The "acutely sick" African child: applying new statistical methods to delineate mortality risks and identify ways to improve management.

Elizabeth Clare George

UCL

Submitted for the degree of Doctor of Philosophy

June 2018

Declaration:

I, Elizabeth Clare George confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm this has been indicated in the thesis.

Signature:

Date:

Thesis Abstract:

The admission burden to paediatric wards in African hospitals is very high, and many children have life-threatening complications of common infectious diseases including malaria. In this setting the Fluid Expansion As Supportive Therapy (FEAST) trial unexpectedly showed that fluid resuscitation (giving a bolus of saline or albumin) to febrile children with impaired perfusion was harmful. This PhD used data from the FEAST trial to answer several important questions: 1) how can children at the highest risk of mortality be identified and prioritised in these low-income settings; 2) were there surrogate markers in bedside measures recorded over admission that could explain some of the detrimental impact of the bolus; and 3) did the bolus increase mortality risk immediately after administration or was there a different mechanism of action?

Prognostic factors for mortality were identified in the FEAST data and a bedside risk score developed. The score was validated in data collected on children admitted to a rural district hospital in Kenya and compared to other risk scores. The heterogeneity of the effect of bolus over levels of different measures, including malaria parasitaemia, was explored. The proportion of treatment effect explained by measures recorded over time was calculated. No one measure was shown to be a suitable surrogate marker, but the impact of the bolus varied across levels of oxygen saturation, and across levels of base excess in those with malaria at baseline.

Further insight into the mechanism by which the bolus had detrimental impact on the children in the FEAST trial was sought by modelling the mortality risk over time. The maximum mortality risk occurred in both arms within the first 2 hours but the risk in the bolus arm was slower to decrease, showing that children were recovering more slowly in the bolus arms compared to the no bolus arm.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to first thank my supervisors, Abdel Babiker, Sarah Walker and Kath Maitland, for all their support, advice and guidance over the years this PhD has taken. They have always been very encouraging and enabled me to prioritise this PhD where necessary, both in terms of workload and funding, of which I'm very appreciative. I would also like to thank and acknowledge colleagues at the MRC CTU at UCL who have been considerate of the fact I've been studying and working simultaneously and also very supportive. In particular, Di Gibb who has encouraged and supported me with advice and guidance, and helped provide support for my time. I've also very much appreciated the listening ear and advice of friends who are studying or have completed doctorates over the years.

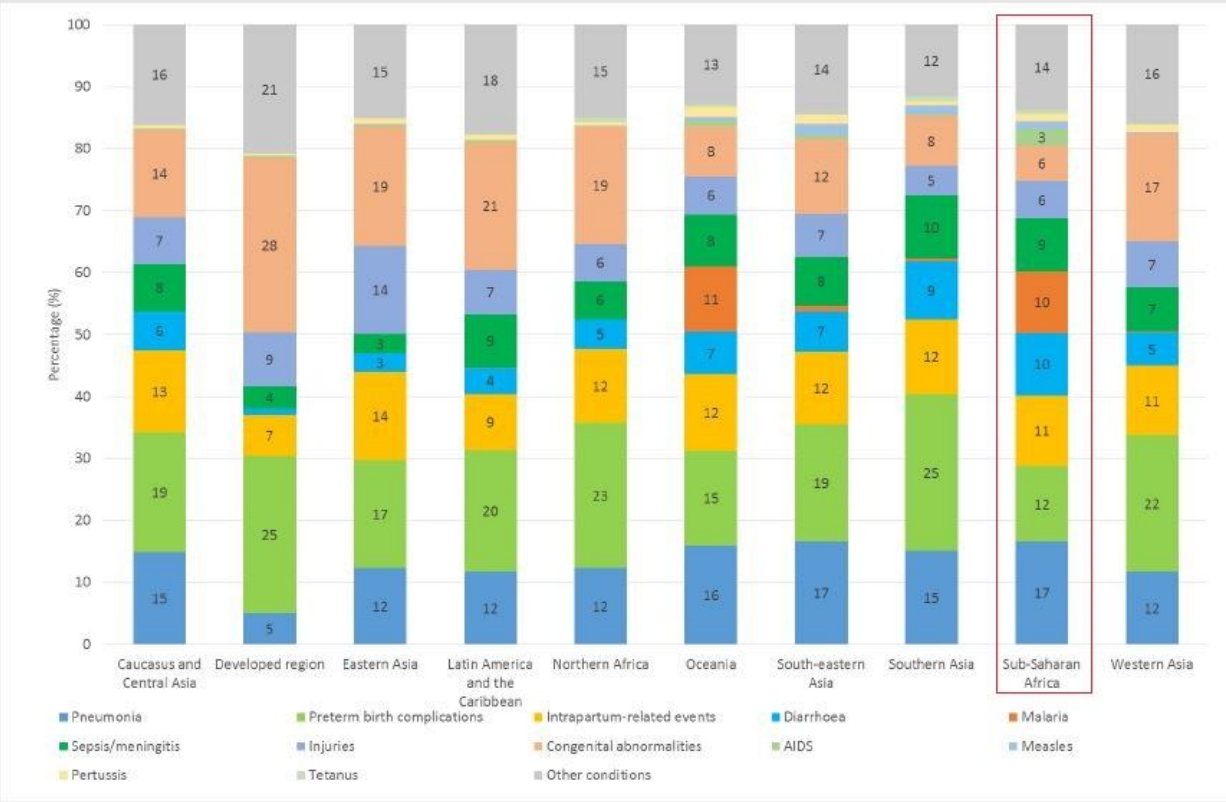
I would like to thank and acknowledge the support and encouragement of my parents who have always shown a great interest in my work and study. I would most especially like to thank my husband who has been incredibly supportive and patient throughout the whole PhD. Last, but not least, I would like to thank my son who is generally very chilled out, despite currently being a toddler, and has pushed me to find a good balance of studying, working and family time.

