological challenges as well as problems with reporting and analysis. Many failed to assess whether the marker had any supplementary value over and above the simple admission data collected by the clinicians at every encounter: age, malignancy, temperature, age-corrected vital statistics and blood count.

To interpret the information on serum markers in a clinically meaningful way, we had to allow for the marked heterogeneity of the results. The quantitative pooling and a qualitative summary of the results suggested that procalcitonin might be a better discriminatory marker than C-reactive protein (CRP) and that IL-6 had a very good ability to predict documented infection. Overall the findings were uncertain and unstable, and only small amounts of new data may alter them substantially. Data for the other markers were too sparse to reasonably be interpreted, although IL-8 had significant potential value.

These reviews have a wide range of rules for the prediction of poor outcomes during episodes of FNP in children and the use of a variety of individual serum markers to predict outcome. None of the rules found has yet been subjected to the extensive geographical and temporal discriminatory validity assessments that mark the highest quality CDR. Many potential difficulties with different outcomes, variable selection and model-building have been identified. The data on serum markers were extremely heterogeneous, and only tentative conclusions could be drawn.

The problems identified are inherent in the attempt to undertake meta-analyses of aggregate data. The limitations of the reporting in published studies mean we do not have access to the exact data distribution or the full range of univariable estimates of predictive power. These issues could be addressed by attempting to collect more detailed summary data from the authors of the original studies, but this would not allow cross-study validation of different rules or attempts at alternative rule-building. To meet these challenges and to maximise the value of the information already collected by these groups and in other cohorts of children with FNP, an IPD meta-analysis will enable us to develop and test new and existing prediction models. This will provide a firmer basis for stratified treatment trials in this common and occasionally fatal complication of therapy.

Rationale for an individual participant data analysis

Individual patient data pooled analysis in therapeutic studies have been developed for two decades to improve the precision and reliability of answers to questions regarding treatment [24,25]. It has more recently been promoted for the synthesis of diagnostic data [26] and prognostic information [27] to improve the quality of answers to important prognostic questions [28] and matters of diagnostic accuracy [29]. These techniques have been applied to real-world clinical data sets [30,31], in which they have clarified existing understanding of particular prognostic variables and enhanced understanding of how different diagnostic tests can be used [32].

The key benefits of prognostic IPD analysis generally can be summarised as follows: (1) Analyses are not restricted to those of the published results or subgroups; (2) analytical techniques, inclusion criteria and outcome definitions can be standardised across studies; (3) larger numbers of data points allow more powerful statistical conclusions to be drawn, including checking modelling assumptions and accounting for missing data at the individual level; (4) IPD can model data more appropriately, such as by analysing continuous variables on continuous scales (unlike in many prognostic studies in which data are reported as categorical variables); (5) analysis can account for clustering (for example, of patients within studies) and correlated information (for example, multiple events per individual); (6) multivariate models can be created across different healthcare settings; (7) data can be reviewed for completeness and accuracy; and (8) the analysis can provide extensive internal cross-validation to guard against data-driven exaggerations of predictive power.

In the Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer (PICNICC) study, the collection and analysis of IPD will provide specific benefits that overcome many of the problems found in the aggregate data meta-analysis. Many of the benefits of IPD analysis are technical, being related to the statistical methods underlying the meta-analysis and the building of predictive models. Although at first sight the failure to address the problems inherent in statistical interpretation may seem to be clinically irrelevant, it has clear and real clinical implications [33]. Other benefits are more obviously clinical; for example, the collection of the different data sets will enable us to clarify and harmonise the different outcomes collected.

One of the primary 'statistical' benefits will be the use of firmly prespecified potential predictor variables built upon the experience of the PICNICC Collaborative and the systematic reviews. This will guard against the development of purely data-driven analyses, which have a tendency to overestimate any predictive value [28].

In the reviews, we found the studies designed to build a CDR used a large number of variables (median = 13, range = 2 to 39) and had a small number of events

(median = 36, range = 4 to 178) with 70% (12 of 16) studies having fewer than 10 events per variable under consideration and no study having more than 14 events per variable. These low event-per-variable ratios render predictive conclusions drawn from the studies unstable and estimates of predictive power overly optimistic [34]. IPD will allow us to consolidate the information and increase greatly the number of events studied from the same number of predictive variables.

The raw data will also allow a detailed analysis of the clustering of events (multiple episodes per patient) and variation at the level of the individual patient. This issue is significant when assessing the problems identified in the aggregate data reviews. Multiple episodes in individual patients were treated primarily as if they came from dissimilar individuals in the 20 CDR and 24 serum marker studies. Four papers explicitly described no adjustment [22,35-37], with 12 undertaking some attempt at assessment. Secondary analysis was performed to assess 'first included case' versus 'all episodes' and 'no significant differences' in three studies [22,38,39] and in nine others in which more advanced statistical modelling was used [6,21,40-46]. In 28 studies, the assessment was unclear.

The functional form of the data regarding *a priori* nonlinear fractional polynomial relationships can be assessed in detail. In no study assessed were clear attempts made to fit nonlinear forms to the data. This is unsurprising, as the development of practical techniques to undertake this effort is very recent [47].

Modern statistical developments in the handling of missing data may enhance the information already acquired. Again, very little information on the assessment and management of missing data was available from the reviews (five CDR studies [21,38,39,42,44] and two serum marker studies [48,49]). Very recent publications of studies in which simulation [50] and surveying practice [51] were used produced workable guidelines for the use of imputation techniques to maximise the value of the data collected. IPD will allow us not only to test existing rules and combine data derived from attempts to examine the rules but also, potentially, to develop a more robust rule for future use worldwide.

Parent and/or caregiver involvement

The development of shared research initiatives involving patients, clinicians and researchers has been a notable change in the practice of clinical research over the past decade [52]. It remains surprising to many researchers, clinicians and patients when they learn that their views are often strikingly different from each others' [53]. A systematic review of studies of the process of research planning and priority setting undertaken by the James Lind Alliance [52] demonstrated that the involvement of patients and parents as well as other caregivers was extremely infrequent.

The PICNICC group has sought to involve parents early in the treatment process. Discussions of the nature of their engagement in the process have so far highlighted that the representatives involved have not wished to be actively involved in the process of reviewing, but to be included in discussions about the nature of, the adverse effects of FNP and that they have been willing to provide their own nonmedical expertise in advancing the project.

The discussion of the nature and extent of patient and caregiver involvement in the PICNICC group will continue as the project develops. Possible opportunities for further involvement include writing commentaries on the study for patients and their families, providing alterative views on ethical questions, making choices regarding risk thresholds and considering how uncertainty and imprecision should be managed.

Methods

Aims

Primary

A primary aim of the project is to undertake an IPD pooled analysis to quantify the risk of adverse clinical outcomes according to clinical variables in children and young people undergoing treatment for malignant disease who present with an episode of FNP; that is, to identify which variables are prognostic and which have the most independent prognostic importance. Another primary aim is to develop and validate a new risk prediction model containing multiple prognostic factors in combination.

Secondary

The secondary aim of this project is to develop and explore practical and methodological issues surrounding the use of pooled IPD analysis in the development of predictive models.

Inclusion and exclusion criteria

Studies will be considered for inclusion in the IPD meta-analysis if they are cohort studies of children and young people presenting with FNP and/or with either prospective or retrospective data collection, including RCT data; if they provide data for all 'essential' predictive variables in more than 50% of included episodes (see 'Core data set and variables' section); and if they provide details of two or more study-defined outcomes in more than 90% of individual episodes of FNP.

Studies will be excluded if they are case series (for example, studies of only 'Gram-negative bacteraemias') and if they did not record data on all 'essential' predictive variables or cannot provide sufficient outcome data.

Studies will be included if they focus on the collection of data from children and young people (between 0 and

24 years old). The purpose of the inclusion criterion of studies of young people up to the age of 24 years is to address a paucity of research on individuals in the 'young adult' age range [54]. Data from individual patients ages 25 years and older will be excluded from this analysis. The median age of inclusion in the 'children's' cohorts examined in our reviews was about 7 years old (ranging from 1 month to 23 years), and the 'adult' study from the Multinational Association for Supportive Care in Cancer group [55,56] has a median age of 52 years (range, 16 to 91 years old).

Identification of potential studies

The initial identification of studies has been through extensive literature searches undertaken as part of the systematic reviews reported briefly in the Additional material at the end of the protocol (see Appendix 1 in Additional file 1 for a list of studies).

The following databases were searched by two independent reviewers to identify potential collaborators: MEDLINE, MEDLINE in-process and other nonindexed citations, Embase, Cumulative Index to Nursing and Allied Health Literature, *Cochrane Database of Systematic Reviews*, Database of Abstracts of Reviews of Effects, Health Technology Assessment Database, Cochrane Central Register of Controlled Trials, Thomson Reuters Conference Proceedings Citation Index-Science and Literatura Latinoamericana y del Caribe en Ciencias de la Salud. The reference lists of relevant systematic reviews and included articles were reviewed for further relevant studies. Published and unpublished studies were sought, and no language restrictions were applied. Non-English-language studies were translated into English. (See Appendix 2 in Additional file 2 for a sample search that we conducted.)

Further analysis of the initial literature searches will be undertaken to identify any published cohorts of FNP patients that may have been excluded from the reviews because a CDR or serum marker was not tested, yet could provide the information essential to being included in the IPD study. In addition to this, open calls for participation have been made via the International Society for Paediatric Oncology Supportive Care Group, the University of York Centre for Reviews and Dissemination website (http://www.york.ac.uk/inst/crd/projects/picnicc_patient.htm), presentations at relevant UK and international conferences, and via the Oncopedia web community of paediatric oncologists (https://www.cure4kids.org/ums/home/index.php?location=%2Fums%2Foncopedia%2F).

Core data set and variables

This IPD meta-analysis will develop a risk stratification model to predict which children and young people have a low risk of adverse outcomes during an episode of FNP. The predictor variables and adverse outcomes sought have been based on our systematic reviews of aggregate data, in which exploratory analysis showed that age, malignant disease state, clinical assessment of circulatory and respiratory compromise, higher body temperatures and bone marrow suppression had explanatory value and reflected clinical experience of the paediatric oncologists.

The following predictor variables are divided into 'essential' and 'desirable' items and can be categorised as (1) patient-related, episode-related clinical variables and (2) patient-related, episode-related laboratory variables:

1. Age
2. Underlying tumour type
3. Marrow involvement and/or remission status
4. Chemotherapy type and time elapsed since last cycle
5. Presence of central venous line
6. Inpatient or outpatient at onset of episode
7. Maximum temperature
8. Antibiotic therapy used
9. Respiratory rate (or compromise)
10. Circulatory parameters (or compromise)
11. Severe mucositis
12. Global assessment of illness severity
13. Haemoglobin
14. Platelet count
15. White blood cell count
16. Neutrophil count
17. Monocyte count
18. CRP
19. Procalcitonin
20. IL-6
21. IL-8

The following are outcomes of primary interest from each episode:

1. Death
2. ICU admission
3. Need for moderate organ support (fluid bolus, oxygen)
4. Clinically documented infections
5. Microbiologically documented infections

Two or more of these outcome measures should be provided for more than 90% of episodes.

If available, we will also collect data on the following:

1. Duration of fever
2. Duration of admission

An example of the initial survey of data available from collaborators is provided in Appendix 3 in Additional file 3. An *a priori* mapping schema linking microbiological and clinical outcome variables into a unified description of 'severe' and 'nonsevere' infections has been developed to assist with unifying outcome definitions.

Providing data

Anonymised deidentified data

Data sets should be anonymised (that is, have all directly identifiable material removed, such as name, address, postal code, record number). A patient identification number should be provided to facilitate communication and data queries. For the purposes of this report, the age of the patient (an indirect identifier) is essential and should be provided [57].

Data format

The data will be accepted by the PICNICC Collaborative in any electronic format, but ideally a 'flat' spreadsheet format (such as Microsoft Excel; Microsoft Corp, Redmond, WA, USA) will be most useful, with one episode per row and variables listed in columns. Each patient should be assigned an in-cohort unique identifier (such as a simple number 1, 2 ... *n*) to highlight repeated episodes in the same patient. A suggestion for coding the variables is provided in Appendix 4 in Additional file 4 and a sample flat file is available on request.

Transfer of data

The data should be transferred to a secure password-protected web server or by pretty good privacy-encrypted email. This permits a secure and identifiable connection to the University of York servers and minimises the possibility of data loss.

Data checking

Simple checks of data integrity will be undertaken prior to analysis. These checks will include sense checking of data (for example, impossibly low presenting temperatures, such as less than 30°C or for second episodes of FNP where the outcome of the first was death), clarifying missing data (that is, ensuring missing data is recorded as 'missing' rather than 'zero') and calculating simple descriptive statistics of 'essential' elements to assess for 'outlier' studies (for example, age, sex, number of episodes per person). Any problems or inconsistencies flagged during these procedures will be discussed with the individual responsible for each study and amended as appropriate by consensus.

Ethical and regulatory considerations

This IPD protocol has been approved in the United Kingdom by the University of York Health Services Research Ethics and Research Governance Committee. Each clinician member of the PICNICC Collaborative is advised to seek country-specific advice regarding the regulations which apply to data shared in this study.

Plan of investigation

Method of analysis

The primary method of analysis for the PICNICC study will be the use of multivariable logistic regression modelling. There are a series of different analytical techniques that can be used to produce rules, including multivariable regression analysis, classification and regression tree (CART) models, discriminant analysis and neural networks. There is no clear evidence that one method is superior to any other [58], and, as multivariable logistic models have the widest clinical understanding and applicability, this method has been selected.

In the primary analysis, data used will be from the first recorded episode for each patient to predict an absence of adverse outcomes due to the individual episode (that is, death, intensive care requirement, medical complication, bacteraemia or other significant bacterial infection). Following the primary analysis, outcome data and predictor variables from subsequent episodes will be analysed to assess the independence or otherwise of these data, and this information will also be included using an appropriate model.

Prospective and retrospective cohorts will be considered separately in the initial analyses on the basis of the hypothesis that there will be a clinically important difference between the two types of studies. If no difference is found, then the data set will be examined as a whole. The prognostic importance of individual variables, both unadjusted and adjusted for other variables (the latter to summarise independent prognostic value), will be summarised for each study.

Assessment of study and data quality

There is very little advice in the literature for assessing the quality of prognostic studies. Altman and Lyman presented suitable criteria that those initiating a primary prognostic study should consider [59], and they suggested that every effort should be made to limit potential biases and to emulate the design standards of a clinical trial. Ideally, the data should be collected prospectively, with little missing data for predictors or outcomes and with predefined hypothesises. We will use these guidelines and those published by Hayden *et al.* [60] to help inform the quality of the IPD obtained. For example, an assessment will be made of the proportion of missing data and the completeness of follow-up. The influence of any studies considered problematic (for example, those with large amounts of missing data or a great deal of incomplete follow-up) on the prediction model will then be considered, resulting in either their exclusion or in sensitivity analyses comparing model estimates when they are included or excluded.

Model development

The model will initially incorporate the simplest predictor variables (malignant diagnosis, age, time since chemotherapy, and maximum recorded temperature) before standard additional variables (such as clinical

assessments of compromise, inpatient or outpatient status, white blood cell counts or other haematological parameters) are added. Further specialist tests (for example, CRP and IL-6 levels) will be added. The type of antibiotic therapy used will always be incorporated into the model as a categorical variable. Potential sources of heterogeneity (for example, in effects of particular variables across studies) will be incorporated as random effects as appropriate. The models will be assessed for improvement in fit by using an information criterion (for example, Akaike's information criterion) with a P-value of < 0.15 used for inclusion. We will use a 15% rather than a 5% level, as we feel this is more conservative and will limit the chance of missing important covariates. At the stage of deciding our final model, however, we will check that the model's predictive accuracy (discriminatory ability) is improved by the inclusion of variables whose significance is between 5% and 15%. If predictive accuracy is not improved, then these variables will be removed.

This approach (of adding specialist tests only after considering the simpler tests) maximises the utility of a model by ensuring that, if extra tests with their additional costs are required, they will add considerable predictive power to existing simpler variables [61]. We will use bootstrapping and shrinkage to adjust for potential overoptimism (bias) in parametric estimates and trends.

Continuous candidate variables will be assessed using the best fitting functional form considering appropriate transformations or fractional polynomials (also assessed using an information criterion) as suggested by previous evidence. Missing data will be examined to define the nature of the 'missingness'. If they are missing at random, then multiple imputation techniques will be used to address these gaps utilising all the other available data [50,51]. The results of these analyses will be compared with a complete case analysis. We will conduct an analysis comparing the new model that we develop with other validated models, for example, that of Santolaya *et al.* [62]. This will provide an opportunity to test these CDRs against data from other geographical areas.

We acknowledge that there may be unforeseen challenges caused by the variations in the data formats available from the different studies. Therefore, we acknowledge that establishing the definitive analysis plan will be an iterative process and may even demand novel methodological developments (see 'Further research opportunities arising from PICNICC' section).

Assessing model performance

An important goal of a prediction model is to classify patients into risk groups. The developed model will produce a risk score for each individual that is based on the patient's own predictor values. We will then use a cutoff value to decide when a risk score is high (such that we predict an adverse outcome) and when it is low (such that we predict a good outcome); this will be our CDR. The calibration of the model will be assessed by classifying children into deciles ordered by predicted risk and considering the agreement between the mean predicted risk and the observed events in each decile. The derived CDR will be cross-validated by comparing the classification of each patient with his or her actual outcome, thus allowing an estimate of the sensitivity and specificity of the prediction model. Next, by varying the chosen cutoff level, we will be able to produce a receiver operating characteristic curve (ROC) summarising the sensitivity and specificity of the predictive rule across the range of cutoffs. The overall discriminatory ability will be summarised as the area under the ROC (AUC ROC) with the 95% confidence interval. The most suitable cutoff level can then also be detected.

Each predictive model will be tested by checking how it performs against the data from all but one of the studies in turn (cross-validation of intrinsic prognostic performance) [63] and by using the bootstrap procedure [64]. This will adjust for overoptimism in the estimation of model performance due to validation in the same data set that was used to develop the model itself.

The improvement in model performance by adding prognostic factors will be assessed by net reclassification improvement. By analysing the difference among the prognostic factors, a shrinkage factor will be calculated and the model will be corrected by this shrinkage factor. Note also that clustering of patients within studies will be accounted for in the model framework.

Validation in new data

We will compare the predicted and observed event rates to assess calibration (as described above) and the AUC ROC to assess discriminatory ability. If new data become available after the formation of the PICNICC Collaborative, they will provide an excellent test bed for the newly proposed model. Such an analysis is outside the initial scope of this project. We will update the model if it shows poor performance to adjust it to the new situation by recalibration or revision methods, depending on discrimination performance. Simple diagnostic test accuracy measures (such as positive and negative predictive values) will be computed for a hypothetical population (with its particular incidence rates) to aid clinical interpretation of the study results that define a low-risk group.

Assessment of publication bias

We do not believe that publication bias will affect the data we obtain. We have sought to retrieve full data from the studies and so have sidestepped many of the

problems of reporting bias. There may remain issues of different outcome collection and different outcome assessment methods, but these will not have been biased by collection and analysis of the predictive data. We have tried to avoid publication bias by making open calls for data which has been collected but not yet published, and we have probably secured three such data sets for analysis. This may be too few to undertake a formal assessment of the difference between the published and unpublished sources. We are also using the data for a purpose different from that used by the original data collectors. We are developing a prediction model, whereas the original researchers are interested only in the prognostic effect of particular variables. Furthermore, by obtaining the IPD, we have obtained outcomes and variables not reported by the original data collectors in any publication. However, to check whether our collection of studies may be affected by publication bias, we will display a funnel plot for each of the variables included in the final model to see whether there is asymmetry (that is, potential publication bias). We will use guidelines for assessing asymmetry recently published in *BMJ* [65].

Publication policy

The main results of the meta-analysis will be published and presented under the PICNICC name, with PICNICC comprising groups supplying data for analysis as well as its advisory group. Any subsequent technical papers which describe innovations in the methodologies used in the meta-analysis will acknowledge the PICNICC Collaborative as the source of the data. The PICNICC Collaborative will disseminate the findings of its research widely at academic conferences and in journal publications, on the University of York website and in lay summaries of the research.

Discussion

Status of the project

Currently, the PICNICC Collaborative has completed study identification and invitation and has collected data derived from 23 data sets from 12 countries, including the Europe-wide European Organisation for Research and Treatment of Cancer studies. No data analyses have yet been undertaken. The opportunity to include data sets for the derivation of a new PICNICC CDR have now closed, but approaches may be made to the authors for consideration of inclusion of further data sets in subsequent validation testing or further refinements of the initiative.

Further research opportunities arising from PICNICC

It is hoped that collaborations developed through the PICNICC project may also lead to a series of international studies to improve patients' experiences and outcomes with regard to infectious complications in cancer. One obvious follow-up study might be the use of the newly derived model in a RCT of alternative management approaches (for example, ambulatory oral antibiotics vs inpatient intravenous antibiotics). Other studies may include the investigation of genetic polymorphisms in determining the outcomes of infectious episodes; the prediction of specific infections which may require different management approaches, such as antibiotic-resistant bacteraemia; or the prediction of the risk of an episode of FNP.

The PICNICC Collaborative will provide data that will prove invaluable in the development of the methodology of IPD meta-analysis for risk prediction. This developmental work, which will be essential to developing the best possible model in PICNICC, is outside the core clinical questions set for the PICNICC Collaborative and will be undertaken as a series of linked projects. The problems to be addressed in developing the methodologies will depend on the nature of the data sets obtained. They may address issues regarding the analysis of missing data, the use of different imputation models, the modelling of multiple-episode data, the relative merits of prospective and retrospectively collected information, the use of alternative modelling techniques (such as CART, structured equation modelling, Bayesian techniques or neural networks), the comparison of episodic and patient-centred analyses and the use of categorical outcome variables. A short methodological protocol will be developed for each methodological investigation prior to commencement.

Additional material

Additional file 1: Appendix 1: Potential IPD Datasets.

Additional file 2: Appendix 2: Search Strategy.

Additional file 3: Data collection survey.

Additional file 4: Suggested coding structure.

Abbreviations

95% CI: 95% confidence interval; 95% CrI; 95% credible interval: CART: classification and regression tree; CDR: clinical decision rule; CRP: C-reactive protein; FNP: febrile neutropenia; IL: interleukin; IPD: individual participant data; RCT: randomised controlled trial.

Acknowledgements

Neil Ranasinghe, Sally Amos and Susan Hay have all specifically helped in their capacity as parents of children who have experienced childhood cancer. A large number of other parents have been highly supportive of the work. The development of the original MRC fellowship proposal was funded through a grant from Candlelighters, the Yorkshire Children's Cancer Charity.
Role of funding source
This research has been funded as part of a Research Training Fellowship by the Medical Research Council (MRC) UK (RSP) and travel expenses for development meetings given to the other authors. The funders reviewed

the overall grant submission but had no influence on question, design or undertaking the research. They had no influence on the decision to submit the manuscript for publication beyond the stipulation that the results of the research must be made publicly accessible within 6 months of final publication and be available through PubMed Central (UK).

The collaboration

The PICNICC collaboration is composed of those who have contributed data, as well as patients and their caregivers through their participation, significantly developed the project. The following are the current PICNICC members: Neil Ranasinghe, Sally Amos and Susan Hay (parent/carer partners); the authors of this article (RSP, AJS, RDR, JCC, SVP and LAS); Roland Ammann (Switzerland); Felix Niggli (Switzerland); David Nadal (Switzerland); Ian Hann (Ireland); Thomas Kühne (Switzerland); Lillian Sung (Canada); Thomas Lehrnbecher (Germany); Arne Simon (Germany); Robert Klaassen (Canada); Hana Hakim (USA); Sarah Alexander (Canada); Karin Meidema and Wim JE Tissing (Netherlands); Julia Chisholm and Rachel Dommett (UK); Elio Castagnola (Italy); Pamela Silva (Chile); Juan Tordecilla (Chile); Maria Spassova (Bulgaria); Glen Stryjewski (USA); Gulsun Tezcan (Turkey); Lidija Kitanovski (Slovenia); and Marianne Paesmann and J Peter Donnelly (EORTC).