1 Introduction

1.1 Child mortality in hospitals in sub-Saharan Africa

Child mortality in sub-Saharan Africa is high and there were an estimated 2.9 million deaths in under-5s in 2015. The most common, individual known causes were pneumonia (16.6%), preterm birth complications (12%) and intrapartum-related events (12%) with malaria (10%) and diarrhoea (10%) also major causes of death in children under 5 as presented in Figure 1.1.1 taken from Liu et al [1].

Figure 1.1.1: Causes of child deaths in different regions of the world (Web-appendix Figure 3 in Liu et al, Lancet 2015) with red box added to highlight sub-Saharan Africa.



A proportion of these child deaths will have occurred in hospital where inpatient mortality rates are high (15-30%) [2-4] but the overall number of paediatric admissions either in children who survived or admissions that lead to death has not been estimated as data has not been consistently collected by health services across countries. The majority of children (90%) presenting to hospital in sub-Saharan Africa arrive with severe forms of common childhood illnesses (especially malaria, diarrhoea, pneumonia, sepsis and meningitis) [2]. Over half the deaths in these children occur within the first 24 hours of admission. This may be due in part

to a delay in getting to the hospital or slow admission care on arrival [5]. For example in Tanzania almost half of the children referred to hospital care take 2 days or longer to reach the hospital and in Guinea-Bissau 16% of acutely sick children die on the way to the hospital or whilst waiting at an outpatient clinic [6].

The high rates of child mortality have been of concern for many years and in the year 2000 specific targets for reducing child mortality were identified by a consensus of experts from the United Nations Secretariat, International Monetary Fund, the Organisation of Economic Cooperation and Development and the World Bank, and set out as part of the 10 Millennium Development Goals (MDG) [7, 8]. Promises were received from countries to commit to reducing the under-5 mortality rate by two thirds from the rates measured in 1990 by 2015. Unfortunately this target was not achieved globally, although 4 million fewer deaths occurred in children <5 years in 2015 compared to 2000. The global annual rate of reduction (ARR) needed to meet the MDG was 4.4% but overall it was 4% between 2000 and 2015. However, average ARR's for measles, neonatal tetanus, HIV/AIDS, meningitis, malaria, diarrhoea, and pneumonia were all >4.4% over the same period.

Programmes and interventions that aimed at reducing child mortality were monitored by an initiative called Countdown to 2015; the findings were summarised in a paper in the Lancet in 2015 by Requejo J et al, *Countdown to 2015 and beyond: fulfilling the health agenda for women and children* [9]. Figure 1.1.2 below is taken from that paper and presents national coverage for a core set of 21 interventions that should be available for all. Using nationally representative household surveys and databases from UNICEF, Save the Children, the UN Population Division and other organisations, they found that key gaps in coverage exist around the time of birth and also with case management of childhood illnesses.

Figure 1.1.2: Coverage of various interventions to improve maternal, newborn and child health.

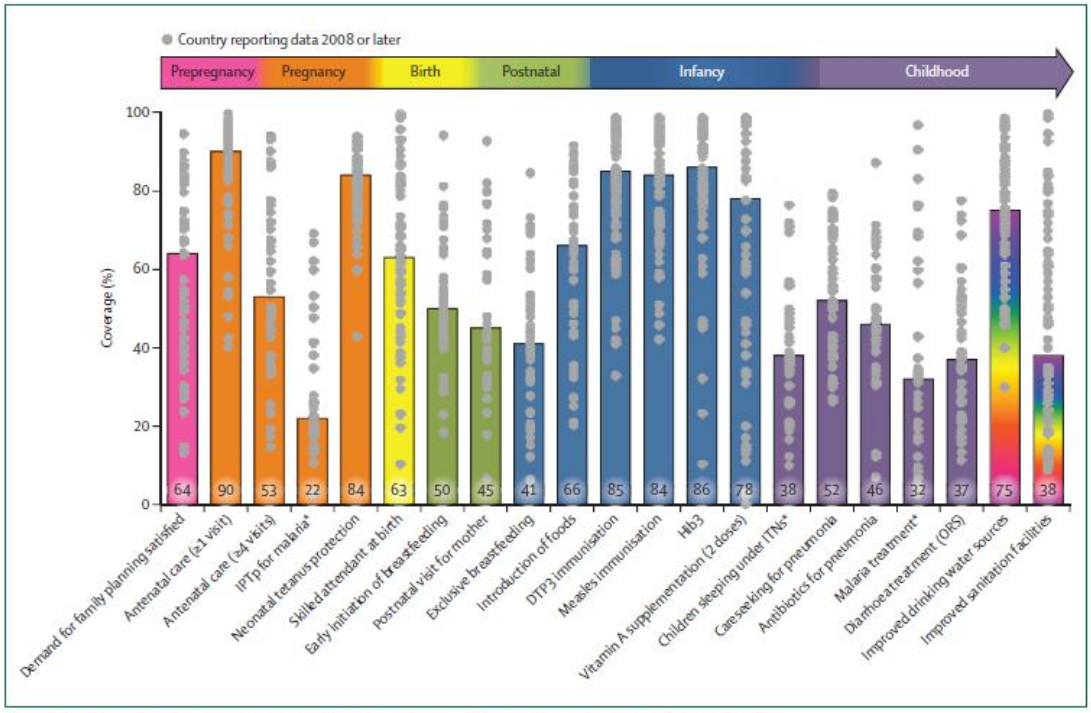


Figure 2: **Median national coverage of selected Countdown interventions, most recent survey, 2008 or later (%)**
The grey dot indicates the point estimate for each country from the most recent available data source. IPTp=intermittent preventive treatment in pregnancy. ITNs=insecticide-treated nets. ORS=oral rehydration salts. *Analysis is based on countries with 75% or more of the population at risk of *P falciparum* transmission and 50% or more cases of malaria caused by *P falciparum*. Data are for 2012. Source: Immunisation rates, WHO, and UNICEF; post-natal visit for mother, Saving Newborn Lives analysis of Demographic and Health Surveys; improved water and sanitation, WHO and UNICEF Joint Monitoring Programme; all other indicators, UNICEF global databases, April 2014, based on Deomgraphic and Health Surveys, Multiple Indicator Cluster Surveys, and other national surveys. Reproduced from Requejo and colleagues.[1]