Author details

[1]Centre for Reviews and Dissemination, Alcuin College, University of York, York, YO10 5DD, UK [2]Department of Health Sciences, 22-28 Princess Road West, University of Leicester, Leicester, LE1 6TP, UK [3]Department of Public Health, Epidemiology & Biostatistics, The Public Health Building, University of Birmingham, Birmingham, B15 2TT, UK [4]Department of Paediatric Oncology, Royal Marsden NHS Foundation Trust, Sutton, Surrey, SM2 5PT, UK [5]Regional Department of Paediatric Haematology/Oncology, Leeds Teaching Hospitals Trust, Leeds, LS1 3EX, UK

Authors' contributions

RSP conceived, developed and drafted the protocol and the systematic reviews referenced herein. He was supported in the clinical details by JCC and SVP. Statistical advice and support were provided by AJS and RDR. The project was overseen and developed with LAS, who also contributed extensively to the practical processes of undertaking an IPD meta-analysis. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 12 November 2011 Accepted: 9 February 2012
Published: 9 February 2012

References

1. Pritchard-Jones K, Kaatsch P, Steliarova-Foucher E, Stiller CA, Coebergh JWW: **Cancer in children and adolescents in Europe: developments over 20 years and future challenges.** *Eur J Cancer* 2006, **42**:2183-2190.
2. Freycon F, Trombert-Paviot B, Casagranda L, Bertrand Y, Plantaz D, Marec-Bérard P: **Trends in treatment-related deaths (TRDs) in childhood cancer and leukemia over time: a follow-up of patients included in the Childhood Cancer Registry of the Rhône-Alpes region in France (ARCERRA).** *Pediatr Blood Cancer* 2008, **50**:1213-1220.
3. Hargrave DR, Hann IM: **Progressive reduction in treatment-related deaths in Medical Research Council childhood lymphoblastic leukaemia trials from 1980 to 1997 (UKALL VIII, X and XI).** *Br J Haematol* 2001, **112**:293-299.
4. Hann I, Viscoli C, Paesmans M, Gaya H, Glauser M: **A comparison of outcome from febrile neutropenic episodes in children compared with adults: results from four EORTC studies. International Antimicrobial Therapy Cooperative Group (IATCG) of the European Organization for Research and Treatment of Cancer (EORTC).** *Br J Haematol* 1997, **99**:580-588.
5. Baorto EP, Aquino VM, Mullen CA, Buchanan GR, DeBaun MR: **Clinical parameters associated with low bacteremia risk in 1100 pediatric oncology patients with fever and neutropenia.** *Cancer* 2001, **92**:909-913.
6. Lucas KG, Brown AE, Armstrong D, Chapman D, Heller G: **The identification of febrile, neutropenic children with neoplastic disease at low risk for bacteremia and complications of sepsis.** *Cancer* 1996, **77**:791-798.
7. Duncan C, Chisholm JC, Freeman S, Riley U, Sharland M, Pritchard-Jones K: **A prospective study of admissions for febrile neutropenia in secondary paediatric units in South East England.** *Pediatr Blood Cancer* 2007, **49**:678-681.
8. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG: **Prognosis and prognostic research: what, why, and how?** *BMJ* 2009, **338**:b375.
9. Royston P, Moons KGM, Altman DG, Vergouwe Y: **Prognosis and prognostic research: developing a prognostic model.** *BMJ* 2009, **338**:b604.
10. Altman DG, Vergouwe Y, Royston P, Moons KGM: **Prognosis and prognostic research: validating a prognostic model.** *BMJ* 2009, **338**:b605.
11. Teuffel O, Amir E, Alibhai SMH, Beyene J, Sung L: **Cost-effectiveness of outpatient management for febrile neutropenia in children with cancer.** *Pediatrics* 2011, **127**:e279-e286.
12. Phillips B, Selwood K, Lane SM, Skinner R, Gibson F, Chisholm JC: **Variation in policies for the management of febrile neutropenia in United Kingdom Children's Cancer Study Group centres.** *Arch Dis Child* 2007, **92**:495-498.
13. Boragina M, Patel H: **Management of febrile neutropenia in pediatric oncology patients: a Canadian survey.** *Pediatr Blood Cancer* 2007, **48**:521-526.
14. Chamberlain JD, Smibert E, Skeen J, Alvaro F: **Prospective audit of treatment of paediatric febrile neutropenia in Australasia.** *J Paediatr Child Health* 2005, **41**:598-603.
15. Stabell N, Nordal E, Stensvold E, Gammelsrud KW, Lund B, Taxt A, Buhring F, Greve-Isdahl M, Fornebo HP, Simonsen GS, Klingenberg C: **Febrile neutropenia in children with cancer: a retrospective Norwegian multicentre study of clinical and microbiological outcome.** *Scand J Infect Dis* 2008, **40**:301-307.
16. Orudjev E, Lange BJ: **Evolving concepts of management of febrile neutropenia in children with cancer.** *Med Pediatr Oncol* 2002, **39**:77-85.
17. Ammann RA, Simon A, de Bont ESJM: **Low risk episodes of fever and neutropenia in pediatric oncology: is outpatient oral antibiotic therapy the new gold standard of care?** *Pediatr Blood Cancer* 2005, **45**:244-247.
18. Chisholm JC, Dommett R: **The evolution towards ambulatory and day-case management of febrile neutropenia.** *Br J Haematol* 2006, **135**:3-16.
19. Phillips B, Wade R, Stewart LA, Sutton AJ: **Systematic review and meta-analysis of the discriminatory performance of risk prediction rules in febrile neutropaenic episodes in children and young people.** *Eur J Cancer* 2010, **46**:2950-2964.
20. Phillips B, Wade R, Westwood M, Riley R, Sutton AJ: **Systematic review and meta-analysis of the value of clinical features to exclude radiographic pneumonia in febrile neutropenic episodes in children and young people.** *J Paediatr Child Health* .
21. Rackoff WR, Gonin R, Robinson C, Kreissman SG, Breitfeld PB: **Predicting the risk of bacteremia in childen with fever and neutropenia.** *J Clin Oncol* 1996, **14**:919-924.
22. Santolaya ME, Alvarez AM, Becker A, Cofré J, Enríquez N, O'Ryan M, Payá E, Pilorget J, Salgado C, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Prospective, multicenter evaluation of risk factors associated with invasive bacterial infection in children with cancer, neutropenia, and fever.** *J Clin Oncol* 2001, **19**:3415-3421.
23. Ammann RA, Bodmer N, Hirt A, Niggli FK, Nadal D, Simon A, Ozsahin H, Kontny U, Kühne T, Popovic MB, Lüthy AR, Aebi C: **Predicting adverse events in children with fever and chemotherapy-induced neutropenia: the prospective multicenter SPOG 2003 FN study.** *J Clin Oncol* 2010, **28**:2008-2014.
24. Stewart LA, Tierney JF: **To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data.** *Eval Health Prof* 2002, **25**:76-97.
25. Riley RD, Lambert PC, Abo-Zaid G: **Meta-analysis of individual participant data: rationale, conduct, and reporting.** *BMJ* 2010, **340**:c221.
26. Khan KS, Bachmann LM, ter Riet G: **Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests.** *Eur J Obstet Gynecol Reprod Biol* 2003, **108**:121-125.
27. Broeze K, Opmeer B, Bachmann L, Broekmans F, Bossuyt P, Coppus S, Johnson N, Khan K, ter Riet G, van der Veen F, van Wely M, Mol B: **Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine.** *BMC Med Res Methodol* 2009, **9**:22.
28. Altman DG, Riley RD: **Primer: an evidence-based approach to prognostic markers.** *Nat Clin Pract Oncol* 2005, **2**:466-472.

29. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM: **Systematic reviews of diagnostic test accuracy.** *Ann Intern Med* 2008, **149**:889-897.
30. Trivella M, Pezzella F, Pastorino U, Harris AL, Altman DG: **Microvessel density as a prognostic factor in non-small-cell lung carcinoma: a meta-analysis of individual patient data.** *Lancet Oncol* 2007, **8**:488-499.
31. Royston P, Parmar MK, Sylvester R: **Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer.** *Stat Med* 2004, **23**:907-926.
32. Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, Mant D, McManus RJ, Holder R, Deeks J, Fletcher K, Qume M, Sohanpal S, Sanders S, Hobbs FD: **Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care.** *Health Technol Assess* 2009, **13**:1-207, iii.
33. Simon R, Altman DG: **Statistical aspects of prognostic factor studies in oncology.** *Br J Cancer* 1994, **69**:979-985.
34. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD: **Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets.** *Stat Med* 2000, **19**:1059-1079.
35. Heney D, Lewis IJ, Evans SW, Banks R, Bailey CC, Whicher JT: **Interleukin-6 and its relationship to C-reactive protein and fever in children with febrile neutropenia.** *J Infect Dis* 1992, **165**:886-890.
36. Kitanovski L, Jazbec J, Hojker S, Gubina M, Derganc M: **Diagnostic accuracy of procalcitonin and interleukin-6 values for predicting bacteremia and clinical sepsis in febrile neutropenic children with cancer.** *Eur J Clin Microbiol Infect Dis* 2006, **25**:413-415.
37. Soker M, Çolpan L, Ece A, Devecioğlu C, Haspolat K: **Serum levels of IL-1β, sIL-2R, IL-6, IL-8, and TNF-α in febrile children with cancer and neutropenia.** *Med Oncol* 2001, **18**:51-57.
38. Ammann RA, Hirt A, Lüthy AR, Aebi C: **Identification of children presenting with fever in chemotherapy-induced neutropenia at low risk for severe bacterial infection.** *Med Pediatr Oncol* 2003, **41**:436-443.
39. Ammann RA, Hirt A, Lüthy AR, Aebi C: **Predicting bacteremia in children with fever and chemotherapy-induced neutropenia.** *Pediatr Infect Dis J* 2004, **23**:61-67.
40. Hann I, Viscoli C, Paesmans M, Gaya H, Glauser M: **A comparison of outcome from febrile neutropenic episodes in children compared with adults: results from four EORTC studies. International Antimicrobial Therapy Cooperative Group (IATCG) of the European Organization for Research and Treatment of Cancer (EORTC).** *Br J Haematol* 1997, **99**:580-588.
41. Rondinelli PIP, Ribeiro KdCB, de Camargo B: **A proposed score for predicting severe infection complications in children with chemotherapy-induced febrile neutropenia.** *J Pediatr Hematol Oncol* 2006, **28**:665-670.
42. Alexander SW, Wade KC, Hibberd PL, Parsons SK: **Evaluation of risk prediction criteria for episodes of febrile neutropenia in children with cancer.** *J Pediatr Hematol Oncol* 2002, **24**:38-42.
43. Spasova MI, Terzieva DD, Tzvetkova TZ, Stoyanova AA, Mumdzhiev IN, Yanev IB, Genev ED: **Interleukin-6, interleukin-8, interleukin-10, and C-reactive protein in febrile neutropenia in children with malignant diseases.** *Folia Medica (Plovdiv)* 2005, **47**:46-52.
44. Klaassen RJ, Goodman TR, Pham B, Doyle JJ: **"Low-risk" prediction rule for pediatric oncology patients presenting with fever and neutropenia.** *J Clin Oncol* 2000, **18**:1012-1019.
45. West DC, Marcin JP, Mawis R, He J, Nagle A, Dimand R: **Children with cancer, fever, and treatment-induced neutropenia: risk factors associated with illness requiring the administration of critical care therapies.** *Pediatr Emerg Care* 2004, **20**:79-84.
46. Oude Nijhuis C, Kamps WA, Daenen SMG, Gietema JA, van der Graaf WTA, Groen HJM, Vellenga E, Ten Vergert EM, Vermeulen KM, de Vries-Hospers HG, de Bont ESJM: **Feasibility of withholding antibiotics in selected febrile neutropenic cancer patients.** *J Clin Oncol* 2005, **23**:7437-7444.
47. Royston P, Sauerbrei W: *Multivariable Model-Building: A Pragmatic Approach to Regression Anaylsis Based on Fractional Polynomials for Modelling Continuous Variables* Chichester, UK: John Wiley & Sons; 2008.
48. Diepold M, Noellke P, Duffner U, Kontny U, Berner R: **Performance of interleukin-6 and interleukin-8 serum levels in pediatric oncology patients with neutropenia and fever for the assessment of low-risk.** *BMC Infect Dis* 2008, **8**:28.
49. Riikonen P, Saarinen UM, Teppo AM, Metsärinne K, Fyhrquist F, Jalanko H: **Cytokine and acute-phase reactant levels in serum of children with cancer admitted for fever and neutropenia.** *J Infect Dis* 1992, **166**:432-436.
50. Marshall A, Altman D, Royston P, Holder R: **Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study.** *BMC Med Res Methodol* 2010, **10**:7.
51. Marshall A, Altman D, Holder R, Royston P: **Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines.** *BMC Med Res Methodol* 2009, **9**:57.
52. Partridge N, Scadding J: **The James Lind Alliance: patients and clinicians should jointly identify their priorities for clinical trials.** *Lancet* 2004, **364**:1923-1924.
53. Tallon D, Chard J, Dieppe P: **Relation between agendas of the research community and the research consumer.** *Lancet* 2000, **355**:2037-2040.
54. Fern L, Davies S, Eden T, Feltbower R, Grant R, Hawkins M, Lewis I, Loucaides E, Rowntree C, Stenning S, Whelan J: **Rates of inclusion of teenagers and young adults in England into National Cancer Research Network clinical trials: report from the National Cancer Research Institute (NCRI) Teenage and Young Adult Clinical Studies Development Group.** *Br J Cancer* 2008, **99**:1967-1974.
55. Klastersky J, Paesmans M, Rubenstein EB, Boyer M, Elting L, Feld R, Gallagher J, Herrstedt J, Rapoport B, Rolston K, Talcott J: **The Multinational Association for Supportive Care in Cancer Risk Index: a multinational scoring system for identifying low-risk febrile neutropenic cancer patients.** *J Clin Oncol* 2000, **18**:3038-3051.
56. Kris MG, Hesketh PJ, Herrstedt J, Rittenberg C, Einhorn L, Grunberg SM, Koeller J, Oliver I, Borjeson S, Ballatori E: **Consensus proposals for the prevention of acute and delayed vomiting and nausea following high-emetic-risk chemotherapy.** *Support Care Cancer* 2005, **13**:85-96.
57. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG: **Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers.** *BMJ* 2010, **340**:c181.
58. Riley RD, Heney D, Jones DR, Sutton AJ, Lambert PC, Abrams KR, Young B, Wailoo AJ, Burchill SA: **A systematic review of molecular and biological tumor markers in neuroblastoma.** *Clin Cancer Res* 2004, **10**:4-12.
59. Altman DG, Lyman GH: **Methodological challenges in the evaluation of prognostic factors in breast cancer.** *Breast Cancer Res Treat* 1998, **52**:289-303.
60. Hayden JA, Côté P, Bombardier C: **Evaluation of the quality of prognosis studies in systematic reviews.** *Ann Intern Med* 2006, **144**:427-437.
61. Altman DG, Lausen B, Sauerbrei W, Schumacher M: **Dangers of using "optimal" cutpoints in the evaluation of prognostic factors.** *J Natl Cancer Inst* 1994, **86**:829-835.
62. Santolaya ME, Alvarez AM, Avilés CL, Becker A, Cofré J, Enríquez N, O'Ryan M, Payá E, Salgado C, Silva P, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Prospective evaluation of a model of prediction of invasive bacterial infection risk among children with cancer, fever, and neutropenia.** *Clin Infect Dis* 2002, **35**:678-683.
63. Altman DG, Royston P: **What do we mean by validating a prognostic model?** *Stat Med* 2000, **19**:453-473.
64. Harrell FE Jr, Lee KL, Mark DB: **Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.** *Stat Med* 1996, **15**:361-387.
65. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, Carpenter J, Rücker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, Higgins JPT: **Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials.** *BMJ* 2011, **343**:d4002.

doi:10.1186/2046-4053-1-8
Cite this article as: Phillips *et al.*: Predicting infectious complications in neutropenic children and young people with cancer (IPD protocol). *Systematic Reviews* 2012 **1**:8.

Appendix 18. Mapping for severe infection

Severe infections:

- Isolation in the blood of any gram-negative organism
- Isolation in the blood of selected gram-positive infections (eg Staph. Aureus, Strep viridans)
- Any significant bacterial isolate from CSF
- Any significant bacterial isolate from lower respiratory tract secretions
- Probable / proven invasive fungal infection according to EORTC definitions Isolation of bacteria from a deep soft-tissue or bone infection
- Mycobacteria
- Malarial parasites
- Serum isolation of significant quantities of adenovirus,VZV, HSV or CMV (as determined by PCR copy number)
- Respiratory virus isolation in the setting of stem cell procedures
- Additionally, clinical site information without definite microbiological confirmation may also indicate:
- Cellulitis or CVC tunnel infection
- Bone / deep soft tissue infection
- Finally, any admission to a critical care facility is considered to be related to severe infection, unless designated clearly as for an alternative cause.

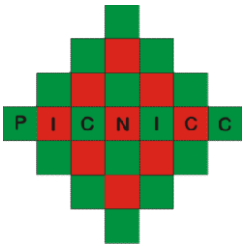
Non-severe infections:

- Isolation of most gram-positive organisms, for example coagulase negative staphylococci, or diptheroids
- Isolated bacterial growth in urine, or scanty yeast
- Respiratory viruses from respiratory secretions without hypoxia
- Superficial skin infections (eg exit site infection without significant cellulitis)
- Viral gastroenteritis
- Additionally, clinical site information without definite microbiological confirmation may also indicate:
- Otitis media
- Tonsillitis
- Superficial skin infection

Appendix 19. Collectable data screening questionnaire

The PICNICC (Predicting Infectious Complications of Neutropenic sepsis In Children with Cancer) Collaborative

Optimizing risk predictive strategies in febrile neutropenic episodes in children and young people undergoing treatment for malignant disease



Checklist of potential data items for the PICNICC study

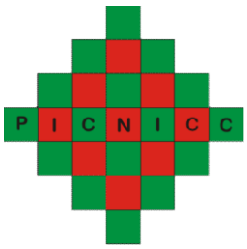
Centre Name/Location	
Principle clinician	
Contact for data queries (if known)	

Please tick (√) each item as to if your database could or could not provide the information.

Item	Can provide	Cannot provide	Unsure
Predictors			
Age			
Underlying tumour type			
Marrow involvement/remission status			
Chemotherapy type and date of last cycle			
Presence of central venous line			
In-patient or out-patient at onset of episode			
Maximum temperature			
Respiratory rate (or compromise)			
Circulatory (or compromise)			
Severe mucositis			
Global assessment of illness severity			
Antibiotic therapy used			

Item	Can provide	Cannot provide	Unsure
Haemoglobin			
Platelet count			
White cell count			
Neutrophil count			
Monocyte count			
CRP			
PCT			
IL6			
IL8			
Outcome			
Death			
Duration of intensive care admission			
Need for moderate organ support (fluid bolus, oxygen)			
Clinically documented infections			
Microbiologically documented infections			
Duration of fever			
Duration of admission			
Other			
Date of episode of FNP			
Age at episode of FNP			

Please return completed form to crd-picnicc@york.ac.uk



Appendix 20. Coding structure.

Age

Actual age at start of episode, in months
999 = unknown

Tumour type

(Diagnosis under treatment, coded as:)
Acute lymphoblastic leukaemia
Acute myeloid leukaemia
Other leukaemia
Hodgkins lymphoma
Non-Hodgkins lymphoma
Low-grade brain tumour (I-II)
High-grade brain tumour (III-IV)
'High risk' neuroblastoma
Other neuroblastoma
Retinoblastoma
Wilm's tumour
Other renal tumour
Hepatoblastoma
Other liver tumor
Osteosarcoma
Ewing's sarcoma
Rhabdomyosarcoma
Other sarcoma
Germ cell/gonadal neoplasm
Carcinoma/melanoma
LCH
Other Please provide separate details of any 'other' diagnoses
99 Unknown

Relapsed/progressive disease

Is this a relapsed/progressive malignancy:

0 = no
1 = yes
9 = unknown

Marrow involvement

Bone marrow involvement at diagnosis:
0 = no
1 = yes
9 = unknown

Remission status
In remission (leukaemia only) or on-treatment or post-treatment (solid tumours)
0 = no
1 = yes
9 = unknown

Chemotherapy type
Specify the most recent chemotherapy cycle (in words/by acronym or explicit numerical coding)
This will require a description of each chemotherapy protocol included from each study.
Specify 'Unknown' if unknown

Time from last chemotherapy cycle
Time (in days) since the start of most recent cycle of chemotherapy. For maintenance/prolonged chemotherapy courses, code as 'ongoing' even if temporarily discontinued.

0 = ongoing
1 ….k = time in days
999 = unknown

Presence of central venous line
0 = no
1 = fully implanted (e.g. Port-a-cath)
2 = external tunnelled (e.g. Hickman)
3 = non-tunnelled line (e.g. PICC line or Vascath)
4 = line present, type unknown
9 = unknown if line present or not

In-patient or out-patient at onset of episode
0 = in-patient
1 = out-patient
9 = unknown

Maximum temperature
Maximum recorded temperature at admission. May be parent-reported or clinician-measured.

To be recorded as an absolute value in ^{0}C to one decimal place.
99.9 = unknown

Respiratory assessment
At initial assessment. To be recorded as an absolute value in breaths/min where given.

1 ….k = respiratory rate (breaths/min)

If data are only available on the presence/absence of respiratory compromise:

777 = no compromise
888 = compromised
999 = unknown

Circulatory assessment - HR
At initial assessment. To be recorded as an absolute value of heart rate in beats/min.

1 ….k = pulse rate (beats/min)
999 = unknown/not recorded

Circulatory assessment – BP systolic
At initial assessment. To be recorded as an absolute value mmHg.

1 ….k = systolic blood pressure (mmHg)
999 = unknown/not recorded

Circulatory assessment – BP diastolic
At initial assessment. To be recorded as an absolute value mmHg.

1 ….k = diastolic blood pressure (mmHg)

If data are only available on the presence/absence of **circulatory compromise**, code here:

777 = no compromise
888 = compromised
999 = unknown

Mucositis
At initial assessment. To be recorded as
0 = none
1 = mild
2 = severe
9 = unknown

Global assessment of illness severity
At initial assessment. To be recorded as
0 = well
1 = mildly unwell
2 = severely unwell
9 = unknown

(If an alternative study-specific system is available, please report and specify coding separately.)

Initial antibiotic therapy
The initial antibiotic therapy used should be reported. This can be done by specify the treatment used (in words/by acronym).
(Will require a description of each chemotherapy protocol included from each study)
Specify 'Unknown' if unknown

The PICNICC Secretariat will recode such information as below:

Initial antibiotic therapy coded as the PRODUCT of individual codes
0 = none
2 = oral antibiotics; quinilone
3 = oral antibiotics; penicillin
5 = oral antibiotics; macrolide
7 = IV antibiotics; cephalosporin
11 = IV antibiotics; carbapenem
13 = IV antibiotics; aminoglycoside
17 = IV antibiotics; piperacillin/tazobactam
19 = IV antibiotics; glycopeptide

(as a series of prime numbers, any number which is coded from them will be unique)

Modification of antibiotic therapy
Modification of antibiotic therapy required
0 = no
1 = yes
9 = unknown

Haemoglobin
At initial assessment. In mg/dL
9999 = unknown

Platelet count
At initial assessment. As count $*10^9$
9999 = unknown

White cell count
At initial assessment. As count $*10^6$
9999 = unknown

Neutrophil count
At initial assessment. As count $*10^6$
9999 = unknown

Monocyte count
At initial assessment. As count *10^6
9999 = unknown

CRP
At initial assessment. In mg/dL
9999 = unknown

PCT
At initial assessment. In mg/mL
9999 = unknown

IL6
At initial assessment. In pg/mL
9999 = unknown

IL8
At initial assessment. In pg/mL
9999 = unknown

Appendix 21. Data manipulation SOPs

Data collection & file naming convention

Individuals responsible for the storage of data will send the files in electronic format to the University of York via 'Dropbox'.

This is a secure, personalised web-based system where documents can be shared with specific individuals via any web browser (www.dropbox.com).

Each contact will be emailed a 'share' request to use for this data-drop

Files will be removed from the 'Dropbox' when received

The data file when retrieved will be stored in the \\projectfs\CRDdata\PICNICC\received filestore

For those persons having difficulty with the Dropbox.com system, a secure email service (using PGP key) is also offered.

The file will be converted to a flat spreadsheet format and renamed according to the convention "GroupAcronym-raw" (e.g. SPROG-raw)

Should two different data sources be supplied by the group, they will be suffixed according to the starting year of the data (e.g. SPROG1999-raw and SPROG2008-raw)

This file will be stored in the \\projectfs\CRDdata\PICNICC\raw filestore

The working copy of the file will be named "GroupAcronym-working-yyy-mm-dd" while being cleaned, recoded and tidied up to fit into the standard format

The final version of the data will be named "GroupAcronym-final"

This file will be stored in the \\projectfs\CRDdata\PICNICC\final filestore

The data will be incorporated into the master data file, named “PICNICC-yyyy-mm-dd”

Log files should be named “GroupAcronym-log”

R-logs should be named “GroupAcronym-Rlog-yyy-mm-dd”

This file will be stored in the \\projectfs\CRDdata\PICNICC\logs filestore

Nature of the data supplied

The files containing data should have:

One row of data per episode

One variable per column, first row a header

Linked with a patient identification number, but anonymised (i.e. have all directly identifiable material removed, such as name, address, postcode, medical number).

A coding sheet, detailing the different columns and codes used in the data file should be provided

Any uncertainties will be resolved, in the first instance, by email communication.

Logging interventions upon data supplied

Log files – plain text record of actions – should be opened and retained for each received data file.

It should record any actions (such as renaming) in the following structure, with the oldest records at the top (head) of the document

DATE: Action

e.g. 14-02-2011: Copy saved ‘Valentine-raw’

14-02-2011: Copy saved ‘Valentine-working-2011-02-14’

14-02-2011: Understood code sheet and data file

It should include emails & phone calls sent & received in order to clarify any data queries.

Any actions within the ‘R’ programme will be recorded in its own log, so all that needs to be recorded is that R was used, and a brief note of what was done (for ease of finding again)

Sense checking and resolution of queries

The first action on opening the raw data file should be to rename the file and save as ‘raw’, according to the naming convention.

A copy should then be opened and renamed ‘working’

The first review should be to confirm that the supplied coding sheet and data file correspond, and that any uncertainties are logged.

If uncertainties exist, then a contact (email) should be undertaken with the data supplier clarifying the nature of the problem and requesting a response to the uncertainty.

This should be undertaken by using a common identifier and pointing out clearly the area of uncertainty

A copy of the isolated lines of queried data may be placed as an XLS worksheet in the ‘dropbox’ if further clarification is needed: this should have the initial datafile’s row/column identified, and the queried calls should be highlighted

The columns should be reordered in line with the PICNICC master data file structure

Presence of the essential variables and outcomes should be verified at the 'column' level

The variable names should be altered according to the conventions of the PICNICC master data file structure

The file should be examined for missing data and recoded as `NA` (in accordance with R data structures)

When any uncertainties at this level are resolved, it should be **still** saved as a **'working copy'**

The data checks which are simplest to be undertaken in Excel (as it's prettier) seem to be

age checking (negative, zero and not older than 9125 days (25yr), consistency of patient DOBs, and sensible diagnosis & age relationships)

episodic checking – listing by age and DOB and then admission date and looking for odd/inconsistent elements (>6m in between FNP episodes)

time-since-chemo checking (negative and >42 days), and looking for consistency with other episodes

white cell indices (making sure not 'zero', and that components eg ANC and AMC are not greater than the total WCC)

The data should be imported into R and the data checking procedures run upon the data as an independent data file, unconnected to the rest of the PICNICC dataset. See section 'R-data-checking'

Subsequent queries and their resolutions should be recorded in the 'flat' & updated working copy file

When all queries have been adequately resolved, the data should be saved as a final form. This is the data which will be imported into the master R data file.

R data checking

All actions in R should be logged and the file saved according to the convention on a sessional basis

As data manipulation in R is difficult, data have been exported as CSV files and undergone checking as per non-R files

Appendix 22. Further detailed information on the IPD data

Table 43: Summary of relevant QUADAS criteria for included studies

Author	Data collection	1	2	3	4	6	7	8	9	11
Alexander [134]	Retrospective	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
EORTC Studies[5]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes
Genoa[258]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Hakim [129]	Retrospective	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Kitanovski [163]	Prospective	Yes	No	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes
Klaassen [26]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lehrnbecher [259]	Prospective	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes
PINE [154]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
RetroBern [79]	Retrospective	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Silva[260]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Spasova [160, 261]	Prospective	Yes	No	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes

SPOG groups [144]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Stryjewski [262]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes
Sung [263]	Prospective	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Tezcan [69]	Retrospective	Yes	No	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear
Tissing [148, 158]	Prospective	Yes	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes

1 = representative patients, 2 = clearly described selection criteria, 3 = whole sample, or a random selection of sample, received reference standard, 4 = all patients received same reference standard, 6 = index test described adequately, 7 = reference standard described adequately, 8 = blinded interpretation of index test results, 9 = blinded interpretation of reference standard results, 11 = adequate reference standard.