The sustainable development goals (SDGs) were developed in 2015 to continue the efforts of countries to improve health (and other factors such as environmental impact) across the world. The child survival targets were included in SDG3: Good health and well-being, and stated that all countries should aim to reduce under 5 mortality rates to <25 per 1000 live births by 2030 [1]. Researchers monitoring interventions and child mortality (as part of an initiative called Countdown to 2030) suggest that for countries with very high mortality the focus should be on infectious causes such as pneumonia, diarrhoea and malaria. There are programmes that are relatively low-cost that can reduce some of these infectious causes including increasing immunization coverage with old and new vaccines, water and sanitation improvement, increase in use of insecticide treated bed nets and improved treatment of sick children within all levels of healthcare facilities. However, despite the introduction of the Haemophilus influenza type b, pneumococcal conjugate (PCV13), and rotavirus vaccines, mortality from pneumonia, malaria and diarrhoea remain high. For example, 25,000 children were estimated to have died from these diseases in Uganda alone in 2015 [1].

Treatment of (and diagnostics for) malaria and care-seeking for pneumonia form part of the key interventions that are being monitored for Countdown to 2030 and healthcare systems

strengthening is important for these interventions. Within improvement of hospital services and healthcare systems, emergency treatment and triage has been highlighted as a key priority area [2]. Reorganisation of the emergency department to improve triage leads to better outcomes; for example, reorganisation of a Malawian hospital along with other interventions led to a reduction of 4-10% in inpatient child mortality [10].

Despite some progress having been made in the overall treatment of critically ill children, mortality rates in hospital remain high and developing more effective emergency care services for children has been identified as crucial to making reductions [6]. Improving these emergency care services need not involve large interventions, instead integration with other hospital services, prioritising the critically ill and introducing simple emergency treatments have been shown to be effective [11].

Prioritising care to the most critically ill as they arrive into emergency departments and clinics at hospital has been an active priority for the World Health Organisation (WHO) since 1998 when a WHO study found numerous significant deficiencies in emergency care and triage [12]. A study in Kenya in 2004 also found that there was a great need to improve emergency care especially for children [3] and that important clinical signs were only recorded a median of 35-76% of the time (for example respiratory rate, mental state, pallor). Thus various guidelines and initiatives have been published and supported in the last 20 years with the aim of improving triage in African hospitals. Simple emergency treatments have also been evaluated for their effectiveness in these settings and results have been incorporated into the guidelines for triage and treatment of critically ill children.

1.1.1 Prioritising the critically ill

Children presenting with severe illness often have multiple disorders [13], so although there had been disease specific programmes (for example for diarrhoea or acute respiratory infections) a more comprehensive approach was needed. It is hard to differentiate between diseases and conditions that have resulted in the child being critically ill at the point of admission and so vertical programmes (i.e specifically for malaria, diarrhoea, acute respiratory illness) needed to be harmonised. This would also enable the most life-threatening aspects of the illness to be tackled in sequence (i.e ABC, Airway, Breathing, Circulation). A comprehensive approach was started in the Integrated Management of Childhood Illnesses (IMCI) guidelines. The IMCI was designed to bring together the clinical management guidelines for a variety of diseases and ensure a combined approach, working inside and outside of hospitals and

including training for health care workers at all levels [14]. The guidelines give methods suitable for low-income countries to prevent and manage leading causes of child morbidity and mortality and are designed to be adapted within each country [15]. They also work under the assumption that extensive laboratory testing is not available. The IMCI guidelines include training documents, manuals, teaching materials and information to improve care in clinics or small hospitals and improve household management of children. However, in 2004 only one district hospital out of 14 surveyed in Kenya was implementing the strategies in the IMCI guidelines [16] and in Uganda only 28% of children fulfilling the IMCI criteria completed referral to a district hospital for inpatient care [17]. One of the documents to come from the IMCI programme which looked to harmonise programmes and guidelines in hospitals was the WHO *Pocket Book of Hospital Care for Children — Guidelines for Management of Common Childhood Illnesses in Rural and District Hospitals with Limited Resources* (the 'pocketbook') [18]. This resource was introduced in 2005 and is available in 20 languages and has been widely distributed across Africa. However, uptake has only been partial in some countries and has been fragmented in others [19]. The pocketbook was revised in 2013 and the new edition is available in three languages. The uptake of the update has not yet been evaluated.

It is acknowledged that guidelines need to be introduced and supported via regular training, and as deficiencies were found in triage and emergency care in particular, a course was developed called Emergency Triage Assessment and Treatment (ETAT). ETAT was first developed in 1999 by an expert group of clinicians to help standardise emergency management and to link on from the methods taught in Advanced Pediatric Life Support (APLS) [12]. It also covered emergency treatments and focussed on the type of children that would be presenting to hospital in low income settings; it was developed by adapting previous triage guidance set up for adults and care of children in high-income settings. ETAT was evaluated in Brazil; trained nurses' assessments performed well compared to the 'gold standard', which was an APLS assessment done by a paediatrician [20]. Since 1999 ETAT has been successfully introduced to hospitals [21], although it has been difficult for researchers to evaluate ETAT directly with specific outcomes as it is often implemented along with other interventions and general improvement in hospital systems [22]. For example, in a hospital in Malawi, triage categories were implemented (as in ETAT) as follows: P1 – for patients requiring immediate life-saving care; P2 for patients requiring urgent care (within about 20-30 minutes); P3 – for patients whose needs are not urgent. This helped improve patient flow and along with training, changes in the lay-out and design of wards, and monitoring, helped to improve triage overall and reduce mortality [10]. However, improved outcomes have not been universally reported with introduction of ETAT: for example, a new paediatric intensive care unit in

Rwanda struggled to implement protocols and measures shown to improve outcomes and had a high mortality rate of 50% [23].

Kenyan researchers have expanded ETAT to create ETAT+ which links to their Clinical Practice Guidelines and has additional information on newborn resuscitation and common causes of serious illness [24]. ETAT has also been adapted for South Africa and was felt to be appropriate for a large busy hospital in Cape Town [25]. Paediatric ETAT guidelines were updated in 2016 by WHO to provide guidance on the most common emergency conditions in children presenting to hospital and included recommendations on oxygen therapy, oxygen flow rates, intravenous fluids, anticonvulsant medicines and diagnostic tests for children with seizures [26]. This update was only partial and WHO plan to address the other areas in future ETAT guideline updates. Paediatric emergency care training and ongoing support for health workers is required for improvements in acute care of critically ill children in low-income countries [27], and this, along with other interventions, is hopefully continuing to enable hospitals to move towards more consistent, good quality care in emergency rooms across the African continent.

1.1.2 Simple emergency treatment

The mainstay of treatment for critically ill children is antibiotics and antimalarials along with adjunctive supportive therapies where available in the hospitals. Antimalarials are generally widely available and since results of a multicountry trial (AQUAMAT) involving 5425 children showed that artesunate is superior to quinine for treating severe malaria in childhood [28], this has become standard of care.

One of the key adjunctive supportive therapies recommended for immediate management of children with shock is fluid resuscitation, which consists of very rapid infusion of fluid intravenously (a bolus). Use of boluses was supported by expert opinion (evidence level 2C)[29] and has been widely used in well-resourced countries (US, Europe, Australia). The aim of this treatment is to improve perfusion and correct shock. This therapy is also part of WHO Guidelines and was written in the Pocket Book of Hospital Care [18] in 2005 for children in severe shock (defined as cold hands and a weak and fast pulse and capillary refill time >3 seconds): fluid resuscitation though is not recommended in children with severe malnutrition. The fluids that could be used for this resuscitation were available in the majority of district hospitals (0.9% saline for example) [3] and the rest of the equipment needed is basic (such as lines and burettes) although burettes were not always available in these settings. The ETAT training includes use of boluses for certain types of children in paediatric shock (in the most

severe cases) - but there was only weak evidence regarding benefits of this supportive therapy in resource-limited settings. Small studies had been undertaken comparing different fluids for fluid resuscitation located in only one hospital [30, 31], and a systematic review of the choice of fluids for resuscitation in children in shock showed none of the trials had mortality as an endpoint and they predominately included only children with malaria or dengue fever [32]; thus, strong evidence was needed for fluid resuscitation in African settings.

1.2 Introduction to datasets used in this thesis

There are two datasets used in this thesis which are described in more detail below. One dataset is from a randomised clinical trial investigating the role of fluid resuscitation in low-income settings. This is the Fluid Expansion as Supportive Therapy (FEAST) trial which took place between 2009 and 2011 [33]. The other dataset is from a cohort study including all admissions to a district hospital in Kilifi on the coast of Kenya from 2002 – 2012 (KEMRI Kilifi admissions dataset).