(5 = index test not part of reference standard: omitted as data permitted unpicking of these, 10 = same clinical data available as in clinical practice: omitted as all data selected were available in clinical practice)

Table 44: Age distribution (in years) across derivation studies

Study ID	Min	Median	Mean	Max
Alexander	1.4	7.3	8.8	25
BaselSPOG	4.1	5.4	6	9.7
BernSPOG	1	6.8	7.6	16
BonnSPOG	1.3	8.1	8.6	18
EORCT-XIV	2.1	19	18	25
EORTC-IX	1	8	9.9	24
EORTC-XI	0.65	13	13	25
EORTC-XII	5.2	21	19	25
Genoa	0.14	5.7	7	20
Hakim	0.2	6	7.8	22
Kitanovski	0.92	6.7	8.5	18
Klaassen	0.48	6.4	7.4	18
Lehrnbecher	0.36	7.5	9	29
PINE	0.15	5.5	6.7	18
RetroBern	0.63	6.8	8	17
Silva	1.2	7.5	7.6	29
Spassova	0.2	7.8	8.5	19
Styjewski	0.42	5	6.8	17
Sung	0.81	6.2	7.9	18
Tezcan	0.25	4.57	6.7	17.65
Tissing	0.52	5.9	7.2	19
ZurichSPOG	1.2	6.8	7.9	17

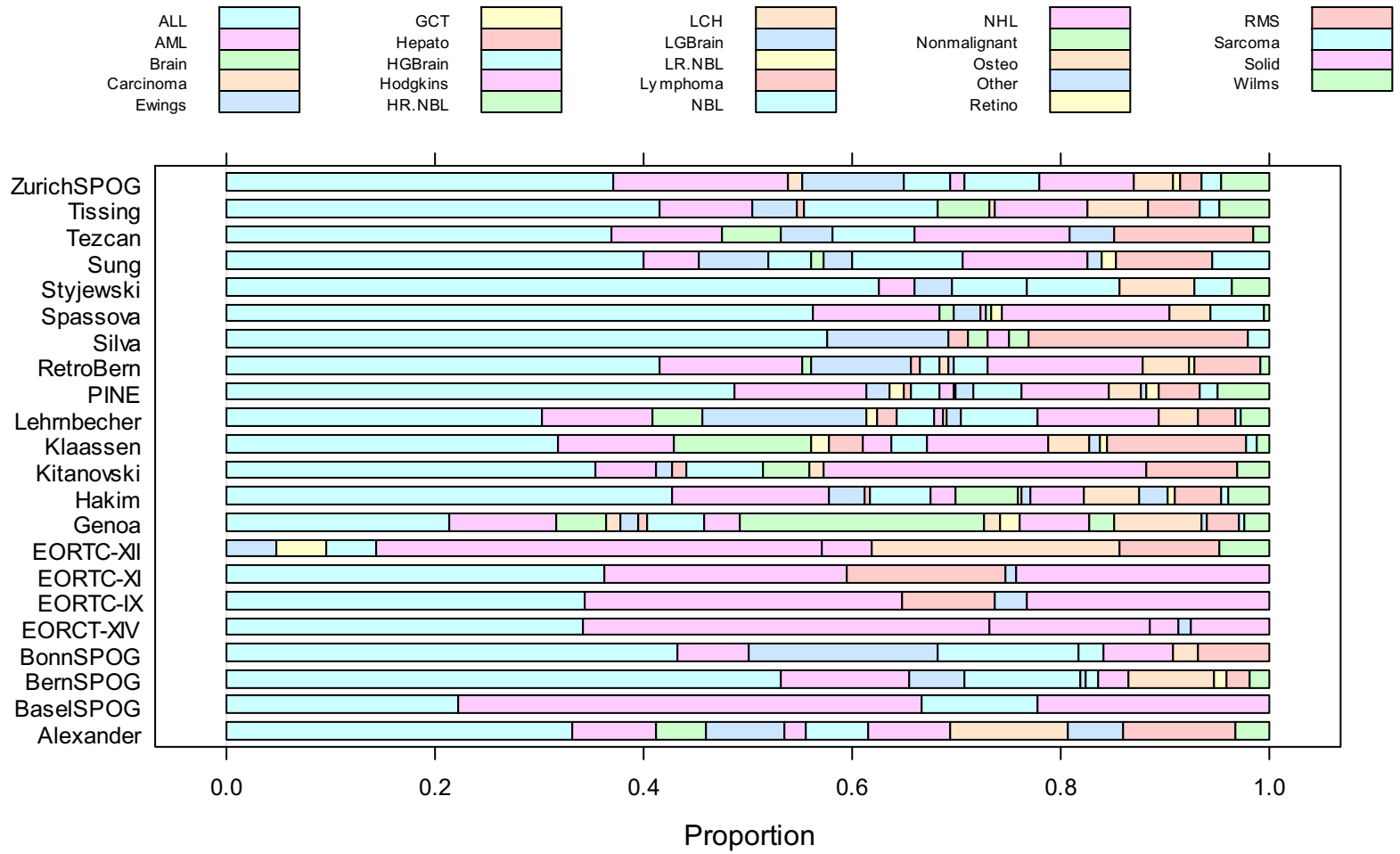


Figure 57: Proportions of diagnoses per episode per study

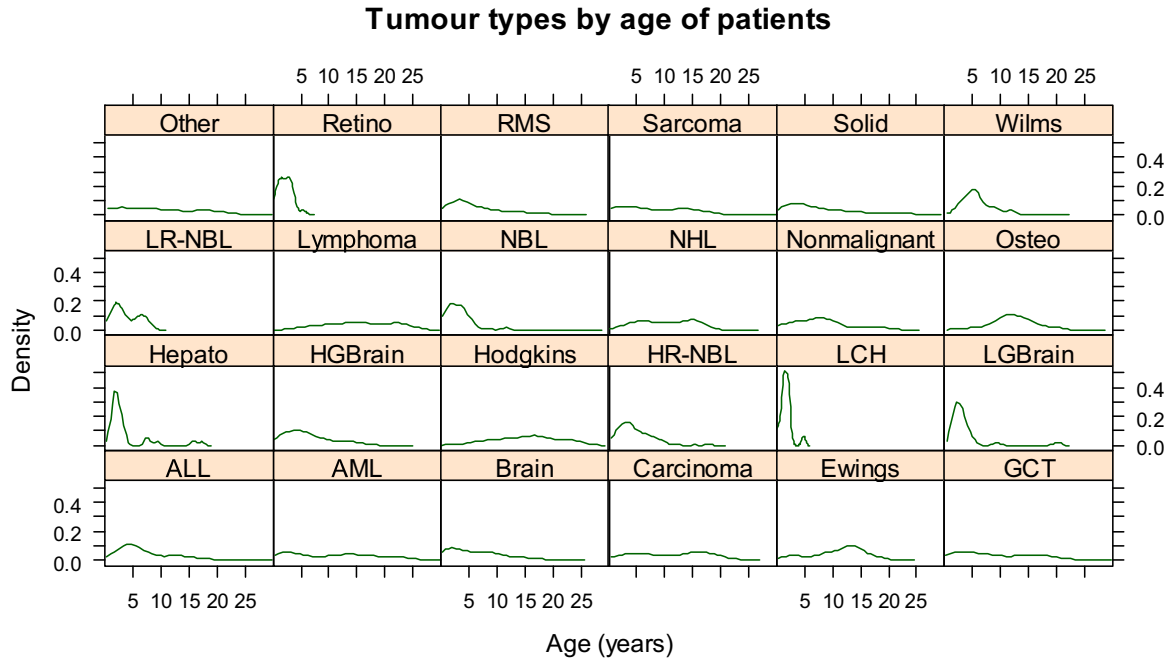


Figure 58: Distribution of tumour type by age

Table 45: Gender distribution per study

Study	Female	Male	NA.s	Proportion Male	M:F Ratio
Alexander	90	97	0	0.52	1.08
BaselSPOG	3	6	0	0.67	2
BernSPOG	100	71	0	0.42	0.71
BonnSPOG	18	26	0	0.59	1.44
EORTC-IX	136	179	0	0.57	1.32
EORTC-XI	127	174	0	0.58	1.37
EORTC-XII	10	11	0	0.52	1.1
Genoa	271	432	0	0.61	1.59
Kitanovski	25	43	0	0.63	1.72
Klaassen	198	233	0	0.54	1.18
Lehrnbecher	124	186	1	0.6	1.5
PINE	374	438	0	0.54	1.17
RetroBern	143	221	0	0.61	1.55

Silva	26	26	0	0.5	1
Spassova	82	117	0	0.59	1.43
Styjewski	27	29	0	0.52	1.07
Sung	29	46	0	0.61	1.59
Tezcan	68	77	0	0.53	1.13
Tissing	127	123	8	0.49	0.97
ZurichSPOG	63	91	0	0.59	1.44

Table 46: Episodes per patient in non-unique-entry studies

Study ID	Min	Median	Mean	Max
Alexander	1	2	2.3	8
BaselSPOG	1	1.5	1.5	2
BernSPOG	1	2	2.7	10
BonnSPOG	1	2	1.9	4
Genoa	1	1	1.6	7
Kitanovski	1	2	2.6	6
Klaassen	1	2	2.4	9
Lehrnbecher	1	2	2.4	10
PINE	1	1	1.5	2
RetroBern	1	2	2.8	12
Silva	1	2	2.5	6
Spassova	1	2	2.5	7
Tezcan	1	2	2.6	10
Tissing	1	2	2.6	14
ZurichSPOG	1	2	2.1	6

Table 47: Proportion of patients who experienced the dichotomous outcomes

	Death	ICU admission	Organ support	Severe infection	Clinically documented infection	Microbiologically documented infection	Blood stream infection
Alexander	1.1%	NA	8.0%	16.6%	NA	14.4%	8.6%
BaselSPOG	0.0%	0.0%	NA	44.4%	NA	44.4%	44.4%
BernSPOG	1.2%	51.5%	NA	NA	NA	32.2%	19.9%
BonnSPOG	0.0%	0.0%	NA	NA	NA	13.6%	13.6%
EORCT-XIV	0.0%	NA	NA	36.9%	10.1%	26.8%	22.1%
EORTC-IX	1.9%	NA	NA	46.7%	16.8%	29.8%	24.4%
EORTC-XI	1.3%	NA	NA	45.8%	19.3%	26.6%	20.6%
EORTC-XII	0.0%	NA	NA	42.9%	23.8%	19.0%	19.0%
Genoa	2.4%	NA	NA	21.5%	6.4%	17.4%	11.2%
Hakim	1.5%	35.7%	11.7%	35.5%	21.4%	23.2%	12.3%
Kitanovski	0.0%	39.7%	11.8%	44.1%	32.4%	26.5%	24.6%
Klaassen	0.7%	6.0%	NA	23.4%	28.3%	28.3%	12.5%
Lehrnbecher	NA	NA	NA	33.1%	22.5%	17.4%	10.6%
PINE	0.4%	19.0%	NA	33.3%	12.1%	39.2%	32.8%
RetroBern	0.5%	NA	NA	50.3%	41.2%	NA	23.9%
Silva	0.0%	9.9%	1.9%	17.3%	67.3%	17.3%	11.5%
Spassova	3.5%	15.6%	NA	56.3%	0.0%	42.7%	42.7%
Styjewski	5.4%	NA	NA	30.4%	NA	28.6%	19.6%
Sung	0.0%	NA	NA	NA	NA	12.0%	9.3%
Tezcan	2.1%	7.2%	NA	63.4%	40.0%	53.1%	39.3%
Tissing	0.8%	32.0%	8.8%	NA	NA	34.4%	23.7%
ZurichSPOG	0.6%	7.1%	NA	17.5%	NA	14.9%	11.7%

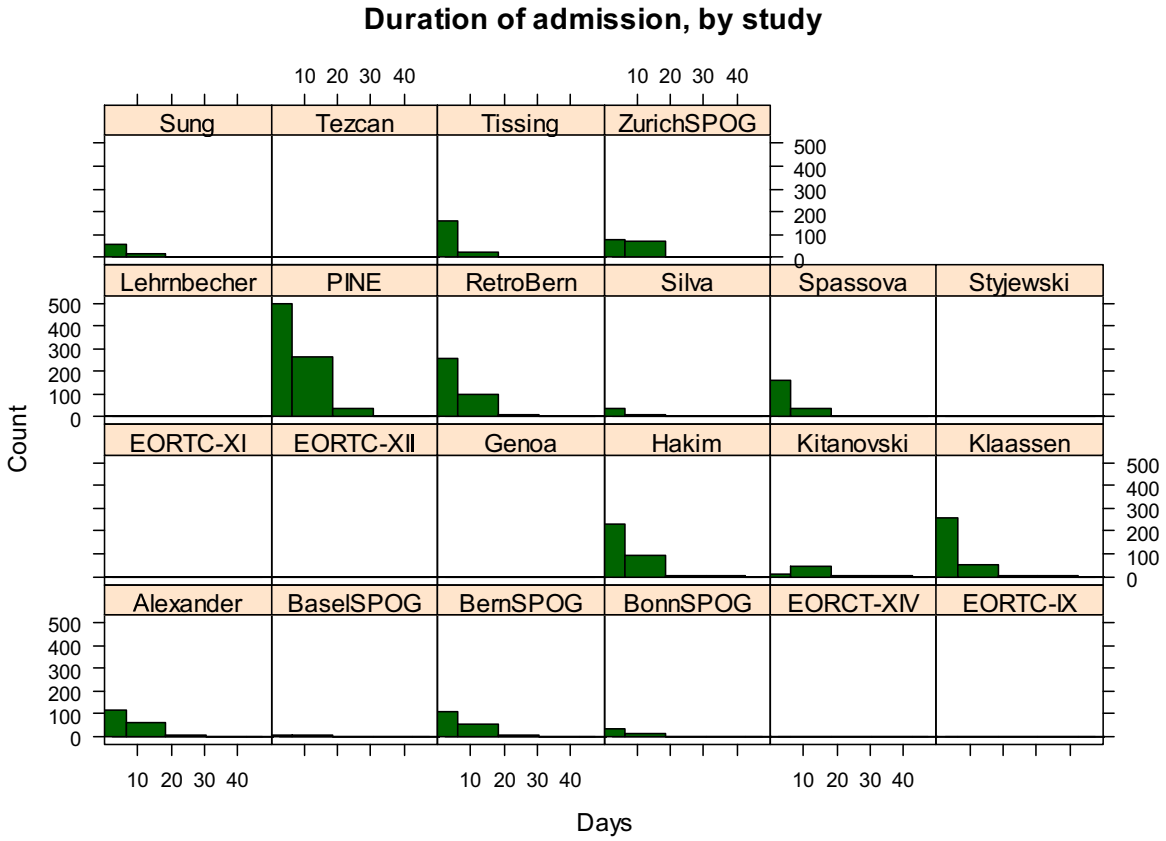


Figure 59: Duration of admission (per study)

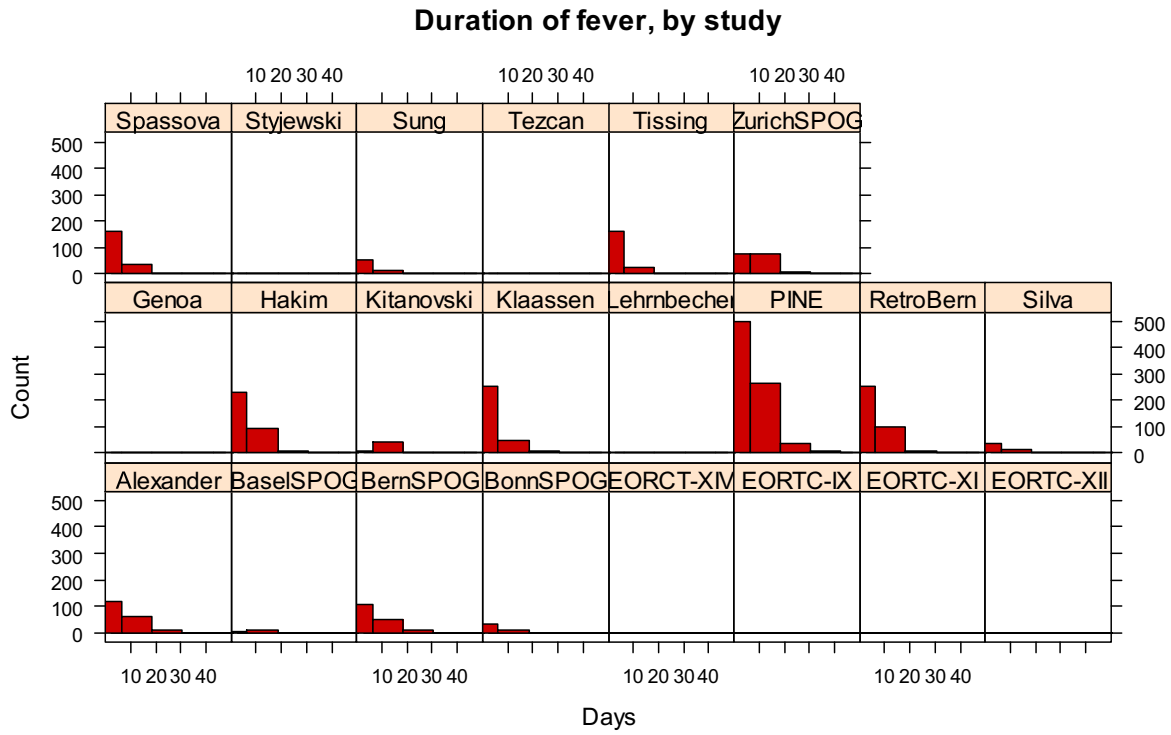


Figure 60: Duration of fever (per study)

Table 48: Days since chemotherapy (by study)

Study ID	Min	Median	Mean	Max
BaselSPOG	12	16	20	40
BernSPOG	2	11	13	66
BonnSPOG	3	14	16	44
Hakim	0	5.1	7.3	510
Kitanovski	1	11	10	24
Lehrnbecher	1	6.9	8	77
RetroBern	2	11	12	44
Silva	0	8	7.3	23
Spassova	1	17	30	310
Sung	2	6.9	14	360
Tezcan	1	1	1.4	3
Tissing	0	5.8	6.6	40
ZurichSPOG	1	12	25	310

Table 49: Temperature (°C) per study

Study ID	Min	Median	Mean	Max
BaselSPOG	37.9	39	38.76	39.4
BernSPOG	36.3	39	38.88	40.6
BonnSPOG	38.1	38.8	38.85	39.8
EORCT-XIV	38	38.6	38.66	40
EORTC-IX	38	38.6	38.71	40.9
EORTC-XI	38	38.6	38.73	40.5
EORTC-XII	38	38.7	38.82	39.5
Hakim	36.1	38.4	38.5	41
Kitanovski	38	38.8	38.8	41
Klaassen	37.9	39	39.07	41.8
RetroBern	36.3	39.1	39.03	41.3
Silva	38	38.5	38.63	40
Spassova	35	38.4	38.5	40.4
Styjewski	37.2	38.3	38.51	40.2
Tissing	37.8	39.5	39.48	41
ZurichSPOG	37	38.6	38.62	40.5

Table 50: Distribution of Hb (g/dL) by study

Study ID	Min	Median	Mean	Max
BaselSPOG	5	6.2	6.3	8.1
BernSPOG	4.4	9.5	9.4	15
BonnSPOG	6.6	10	10	15
EORCT-XIV	4.3	8.7	8.9	15
EORTC-IX	3.5	9.4	9.5	17
EORTC-XI	5.6	9.6	9.7	16
EORTC-XII	4.7	9.1	8.9	13
Hakim	5	8.8	8.9	13
Kitanovski	5.6	9.3	9.3	14
Klaassen	4.6	8.6	8.6	15
RetroBern	3.4	8.8	8.9	14
Sung	5.4	8.9	9	14
Tissing	3.9	8.3	8.4	17
ZurichSPOG	4.7	7.6	7.9	13

Table 51: Distribution of platelet count (x10^9) per study

Study ID	**Log-Min**	**Log-Median**	**Log-Mean**	**Log-Max**	**Natural min**	**Natural median**	**Natural mean**	**Natural max**
BaselSPOG	1.792	2.996	3.324	5.112	6	20	49.22	166
BernSPOG	0	3.296	3.261	5.927	1	27	56.19	375
BonnSPOG	0	3.433	3.589	5.561	1	31	60	260
EORCT-XIV	0	3.258	3.263	6.073	1	26	40.91	434
EORTC-IX	0	3.466	3.474	6.021	1	32	51.12	412
EORTC-XI	0.6931	3.511	3.526	6.461	2	33.5	57.83	640
EORTC-XII	2.303	4.137	4.202	5.905	10	63	107.8	367
Hakim	0	4.016	4.022	6.422	1	55.5	98.19	615
Kitanovski	0	3.636	3.464	6.215	1	38	74.99	500
Klaassen	0	4.111	3.93	6.477	1	61	93.79	650
RetroBern	0	2.89	3.032	6.447	1	18	56.1	631
Silva	0.6931	4.483	4.376	5.875	2	88.5	120.3	356
Styjewski	0.6931	4.007	3.877	6.438	2	55	91.22	625

Sung	1.099	4.443	4.246	6.227	3	85	117	506
Tissing	0	3.761	3.375	6.709	1	43	66.97	820
ZurichSPOG	0	3.892	3.796	6.524	1	49	76.05	681

Table 52: Distribution of white cell counts (WCC x10^6) per study

Study ID	Log-Min	Log-Median	Log-Mean	Log-Max	Natural min	Natural median	Natural mean	Natural max
Alexander	-2.813	-0.6162	-0.5945	2.917	0.06	0.54	0.9297	18.48
BaselSPOG	-2.12	-1.238	-0.9462	0.1398	0.12	0.29	0.5022	1.15
BernSPOG	-3.507	-0.9163	-1.077	1.308	0.03	0.4	0.5411	3.7
BonnSPOG	-2.303	-0.1054	-0.2515	1.03	0.1	0.9	0.937	2.8
EORCT-XIV	-4.605	-1.204	-1.073	5.226	0.01	0.3	2.581	186
EORTC-IX	-2.303	-0.9163	-0.8323	3.493	0.1	0.4	1.023	32.9
EORTC-XI	-2.303	-0.6931	-0.8024	3.699	0.1	0.5	1.163	40.4
EORTC-XII	-2.303	-0.734	-0.6643	0.7885	0.1	0.48	0.8038	2.2
Hakim	-2.303	-0.5108	-0.6391	2.617	0.1	0.6	0.8421	13.7
Kitanovski	-2.303	-0.5108	-0.5823	2.14	0.1	0.6	0.9912	8.5
Klaassen	-2.303	-0.5108	-0.581	2.639	0.1	0.6	0.8494	14
Lehrnbecher	-3.219	-0.1054	-0.32	2.934	0.04	0.9	1.22	18.8
RetroBern	-2.303	-0.6931	-0.7393	1.482	0.1	0.5	0.7344	4.4
Silva	-1.609	-0.0526	-0.1379	1.335	0.2	0.95	1.076	3.8
Tissing	-2.996	-0.9163	-1.197	1.065	0.05	0.4	0.5677	2.9
ZurichSPOG	-4.605	-1.036	-1.011	1.411	0.01	0.355	0.5777	4.1

Table 53: Distribution of absolute neutrophil counts (ANC; cells/cubic mm), by study

Study ID	Log-Min	Log-Median	Log-Mean	Log-Max	Natural min	Natural median	Natural mean	Natural max
Alexander	0	3.689	3.195	6.209	1	40	95.78	497
BaselSPOG	4.382	4.382	4.811	5.67	80	80	150	290
BernSPOG	0	2.303	2.363	6.215	1	10	82.68	500
BonnSPOG	0	3.589	3.343	3.912	1	36.2	33.46	50
EORCT-XIV	0	2.996	2.723	6.685	1	20	81.46	800
EORTC-IX	0	0	1.776	6.888	1	1	87.79	980
EORTC-XI	0	3.401	2.821	6.894	1	30	114.6	986
EORTC-XII	0	3.916	3.364	6.729	1	51.5	187.6	836

Genoa	0	3.912	3.264	8.716	1	50	172.4	6100
Hakim	0	0	2.256	6.215	1	1	85.33	500
Kitanovski	0	1.589	1.993	6.867	1	5	69.6	960
Klaassen	0	3.689	3.108	6.856	1	40	119.9	950
Lehrnbecher	0	5.58	5.053	9.035	1	265	497.9	8390
PINE	0	4.605	4.083	9.036	1	100	326.4	8400
RetroBern	0	4.094	3.553	6.824	1	60	178	920
Silva	0	2.996	2.782	7.601	1	20	184.9	2000
Spassova	0	4.605	3.659	6.908	1	100	161.1	1000
Sung	0	4.552	4.194	6.856	1	95	203	950
Tezcan	0	4.277	4.419	6.908	1	72	222.9	1000
Tissing	0	4.605	4.891	10.31	1	100	489.6	30000
ZurichSPOG	0	1.151	2.327	6.215	1	5.5	83.61	500

Table 54: Distribution of absolute monocyte count (AMC; cells / cubic mm) per study

Study ID	Log-Min	Log-Median	Log-Mean	Log-Max	Natural min	Natural median	Natural mean	Natural max
Alexander	0	3.718	3.477	7.652	1	41.2	125.2	2105
BaselSPOG	0	1.099	1.134	2.303	1	3	4.667	10
BernSPOG	0	0.6931	1.042	4.394	1	2	7.223	81
BonnSPOG	0	1.609	1.702	3.912	1	5	9.608	50
Hakim	0	4.357	3.568	7.473	1	78	141	1760
Kitanovski	0	1.946	2.286	7.153	1	7	98.28	1278
Klaassen	0	1.386	1.58	5.517	1	4	15.92	249
RetroBern	0	2.079	1.976	5.159	1	8	20.63	174
Spassova	0	2.303	2.26	5.704	1	10	53.34	300
Styjewski	0	3.515	2.786	6.109	1	33.6	67.04	450
Tezcan	0	3.219	3.361	8.208	1	25	179	3672
ZurichSPOG	0	0	1.146	5.635	1	1	13.52	280

Table 55: C-reactive protein (CRP) values (mg/dL) per study

Study ID	Log-Min	Log-Median	Log-Mean	Log-Max	Natural min	Natural median	Natural mean	Natural max
BaselSPOG	1.099	2.835	2.884	5.081	3	19.5	38.63	161
BernSPOG	1.099	3.829	3.744	5.849	3	46	66.05	347
BonnSPOG	0.6931	3.497	3.408	5.124	2	33	42.36	168
Kitanovski	1.856	3.892	3.783	5.568	6.4	49	64.06	262
Lehrnbecher	0	3.198	2.972	5.659	1	24.5	42.84	287
RetroBern	0	3.871	3.617	5.771	1	48	64.04	321
Silva	0	3.02	2.87	4.394	1	20.5	25.77	81
Spassova	0	3.258	3.139	5.932	1	26	46.54	377
Tezcan	0	3.848	3.797	6.064	1	46.9	80.15	430
Tissing	0.8329	3.807	3.646	6.023	2.3	45	63.42	413
ZurichSPOG	1.386	3.638	3.553	6.246	4	38	61.97	516

Table 56: Further biomarker distributions

PCT (ng/ml)	Log-Min	Log-Median	Log-Mean	Log-Max	Natural min	Natural median	Natural mean	Natural max
Kitanovski	-1.273	-0.5978	-0.3852	2.351	0.28	0.55	1.036	10.5
Styjewski	-0.9416	0.9126	1.437	6.306	0.39	2.495	32.93	547.7
IL-6 (pg/ml)	**Log-Min**	**Log-Median**	**Log-Mean**	**Log-Max**	**Natural min**	**Natural median**	**Natural mean**	**Natural max**
Kitanovski	1.504	4.508	4.511	5.991	4.5	90.75	167.4	400
Lehrnbecher	0	3.978	4.059	9.418	1	53.4	216.1	12310
Spassova	1.131	3.761	3.676	4.605	3.1	43	55.98	100
Styjewski	0	2.35	2.372	6.392	1	10.5	40.68	597
IL-8 (pg/ml)	**Log-Min**	**Log-Median**	**Log-Mean**	**Log-Max**	**Natural min**	**Natural median**	**Natural mean**	**Natural max**
BonnSPOG	2.175	4.58	4.435	6.418	8.8	97.55	110.1	613

Lehrnbecher	-1.386	5.178	4.87	8.567	0.25	177.4	306.5	5256
Spassova	2.398	4.035	4.324	6.892	11	56.65	184.9	984.5
Styjewski	0	2.944	2.925	5.628	1	19	46.61	278
Tissing	2.197	4.525	4.59	7.703	9	92.5	212.9	2214

Appendix 23. R-code for IPD models

Full hierarchical model

```
glmer(data[,outcome] ~ factor(study.id) + data[,predictor] + (data[,predictor]-1|study.id) +
(1|picnicc.id), data=data, na.action(na.exclude), family=binomial)
```

Reduced hierarchical model

```
glmer(data[,outcome] ~ factor(study.id) + data[,predictor] + (data[,predictor]-1|study.id), data=data,
na.action(na.exclude), family=binomial)
```

Fixed effects model

```
glm(data[,outcome] ~ factor(study.id) + data[,predictor], data=data, na.action(na.exclude),
family=binomial)
```

Appendix 24. Hierarchical model comparing estimates with all vs. multi-episode data