1.2.1 Fluid Expansion As Supportive Therapy (FEAST) trial

Given the lack of good evidence for bolus resuscitation described above, a trial was developed by collaborators from the KEMRI Wellcome Trust, Imperial College London and the Medical Research Council Clinical Trials Unit (MRC CTU) to investigate the effect of immediate fluid resuscitation on child survival and what fluid should be used in low-income settings.

Children, aged 60 days to 12 years, with severe febrile illness (impaired consciousness (prostration or coma) and/or respiratory distress (increased work of breathing)) plus clinical evidence of impaired perfusion (one of capillary refill time (CRT)>2 seconds, lower limb temperature gradient, weak radial pulse volume or severe tachycardia) at 6 hospitals in Kenya, Tanzania and Uganda were enrolled into two strata according to systolic blood pressure. Stratum A included 3141 children without severe hypotension who were randomised to immediate bolus of 20ml/kg (subsequently increased to 40ml/kg after protocol amendment; of note, 40-60 mls/kg boluses are commonly used in high-income countries) of 5% albumin (albumin-bolus: 1050 children) or 0.9% saline (saline-bolus: 1047 children), or no-bolus (control, maintenance fluids 4ml/kg/hour, according to national guidelines: 1044 children). The protocol stipulated that the saline-bolus and albumin-bolus arms, but not the control arm, should receive an additional 20mls/kg bolus at one hour if impaired perfusion persisted. In all three arms, beyond one-hour, further 40ml/kg boluses of study fluid (saline for the control arm) were only prescribed if severe hypotension developed (see definition below). Stratum B included 29 children with FEAST entry criteria plus severe hypotension (defined as systolic blood pressure <50mmHg if <12m; <60mmHg if 1-5y; <70mmHg if > 5y) who were randomised to albumin or saline boluses 40-60mls/kg only. All trial patients received intravenous antibiotics if required, antimalarial drugs (for those with falciparum malaria) and IV

maintenance fluids (2.5-4mls/kg/hour as per national guidelines) until the child was able to retain oral fluids. Antipyretics, anticonvulsants and treatment for hypoglycaemia (blood sugar <2.5mmols/L) were administered according to nationally agreed protocols. Children with a haemoglobin <5g/dl were transfused with 20mls/kg of whole blood over 4 hours.

1.2.1.1 Trial Sites

The children were enrolled from six sites (hospitals) in three countries (Kenya, Tanzania and Uganda). The sites varied in size and experience with clinical research and are described below (number in brackets indicating the number of children enrolled in the trial).

Kenya – Kilifi District Hospital (216)

Kilifi Hospital is a district hospital on the Kenyan coast that is linked to the Kenya Medical Research Institute Wellcome Trust programme of research and has been involved in studies and clinical trials for over 20 years [34]. It is described in more detail in the section below covering the Kilifi Admissions dataset.

Tanzania – Teule District Hospital (97)

Teule Hospital is a district hospital in North East Tanzania where there is a high prevalence of malaria. The hospital is described as "a district hospital serving a rural population of approximately 277 000 people with a mortality in children under 5 of 165 per 1000" [35].

Uganda – Mbale Regional Referral Hospital (1240)

Mbale Hospital is a large regional referral hospital in Eastern Uganda. Malaria occurs throughout the year and is consistent with high stable transmission. The hospital has a 95 bed general paediatric department with 45 beds in a Pediatric Acute Care Unit. The department admits approximately 21000 children annually. The FEAST trial was the first major research project that had been undertaken in the pediatric ward of the hospital.

Uganda – Mulago Regional Referral Hospital (750)

This hospital is the national referral and teaching hospital of Makerere University Medical School, Kampala, Uganda, situated in the central region of Uganda. It receives patients from the capital city, Kampala, and referrals from the rest of the country. It has approximately 1500 beds and a previous study documented a 4.2% case-fatality rate among 23,342 children with malaria [36]. The paediatric unit admits between 30 and 80 patients daily. Mulago hospital also

functions as the primary health facility for the surrounding population and serves both peri-urban and rural populations [37].