	All episodes with clustering						Multi-episodes only with clustering						**Variation**						
Predictor name	OR	beta	SE-beta	Individual-level SD	Study-level SD	p-value (beta)	OR	beta	SE-beta	Individual-level SD	Study-level SD	p-value (beta)	beta variation	relative to s.err	p-value for beta diff	*	ind SD variation	glass' effect size	
sex - F	1	0					1	0					0						
sex - M	0.98	-0.022	0.084	0.87	1.70E-05	0.8	0.92	-0.08	0.15	1	0	0.59	0.058	0.690	0.490	-	-0.130	-0.149	-
age.days	1	2.60E-05	2.00E-05	0.85	0	0.19	1	-5.90E-05	4.00E-05	0.99	0	0.14	0.000085	4.250	0.000	*	-0.140	-0.165	-
marrow	1.4	0.34	0.38	2.1	0	0.37	1.1	0.14	0.42	2.1	5.60E-07	0.75	0.2	0.526	0.599	-	0.000	0.000	-
relapse	1.5	0.39	0.19	1.2	0	0.038	1.8	0.59	0.24	1.2	0	0.015	-0.2	1.053	0.293	-	0.000	0.000	-
chemo.intensity - Low	1	0					1	0					0				0		
chemo.intensity - Standard	2.5	0.93	0.2	1.1	0	3.30E-06	3.5	1.3	0.63	1.3	0.05	0.045	-0.37	1.850	0.064		-0.200	-0.182	-
chemo.intensity - HSCT	0.99	-0.0062	0.12	1.1	0	0.96	0.72	-0.33	0.37	1.3	0.043	0.37	0.3238	2.698	0.007	*	-0.200	-0.182	-
cvl	1.4	0.37	0.28	1.1	0.027	0.19	1.1	0.06	0.29	1.1	2.60E-06	0.83	0.31	1.107	0.268	-	0.000	0.000	-
cvl.type - None	1	0					1	0					0				0		
cvl.type - Port	0.98	-0.021	0.85	4.7	0.26	0.98	0.96	-0.04	0.49	2	0.00012	0.94	0.019	0.022	0.982	-	2.700	0.574	+
cvl.type - Hickman	1.8	0.6	0.76	4.7	0.061	0.43	1.9	0.63	0.52	2	6.60E-06	0.23	-0.03	0.039	0.969		2.700	0.574	+

cvl.type - Untunnelled	12	2.5	1.9	4.7	0.68	0.2	14	2.6	2.7	2	8.10E-06	0.33	-0.1	0.053	0.958	-	2.700	0.574	+
out.patient	0.62	-0.48	0.16	1.1	2.00E-07	0.0028	1.1	0.11	0.3	1.2	0	0.71	-0.59	3.688	0.000	*	-0.100	-0.091	-
temp	2.2	0.81	0.098	1.1	0	2.50E-16	2.2	0.77	0.15	1.4	0	2.00E-07	0.04	0.408	0.683	-	-0.300	-0.273	+
shock	3.5	1.2	0.37	1	0.77	0.00093	2.7	1	0.69	1.4	1.5	0.15	0.2	0.541	0.589	-	-0.400	-0.400	+
sys	1	-0.0013	0.011	0.25	0	0.91	1	0.0018	0.012	0.21	0	0.89	-0.0031	0.282	0.778	-	0.040	0.160	-
dia	0.99	-0.0095	0.013	0.28	0	0.46	0.99	-0.0078	0.014	0.26	0	0.58	-0.0017	0.131	0.896	-	0.020	0.071	-
mucositis	0.83	-0.18	0.076	1.3	0	0.017	0.81	-0.21	0.089	1.4	2.80E-07	0.018	0.03	0.395	0.693	-	-0.100	-0.077	-
severe.mucositis	0.68	-0.38	0.19	1.1	2.60E-05	0.045	0.62	-0.48	0.28	1.4	3.20E-06	0.085	0.1	0.526	0.599	-	-0.300	-0.273	+
severe.unwell	2.5	0.91	0.12	1	0	1.50E-13	1.9	0.65	0.2	1.3	0	0.0012	0.26	2.167	0.030	*	-0.300	-0.300	+
hb	1	0.047	0.029	0.45	0	0.1	1.1	0.061	0.047	0.5	0	0.2	-0.014	0.483	0.629	-	-0.050	-0.111	-
ln.plt	0.82	-0.19	0.038	0.41	0	3.00E-07	0.82	-0.2	0.06	0.45	0	0.00091	0.01	0.263	0.792	-	-0.040	-0.098	-
ln.wcc	0.76	-0.27	0.076	0.51	0.22	0.00032	0.7	-0.36	0.11	0.6	0.22	0.00067	0.09	1.184	0.236	-	-0.090	-0.176	-
ln.anc	0.97	-0.033	0.0054	0.92	0	1.40E-09	0.95	-0.048	0.011	1.1	0.013	4.50E-06	0.015	2.778	0.005	*	-0.180	-0.196	-
ln.amc	0.95	-0.053	0.0097	1.3	0.01	4.60E-08	0.95	-0.053	0.012	1.3	0.0097	1.00E-05	0	0.000	1.000	-	0.000	0.000	-
ln.crp	1	0.024	0.034	1.3	0	0.48	1.1	0.07	0.048	1.3	0	0.14	-0.046	1.353	0.176	-	0.000	0.000	-
ln.pct	1.9	0.65	0.18	6.20E-06	0	0.00033													
ln.IL-6	2.4	0.87	0.12	1.1	0	2.80E-12	2.9	1.1	0.18	1.1	0	1.10E-09	-0.23	1.917	0.055	-	0.000	0.000	-

ln.IL-8	2	0.7	0.13	1	0	2.50E-08	2	0.67	0.16	1	0	1.60E-05	0.03	0.231	0.817	-	0.000	0.000	-

Appendix 25. Reduced hierarchical model comparing all vs. single-episode estimates

	All episodes, no clustering					1 episode per patient					Variation			
Predictor name	OR	beta	SE-beta	Study-level SD	p-value	OR	beta	SE-beta	Study-level SD	p-value	beta variation	relative to s.error	p-value	
sex - F	1	0				1	0				0			
sex - M	0.97	-0.029	0.071	0	0.68	0.95	-0.05	8.30E-02	7.70E-07	0.54	0.021	0.296	0.767	-
age.days	1	2.20E-05	1.70E-05	1.20E-13	0.21	1	2.50E-05	1.90E-05	0	0.19	-3E-06	0.176	0.860	-
marrow	1.5	0.41	0.25	4.70E-06	0.095	1.1	0.11	3.40E-01	4.90E-06	0.73	0.3	1.200	0.230	-
relapse	1.4	0.36	0.18	0.27	0.042	1.3	0.27	1.90E-01	3.50E-08	0.15	0.09	0.500	0.617	-
chemo.intensity - Low	1	0				1	0				0			
chemo.intensity - Standard	2	0.7	0.23	0.033	0.002	1.5	0.43	0.27	0.058	0.12	0.27	1.174	0.240	-
chemo.intensity - HSCT	1.1	0.059	0.14	0.021	0.67	1.3	0.27	0.17	0.028	0.11	-0.211	1.507	0.132	-
cvl	1.3	0.28	0.24	0.021	0.25	1.3	0.23	2.30E-01	2.70E-08	0.32	0.05	0.208	0.835	-
cvl.type - None	1	0				1	0				0			
cvl.type - Port	1.1	0.061	0.24	0	0.8	1.6	0.46	0.35	0	0.19	-0.399	1.663	0.096	-
cvl.type - Hickman	1.4	0.35	0.25	0	0.17	2.2	0.77	0.37	0	0.037	-0.42	1.680	0.093	-
cvl.type - Untunnelled	3.1	1.1	0.59	0	0.054	5.1	1.6	0.65	0	0.013	-0.5	0.847	0.397	-

out.patient	0.7	-0.36	0.14	9.10E-08	0.0078	0.62	-0.48	1.60E-01	3.20E-08	0.0024	0.12	0.857	0.391	-
temp	1.9	0.64	0.084	0	2.60E-14	2	0.7	0.1	0.00E+00	3.70E-12	-0.06	0.714	0.475	-
shock	2.8	1	0.32	0.63	0.0013	1.9	0.63	0.34	0.47	0.066	0.37	1.156	0.248	-
sys	1	-0.0011	0.011	0	0.92	1	0.019	0.016	0	0.24	-0.0201	1.827	0.068	-
dia	0.99	-0.0092	0.013	0	0.47	1	0.012	0.018	0	0.5	-0.0212	1.631	0.103	-
mucositis	0.89	-0.11	0.058	0	0.052	0.85	-0.17	8.80E-02	6.60E-07	0.057	0.06	1.034	0.301	-
severe.mucositis	0.76	-0.28	0.16	3.50E-07	0.078	0.71	-0.35	2.00E-01	1.30E-06	0.091	0.07	0.438	0.662	-
severe.unwell	2.2	0.79	0.11	0	2.20E-13	2.3	0.84	0.13	0.00E+00	4.00E-11	-0.05	0.455	0.649	-
hb	1	0.044	0.028	0	0.11	1	0.047	0.033	0	0.15	-0.003	0.107	0.915	-
ln.plt	0.83	-0.19	0.037	0	3.00E-07	0.81	-0.21	0.045	0.00E+00	4.50E-06	0.02	0.541	0.589	-
ln.wcc	0.78	-0.25	0.072	0.21	0.00041	0.84	-0.17	0.053	0.12	0.001	-0.08	1.111	0.267	-
ln.anc	0.97	-0.027	0.0048	4.10E-07	9.90E-09	0.97	-0.027	0.0055	0.00E+00	6.60E-07	0			
ln.amc	0.96	-0.045	0.0098	0.018	4.40E-06	0.96	-0.046	9.30E-03	2.20E-07	8.60E-07	0.001	0.102	0.919	-
ln.crp	1	0.017	0.028	0	0.54	1	0.00016	0.038	0	1	0.01684	0.601	0.548	-
ln.pct	1.9	0.65	0.18	0	0.00033	1.8	0.58	0.18	0	0.0017	0.07	0.389	0.697	-
ln.IL-6	2.1	0.75	0.11	0	4.40E-12	1.7	0.51	0.13	0	0.00011	0.24	2.182	0.029	*
ln.IL-8	1.8	0.61	0.11	0	1.70E-08	1.7	0.54	0.14	0	0.00018	0.07	0.636	0.525	-

Appendix 26. Comparing full hierarchical model vs. fixed effects model estimates

	Full heirarchical		Fixed effect			Full vs Fixed			
Predictor name	beta	SE-beta	beta	SE-beta		diff beta	as prop se	diff p-val	p<0.05
sex - F	0		0						
sex - M	-0.022	0.084	-0.041	0.07		0.019	0.22619	0.821	-
age.days	0.000026	0.00002	0.000015	0.000017		0.000011	0.55	0.582	-
marrow	0.34	0.38	0.41	0.25		-0.07	0.184211	0.854	-
remission	0.086	0.25	-0.023	0.14		0.109	0.436	0.663	-
relapse	0.39	0.19	0.35	0.14		0.04	0.210526	0.833	-
chemo.intensity - Low	0		0			0			-
chemo.intensity - Standard	0.93	0.2	0.78	0.17		0.15	0.75	0.453	-
chemo.intensity - HSCT	-0.0062	0.12	0.0055	0.1		-0.0117	0.0975	0.922	-
chemo.time	-0.0035	0.0039	-0.0019	0.0029		-0.0016	0.410256	0.682	-
cvl	0.37	0.28	0.15	0.18		0.22	0.785714	0.432	-
cvl.type - None	0		0			0			-
cvl.type - Port	-0.021	0.85	0.061	0.24		-0.082	0.096471	0.923	-

cvl.type - Hickman	0.6	0.76	0.35	0.25		0.25	0.328947	0.742	-
cvl.type - Untunnelled	2.5	1.9	1.1	0.59		1.4	0.736842	0.461	-
out.patient	-0.48	0.16	-0.36	0.14		-0.12	0.75	0.453	-
temp	0.81	0.098	0.64	0.084		0.17	1.734694	0.083	-
resp.rate	-0.038	0.026	-0.031	0.022		-0.007	0.269231	0.788	-
resp.compromise	0.22	0.32	0.028	0.17		0.192	0.6	0.549	-
pulse.rate	0.002	0.0055	0.0015	0.0054		0.0005	0.090909	0.928	-
shock	1.2	0.37	0.88	0.18		0.32	0.864865	0.387	-
sys	-0.0013	0.011	-0.0011	0.011		-0.0002	0.018182	0.985	-
dia	-0.0095	0.013	-0.0092	0.013		-0.0003	0.023077	0.982	-
mucositis	-0.18	0.076	-0.11	0.058		-0.07	0.921053	0.357	-
severe.mucositis	-0.38	0.19	-0.28	0.16		-0.1	0.526316	0.599	-
severe.unwell	0.91	0.12	0.79	0.11		0.12	1	0.317	-
hb	0.047	0.029	0.044	0.028		0.003	0.103448	0.918	-
ln.plt	-0.22	0.04	-0.22	0.039		0			-
ln.wcc	-0.37	0.057	-0.33	0.044		-0.04	0.701754	0.483	-
ln.anc	-0.095	0.017	-0.08	0.015		-0.015	0.882353	0.378	-

ln.amc	-0.28	0.042	-0.22	0.034		-0.06	1.428571	0.153	-
ln.crp	0.064	0.067	0.063	0.055		0.001	0.014925	0.988	-
ln.pct	0.65	0.18	0.65	0.18		0			-
ln.IL-6	0.87	0.12	0.75	0.11		0.12	1	0.317	-
ln.IL-8	0.7	0.13	0.61	0.11		0.09	0.692308	0.489	-

Appendix 27. Further within and between study results for univariate predictors

Gender

	Predictor name	OR	beta	SE-beta	p-value
Alexander	Male	0.84	-0.17	0.42	0.68
BaselSPOG	Male	2	0.69	1.5	0.64
BernSPOG	Male	1.6	0.46	0.33	0.17
BonnSPOG	Male	0.1	-2.3	1.1	0.049
EORTC-IX	Male	0.92	-0.087	0.25	0.72
EORTC-XI	Male	1.9	0.63	0.28	0.022
EORTC-XII	Male	0.89	-0.12	1.1	0.92
Genoa	Male	0.75	-0.29	0.2	0.15
Kitanovski	Male	0.64	-0.44	0.56	0.43
Klaassen	Male	1.2	0.14	0.22	0.51
Lehrnbecher	Male	0.96	-0.037	0.3	0.9
PINE	Male	0.95	-0.053	0.14	0.71
Silva	Male	2.3	0.83	0.77	0.28
Spassova	Male	0.78	-0.25	0.29	0.39
Styjewski	Male	0.9	-0.1	0.59	0.87
Sung	Male	0.76	-0.27	0.72	0.7
Tezcan	Male	0.91	-0.099	0.33	0.77
Tissing	Male	0.72	-0.33	0.27	0.22
ZurichSPOG	Male	0.88	-0.12	0.46	0.79
IPD estimate	**Male**	**0.96**	**-0.041**	**0.07**	**0.56**

Marrow involvement

	Predictor name	OR	beta	SE-beta	p-value
BaselSPOG	Marrow Involved	2.40E-08	-18	4000	1
BernSPOG	Marrow Involved	2	0.7	0.52	0.17
Kitanovski	Marrow Involved	2.3	0.85	0.57	0.14
Spassova	Marrow Involved	1.5	0.37	0.35	0.28
ZurichSPOG	Marrow Involved	1.20E-07	-16	1400	0.99
IPD estimate	**Marrow Involved**	**1.5**	**0.41**	**0.25**	**0.095**

Out patient status

	Predictor name	OR	beta	SE-beta	p-value
BaselSPOG	Out-patient	71000000.00	18	4000	1
BernSPOG	Out-patient	0.43	-0.85	0.41	0.036
BonnSPOG	Out-patient	22000000.00	17	2500	0.99
EORCT-XIV	Out-patient	0.10	-2.3	1	0.028
EORTC-IX	Out-patient	0.45	-0.79	0.28	0.0046
EORTC-XI	Out-patient	0.71	-0.34	0.28	0.22
EORTC-XII	Out-patient	0.42	-0.88	1.1	0.44
Kitanovski	Out-patient	2.3	0.85	0.64	0.18
Silva	Out-patient	0.19	-1.7	1.5	0.26
Spassova	Out-patient	1.30	0.24	0.33	0.47
Sung	Out-patient	2200000	15	1700	0.99
ZurichSPOG	Out-patient	1.4	0.3	0.59	0.61
IPD estimate		**0.7 (0.53 to 0.91)**	**-0.36**	**0.14**	**0.0078**

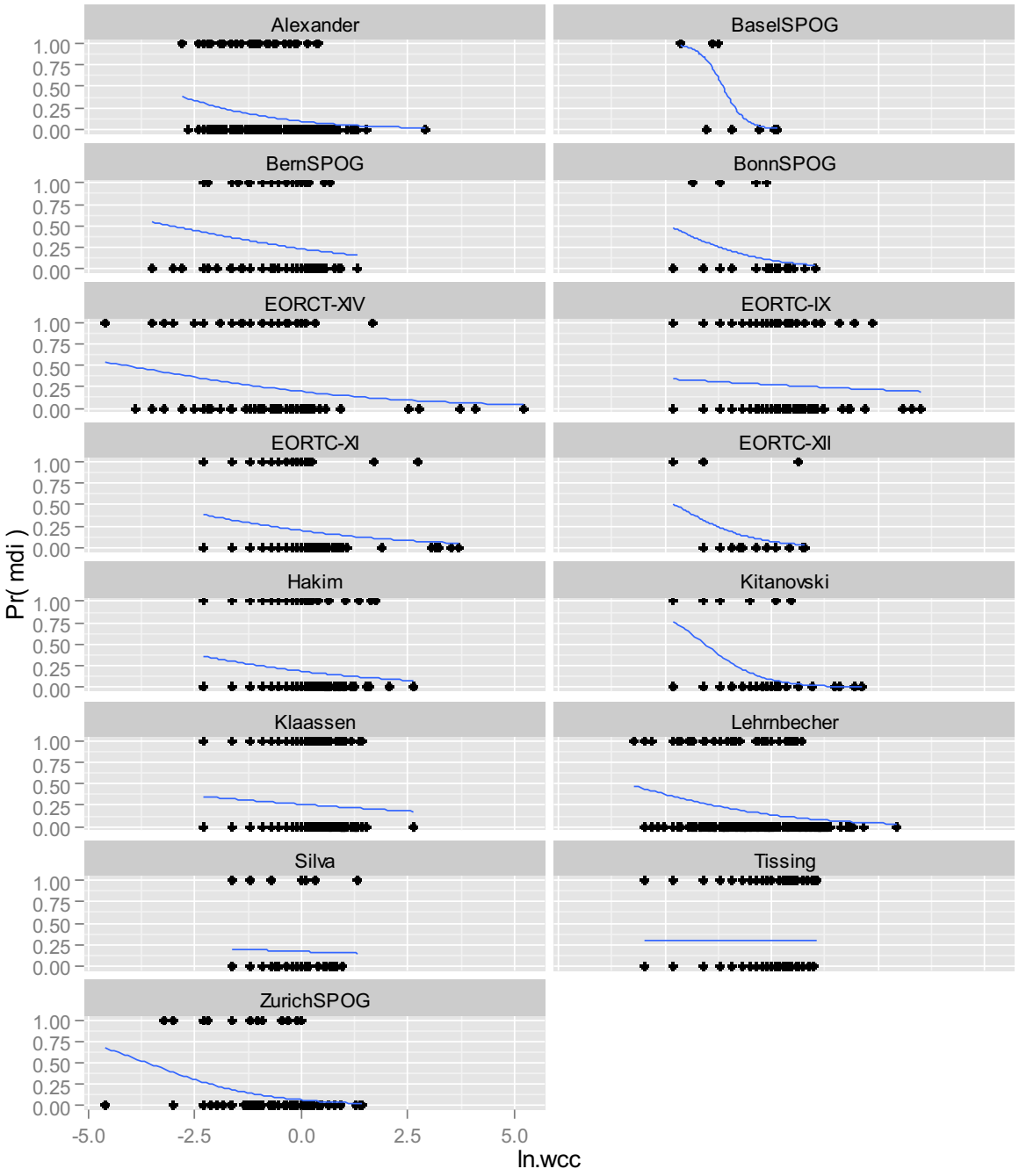


Figure 61: Association between probability of MDI and ln(WCC) by study

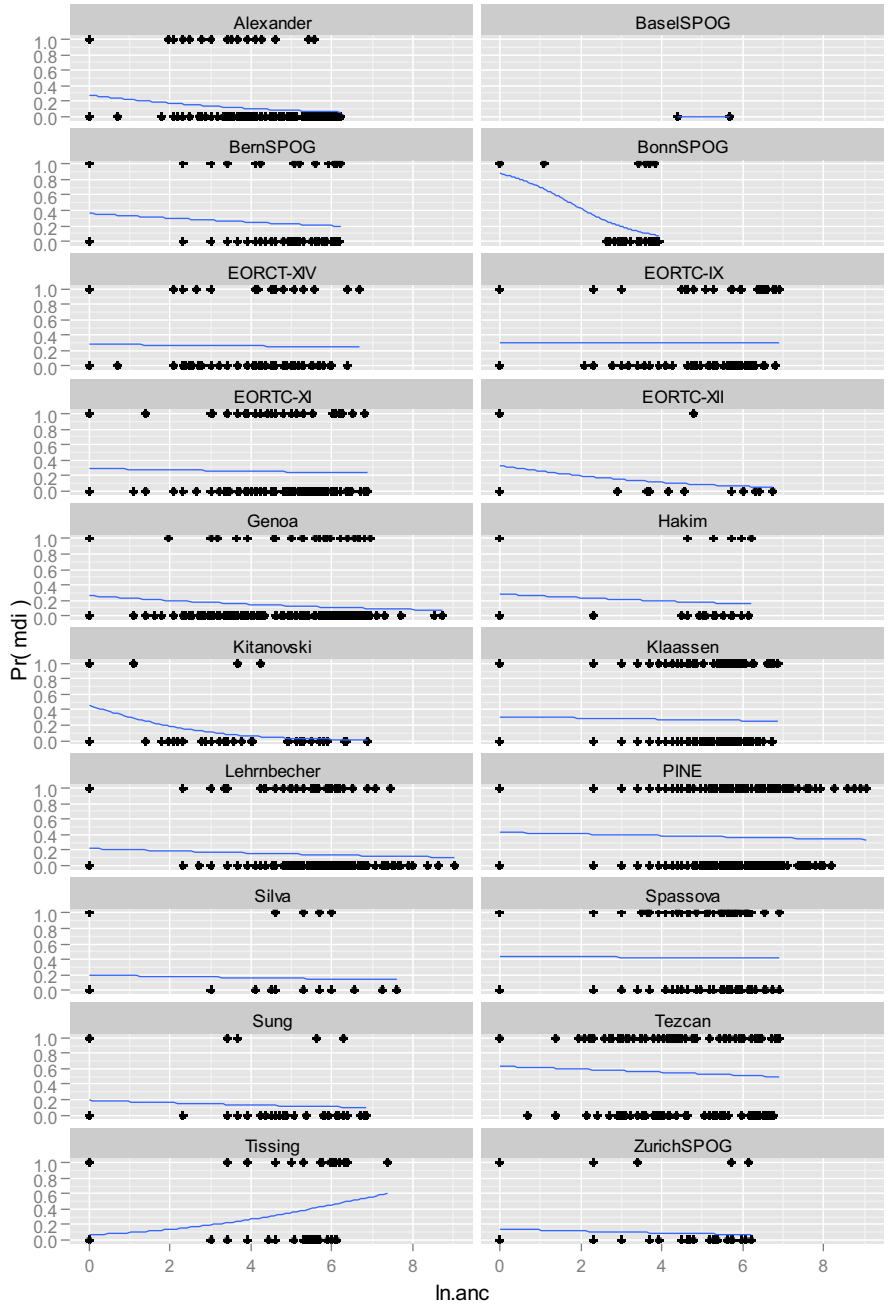


Figure 62: Association between probability of MDI and ln(ANC) by study

Tumour type

Table 57: Individual study p-values for parameter estimates of tumour type

	Genoa	Alexander	Basel SPOG	Bern SPOG	Bonn SPOG	EORTC XIV	EORTC XI	EORTC IX	EORTC XII	Hakim	Kitanovski	Klaassen	Lehrnbecher	PINE	Silva	Spassova	Styjewski	Sung	Tezcan	Tissing	Zurich SPOG
ALL (referent)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AML	0.57	0.0056	1	0.034	0.31	0.83	0.41	0.026	.	0.00096	0.11	0.93	0.09	0.037	.	0.019	1	1	0.98	0.99	0.049
Brain	0.28	0.88	.	.	.	.	.	.	.	.	.	0.42	0.64	.	.	0.99	.	.	0.035	.	.
Carcinoma	0.23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.18
Ewings	0.99	0.55	.	0.55	1	.	.	.	.	0.82	1	.	0.028	0.43	0.99	0.2	1	1	0.9	0.99	0.99
GCT	1	.	.	.	.	.	.	.	1	.	.	1	0.62	0.73	.	.	.	.	.	.	.
Hepato	0.99	.	.	.	.	.	.	.	.	0.36	1	0.57	0.99	0.98	1	.	.	.	.	0.85	.
HGBrain	0.014	.	.	0.63	0.21	.	.	.	1	0.99	0.2	.	0.81	0.046	.	.	1	1	.	0.31	0.88
Hodgkins	0.051	0.99	.	.	.	0.93	.	.	1	0.8	.	0.39	0.62	0.2	.	0.99	.	.	.	.	1
HR-NBL	0.00011	.	.	.	.	.	.	.	.	0.91	0.88	.	.	.	1	0.99	.	1	.	0.99	.
LCH	0.92	.	.	.	.	.	.	.	.	1	1	.	1	0.99	.	.	.	.	.	1	.

LGBrain	.	.	.	0.99	.	.	.	.	.	0.99	.	.	0.86	0.98	.	.	.	1	.	.	.
LR-NBL	0.99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.99	.	.	.	.	.
NBL	.	0.72	1	0.6	1	.	.	.	.	.	.	0.69	0.68	0.038	.	.	0.51	0.77	0.078	.	0.6
NHL	0.0081	0.2	1	0.71	1	0.99	.	.	1	0.82	0.43	0.95	0.19	0.65	1	0.092	.	0.7	0.2	0.27	1
Nonmalig.nt	0.66	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.
Osteo	0.0018	0.87	.	0.089	1	.	.	.	1	0.66	.	0.35	0.99	0.31	.	0.084	0.98	.	.	0.23	1
Other	0.87	0.8	.	.	.	0.99	0.99	0.56	.	0.4	.	0.37	.	0.49	.	.	.	1	0.23	.	.
Retino	.	.	.	0.6	.	.	.	.	.	1	.	0.85	.	0.17	.	.	.	1	.	.	1
RMS	0.08	0.34	.	0.46	0.31	.	.	.	1	0.62	0.99	0.55	0.81	0.14	0.9	.	.	0.88	0.36	0.39	0.32
Solid	.	.	.	.	.	0.45	0.37	0.37	.	.	.	.	.	.	.	.	.	.	.	.	.
Sarcoma	0.37	.	.	.	.	.	.	.	.	0.36	.	0.88	1	0.11	1	0.28	1	1	.	0.73	0.32
Wilms	0.4	0.99	.	0.99	.	.	.	.	1	0.99	1	0.58	0.99	0.48	.	0.99	1	.	0.79	0.23	0.26

Highlighted cells are significant ($p<0.05$)

Relapsed disease

Table 58: Association between probability of MDI and relapsed disease

	Predictor name	OR	beta	SE-beta	p-value
BaselSPOG	Relapse	5.00E-01	-0.69	1.5	0.64
BernSPOG	Relapse	0.94	-0.066	0.37	0.86
BonnSPOG	Relapse	1.40E-07	-16	2800	1
Genoa	Relapse	1.4	0.3	0.23	0.18
Kitanovski	Relapse	2.50E+00	0.91	0.6	0.13
Lehrnbecher	Relapse	4.3	1.5	0.38	0.00012
Spassova	Relapse	1.6	0.48	0.58	0.4
Sung	Relapse	0.7	-0.36	1.1	0.75
Tissing	Relapse	0.91	-0.098	0.34	0.77
ZurichSPOG	Relapse	1.6	0.49	0.56	0.38
IPD estimate	**Relapse**	**1.4 (1.07 to 1.86)**	**0.35**	**0.14**	**0.012**

Cardiovascular compromise

Table 59: Association of shock with MDI

	Predictor name	OR	beta	SE-beta	p-value
BernSPOG	Shock	1.40	0.36	0.76	0.63
BonnSPOG	Shock	0.00	-16	2800	1
Hakim	Shock	3.50	1.2	0.83	0.13
Kitanovski	Shock	1.70E+07	17	1500	0.99
Klaassen	Shock	0.84	-0.18	0.59	0.77
PINE	Shock	2.4	0.88	0.26	0.00074
Spassova	Shock	1.8	0.57	0.5	0.25
Sung	Shock	6	1.8	1	0.072
Tezcan	Shock	4.5	1.5	1.1	0.18
IPD estimate		**2.4 (1.69 to 3.43)**	**0.88**	**0.18**	**1.6E-06**

Outpatient status

Presenting from outside the hospital was associated with a reduced chance on MDI in the IPD analysis, with an estimated OR 0.7 (95% CI 0.53 to 0.91).