Uganda – Soroti Regional Referral Hospital (633)

A large regional referral hospital also in Eastern Uganda. This hospital has a 62 bed paediatric ward for which the trial was the first major research project undertaken [34].

Uganda – St Mary's Hospital (Lacor/Gulu) (234)

St. Mary's Hospital in Lacor is a private, non-profit, 450-bedded hospital in Gulu District, Northern Uganda. The hospital has nine inpatient wards, of which one is a paediatric ward, and admits approximately 35,000 patients per year and serves a further 250,000 outpatients each year [38].

A good description of two of the FEAST sites was given in Molyneux S et al's paper on emergency consent procedures in the FEAST trial [34]. It compares the most research experienced of the sites, Kilifi district hospital, with Soroti hospital which was a new research site.

Figure 1.2.1: Description of two FEAST trial sites from Molyneux S et al, PloS One 2013.

Table 4. Key features of the two FEAST trial sites.

Site	Kilifi District Hospital (KDH), Kenya	Soroti Regional Referral Hospital (SRRH), Uganda
Size of paediatric wards	42 beds	62 beds
Experience conducting trials	Clinical trials conducted for over 20 years	First major research activity
Community engagement activities	Coordinated community engagement activities focusing on the institution and (where appropriate) specific studies	No formalised community engagement strategies for research
Employment and training of staff	Most staff within pre-existing clinical research group	Most staff involved in the trial trained and recruited specifically
Refusal rates for trial	High relative to other sites	Low relative particularly to KDH
Hospital user charges and other costs (a)	National exemption of charges for under five year olds, but little policy adherence. Food is provided for patients and most basic consumables available at a cost.	National exemption of charges for all admissions, but little policy adherence. Relatives provide and cook food in hospital grounds, and purchase many consumables needed (e.g. gloves, intravenous lines, some medicines) from local shops.
Trial inputs into study site	Established clinical programme as a robust platform for clinical trials (b). Also, trial provided extra personnel for emergency triage and patient monitoring, additional diagnostic tests and service-wide training in emergency care.	Basic maintenance and painting of paediatric wards, emergency and triage equipment and training for all staff, and employment of additional personnel.
Trial benefits for individuals (d)	Close observation, treatment of new illnesses identified during admission, and free treatment of minor illnesses post discharge up until the 28-day follow-up visit.	As with KDH, every trial participant was allocated a dedicated 'clinical trial pack'(c).

(a) See [32,33] for information on lack of adherence to user fee policies.
(b) Includes substantial support to the hospital for medical personnel (doctors, clinical officers, nurses and ward assistants); paediatric drugs, devices and equipment. Research funds also support the construction and running of an 8-bed paediatric high dependency ward available to all paediatric admissions, regardless of research involvement.
(c) Including cannulae, syringes, infusion sets, antibiotics, anti-malarials and blood testing consumables. Not needed in KDH where such support is provided to all inpatients in HDU.
(d) The usual hospital admission fees for the participants were not waived during the trial, in an effort to ensure that parents did not feel obliged to join the trial to save money. However, in Soroti only a few patients in a semi-private room are charged fees by the hospital.
doi:10.1371/journal.pone.0054894.t004

Training took place at all sites prior to the start of the trial including Good Clinical Practice (GCP) for all staff involved in the trial, as well as aspects of the ETAT training programme. A clear triage system was not in place at some sites and so introducing this process was also part of preparing the sites for taking part in the trial. The introduction of these processes was one possible reason for the overall mortality rate in the trial being lower than expected [21], highlighting the importance that improvements in basic clinical care had in the progress that was made towards the MDG.

1.2.1.2 Trial Eligibility – inclusion and exclusion criteria

The eligibility criteria of the FEAST trial were designed to be generalisable to acutely sick children presenting to hospital across Africa, and thus to cover a wide range of diagnoses. They were thus not specific to malaria or pneumonia or other illnesses but would identify children in shock for which immediate fluid resuscitation was hypothesised to be most beneficial. The three key elements of the eligibility criteria were for the child to have a history of fever or abnormal temperature, have severe illness and clinical evidence of impaired perfusion. These conditions were defined as follows:

- Abnormal temperature: Pyrexia (≥37.5°C) or Hypothermia (<36°C).
- Severe illness: one or more of impaired consciousness (prostration or coma) and respiratory distress
- Impaired perfusion: one or more of the following: capillary refill time >2 seconds, lower limb temperature gradient, weak radial pulse volume, severe tachycardia.