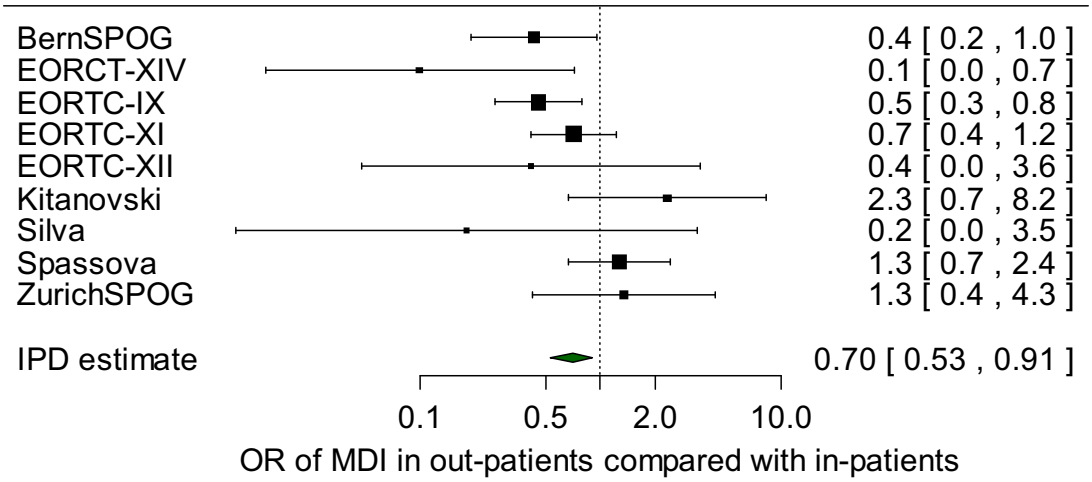


Figure 63: OR of MDI by hospitalisation status

Clinical recognition of 'severely unwell'

Table 60: Association of 'severely unwell' appearance with MDI

	Predictor name	OR	beta	SE-beta	p-value	Prevalence of 'unwell'
Alexander	Severely unwell	1.60	0.48	0.46	0.3	0.22
BernSPOG	Severely unwell	2.40	0.87	0.45	0.051	0.14
BonnSPOG	Severely unwell	7.40	2	1.5	0.18	0.05
Hakim	Severely unwell	1.90	0.65	0.28	0.022	0.25
Kitanovski	Severely unwell	52000000.00	18	1400	0.99	0.04
Klaassen	Severely unwell	2	0.71	0.23	0.0022	0.27
Lehrnbecher	Severely unwell	4.1	1.4	0.53	0.0079	0.05
PINE	Severely unwell	1.8	0.6	0.2	0.0033	0.14
Spassova	Severely unwell	2.5	0.9	0.34	0.0084	0.23
Styjewski	Severely unwell	2500000000.00	22	3200	0.99	0.20
Sung	Severely unwell	1.2	0.14	0.72	0.84	0.41
ZurichSPOG	Severely unwell	1.6	0.45	0.47	0.33	0.30
IPD estimate		**2.2 (1.77 to 2.33)**	**0.79**	**0.11**	**2.2E-13**	**0.16**

Age

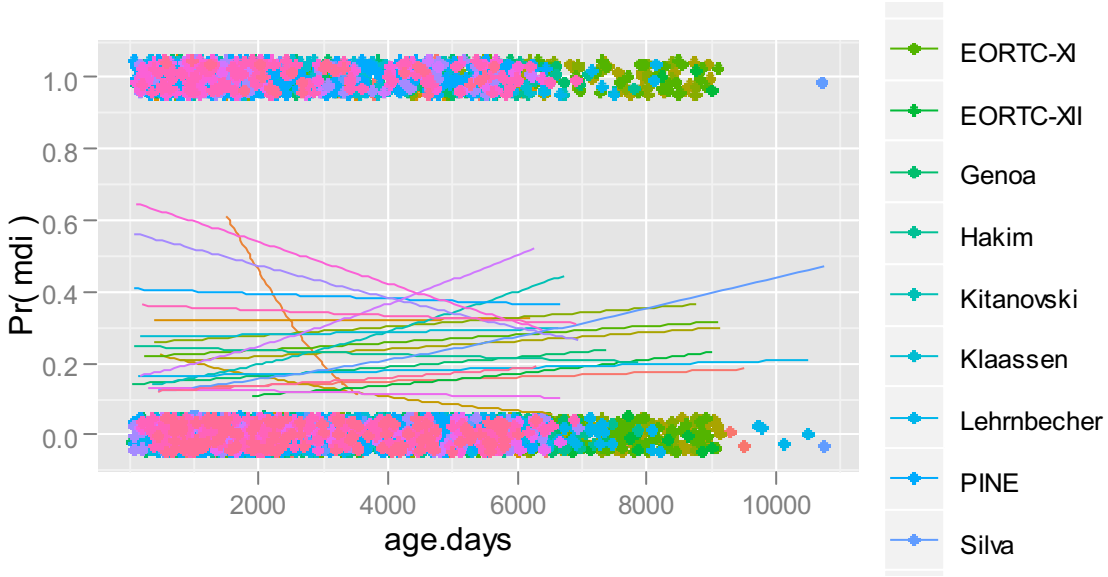


Figure 64: Age (years) and probability of MDI, relationships superimposed

Table 61: Age (in years) and risk of MDI

	Predictor name	OR	95% CI	Beta (age.days)	SE-beta	p-value
Alexander	Age (years)	1.02	0.95 to 1.09	0.000045	0.0001	0.66
BaselSPOG	Age (years)	0.65	0.27 to 1.52	-0.0012	0.0012	0.29
BernSPOG	Age (years)	1.00	0.94 to 1.07	5.4E-06	0.000095	0.95
BonnSPOG	Age (years)	0.91	0.75 to 1.12	-0.00025	0.00028	0.38
EORCT-XIV	Age (years)	1.02	0.96 to 1.09	0.00006	0.000085	0.48
EORTC-IX	Age (years)	1.02	0.99 to 1.06	0.000059	0.000045	0.19
EORTC-XI	Age (years)	1.02	0.99 to 1.06	0.000057	0.000049	0.25

EORTC-XII	Age (years)	1.05	0.85 to 1.3	0.00013	0.0003	0.67
Genoa	Age (years)	1.03	0.99 to 1.07	0.000086	0.000054	0.11
Hakim	Age (years)	0.99	0.94 to 1.03	-3.2E-05	0.000064	0.62
Kitanovski	Age (years)	1.10	0.98 to 1.22	0.00025	0.00015	0.087
Klaassen	Age (years)	1.01	0.96 to 1.05	0.000018	0.000063	0.77
Lehrnbecher	Age (years)	1.01	0.97 to 1.06	0.00003	0.000064	0.64
PINE	Age (years)	0.99	0.96 to 1.02	-2.9E-05	0.000045	0.52
Silva	Age (years)	1.07	0.96 to 1.19	0.00018	0.00015	0.24
Spassova	Age (years)	0.93	0.88 to 0.99	-0.00019	0.000076	0.014
Styjewski	Age (years)	1.11	0.99 to 1.24	0.00028	0.00016	0.088
Sung	Age (years)	0.99	0.86 to 1.13	-0.00004	0.00019	0.83
Tezcan	Age (years)	0.92	0.86 to 0.99	-0.00023	0.0001	0.024
Tissing	Age (years)	0.99	0.94 to 1.04	-3.4E-05	0.000075	0.65
ZurichSPOG	Age (years)	1.03	0.94 to 1.13	0.000089	0.00013	0.51
IPD estimate		**1.01**	**0.99 to 1.02**	**0.000015**	**0.000017**	**0.38**

As with temperature, an analysis was undertaken to compare the common Box-Tidwell transformations of [-2, -1, -0.5, log, 0.5, 1 and 2]. These worsened the AIC values. Splines with df=2,3,4 were also assessed. A single knot (at 6.8yrs) produced a statistically significant (p=0.0018)

small decrease in AIC (5449 cf 5442), two knots at 4.4y and 10.8y had borderline significance (p=0.05) and a very small further decrease in AIC (5440).

Mucositis

Table 62: Study level association of graded mucositis with MDI

	Predictor name	OR	beta	SE-beta	p-value
BaselSPOG	Mucositis (per grade)	0.00	-18.00	5200.00	1.00
BernSPOG	Mucositis (per grade)	0.78	-0.25	0.16	0.10
BonnSPOG	Mucositis (per grade)	0.66	-0.42	0.64	0.51
Hakim	Mucositis (per grade)	1.10	0.08	0.17	0.66
Kitanovski	Mucositis (per grade)	0.99	-0.01	0.26	0.98
Klaassen	Mucositis (per grade)	0.91	-0.09	0.16	0.57
Lehrnbecher	Mucositis (per grade)	0.75	-0.29	0.16	0.07
Spassova	Mucositis (per grade)	1.00	0.05	0.11	0.69
Sung	Mucositis (per grade)	0.73	-0.32	0.53	0.55
ZurichSPOG	Mucositis (per grade)	0.80	-0.22	0.20	0.26
IPD estimate		**0.89 (0.80 to 1.00)**	**-0.11**	**0.058**	**0.052**

Table 63: Study level associations of severe mucositis with MDI

	Predictor name	OR	beta	SE-beta	p-value
BaselSPOG	Severe mucositis	0.00	-18.00	4000.00	1.00
BernSPOG	Severe mucositis	0.51	-0.67	0.53	0.21
BonnSPOG	Severe mucositis	0.00	-15.00	2400.00	1.00
Hakim	Severe mucositis	1.30	0.26	0.54	0.64
Kitanovski	Severe mucositis	0.71	-0.34	0.72	0.63
Klaassen	Severe mucositis	0.78	-0.25	0.52	0.64
Lehrnbecher	Severe mucositis	0.41	-0.89	0.62	0.15
Spassova	Severe mucositis	1.10	0.12	0.33	0.72
Sung	Severe mucositis	4.60	1.50	0.74	0.04
ZurichSPOG	Severe mucositis	0.50	-0.70	0.78	0.37
PINE	Severe mucositis	0.54	-0.61	0.32	0.05
IPD estimate		**0.76 (0.55 to 1.03)**	**-0.28**	**0.16**	**0.078**

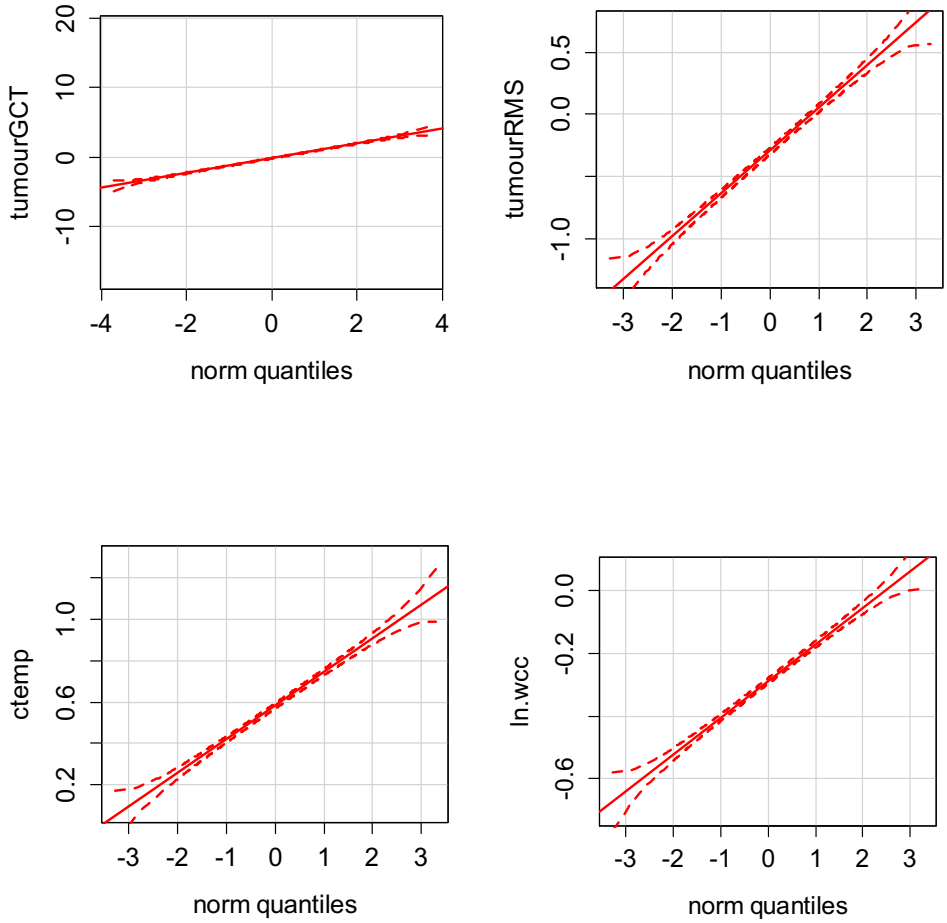
Shaded row shows single added dataset beyond Table 62

Appendix 28. Detail of final multivariate model

Final model beta-estimates and standard errors

Item	Estimate	Std. Error
(Intercept)	-3.996	816.492
tumourAML	0.655	0.255
tumourBrain	-0.457	0.385
tumourCarcinoma	16.668	1455.398
tumourEwings	-0.642	0.662
tumourGCT	-0.069	0.876
tumourHepato	0.475	0.57
tumourHGBrain	-0.345	0.462
tumourHodgkins	-0.408	0.701
tumourHR-NBL	0.921	0.661
tumourLCH	-14.096	1025.44
tumourLGBrain	-14.157	677.944
tumourNBL	0.472	0.495
tumourNHL	-0.471	0.317
tumourOsteo	-1.193	0.566
tumourOther	0.797	0.768
tumourRetino	0.547	0.856
tumourRMS	-0.244	0.319
tumourSarcoma	0.188	0.821
tumourWilms	-0.491	0.663
ctemp (temperature - 37°C)	0.566	0.144
severe.unwell (TRUE)	0.786	0.193
hb (g/dL)	0.18	0.05
ln (wcc x10^6)	-0.299	0.101
ln (amc /mm^3)	-0.209	0.057

QQ quantile plots for selected bootstrap estimates



Bootstrapped calibration plot

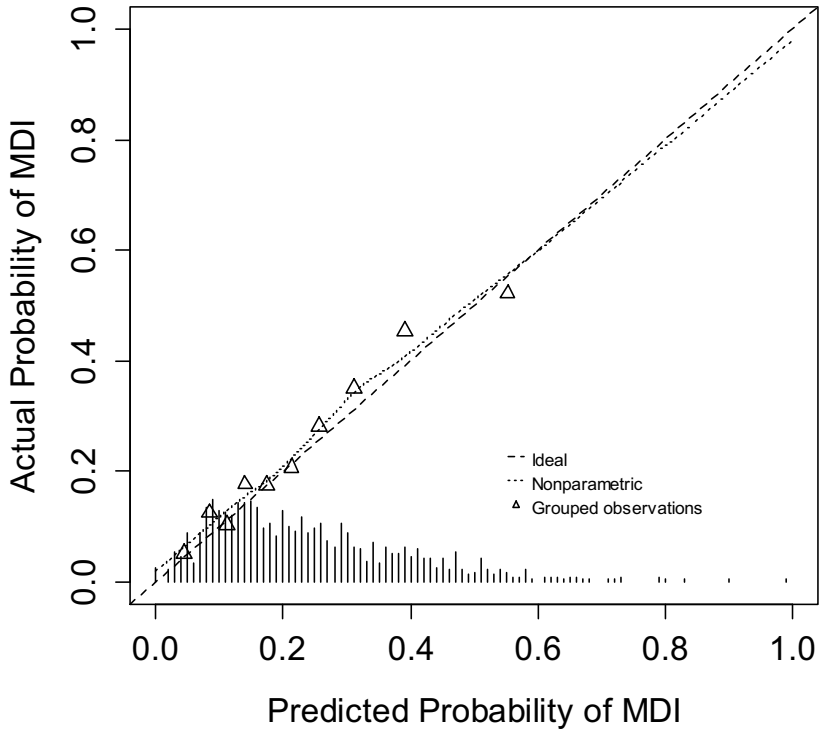
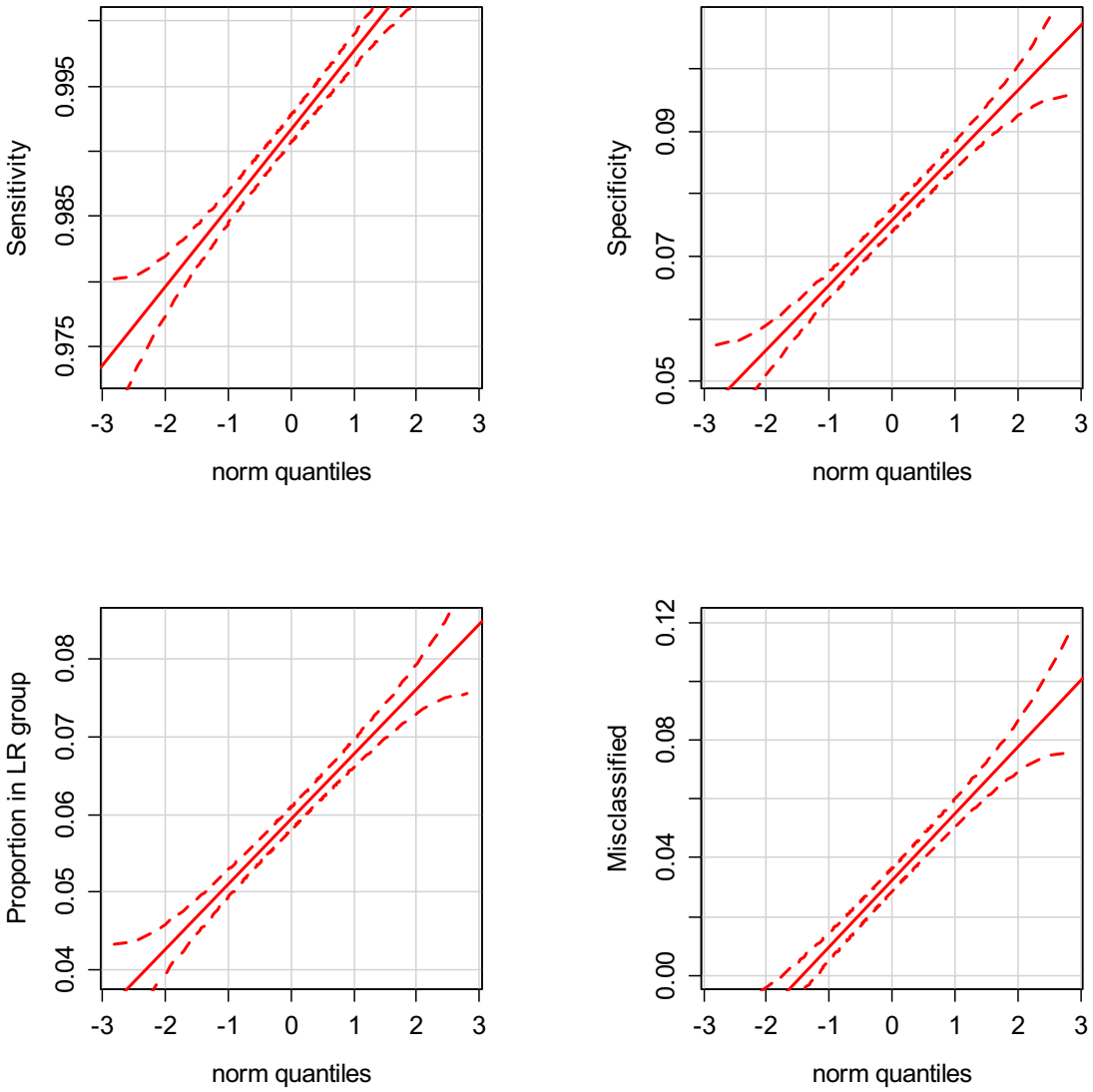
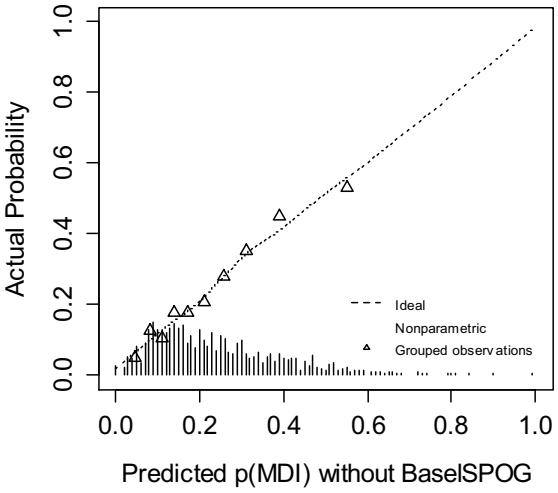


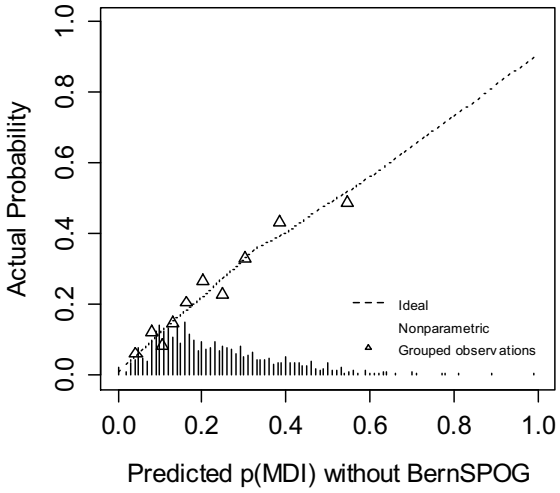
Figure 65: Calibration plot using bootstrapped estimates

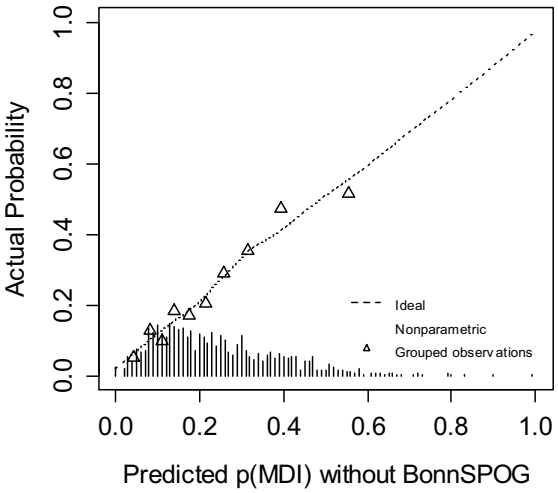
QQ quantile plots for discrimination bootstrap

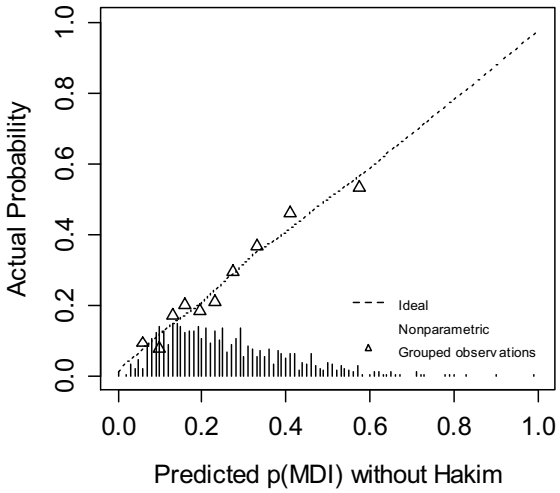


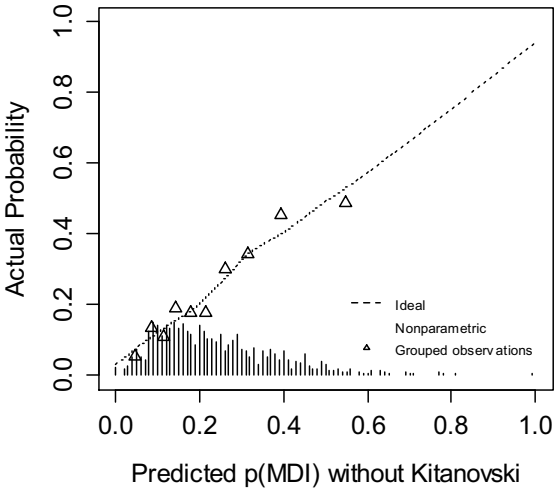
Leave on out calibration plots

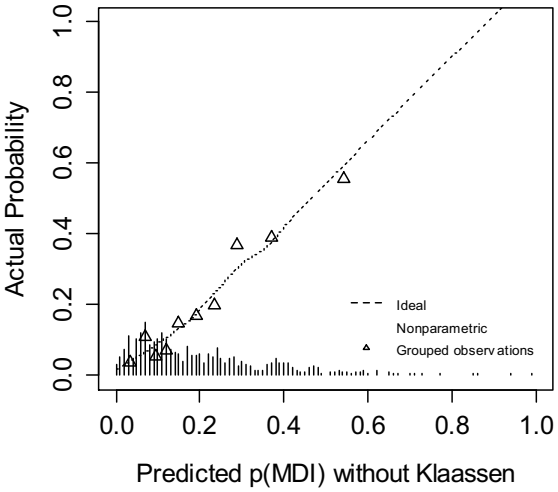


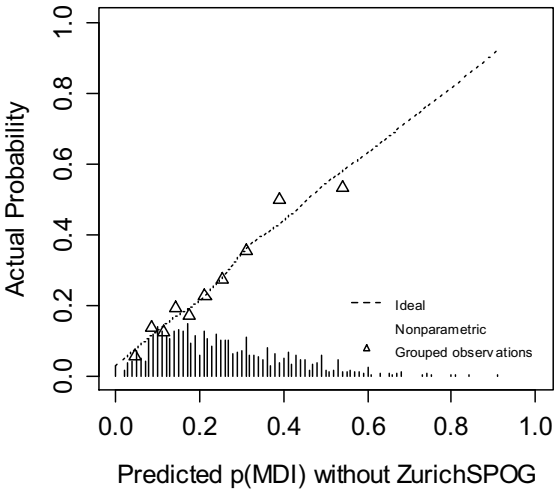







Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Ideal
Nonparametric
Grouped observations
Predicted p(MDI) without Kitanovski


Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Ideal
Nonparametric
Grouped observations
Predicted p(MDI) without Klaassen


Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Ideal
Nonparametric
Grouped observations
Predicted p(MDI) without ZurichSPOG

Comparison of included and excluded studies univariate estimates

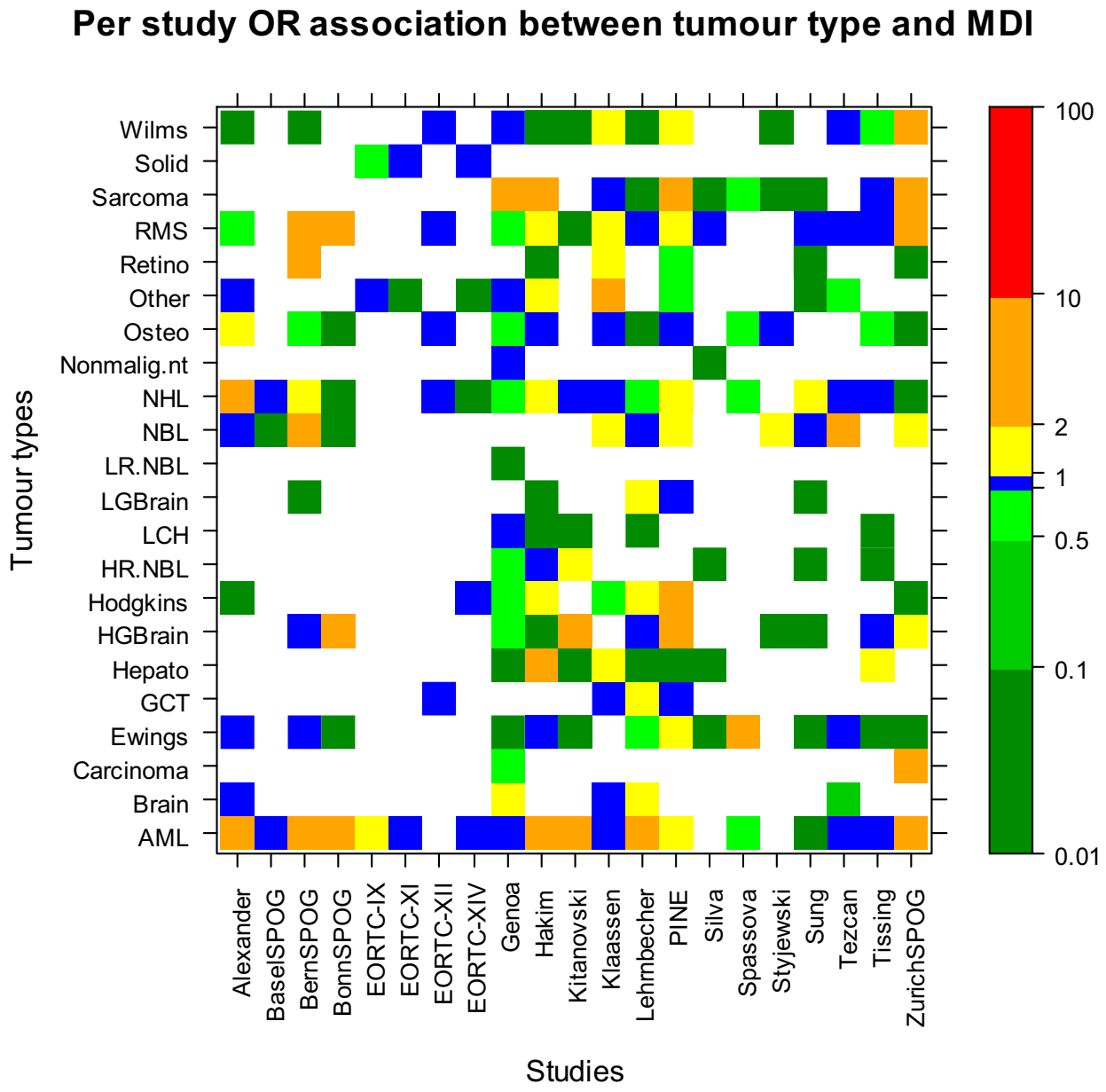
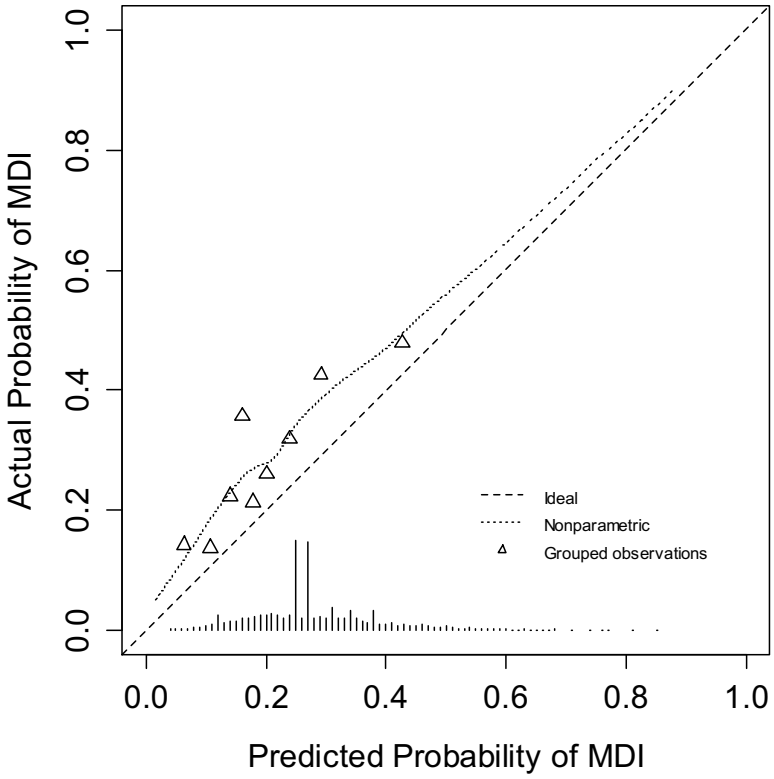


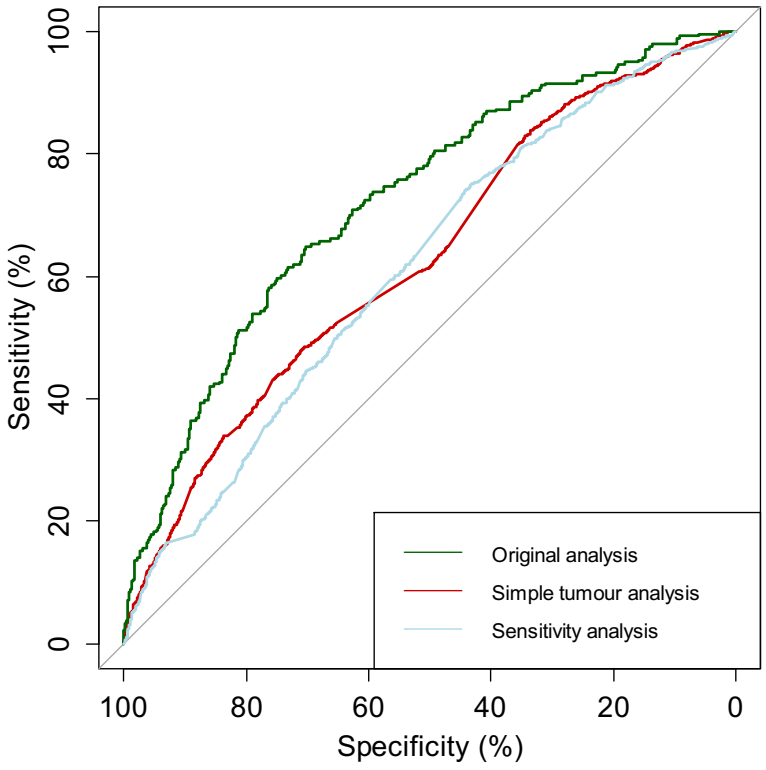
Figure 66: Comparison of included and excluded studies univariate estimates of tumour-type predictive value

Table 64: Comparison of included and excluded studies estimates of predictive value of Hb

Study	OR (Hb)	95% CI	p-value
BaselSPOG	0.92	0.71 to 1.19	0.49
BernSPOG	3.6	0.53 to 25.53	0.2
BonnSPOG	1.1	0.9 to 1.27	0.48
Hakim	0.88	0.48 to 1.62	0.68
Kitanovski	1.1	0.95 to 1.36	0.16
Klaassen	1.1	0.77 to 1.45	0.72
ZurichSPOG	1.1	0.96 to 1.3	0.16
EORCT-XIV	0.79	0.49 to 1.26	0.32
EORTC-IX	1.2	0.98 to 1.52	0.06
EORTC-XI	0.98	0.87 to 1.12	0.81
EORTC-XII	0.85	0.72 to 1.01	0.06
Sung	1.1	0.99 to 1.28	0.07
Tissing	1.2	0.8 to 1.75	0.39

Calibration and discrimination plots for tumour simplified model





Glossary

Adverse event	Detrimental change in health, or side effect, occurring in a patient receiving the treatment.
Afebrile	No fever, normal body temperature.
Aggregate data	Data collected relating to the average values of episodes or events in clinical research studies, sometimes arranged in subgroups, but not referring to individual participants
Anti microbial Therapy	Treatment of infectious disease using agents that either kill microbes or otherwise interfere with microbial growth
Antibiotic resistance	Resistance of a microorganism to an antimicrobial medicine to which it was previously sensitive.
Bacterial infection	Occurs when harmful bacteria enters the body and multiply, causing unpleasant symptoms and/or an adverse event.
Bias	Deviation from the truth
Biomarkers	In this setting, serum (blood-derived) markers of inflammation and infection
Bivariate	Using two variables (cf multivariate)
Bootstrapping	A mathematical technique where a repeated set of analyses are performed on a new collection of data, which have been created by radomly choosing items of the original data, including the posibilty of selecting one item more than once. An internal validation technique.
Calibration	The extent to which the numerical risk predictions from a model agree with the observed (actual) outcomes
Cart	Classification and regression tree - a different approach to discriminatory reasoning than regression analysis
Clinical decision rule (CDR)	A clinical tool designed to be used at the bedside to assist clinical decision making
Clinically documented infection	An infection which has been diagnosed by the use of careful observation and physical examination of a patient.
Clinically relevant	An outcome or event which has a direct relevance to a patient's health status, or which is important in modifying which treatment is received or how it is delivered.
Clostridium difficile	A type of bacteria that lives within the gut which can produce toxins (poisons), which cause illness such as diarrhoea and fever
Co-efficient	The amount of predictive power a covariate has in a predictive model
Covariate	A variable in a prediction model which may be useful in making a prediction more accurate
Crd	Centre for Reviews and Dissemination
C-reactive protein (crp)	A protein that is produced by the liver and found in the blood. May be raised by a variety of problems, including infection.
Critical care	Facilities within a hospital to look after patients whose conditions are life-threatening and need constant close monitoring and support from equipment and medication to keep normal body functions.
Ctc	Common Toxicity Criteria - well documented grading system for adverse effects used in many cancer studies
Discrimination	The ability of a predictive model to separate patients at lower and higher risk of an outcome
Documented infection	An infection which has been diagnosed by clinical examination, or by the detection of pathogenic organisms.
Domain validation	Checking the CDR works when undertaken in a different location, but different clinical setting such as secondary rather than tertiary care

Eortc	European Organisation for Research and Treatment in Cancer
Extrapolation	In data analysis, predicting the value of a parameter outside the range of observed values.
False negative	A result that appears negative but should have been positive, i.e. A test failure
False positive	A result that appears positive but should have been negative, i.e. A test failure.
Febrile neutropenia (fn)	The development of fever, often with other signs of infection, in a patient with neutropenia,
Fever	A raise in body temperature above normal range.
Fixed effect covariate	The same effect is present across each study, and any differences are due to chance sampling
Geographical validation	Checking the CDR works when undertaken in a different location, but similar clinical setting
Granulocyte colony stimulating factor	A type of protein that stimulates the bone marrow to make white blood cells (granulocytes).
Granulocyte macrophage colony stimulating factor	A type of protein that stimulates the bone marrow to make white blood cells (granulocytes and monocytes)
Heterogeneity	A term used to describe the amount of difference of results or effects.
Heuristics	Shortcuts or rules-of-thumb, applied in a variety of situations, for example diagnostic thinking or in shrinkage techniques in statistics
Hierarchical logistic regression	A type of logistic regression technique where the structure of the data (for example, episodes occuring in patients in studies) is explicitly considered to assess if the 'structure' (for example, patient or study) affects how effective the predictor is
Hierarchical summary receiver operator curve (HSROC)	The average ROC curve derived from the individual curves produced from multiple studies in a diagnostic meta-analysis
Homogeneity	A term used to describe the amount of similarity of results or effects
Imputation	Mathematical approach to use values which have been derived from some other source (e.g. Group averages)
Individual participant data (IPD).	Data collected relating to individual episodes or events in clinical research studies (cf. Aggregate data)
Infection	The growth of a pathological organism within the body.
Inflammatory markers	Proteins or other molecules which are raised by inflammatory processes in the body and can be measured, usually by blood tests
Information criterion	A value representing how poorly the data fit a statistical model, with lower numbers indicating a better fit, e.g. Deviance Information Criterion (DIC) or Akaike's Information Criterion (AIC)
Internal validation techniques	Mathematical techniques to test for the likely truth of models (e.g. Predictive models) using the same set of data
Leave-one-out cross-validation	A mathematical technique where a repeated set of analyses are performed on a new collection of data, created by using teh original data set but removing the items contributed by each of the studies in turn, to assess if the results are similar across each grouping. An interval validation technique.

Life threatening infection	An infection which may cause death.
Linear relationship.	(Within regression) The relationship between predictor and outcome variable is a staight line; doubling one quantity doubles the other
Logistic regression	A type of regression techique where the outcome is binary (yes/no)
Low risk	To be safe or without problems.
Meta-analysis	A method of summarising previous research by reviewing and combining the results of a number of different clinical studies
Microbiologically documented infection (mdi)	An infection which has been diagnosed by the detection of pathogenic organisms.
Missing at random (MAR)	Data are missing but related to other known and measured factors
Missing completely at random (MCAR)	Data are missing for no reason but random chance
Missing not at random (MNAR)	Data are missing but related to unknown and unmeasured factors
Monocyte count	The amount of monocytes in blood. Monocytes are a type of white blood cell
Morbidity	A diseased condition or state.
Mortality	Death
Multicollinearity	A mathematical description of the close relationship between quantities which may lead to inaccurate conclusions if not accounted for
Multivariable	Using more than one predictive variable (cf multivariate), usually more than two
Mutivariate	Predicting more than one outcome variable (cf multivariable)
Negative predictive value	Proportion of people with a negative test result who truly didn't have disease
Net reclassification improvement (NRI)	Measure of the overall 'benefit' of a new classification model
Neutropenia	An abnormally low number of neutrophils, the most important type of white blood cell to fight off bacterial infections.
Neutropenic sepsis	An abnormal decrease in the number of neutrophils in the blood together with infection.
Neutrophil	A type of white blood cell, important in fighting off particularly bacterial infections.
Neutrophil count	This test measures the number of neutrophils in blood. Neutrophils are a type of white blood cell
Non-linear form	(Within regression) The relationship between predictor and outcome variable is not staight line; doubling one quantity does not double the other
Odds ratio	A measure of treatment effectiveness. The odds of an event happening in the intervention group, divided by the odds of it happening in the control group. The 'odds' is the ratio of non-events to events.
Outcome	An end result; a consequence.
PGP-encrypted	Electronic communication using a highly secure shared-passkey encryption
Picnicc	Predicting Infectious complications In Children with Cancer
Positive predictive value	Proportion of people with a postive test result who did have disease
Predictive	Mathematical models using data to estaime a probability (chance) of a specific outcome

models	or outcomes
Primary prophylaxis	A preventative intervention administered in all cycles of chemotherapy.
Prognostic study	A study that examines selected predictive variables, or risk factors, and assesses their influence on the outcome of a disease.
Prospective study	A study in which people are entered into research and then followed up over a period of time with future events recorded as they happen.
Publication bias	Also known as reporting bias. A bias caused by only a subset of all the relevant data being available. The publication of research can depend on the nature and direction of the study results. Studies in which an intervention is not found to be effective are sometimes not published. Because of this, systematic reviews that fail to include unpublished studies may overestimate the true effect of an intervention. In addition, a published report might present a biased set of results (e.g. Only outcomes or sub-groups where a statistically significant difference was found.
Qualitative study	A study used to explore and understand peoples' beliefs, experiences, attitudes, behaviour and interactions.
Quality adjusted life years (qalys)	A measure of health outcome which looks at both length of life and quality of life. QALYS are calculated by estimating the years of life remaining for a patient following a particular care pathway and weighting each year with a quality of life score (on a 0 to 1 scale). One QALY is equal to 1 year of life in perfect health, or 2 years at 50% health, and so on
Quality of life	An overall appraisal of well being.
Radiotherapy	A treatment for cancer that uses high energy ionising radiation to kill cells.
Random effect covariate	The effects come from a normal distribution of true effects; the estimates are both different by chance and real differences in true effect between studies
Randomised controlled trials (rcts)	A clinical trial in which subjects are randomised to different groups for the purpose of studying the effect of a new intervention, for example a drug or other therapy.
Rcpch	Royal College of Paediatrics and Child Health
Receiver operator curve (ROC)	A curve describing the relationship between the sensitivity and specificity of a test
Regression	A mathematical technique which relates one (or more) measured variables to an outcome variable
Relative risk (also known as risk ratio)	The ratio of risk in the intervention group to the risk in the control group. The risk (proportion, probability or rate) is the ratio of people with an event in a group to the total in the group. A relative risk (RR) of 1 indicates no difference between comparison groups. For undesirable outcomes, an RR that is less than 1 indicates that the intervention was effective in reducing the risk of that outcome
Retrospective data	Data that deals with the present/past and does not involve studying future events.
Risk	The chance of an adverse outcome happening.
Risk assessment tool	A tool, usually a score from pieces of information given by patients, blood tests and examination finding, which is used to assess a patient's risk of a particular outcome.
Risk stratification	The process of grouping people into categories with different probabilities of a specific, usually adverse, outcome
Sensitivity	The proportion of individuals who have disease correctly identified by the study test
Sepsis	The body's response to an infection
Septic shock	Septic shock is a medical emergency caused by decreased tissue perfusion and oxygen delivery as a result of severe infection and sepsis,
Serious bacterial	A bacterial infection with a high chance of causing siginifcant morbidity or death

infection	
Severe sepsis	A life-threatening form of sepsis
Short-term mortality	Death within a short period of time, for instance 30 days from onset of fever.
Shrinkage	A mathematical technique used to improve the chances of a predictive model being accurate in practice, related to internal validation techniques
Siop	International Society of Paediatric Oncology
Specificity	The proportion of individuals who do not have a disease and who are correctly identified by the study test.
Statistical significance	A mathematical concept, to be understood as 'unlikely to be due to chance'
Stem cell transplant	A procedure that replaces the cells in a patient which make blood. (Haemopoietic stem cell transplant.)
Step down	Decrease or reduction in treatment or medication.
Systematic review	A review of the literature done to answer a defined question often using quantitative methods to summarise the results.
Temporal validation	Checking the CDR works when undertaken at a different point in time
Transformation	A mathematical technique of consistently modifiying varaibels, for example taking the logarithm or square-root of a quantity
Treatment failure	Unsuccessful results or consequences of treatments used in combating disease.
True negative	When testing for a condition or disease, this result confirms the absence of the condition in an individual who genuinely does not have the condition in question. (Contrast with false negative (see above) where the test may incorrectly indicate that the individual is free from the condition being investigated. The condition is present but not detected by the test.).
True positive	When testing for a condition or disease, this result confirms the presence of the condition in question in individuals who have it. (Compare with false positive where the test may incorrectly indicate that the individual has a condition, but in fact they do not.)
Tunnel infection	A device-related infection seen in central venous access devices, related to the tube as it passes beneath the skin.

References

1. Moscucci O, Herring R, Berridge V: **Networking Health Research in Britain: The Post-War Childhood Leukaemia Trials.** *Twentieth Century Brit Hist* 2009, **20:**23-52.
2. Pritchard-Jones K, Kaatsch P, Steliarova-Foucher E, Stiller CA, Coebergh JWW: **Cancer in children and adolescents in Europe: Developments over 20 years and future challenges.** *European Journal of Cancer* 2006, **42:**2183-2190.
3. Freycon F, Trombert-Paviot B: **Trends in treatment-related deaths (TRDs) in childhood cancer and leukemia over time: A follow-up of patients included in the childhood cancer registry of the Rhone-Alpes region in France (ARCERRA).** *Pediatric Blood & Cancer* 2008, **50:**1213-1220.

4. Hargrave DR, Hann IM: **Progressive reduction in treatment-related deaths in Medical Research Council childhood lymphoblastic leukaemia trials from 1980 to 1997 (UKALL VIII, X and XI).** *British Journal of Haematology* 2001, **112:**293-299.
5. Hann I, Viscoli C, Paesmans M, Gaya H, Glauser M: **A comparison of outcome from febrile neutropenic episodes in children compared with adults: results from four EORTC studies. International Antimicrobial Therapy Cooperative Group (IATCG) of the European Organization for Research and Treatment of Cancer (EORTC).** *British Journal of Haematology* 1997, **99:**580-588.
6. Bate J, Ladhani S, Sharland M, Chisholm J, Lamagni T, Ramsay M, Johnson A, Pebody R: **Infection-related mortality in children with malignancy in England and Wales, 2003-2005.** *Pediatric Blood & Cancer* 2009.
7. Bodey GP, Buckley M, Sathe YS, Freireich EJ: **Quantitative relationships between circulating leukocytes and infection in patients with acute leukemia.** *Annals of Internal Medicine* 1966, **64:**328-340.
8. Schimpff SC: **Empiric antibiotic therapy for granulocytopenic cancer patients.** *The American journal of medicine* 1986, **80:**13-20.
9. Hartel C, Deuster M, Lehrnbecher T, Schultz C: **Current approaches for risk stratification of infectious complications in pediatric oncology.** *Pediatr Blood Cancer* 2007.
10. Wicki S, Keisker A, Aebi C, Leibundgut K, Hirt A, Ammann RA: **Risk prediction of fever in neutropenia in children with cancer: a step towards individually tailored supportive therapy?** *Pediatric Blood & Cancer* 2008, **51:**778-783.
11. Petrilli AS, Melaragno R, Bianchi A, Kusano E, Barros KV, Silva AA: **[Fever and neutropenia in children with cancer: a new therapeutic proposal].** *Amb; Revista Da Associacao Medica Brasileira* 1991, **37:**173-180.
12. Baorto EP, Aquino VM, Mullen CA, Buchanan GR, DeBaun MR: **Clinical parameters associated with low bacteremia risk in 1100 pediatric oncology patients with fever and neutropenia.** *Cancer* 2001, **92:**909-913.
13. Klastersky J, Paesmans M, Rubenstein EB, Boyer M, Elting L, Feld R, Gallagher J, Herrstedt J, Rapoport B, Rolston K, Talcott J: **The Multinational Association for Supportive Care in Cancer risk index: A multinational scoring system for identifying low-risk febrile neutropenic cancer patients.** *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 2000, **18:**3038-3051.
14. Vidal L, Paul M, Ben-Dor I, Pokroy E, Soares-Weiser K, Leibovici L: **Oral versus intravenous antibiotic treatment for febrile neutropenia in cancer patients.** *Cochrane Database of Systematic Reviews (Online)* 2004**:**CD003992-CD003992.
15. Innes H, Billingham L, Gaunt C, Steven N, Marshall E: **Management of febrile neutropenia in the United Kingdom: time for a national trial?** *BRITISH JOURNAL OF CANCER* 2005, **93:**1324-1328.
16. Sung L, Phillips R, Lehrnbecher T: **Time for paediatric febrile neutropenia guidelines - children are not little adults.** *Eur J Cancer* 2011, **47:**811-813.
17. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS: **Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group.** *JAMA* 2000, **284:**79-84.
18. Rackoff WR, Gonin R, Robinson C, Kreissman SG, Breitfeld PB: **Predicting the risk of bacteremia in childen with fever and neutropenia.** *J Clin Oncol* 1996, **14:**919-924.
19. Ammann RA, Hirt A, Lüthy AR, Aebi C: **Identification of children presenting with fever in chemotherapy-induced neutropenia at low risk for severe bacterial infection.** *Medical and Pediatric Oncology* 2003, **41:**436-443.
20. Baorto EP, Aquino VM, Mullen CA, Buchanan GR, DeBaun MR: **Clinical parameters associated with low bacteremia risk in 1100 pediatric oncology patients with fever and neutropenia.** *Cancer* 2001, **92:**909-913.

21. Klaassen RJ, Goodman TR, Pham B, Doyle JJ: **"Low-Risk" Prediction Rule for Pediatric Oncology Patients Presenting With Fever and Neutropenia.** *J Clin Oncol* 2000, **18:**1012-.
22. Santolaya ME, Alvarez AM, Avil‚s CL, Becker A, Cofr, J, Enr¡quez N, O'Ryan M, Pay E, Salgado C, Silva P, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Prospective evaluation of a model of prediction of invasive bacterial infection risk among children with cancer, fever, and neutropenia.** *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 2002, **35:**678-683.
23. Alexander SW, Wade KC, Hibberd PL, Parsons SK: **Evaluation of risk prediction criteria for episodes of febrile neutropenia in children with cancer.** *J Pediatr Hematol Oncol* 2002, **24:**38-42.
24. Gala Peralta S, Cardesa Salzman T, Garcia Garcia JJ, Estella Aguado J, Gene Giralt A, Luaces Cubells C: **[Bacteraemia risk criteria in the paediatric febrile neutropenic cancer patient].** *Clinical & Translational Oncology: Official Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico* 2005, **7:**165-168.
25. Lucas KG, Brown AE, Armstrong D, Chapman D, Heller G: **The identification of febrile, neutropenic children with neoplastic disease at low risk for bacteremia and complications of sepsis.** *Cancer* 1996, **77:**791-798.
26. Klaassen RJ, Goodman TR, Pham B, Doyle JJ: **"Low-risk" prediction rule for pediatric oncology patients presenting with fever and neutropenia.** *J Clin Oncol* 2000, **18:**1012-1019.
27. Santolaya M, Cofre J, Beresi V: **C-reactive protein: a valuable aid for the management of febrile children with cancer and neutropenia.** *Clin Infect Dis* 1994, **18:**589-595.
28. Nijhuis CO, Kamps WA, Daenen SMG, Gietema JA, van der Graaf WTA, Groen HJM, Vellenga E, ten Vergert EM, Vermeulen KM, de Vries-Hospers HG, de Bont ESJM: **Feasibility of Withholding Antibiotics in Selected Febrile Neutropenic Cancer Patients.** *J Clin Oncol* 2005, **23:**7437-7444.
29. Lehrnbecher T, Fleischhack G, Hanisch M, et al.: **Circluating levels and promoter polymorphisms of interleukins-6 and 8 in pediatric cancer patients with fever and neutropenia.** *Hematologica* 2004, **89:**234-236.
30. Stryjewski GR, Nylen ES, Bell MJ, et al.: **Interleukin-6, interleukin-8, and a rapid and sensitive assay for calcitonin precursors for the determination of bacterial sepsis in febrile neutropenic children.** *Pediatr Crit Care Med* 2005, **6:**129-135.
31. Tillett WS, Francis T: **SEROLOGICAL REACTIONS IN PNEUMONIA WITH A NON-PROTEIN SOMATIC FRACTION OF PNEUMOCOCCUS.** *The Journal of Experimental Medicine* 1930, **52:**561-571.
32. McWilliam S, Riordan A: **How to use: C-reactive protein.** *Arch Dis Child Educ Pract Ed* 2010, **95:**55-58.
33. Santolaya ME, Alvarez AM, Aviles CL, Becker A, Cofre J, Cumsille MA, O'Ryan ML, Paya E, Salgado C, Silva P, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Early hospital discharge followed by outpatient management versus continued hospitalization of children with cancer, fever, and neutropenia at low risk for invasive bacterial infection.** *J Clin Oncol* 2004, **22:**3784-3789.
34. Paganini H, Rodriguez-Brieshcke T, Zubizarreta P, Latella A, Firpo V, Casimir L, Armada A, Fernandez C, Caceres E, Debbag R: **Oral ciprofloxacin in the management of children with cancer with lower risk febrile neutropenia.** *Cancer* 2001, **91:**1563-1567.
35. Shenep JL, Flynn PM, Baker DK, Hetherington SV, Hudson MM, Hughes WT, Patrick CC, Roberson PK, Sandlund JT, Santana VM, Sixbey JW, Slobod KS: **Oral cefixime is similar to continued intravenous antibiotics in the empirical treatment of febrile neutropenic children with cancer.** *Clin Infect Dis* 2001, **32:**36-43.
36. Petrilli AS, Dantas LS, Campos MC, Tanaka C, Ginani VC, Seber A: **Oral ciprofloxacin vs. intravenous ceftriaxone administered in an outpatient setting for fever and neutropenia in**

low-risk pediatric oncology patients: randomized prospective trial. *Med Pediatr Oncol* 2000, **34:**87-91.
37. Oude Nijhuis C, Kamps WA, Daenen SMG, Gietema JA, van der Graaf WTA, Groen HJM, Vellenga E, Ten Vergert EM, Vermeulen KM, de Vries-Hospers HG, de Bont ESJM: **Feasibility of withholding antibiotics in selected febrile neutropenic cancer patients.** *Journal of Clinical Oncology* 2005, **23:**7437-7444.
38. Sung L, Feldman BM, Schwamborn G, Paczesny D, Cochrane A, Greenberg ML, Maloney AM, Hendershot EI, Naqvi A, Barrera M, Llewellyn-Thomas HA: **Inpatient versus outpatient management of low-risk pediatric febrile neutropenia: measuring parents' and healthcare professionals' preferences.** *J Clin Oncol* 2004, **22:**3922-3929.
39. James RM, Kinsey SE: **The investigation and management of chronic neutropenia in children.** *Archives of Disease in Childhood* 2006, **91:**852-858.
40. Sung L, Nathan PC, Alibhai SMH, Tomlinson GA, Beyene J: **Meta-analysis: Effect of Prophylactic Hematopoietic Colony-Stimulating Factors on Mortality and Outcomes of Infection.** *Ann Intern Med* 2007, **147:**400-411.
41. Sung L, Beyene J, Hayden J, Nathan PC, Lange B, Tomlinson GA: **A Bayesian meta-analysis of prophylactic granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor in children with cancer.** *American Journal of Epidemiology* 2006, **163:**811-817.
42. National Institute of Clinical Excellence: **Urinary tract infection: diagnosis, treatment and long-term management of urinary tract infection in children.** *Urinary tract infection: diagnosis, treatment and long-term management of urinary tract infection in children* 2007.
43. Scottish Intercollegiate Guidelines Network: **SIGN 34 - Management of sore throat and indications for tonsillectomy.** *SIGN 34 - Management of sore throat and indications for tonsillectomy* 1999.
44. Joint Formulary Committee: *British National Formulary. 56.* London: British Medical Association and Royal Pharmaceutical Society of Great Britain; 2008.
45. Dellinger EP, Hausmann SM, Bratzler DW, Johnson RM, Daniel DM, Bunt KM, Baumgardner GA, Sugarman JR: **Hospitals collaborate to decrease surgical site infections.** *American Journal of Surgery* 2005, **190:**9-15.
46. Gafter-Gvili A, Fraser A, Paul M, van de Wetering M, Kremer L, Leibovici L: **Antibiotic prophylaxis for bacterial infections in afebrile neutropenic patients following chemotherapy.** *Cochrane Database Syst Rev* 2005, **19:**CD004386.
47. Reuter S, Kern WV, Sigge A, D"hner H, Marre R, Kern P, von Baum H: **Impact of fluoroquinolone prophylaxis on reduced infection-related mortality among patients with neutropenia and hematologic malignancies.** *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 2005, **40:**1087-1093.
48. Kern WV, Klose K, Jellen-Ritter AS, Oethinger M, Bohnert J, Kern P, Reuter S, von Baum H, Marre R: **Fluoroquinolone resistance of Escherichia coli at a cancer center: epidemiologic evolution and effects of discontinuing prophylactic fluoroquinolone use in neutropenic patients with leukemia.** *European Journal of Clinical Microbiology & Infectious Diseases:* 2005, **24:**111-118.
49. Andre FE, Booy R, Bock HL, Clemens J, Datta SK, John TJ, Lee BW, Lolekha S, Peltola H, Ruff TA, Santosham M, Schmitt HJ: **Vaccination greatly reduces disease, disability, death and inequity worldwide.** *Bulletin of the World Health Organization* 2008, **86:**140-146.
50. Goossen GM, Kremer LC, van de Wetering MD: **Influenza vaccination in children being treated with chemotherapy for cancer.** *Cochrane Database Syst Rev* 2009**:**CD006484.
51. Boragina M, Patel H: **Management of febrile neutropenia in pediatric oncology patients: A Canadian survey.** *Pediatric Blood & Cancer* 2007, **48:**521-526.
52. Sartori AMC: **A review of the varicella vaccine in immunocompromised individuals.** *International Journal of Infectious Diseases* 2004, **8:**259-270.

53. Phillips B, Selwood K, Lane SM, Skinner R, Gibson F, Chisholm JC: **Variation in policies for the management of febrile neutropenia in United Kingdom Children's Cancer Study Group centres.** *Arch Dis Child* 2007, **92:**495-498.
54. Chamberlain JD, Smibert E, Skeen J, Alvaro F: **Prospective audit of treatment of paediatric febrile neutropenia in Australasia.** *Journal of Paediatrics and Child Health* 2005, **41:**598-603.
55. Stabell N, Nordal E, Stensvold E, Gammelsrud KW, Lund B, Taxt A, Buhring F, Greve-Isdahl M, Fornebo HP, Simonsen GS, Klingenberg C: **Febrile neutropenia in children with cancer: A retrospective Norwegian multicentre study of clinical and microbiological outcome.** *Scandinavian Journal of Infectious Diseases* 2007,**:**1 - 7.
56. Orudjev E, Lange BJ: **Evolving concepts of management of febrile neutropenia in children with cancer.** *Medical and Pediatric Oncology* 2002, **39:**77-85.
57. Ammann RA, Simon A, de Bont ESJM: **Low risk episodes of fever and neutropenia in pediatric oncology: Is outpatient oral antibiotic therapy the new gold standard of care?** *Pediatric Blood & Cancer* 2005, **45:**244-247.
58. Chisholm JC, Dommett R: **The evolution towards ambulatory and day-case management of febrile neutropenia.** *Br J Haematol* 2006, **135:**3-16.
59. Bachmann LM, Kolb E, Koller MT, Steurer J, Riet Gt: **Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review.** *BMJ* 2003, **326:**417.
60. Goodacre S, Sutton AJ, Sampson FC: **Meta-analysis: The value of clinical assessment in the diagnosis of deep venous thrombosis.** *Annals of Internal Medicine* 2005, **143:**129-139.
61. Taekema HC, Landham PR, Maconochie I: **Towards evidence based medicine for paediatricians. Distinguishing between transient synovitis and septic arthritis in the limping child: how useful are clinical prediction tools?** *Arch Dis Child* 2009, **94:**167-168.
62. Bleeker SE, Derksen-Lubsen G, Grobbee DE, Donders AR, Moons KG, Moll HA: **Validating and updating a prediction rule for serious bacterial infection in patients with fever without source.** *Acta Paediatr* 2007, **96:**100-104.
63. Dubos F, Martinot A, Gendrel D, Breart G, Chalumeau M: **Clinical decision rules for evaluating meningitis in children.** *Curr Opin Neurol* 2009, **22:**288-293.
64. Lynch T, Platt R, Gouin S, Larson C, Patenaude Y: **Can we predict which children with clinically suspected pneumonia will have the presence of focal infiltrates on chest radiographs?** *Pediatrics* 2004, **113:**e186-189.
65. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM: **Validation, updating and impact of clinical prediction rules: A review.** *Journal of Clinical Epidemiology* 2008, **61:**1085-1094.
66. Selby PJ, Powles RL, Easton D, Perren TJ, Stolle K, Jameson B, Fiddian AP, Tryhorn Y, Stern H: **The prophylactic role of intravenous and long-term oral acyclovir after allogeneic bone marrow transplantation.** *Br J Cancer* 1989, **59:**434-438.
67. **Oxford English Dictionary.** *Oxford English Dictionary* Second Edition edition. City: Oxford University Press; 1989.
68. Castagnola E, Cesaro S, Giacchino M, Livadiotti S, Tucci F, Zanazzo G, Caselli D, Caviglia I, Parodi S, Rondelli R, Cornelli PE, Mura R, Santoro N, Russo G, De Santis R, Buffardi S, Viscoli C, Haupt R, Rossi MR: **Fungal infections in children with cancer: a prospective, multicenter surveillance study.** *Pediatr Infect Dis J* 2006, **25:**634-639.
69. Tezcan G, Kupesiz A, Ozturk F, Ogunc D, Gultekin M, Yesilipek A, Hazar V: **Episodes of fever and neutropenia in children with cancer in a tertiary care medical center in Turkey.** *Pediatric Hematology & Oncology* 2006, **23:**217-229.
70. Andre M, Odenholt I, Schwan A, Axelsson I, Eriksson M, Hoffman M, Molstad S, Runehagen A, Lundborg CSl, Wahlstrom R: **Upper respiratory tract infections in general practice: diagnosis, antibiotic prescribing, duration of symptoms and use of diagnostic tests.** *Scandinavian Journal of Infectious Diseases* 2002, **34:**880-886.
71. Kyzas PA, Loizou KT, Ioannidis JPA: **Selective Reporting Biases in Cancer Prognostic Factor Studies.** *J Natl Cancer Inst* 2005, **97:**1043-1055.

72. Ellenberg JH, Nelson KB: **Sample selection and the natural history of disease. Studies of febrile seizures.** *JAMA: The Journal of the American Medical Association* 1980, **243:**1337-1340.
73. Parkes J, Kerr C, McDowell BC, Cosgrove AP: **Recruitment bias in a population-based study of children with cerebral palsy.** *Pediatrics* 2006, **118:**1616-1622.
74. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J: **Sources of Variation and Bias in Studies of Diagnostic Accuracy: A Systematic Review.** *Ann Intern Med* 2004, **140:**189-202.
75. Wood L, Egger M, Gluud LL, Schulz KF, J ni P, Altman DG, Gluud C, Martin RM, Wood AJG, Sterne JAC: **Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study.** *BMJ* 2008, **336:**601-605.
76. Kyzas PA, Denaxa-Kyza D, Ioannidis JPA: **Quality of Reporting of Cancer Prognostic Marker Studies: Association With Reported Prognostic Effect.** *J Natl Cancer Inst* 2007, **99:**236-243.
77. Young B, Dixon-Woods M, Findlay M, Heney D: **Parenting in a crisis: conceptualising mothers of children with cancer.** *Soc Sci Med* 2002, **55:**1835-1847.
78. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, Burchill SA: **Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future.** *Br J Cancer* 2003, **88:**1191-1198.
79. Ammann RA, Hirt A, Luthy AR, Aebi C: **Identification of children presenting with fever in chemotherapy-induced neutropenia at low risk for severe bacterial infection.** *Medical & Pediatric Oncology* 2003, **41:**436-443.
80. Altman DG, Lausen B, Sauerbrei W, Schumacher M: **Dangers of using "optimal" cutpoints in the evaluation of prognostic factors.** *J Natl Cancer Inst* 1994, **86:**829-835.
81. Harrell FE, Jr., Lee KL, Mark DB: **Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.** *Stat Med* 1996, **15:**361-387.
82. Sun GW, Shook TL, Kay GL: **Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis.** *J Clin Epidemiol* 1996, **49:**907-916.
83. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD: **Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets.** *Stat Med* 2000, **19:**1059-1079.
84. Altman DG, Royston P: **What do we mean by validating a prognostic model?** *Stat Med* 2000, **19:**453-473.
85. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD: **Internal validation of predictive models: efficiency of some procedures for logistic regression analysis.** *J Clin Epidemiol* 2001, **54:**774-781.
86. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y: **Updating methods improved the performance of a clinical prediction model in new patients.** *J Clin Epidemiol* 2008, **61:**76-86.
87. Moons KG, Harrell FE, Steyerberg EW: **Should scoring rules be based on odds ratios or regression coefficients?** *J Clin Epidemiol* 2002, **55:**1054-1055.
88. Dawes RM: **The robust beauty of improper linear models in decision making.** *American Psychologist* 1979, **34:**571-582.
89. Gigerenzer G: *Reckoning with Risk: Learning to Live with Uncertainty.* Penguin; 2003.
90. Wardle FJ, Collins W, Pernet AL, Whitehead MI, Bourne TH, Campbell S: **Psychological Impact of Screening for Familial Ovarian Cancer.** *J Natl Cancer Inst* 1993, **85:**653-657.
91. Lee MT, Piomelli S, Granger S, Miller ST, Harkness S, Brambilla DJ, Adams RJ: **Stroke Prevention Trial in Sickle Cell Anemia (STOP): extended follow-up and final results.** *Blood* 2006, **108:**847-852.

92. Montgomery K: *How Doctors Think: Clinical judgment and the practice of medicine: Clinical Judgement and the Practice of Medicine.* OUP USA; 2005.
93. Norman G: **Research in clinical reasoning: past history and current trends.** *Medical Education* 2005, **39:**418-427.
94. Elstein AS, Schwarz A: **Evidence base of clinical diagnosis: Clinical problem solving and diagnostic decision making: selective review of the cognitive literature.** *Bmj* 2002, **324:**729-732.
95. Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, Scott C, Perera R: **Diagnostic strategies used in primary care.** *BMJ* 2009, **338:**b946.
96. Graber ML, Franklin N, Gordon R: **Diagnostic Error in Internal Medicine.** *Arch Intern Med* 2005, **165:**1493-1499.
97. Redelmeier DA: **The Cognitive Psychology of Missed Diagnoses.** *Ann Intern Med* 2005, **142:**115-120.
98. **House. M.D** [http://www.fox.com/house/showInfo/]
99. Doyle AC: *The Adventures of Sherlock Holmes.* Penguin Classics; 2007.
100. Freeman RA: *John Thorndyke's Cases related by Christopher Jervis and edited by R. Austin Freeman.* 2004.
101. Dawes RM, Faust D, Meehl PE: **Clinical Versus Actuarial Judgment.** *Science* 1989, **243:**1668-1674.
102. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C: **Clinical Versus Mechanical Prediction: A Meta-Analysis.** *Psychological Assessment* 2000, **12:**19-30.
103. Quezada G, T S, Chan K, Rolston KV, Mullen CA: **Medical and non-medical barriers to outpatient treatment of fever and neutropenia in children with cancer.** *Pediatric Blood & Cancer* 2007, **48:**273-277.
104. Lee TH, Pearson SD, Johnson PA, Garcia TB, Weisberg MC, Guadagnoli E, Cook EF, Goldman L: **Failure of Information as an Intervention to Modify Clinical Management: A Time-Series Trial in Patients with Acute Chest Pain.** *Ann Intern Med* 1995, **122:**434-437.
105. Bessen T, Clark R, Shakib S, Hughes G: **A multifaceted strategy for implementation of the Ottawa ankle rules in two emergency departments.** *BMJ* 2009, **339:**b3056-b3056.
106. Gaddis GM, Greenwald P, Huckson S: **Toward improved implementation of evidence-based clinical algorithms: clinical practice guidelines, clinical decision rules, and clinical pathways.** *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine* 2007, **14:**1015-1022.
107. Roukema J, Steyerberg EW, Lei Jvd, Moll HtA: **Randomized Trial of a Clinical Decision Support System: Impact on the Management of Children with Fever without Apparent Source.** *Journal of the American Medical Informatics Association : JAMIA* 2008, **15**.
108. Gifford DR, Holloway RG, Vickrey BG: **Systematic Review of Clinical Prediction Rules for Neuroimaging in the Evaluation of Dementia.** *Arch Intern Med* 2000, **160:**2855-2862.
109. Riley RD, Heney D, Jones DR, Sutton AJ, Lambert PC, Abrams KR, Young B, Wailoo AJ, Burchill SA: **A Systematic Review of Molecular and Biological Tumor Markers in Neuroblastoma.** *Clin Cancer Res* 2004, **10:**4-12.
110. Hess EP, Thiruganasambandamoorthy V, Wells GA, Erwin P, Jaffe AS, Hollander JE, Montori VM, Stiell IG: **Diagnostic accuracy of clinical prediction rules to exclude acute coronary syndrome in the emergency department setting: a systematic review.** *CJEM: Canadian Journal of Emergency Medical Care* 2008, **10:**373-382.
111. Altman DG, Riley RD: **Primer: an evidence-based approach to prognostic markers.** *Nature Clinical Practice Oncology* 2005, **2:**466-472.
112. Planche T, Aghaizu A, Holliman R, Riley P, Poloniecki J, Breathnach A, Krishna S: **Diagnosis of Clostridium difficile infection by toxin detection kits: a systematic review.** *Lancet Infect Dis* 2008, **8:**777-784.

113. Phillips B, Wade R, Stewart LA, Sutton AJ: **Systematic review and meta-analysis of the discriminatory performance of risk prediction rules in febrile neutropaenic episodes in children and young people.** *Eur J Cancer* 2010, **46:**2950-2964.
114. Phillips RS, Wade R, Lehrnbecher T, Stewart LA, Sutton AJ: **Systematic review and meta-analysis of the value of initial biomarkers in predicting adverse outcome in febrile neutropenic episodes in children and young people with cancer.** *BMC Med* 2012, **10:**6.
115. Phillips RS, Lehrnbecher T, Alexander S, Sung L: **Updated systematic review and meta-analysis of the performance of risk prediction rules in children and young people with febrile neutropenia.** *PLoS One* 2012, **7:**e38300.
116. Haeusler GM, Carlesse F, Phillips RS: **An Updated Systematic Review and Meta-Analysis of the Predictive Value of Serum Biomarkers in the Assessment of Fever during Neutropenia in Children with Cancer.** *Pediatr Infect Dis J* 2013, **32:**e390-396.
117. Phillips B, Wade R, Westwood M, Riley R, Sutton AJ: **Systematic review and meta-analysis of the value of clinical features to exclude radiographic pneumonia in febrile neutropenic episodes in children and young people.** *J Paediatr Child Health* 2012, **48:**641-648.
118. Centre for Reviews and Dissemination: *Systematic review: CRD's guidance for undertaking reviews in health care.* York: University of York; 2009.
119. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, Bossuyt PM: **Empirical evidence of design-related bias in studies of diagnostic tests.** *JAMA* 1999, **282:**1061-1066.
120. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J: **The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.** *BMC Medical Research Methodology* 2003, **3:**25-25.
121. Centre for Reviews and Dissemination: **Systematic Reviews of Clinical Tests.** In *Systematic review: CRD's guidance for undertaking reviews in health care.* Edited by Centre for Reviews and Dissemination. York: University of York; 2009: 109 - 156
122. Harbord R: **METANDI: Stata module to perform meta-analysis of diagnostic accuracy.** In *Book METANDI: Stata module to perform meta-analysis of diagnostic accuracy* (Editor ed.^eds.), S456932 edition. City: Department of Social Medicine, University of Bristol; 2008.
123. Lunn D, Thomas A, Best N, Spiegelhalter D: **WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility.** *Statistics and Computing* 2000, **10:**337, 325-337, 325.
124. Dukic V, Gatsonis C: **Meta-analysis of Diagnostic Test Accuracy Assessment Studies with Varying Number of Thresholds.** *Biometrics* 2003, **59:**936-946.
125. Centre for Reviews and Dissemination: **Core Principles and Methods for Conducting a Systematic Review of Health Interventions.** In *Systematic review: CRD's guidance for undertaking reviews in health care.* Edited by Centre for Reviews and Dissemination. York: University of York; 2009: 1 - 91
126. Phillips B, Stewart LA, Sutton AJ: **'Cross hairs' plots for diagnostic meta-analysis.** *Research Synthesis Methods* 2010, **1:**308-315.
127. Doebler P: **mada: Meta-Analysis of Diagnostic Accuracy.** In *Book mada: Meta-Analysis of Diagnostic Accuracy* (Editor ed.^eds.), 0.5.1 edition. City: CRAN; 2011.
128. **Neutropenic sepsis** [http://guidance.nice.org.uk/CG/Wave23/11]
129. Lehrnbecher T, Phillips R, Alexander S, Alvaro F, Carlesse F, Fisher B, Hakim H, Santolaya M, Castagnola E, Davis BL, Dupuis LL, Gibson F, Groll AH, Gaur A, Gupta A, Kebudi R, Petrilli S, Steinbach WJ, Villarroel M, Zaoutis T, Sung L: **Guideline for the management of fever and neutropenia in children with cancer and/or undergoing hematopoietic stem-cell transplantation.** *J Clin Oncol* 2012, **30:**4427-4438.
130. Macher E, Dubos F, Garnier N, Delebarre M, De Berranger E, Thebaud E, Mazingue F, Leblond P, Martinot A: **Predicting the risk of severe bacterial infection in children with chemotherapy-induced febrile neutropenia.** *Pediatr Blood Cancer* 2010, **55:**662-667.

131. Badiei Z, Khalesi M, Alami MH, Kianifar HR, Banihashem A, Farhangi H, Razavi AR: **Risk factors associated with life-threatening infections in children with febrile neutropenia: a data mining approach.** *JOURNAL OF PEDIATRIC HEMATOLOGY/ONCOLOGY* 2011, **33:**e9-e12.
132. Delebarre M, Dubos F, Macher E, Garnier N, Mazingue F, De Berranger E, Nelken B, Thebaud E, Leblond P, Martinot A: **Identifying high-risk patients for severe infection in children with chemotherapy-induced febrile neutropenia: A new decision rule.** *Pediatric Blood & Cancer* 2010, **55:**823.
133. Adcock KG, Akins RL, Farrington EA: **Evaluation of empiric vancomycin therapy in children with fever and neutropenia.** *Pharmacotherapy* 1999, **19:**1315-1320.
134. Alexander SW, Wade KC, Hibberd PL, Parsons SK: **Evaluation of risk prediction criteria for episodes of febrile neutropenia in children with cancer.** *Journal of Pediatric Hematology/Oncology* 2002, **24:**38-42.
135. Jones GR, Konsler GK, Dunaway RP, Pusek SN: **Infection risk factors in febrile, neutropenic children and adolescents.** *Pediatric Hematology & Oncology* 1996, **13:**217-229.
136. Madsen K, Rosenman M, Hui S, Breitfeld PP: **Value of electronic data for model validation and refinement: bacteremia risk in children with fever and neutropenia.** *Journal of Pediatric Hematology/Oncology* 2002, **24:**256-262.
137. Paganini HR, Aguirre C, Puppa G, Garbini C, Javier RG, Ensinck G, Vratnica C, Flynn L, Iacono M, Zubizarreta P, Febrile Neutropenia Study G: **A prospective, multicentric scoring system to predict mortality in febrile neutropenic children with cancer.** *Cancer* 2007, **109:**2572-2579.
138. Rondinelli PIP, Ribeiro KdCB, de Camargo B: **A proposed score for predicting severe infection complications in children with chemotherapy-induced febrile neutropenia.** *Journal of Pediatric Hematology/Oncology* 2006, **28:**665-670.
139. Santolaya ME, Alvarez AM, Becker A, Cofre J, Enriquez N, O'Ryan M, Paya E, Pilorget J, Salgado C, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Prospective, multicenter evaluation of risk factors associated with invasive bacterial infection in children with cancer, neutropenia, and fever.** *Journal of Clinical Oncology* 2001, **19:**3415-3421.
140. West DC, Marcin JP, Mawis R, He J, Nagle A, Dimand R: **Children with cancer, fever, and treatment-induced neutropenia: risk factors associated with illness requiring the administration of critical care therapies.** *Pediatric Emergency Care* 2004, **20:**79-84.
141. Rackoff WR, Gonin R, Robinson C, Kreissman SG, Breitfeld PB: **Predicting the risk of bacteremia in childen with fever and neutropenia.** *Journal of Clinical Oncology* 1996, **14:**919-924.
142. Riikonen P, Jalanko H, Hovi L, Saarinen UM: **Fever and neutropenia in children with cancer: Diagnostic parameters at presentation.** *Acta Paediatrica, International Journal of Paediatrics* 1993, **82:**271-275.
143. Rojo LC, Rodriguez ZN, Tordecilla CJ: **Low risk febrile neutropenia in oncological pediatric patients: Clinical experience. [Spanish].** *Revista Chilena de Pediatria* 2008, **79:**157-162.
144. Ammann RA, Bodmer N, Hirt A, Niggli FK, Nadal D, Simon A, Ozsahin H, Kontny U, Kuhne T, Popovic MB, Luthy AR, Aebi C: **Predicting adverse events in children with fever and chemotherapy-induced neutropenia: the prospective multicenter SPOG 2003 FN study.** *J Clin Oncol* 2010, **28:**2008-2014.
145. Mian A, Prodhan P, Bhutta A, Watkins B: **Clinical variables and acute phase reactants at initial hospitalization as predictors of admission to intensive care unit (ICU) among febrile neutropenic children.** *Crit Care Med* 2009, **37:**A344.
146. Hakim H, Flynn PM, Srivastava DK, Knapp KM, Li C, Okuma J, Gaur AH: **Risk prediction in pediatric cancer patients with fever and neutropenia.** *PEDIATRIC INFECTIOUS DISEASE JOURNAL* 2010, **29:**53-59.
147. Agyeman P, Aebi C, Hirt A, Niggli FK, Nadal D, Simon A, Ozsahin H, Kontny U, Kuhne T, Beck Popovic M, Leibundgut K, Bodmer N, Ammann RA: **Predicting bacteremia in children with**

cancer and fever in chemotherapy-induced neutropenia: results of the prospective multicenter SPOG 2003 FN study. *Pediatr Infect Dis J* 2011, **30:**e114-119.
148. Miedema KG, de Bont ES, Oude Nijhuis CS, van Vliet D, Kamps WA, Tissing WJ: **Validation of a new risk assessment model for predicting adverse events in children with fever and chemotherapy-induced neutropenia.** *J Clin Oncol* 2011, **29:**e182-184; author reply e185.
149. Katz JA, Bash R, Rollins N, Cash J, Buchanan GR: **The yield of routine chest radiography in children with cancer hospitalized for fever and neutropenia.** *Cancer* 1991, **68:**940-943.
150. Korones DN, Hussong MR, Gullace MA: **Routine chest radiography of children with cancer hospitalized for fever and neutropenia: Is it really necessary?** *Cancer* 1997, **80:**1160-1164.
151. Feusner J, Cohen R, O'Leary M, Beach B: **Use of routine chest radiography in the evaluation of fever in neutropenic pediatric oncology patients.** *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 1988, **6:**1699-1702.
152. Renoult E, Buteau C, Turgeon N, Moghrabi A, Duval M, Tapiero B: **Is routine chest radiography necessary for the initial evaluation of fever in neutropenic children with cancer?[see comment].** *Pediatric Blood & Cancer* 2004, **43:**224-228.
153. Ammann RA, Hirt A, Luthy AR, Aebi C: **Predicting bacteremia in children with fever and chemotherapy-induced neutropenia.** *Pediatr Infect Dis J* 2004, **23:**61-67.
154. Dommett R, Geary J, Freeman S, Hartley J, Sharland M, Davidson A, Tulloh R, Taj M, Stoneham S, Chisholm JC: **Successful introduction and audit of a step-down oral antibiotic strategy for low risk paediatric febrile neutropaenia in a UK, multicentre, shared care setting.** *European Journal of Cancer* 2009, **45:**2843-2849.
155. Miedema KGE, de Bont ESJM, Oude Nijhuis CSM, van Vliet D, Kamps WA, Tissing WJE: **Validation of a new risk assessment model for predicting adverse events in children with fever and chemotherapy-induced neutropenia.** *J Clin Oncol* 2011, **29:**e182-184; author reply e185.
156. Lehrnbecher T, Fleischhack G, Hanisch M, Deinlein F, Simon A, Bernig T, Chanock SJ, Klingebiel T: **Circulating levels and promoter polymorphisms of interleukins-6 and 8 in pediatric cancer patients with fever and neutropenia.** *Haematologica* 2004, **89:**234-236.
157. Santolaya ME, Alvarez AM, Aviles CL, Becker A, King A, Mosso C, O'Ryan M, Paya E, Salgado C, Silva P, Topelberg S, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Predictors of severe sepsis not clinically apparent during the first twenty-four hours of hospitalization in children with cancer, neutropenia, and fever: a prospective, multicenter trial.** *The Pediatric infectious disease journal* 2008, **27:**538-543.
158. de Bont ES, Vellenga E, Swaanenburg JC, Fidler V, Visser-van Brummen PJ, Kamps WA: **Plasma IL-8 and IL-6 levels can be used to define a group with low risk of septicaemia among cancer patients with fever and neutropenia.** *British Journal of Haematology* 1999, **107:**375-380.
159. Katz JA, Mustafa MM, Bash RO, Cash JV, Buchanan GR: **Value of C-reactive protein determination in the initial diagnostic evaluation of the febrile, neutropenic child with cancer.** *Pediatric Infectious Disease Journal* 1992, **11:**708-712.
160. Spasova MI, Terzieva DD, Tzvetkova TZ, Stoyanova AA, Mumdzhiev IN, Yanev IB, Genev ED: **Interleukin-6, interleukin-8, interleukin-10, and C-reactive protein in febrile neutropenia in children with malignant diseases.** *Folia Medica (Plovdiv)* 2005, **47:**46-52.
161. Santolaya ME, Alvarez AM, Aviles CL, Becker A, King A, Mosso C, O'Ryan M, Paya E, Salgado C, Silva P, Topelberg S, Tordecilla J, Varas M, Villarroel M: **Predictors of severe sepsis not clinically apparent during the first twenty-four hours of hospitalization in children with cancer, neutropenia, and fever: a prospective, multicenter trial.** *Pediatric Infectious Disease Journal* 2008, **27:**538-543.
162. Richardson MW, Grewal SS, Visintainer PF: **Emesis predicts bacteremia in immunocompromised children with central venous catheters and fever.** 2009, **115:**3335-3340.

163. Kitanovski L, Jazbec J, Hojker S, Gubina M, Derganc M: **Diagnostic accuracy of procalcitonin and interleukin-6 values for predicting bacteremia and clinical sepsis in febrile neutropenic children with cancer.** *European Journal of Clinical Microbiology & Infectious Diseases* 2006, **25:**413-415.
164. Soker M, Colpan L, Ece A, Devecioglu C, Haspolat K: **Serum levels of IL-1 beta, sIL-2R, IL-6, IL-8, and TNF-alpha in febrile children with cancer and neutropenia.** *Medical Oncology* 2001, **18:**51-57.
165. Reitman A, Pisk R, Gates J, Ozeran JD: **Serial procalcitonin levels for bacteremia in pediatric patients with fever and severe neutropenia** 24th Annual Meeting of the American Society of Pediatric Hematology Oncology, ASPHO 2011
166. Cost, C.**Utilizing cytokines to predict bacteremia in 120 pediatric patients with febrile neutropenia .** 24th Annual Meeting of the American Society of Pediatric Hematology Oncology, ASPHO 2011
167. Miedema KGE, de Bont ESJM, Elferink RFMO, van Vliet MJ, Nijhuis CSMO, Kamps WA, Tissing WJE: **The diagnostic value of CRP, IL-8, PCT, and sTREM-1 in the detection of bacterial infections in pediatric oncology patients with febrile neutropenia**. *Supportive care in cancer* 2011, **19**:1593-1600.
168. Diepold M, Noellke P, Duffner U, Kontny U, Berner R: **Performance of Interleukin-6 and Interleukin-8 serum levels in pediatric oncology patients with neutropenia and fever for the assessment of low-risk.** *BMC Infectious Diseases* 2008, **8:**28.
169. Riikonen P, Saarinen UM, Teppo AM, Metsarinne K, Fyhrquist F, Jalanko H: **Cytokine and acute-phase reactant levels in serum of children with cancer admitted for fever and neutropenia.** *Journal of Infectious Diseases* 1992, **166:**432-436.
170. Barnes C, Ignjatovic V, Newall F, Carlin J, Ng F, Hamilton S, Ashley D, Waters K, Monagle P: **Change in serum procalcitonin (deltaPCT) predicts the clinical outcome of children admitted with febrile neutropenia.** *Br J Haematol* 2002, **118:**1197-1198.
171. Lehrnbecher T, Venzon D, de Haas M, Chanock SJ, Kuhl J: **Assessment of measuring circulating levels of interleukin-6, interleukin-8, C-reactive protein, soluble Fc gamma receptor type III, and mannose-binding protein in febrile children with cancer and neutropenia.** *CLINICAL INFECTIOUS DISEASES* 1999, **29:**414-419.
172. Santolaya ME, Alvarez AM, Aviles CL, Becker A, Mosso C, O'Ryan M, Paya E, Salgado C, Silva P, Topelberg S, Tordecilla J, Varas M, Villarroel M, Viviani T: **Admission clinical and laboratory factors associated with death in children with cancer during a febrile neutropenic episode.** *Pediatric Infectious Disease Journal* 2007, **26:**794-798.
173. Hatzistilianou M, Rekleity A, Athanassiadou F, DeLutiis MA, Conti P, Catriu D: **Serial procalcitonin responses in infection of children with secondary immunodeficiency.** *Clinical & Investigative Medicine - Medecine Clinique et Experimentale* 2007, **30:**E75-85.
174. Kharya G, Dinand V, Prakash A, Kahlon D, Kalra M, Radhakrishnan N, Sachdeva A, Yadav SP: **Interleukin 6 and procalcitonin are good early markers of sepsis in children with febrile neutropenia.** 42nd Congress of the International Society of Pediatric Oncology, SIOP 2010
175. Schroder H, Lodahl D: **Procalcitonin adds to diagnosis, but does not reduce initial antibiotics in febrile neutropenic children**. *Danish Medical Bulletin* 2011, **58**(3):A4233
176. El-Maghraby SM, Moneer MM, Ismail MM, Shalaby LM, El-Mahallawy HA: **The diagnostic value of C-reactive protein, interleukin-8, and monocyte chemotactic protein in risk stratification of febrile neutropenic children with hematologic malignancies.** *Journal of Pediatric Hematology/Oncology* 2007, **29:**131-136.
177. Hodge G, Osborn M, Hodge S, Nairn J, Tapp H, Kirby M, Sepulveda H, Morgan E, Revesz T, Zola H: **Rapid simultaneous measurement of multiple cytokines in childhood oncology patients with febrile neutropenia: Increased interleukin (IL)-8 or IL-5 correlates with culture-positive infection [3].** *British Journal of Haematology* 2006, **132:**247-248.

178. Hodge G, Osborn M, Hodge S, Nairn J, Tapp H, Kirby M, Sepulveda H, Morgan E, Revesz T, Zola H: **Rapid simultaneous measurement of multiple cytokines in childhood oncology patients with febrile neutropenia: Increased interleukin (IL)-8 or IL-5 correlates with culture-positive infection.** *British Journal of Haematology* 2006, **132:**247-248.
179. Secmeer G, Devrim I, Kara A, Ceyhan M, Cengiz B, Kutluk T, Buyukpamukcu M, Yetgin S, Tuncer M, Uludag AK, Tezer H, Yildirim I: **Role of procalcitonin and CRP in differentiating a stable from a deteriorating clinical course in pediatric febrile neutropenia.** *Journal of Pediatric Hematology/Oncology* 2007, **29:**107-111.
180. Avabratha KS, Rau ATK, Venkataravanamma P, Rau A: **Significance of C-reactive protein during febrile neutropenia in pediatric malignancies**. *Indian pediatrics* 2009, **46**(gm2, 2985062r):797-799]
181. Dylewska K, Wysocki M, Odrowaz-Sypniewska G, Kurylak A, Jankowska A: **Proinflammatory (IL-6, TNF-alpha) and antiinflammatory (IL-10) cytokines in children with cancer during myelosupression. [Polish].** *Pediatria Polska* 2005, **80:**1091-1095.
182. Dylewska K, Wysocki M, Odrowaz-Sypniewska G, Kurylak A, Jankowska A: **Procalcitonin and C-reactive protein in febrile children with cancer during neutropenia. [Polish].** *Pediatria Polska* 2005, **80:**175-179.
183. Heney D, Lewis IJ, Evans SW, Banks R, Bailey CC, Whicher JT: **Interleukin-6 and its relationship to C-reactive protein and fever in children with febrile neutropenia.** *Journal of Infectious Diseases* 1992, **165:**886-890.
184. Hitoglou-Hatzi S, Hatzistilianou M, Gougoustamou D, Rekliti A, Agguridaki C, Athanassiadou F, Frydas S, Kotsis A, Catriu D: **Serum adenosine deaminase and procalcitonin concentrations in neutropenic febrile children with acute lymphoblastic leukaemia.** *Clinical & Experimental Medicine* 2005, **5:**60-65.
185. Hatzistilianou M, Rekliti A, Athanassiadou F, Catriu D: **Procalcitonin as an early marker of bacterial infection in neutropenic febrile children with acute lymphoblastic leukemia.** *Inflammation research* 2010, **59:**339-347.
186. Mian A, Prodhan P, Bhutta A, Watkins B: **Clinical variables and acute phase reactants at initial hospitalization as predictors of admission to intensive care unit (ICU) among febrile neutropenic children**. *39th Critical Care Congress of the Society of Critical Care Medicine* 2009
187. Nishikawa T, Okamoto Y, Kodama Y, Tanabe T, Shinkoda Y, Kawano Y: **Serum derivative of reactive oxygen metabolites (d-ROMs) in pediatric hemato-oncological patients with neutropenic fever**. *Pediatric blood & cancer* 2010, **55**(101186624):91-94]
188. Asturias EJ, Corral JE, Quezada J: **Evaluation of six risk factors for the development of bacteremia in children with cancer and febrile neutropenia**. *Current oncology* 2010, **17**:59-63]
189. Reitman A, Pisk R, Gates III J, Ozeran J: **Serial procalcitonin levels for bacteremia in pediatric patients with fever and severe neutropenia. .** In *Book Serial procalcitonin levels for bacteremia in pediatric patients with fever and severe neutropenia.* (Editor ed.^eds.). City: Pediatr Blood Cancer; 2010.
190. Delebarre M, Dubos F, Macher E, Garnier N, Mazingue F, De Berranger E, Nelken B, Thebaud E, Leblond P, Martinot A: **Identifying high-risk patients for severe infection in children with chemotherapy-induced febrile neutropenia: A new decision rule**. *Pediatric Blood and Cancer* 2010, **55**:823
191. Mian A, Prodhan P, Bhutta A, Watkins B: **Clinical variables and acute phase reactants at initial hospitalization as predictors of admission to intensive care unit (ICU) among febrile neutropenic children**. *39th Critical Care Congress of the Society of Critical Care Medicine* 2009]

192. Hatzistilianou M, Rekliti A, Athanassiadou F, Catriu D: **Procalcitonin as an early marker of bacterial infection in neutropenic febrile children with acute lymphoblastic leukemia.** *Inflammation research* 2010, **59:**339-347.
193. Lodahl D, Schroder H: **Procalcitonin adds to diagnosis, but does not reduce initial antibiotics in febrile neutropenic children.** *Danish medical bulletin* 2011, **58:**A4233.
194. Miedema KGE, de Bont ESJM, Elferink RFMO, van Vliet MJ, Nijhuis CSMO, Kamps WA, Tissing WJE: **The diagnostic value of CRP, IL-8, PCT, and sTREM-1 in the detection of bacterial infections in pediatric oncology patients with febrile neutropenia.** *Supportive Care in Cancer* 2011, **19:**1593-1600.
195. Hatzistilianou M, Rekleity A, Athanassiadou F, DeLutiis MA, Conti P, Catriu D: **Serial procalcitonin responses in infection of children with secondary immunodeficiency.** *Clin Invest Med* 2007, **30:**E75-85.
196. Margolis P, Gadomski A: **The rational clinical examination. Does this infant have pneumonia?** *JAMA* 1998, **279:**308-313.
197. Vittinghoff E, McCulloch CE: **Relaxing the rule of ten events per variable in logistic and Cox regression.** *Am J Epidemiol* 2007, **165:**710-718.
198. Royston P, Sauerbrei W: *Multivariable Model - Building: A Pragmatic Approach to Regression Anaylsis Based on Fractional Polynomials for Modelling Continuous Variables.* John Wiley & Sons Inc (E); 2008.
199. Watson RS, Carcillo JA, Linde-Zwirble WT, Clermont G, Lidicker J, Angus DC: **The epidemiology of severe sepsis in children in the United States.** *Am J Respir Crit Care Med* 2003, **167:**695-701.
200. Donders AR, van der Heijden GJ, Stijnen T, Moons KG: **Review: a gentle introduction to imputation of missing values.** *J Clin Epidemiol* 2006, **59:**1087-1091.
201. Kenward MG, Carpenter J: **Multiple imputation: current perspectives.** *Statistical Methods in Medical Research* 2007, **16:**199-218.
202. Phillips R, Hancock B, Graham J, Bromham N, Jin H, Berendse S: **Prevention and management of neutropenic sepsis in patients with cancer: summary of NICE guidance.** *BMJ* 2012, **345**.
203. Stewart LA, Tierney JF: **To IPD or not to IPD?: Advantages and Disadvantages of Systematic Reviews Using Individual Patient Data.** *Eval Health Prof* 2002, **25:**76-97.
204. Riley RD, Lambert PC, Abo-Zaid G: **Meta-analysis of individual participant data: rationale, conduct, and reporting.** *Bmj* 2010, **340:**c221-.
205. Khan KS, Bachmann LM, ter Riet G: **Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests.** *Eur J Obstet Gynecol Reprod Biol* 2003, **108:**121 - 125.
206. Broeze K, Opmeer B, Bachmann L, Broekmans F, Bossuyt P, Coppus S, Johnson N, Khan K, ter Riet G, van der Veen F, van Wely M, Mol B: **Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine.** *BMC Medical Research Methodology* 2009, **9:**22.
207. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM: **Systematic Reviews of Diagnostic Test Accuracy.** *Annals of Internal Medicine* 2008, **149:**889-897.
208. Trivella M, Pezzella F, Pastorino U, Harris AL, Altman DG: **Microvessel density as a prognostic factor in non-small-cell lung carcinoma: a meta-analysis of individual patient data.** *Lancet Oncol* 2007, **8:**488-499.
209. Royston P, Parmar MK, Sylvester R: **Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer.** *Stat Med* 2004, **23:**907-926.
210. Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, Mant D, McManus RJ, Holder R, Deeks J, Fletcher K, Qume M, Sohanpal S, Sanders S, Hobbs FD: **Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of**

implications of different diagnostic strategies in primary care. *Health Technol Assess* 2009, **13:**1-207, iii.
211. Simon R, Altman DG: **Statistical aspects of prognostic factor studies in oncology.** *Br J Cancer* 1994, **69:**979-985.
212. Kyzas PA, Loizou KT, Ioannidis JP: **Selective reporting biases in cancer prognostic factor studies.** *J Natl Cancer Inst* 2005, **97:**1043-1055.
213. Ioannidis JPA, Panagiotou OA: **Comparison of Effect Sizes Associated With Biomarkers Reported in Highly Cited Individual Articles and in Subsequent Meta-analyses.** *JAMA: The Journal of the American Medical Association*, **305:**2200-2210.
214. Look MP, van Putten WLJ, Duffy MJ, Harbeck N, Christensen IJ, Thomssen C, Kates R, Spyratos F, Ferno M, Eppenberger-Castori S, Sweep CGJF, Ulm K, Peyrat J-P, Martin P-M, Magdelenat H, Brunner N, Duggan C, Lisboa BW, Bendahl P-O, Quillien V, Daver A, Ricolleau G, Meijer-van Gelder ME, Manders P, Fiets WE, Blankenstein MA, Broet P, Romain S, Daxenbichler G, et al: **Pooled Analysis of Prognostic Impact of Urokinase-Type Plasminogen Activator and Its Inhibitor PAI-1 in 8377 Breast Cancer Patients.** *J Natl Cancer Inst* 2002, **94:**116-128.
215. Fibrinogen Studies C: **Plasma Fibrinogen Level and the Risk of Major Cardiovascular Diseases and Nonvascular Mortality: An Individual Participant Meta-analysis.** *JAMA* 2005, **294:**1799-1809.
216. Higgins J, Green S: **6: ASSESSMENT OF STUDY QUALITY.** In *Cochrane Handbook for Systematic Reviews of Interventions 425.*
217. Partridge N, Scadding J: **The James Lind Alliance: patients and clinicians should jointly identify their priorities for clinical trials.** *Lancet* 2004, **364:**1923-1924.
218. Tallon D, Chard J, Dieppe P: **Relation between agendas of the research community and the research consumer.** *Lancet* 2000, **355:**2037-2040.
219. Kris MG, Hesketh PJ, Herrstedt J, Rittenberg C, Einhorn L, Grunberg SM, Koeller J, Oliver I, Borjeson S, Ballatori E: **Consensus proposals for the prevention of acute and delayed vomiting and nausea following high-emetic-risk chemotherapy.** *Supportive Care in Cancer* 2005, **13:**85-96.
220. Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration: **Reducing Uncertainties About the Effects of Chemoradiotherapy for Cervical Cancer: A Systematic Review and Meta-Analysis of Individual Patient Data From 18 Randomized Trials.** *Journal of Clinical Oncology* 2008, **26:**5802-5812.
221. Vale CL, Thompson LC, Murphy C, Forcat S, Hanley B: **Involvement of consumers in studies run by the Medical Research Council Clinical Trials Unit: results of a survey.** *Trials* 2012, **13:**9.
222. [http://www.ctu.mrc.ac.uk/cochrane/ipdmg/faq.asp#faq28]
223. NCRI, National Cancer Intelligence Network, onCore UK: **Summary of responses to consultation on 'Access to Samples and Data for Cancer Research'.** In *Book Summary of responses to consultation on 'Access to Samples and Data for Cancer Research'* (Editor ed.^eds.). City; 2009.
224. **National Information: Germany** [http://www.eurecnet.org/information/germany.html]
225. **ECC Frequently Asked Questions** [http://www.nigb.nhs.uk/ecc/eccfrequently]
226. MRC: **Personal Information In Medical Research.** *Personal Information In Medical Research* 2 edition. Medical Research Council; 2003.
227. National Heath and Medical Research Council: **National Statement on Ethical Conduct in Human Research.** *National Statement on Ethical Conduct in Human Research* Commonwealth of Australia; 2007.
228. **Ethical Guidelines for Observational Studies** [http://www.ethicscommittees.health.govt.nz/moh.nsf/indexcm/ethics-resources-observationalstudies]

229. Health Canada Research Ethics Board: **Ethics Review of Research Involving Humans - Administrative Policy and Procedures Manual.** *Ethics Review of Research Involving Humans - Administrative Policy and Procedures Manual* Department of Health; 2009.
230. Office for Human Research Protections: **TITLE 45, Code Of Federal Regulations, PART 46, Protection Of Human Subjects.** *TITLE 45, Code Of Federal Regulations, PART 46, Protection Of Human Subjects* U.S. Department of Health & Human Services; 2009.
231. Phillips R, Sutton A, Riley R, Chisholm J, Picton S, Stewart L, tPC: **Predicting infectious complications in neutropenic children and young people with cancer (IPD Protocol).** *Systematic Reviews* 2012, **1:**8.
232. Fern L, Davies S, Eden T, Feltbower R, Grant R, Hawkins M, Lewis I, Loucaides E, Rowntree C, Stenning S, Whelan J: **Rates of inclusion of teenagers and young adults in England into National Cancer Research Network clinical trials: report from the National Cancer Research Institute (NCRI) Teenage and Young Adult Clinical Studies Development Group.** *Br J Cancer* 2008, **99:**1967-1974.
233. Kazak AE, Hocking MC, Ittenbach RF, Meadows AT, Hobbie W, Derosa BW, Leahey A, Kersun L, Reilly A: **A revision of the intensity of treatment rating scale: Classifying the intensity of pediatric cancer treatment.** *Pediatr Blood Cancer* 2011.
234. Fleming S, Thompson M, Stevens R, Heneghan C, Pluddemann A, Maconochie I, Tarassenko L, Mant D: **Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies.** *Lancet* 2011, **377:**1011-1018.
235. Goldstein B, Giroir B, Randolph A: **International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics.** *Pediatr Crit Care Med* 2005, **6:**2-8.
236. Haque IU, Zaritsky AL: **Analysis of the evidence for the lower limit of systolic and mean arterial pressure in children.** *Pediatr Crit Care Med* 2007, **8:**138-144.
237. Massaro K, Costa S, Leone C, Chamone D: **Procalcitonin (PCT) and C-reactive Protein (CRP) as severe systemic infection markers in febrile neutropenic adults.** *BMC Infectious Diseases* 2007, **7:**137.
238. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG: **Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers.** *Bmj* 2010, **340:**c181-.
239. Riley RD, Steyerberg EW: **Meta-analysis of a binary outcome using individual participant data and aggregate data.** *Research Synthesis Methods* 2010, **1:**2-19.
240. Altman DG, Lyman GH: **Methodological challenges in the evaluation of prognostic factors in breast cancer.** *Breast Cancer Res Treat* 1998, **52:**289-303.
241. Hayden JA, Cote P, Bombardier C: **Evaluation of the quality of prognosis studies in systematic reviews.** *Ann Intern Med* 2006, **144:**427-437.
242. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C: **Assessing bias in studies of prognostic factors.** *Ann Intern Med* 2013, **158:**280-286.
243. Steyerberg EW: *Clinical prediction models : a practical approach to development, validation, and updating.* New York: Springer; 2009.
244. Pencina MJ, D' Agostino RB, Vasan RS: **Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond.** *Statistics in Medicine* 2008, **27:**157-172.
245. Vickers AJ, Elkin EB: **Decision Curve Analysis: A Novel Method for Evaluating Prediction Models.** *Medical Decision Making* 2006, **26:**565-574.
246. Phillips B, Ranasinghe N, Stewart LA: **Ethical and regulatory considerations in the use of individual participant data for studies of disease prediction.** *Archives of Disease in Childhood* 2013.
247. Wilkinson L, Friendly M: **The History of the Cluster Heat Map.** *The American Statistician* 2009, **63:**179-184.

248. Pearce MS, Parker L: **Childhood cancer registrations in the developing world: Still more boys than girls.** *International Journal of Cancer* 2001, **91:**402-406.
249. Bruel AVd, Thompson M, Buntinx F, Mant D: **Clinicians' gut feeling about serious infections in children: observational study.** *BMJ* 2012, **345**.
250. Phillips R, Hancock B, Graham J, Bromham N, Jin H, Berendse S: **Prevention and management of neutropenic sepsis in patients with cancer: summary of NICE guidance.** *BMJ* 2012, **345:**e5368.
251. Phillips R, Skinner R, Chisholm JC: **Treating low-risk febrile neutropenia: Jenny's story.** *Arch Dis Child* 2007, **92:**7-8.
252. Resche-Rigon M, White IR, Bartlett JW, Peters SA, Thompson SG: **Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data.** *Stat Med* 2013, **32:**4890-4905.
253. Burgess S, White IR, Resche-Rigon M, Wood AM: **Combining multiple imputation and meta-analysis with individual participant data.** *Stat Med* 2013, **32:**4499-4514.
254. Ahmed I, Debray TP, Moons KG, Riley RD: **Developing and validating risk prediction models in an individual participant data meta-analysis.** *BMC Med Res Methodol* 2014, **14:**3.
255. Manji A, Beyene J, Dupuis LL, Phillips R, Lehrnbecher T, Sung L: **Outpatient and oral antibiotic management of low-risk febrile neutropenia are effective in children--a systematic review of prospective trials.** *Support Care Cancer* 2012, **20:**1135-1145.
256. Shimasaki N, Mori T, Samejima H, Sato R, Shimada H, Yahagi N, Torii C, Yoshihara H, Tanigawara Y, Takahashi T, Kosaki K: **Effects of methylenetetrahydrofolate reductase and reduced folate carrier 1 polymorphisms on high-dose methotrexate-induced toxicities in children with acute lymphoblastic leukemia or lymphoma.** *J Pediatr Hematol Oncol* 2006, **28:**64-68.
257. Phillips R, Wade R, Riley R, Sutton A, Stewart LA: **Systematic review and meta-analysis of the value of clinical features to exclude radiographic pneumonia in febrile neutropenic episodes in children and young people.** *Journal of Paediatrics and Child Health* 2011.
258. Castagnola E, Fontana V, Caviglia I, Caruso S, Faraci M, Fioredda F, Garre ML, Moroni C, Conte M, Losurdo G, Scuderi F, Bandettini R, Toma P, Viscoli C, Haupt R: **A prospective study on the epidemiology of febrile episodes during chemotherapy-induced neutropenia in children with cancer or after hemopoietic stem cell transplantation.** *Clin Infect Dis* 2007, **45:**1296-1304.
259. Lehrnbecher T, Venzon D, de Haas M, et al.: **Assessment of measuring circulation levels of interleukin-6, interleukin-8, C-reactive protein, soluable Fc gamma receptor type III, and mannose-binding protein in febrile children with cancer and neutropenia.** *Clin Infect Dis* 1999, **29:**414-419.
260. Santolaya ME, Alvarez AM, Aviles CL, Becker A, Cofre J, Enriquez N, O'Ryan M, Paya E, Salgado C, Silva P, Tordecilla J, Varas M, Villarroel M, Viviani T, Zubieta M: **Prospective evaluation of a model of prediction of invasive bacterial infection risk among children with cancer, fever, and neutropenia.** *CLINICAL INFECTIOUS DISEASES* 2002, **35:**678-683.
261. Spasova MI, Grudeva-Popova JG, Kostyanev SS, Genev ED, Stoyanova AA, Kirina VI, Moumdjiev IN: **Risk index score for bacteremia in febrile neutropenic episodes in children with malignancies.** *Journal of BUON* 2009, **14:**411-418.
262. Stryjewski GR, Nylen ES, Bell MJ, Snider RH, Becker KL, Wu A, Lawlor C, Dalton H: **Interleukin-6, interleukin-8, and a rapid and sensitive assay for calcitonin precursors for the determination of bacterial sepsis in febrile neutropenic children.** *Pediatric Critical Care Medicine* 2005, **6:**129-135.
263. Ramphal R, Grant RM, Dzolganovski B, Constantin J, Tellier R, Allen U, Weitzman S, Matlow A, Petric M, Sung L: **Herpes simplex virus in febrile neutropenic children undergoing chemotherapy for cancer: a prospective cohort study.** *Pediatr Infect Dis J* 2007, **26:**700-704.