These eligibility criteria were informed by studies undertaken at Kilifi District Hospital, and were a more broad definition of shock compared with the WHO shock criteria (cold hands, weak and fast pulse, capillary refill time >3 seconds) which had identified very few children in the studies at Kilifi; also less than 3% of children in a study in Brazil met this criteria [20].

Children were excluded if they had severe acute malnutrition, gastroenteritis, non-infectious causes of severe illness (such as trauma, burns or intoxication) and in situations where

intravascular expansion was contra-indicated or the child had already received volume expansion.

If children had hypotension defined as systolic blood pressure < 50mmHg for children <12 months, <60mmHg for children 12month to 5 years, or <70mmHg for children >5 years old in addition to the criteria above, then they were enrolled into a different part of the FEAST trial (referred to as FEAST B) and were separately randomised between the two immediate fluid resuscitation strategies (0.9% Saline or 5% HAS) (no maintenance fluids group). This was applicable to only 29/3170 children enrolled.

The eligibility criteria were all clinical and possible to measure and record quickly and efficiently at the bedside, reflecting the need to randomise quickly in emergency situations and in settings where waiting for laboratory results is not an option. The exclusion criteria were designed as fluid management is not applicable for management of children with diarrhoea and burns, and is widely debated for children with malnutrition.

Data was collected on standardised CRFs for all required trial information, and on trial specific source documentation for additional observations and clinical notes. All CRF information was entered into a database, as well as the bedside observations from source documentation made from clinical reviews at specific time points (1, 4, 8, 24 hours) up to 48 hours from randomisation. The primary endpoint was mortality at 48 hours after randomisation, and secondary endpoints were mortality at 4 weeks, neurological sequelae at 4 and 24 weeks and adverse events potentially related to fluid resuscitation. There were two primary comparisons planned: saline bolus compared to control, and albumin bolus compared to saline bolus. The sample size was initially calculated as 2800 assuming a risk of 15% in the control group and giving a power of 80% to detect a 33% reduction relative reduction in the saline bolus arm compared to control, and 40% reduction for the albumin bolus arm compared to saline. The sample size was increased to 3600 following a lower than expected mortality rate in the control arm but the trial was stopped early following a recommendation to the Trial Steering Committee by the Independent Data Monitoring Committee [39].

1.2.1.3 Trial results

The trial enrolled 3170 children over 2 years and found increased mortality in the immediate fluid resuscitation 'bolus' arms compared to the standard of care, with 111/1050 (10.6%) in the albumin-bolus arm, 110/1047 (10.5%) in the saline-bolus arm and 76/1044 (7.3%) in the

control groups respectively dying within 48 hours [33]. This gave a relative risk for saline-bolus vs control of 1.44 (95% Confidence Interval (CI) 1.09-1.90), albumin-bolus vs saline-bolus 1.01 (95% CI 0.78-1.29) and for any bolus vs control 1.45 (95% CI 1.13 – 1.86). The increased mortality was also found at 28 days with 12.2%, 12.0% and 8.7% having died in the three groups respectively. The 48-hour mortality in FEAST B was similar in both arms but higher overall compared to the main trial, at 9/13 (67%) for the albumin-bolus arm and 9/16 (56%) in the saline-bolus arm.

Follow up in the trial was very good with only 17/3170 (0.5%) lost to follow up by 48 hours and 70/3170 (2.2%) by 28 days. The median age of children enrolled was 24 months (IQR 13-38) and the majority (52%) had more than one feature of impaired perfusion. Malaria was present in 57% of enrolled children, and Human Immunodeficiency Virus (HIV) present in 4% of those tested (106/2483). Other common diagnoses were severe anaemia (given as a diagnosis in 43% of children), lower respiratory tract infection (LRTI) (42%) and sepsis/septicaemia (15%). Of note, sepsis was not recognised as a specific entity in WHO guidelines and so may have been infrequently diagnosed. Also, previous studies have shown that severe LRTI according to WHO criteria could be metabolic acidosis due to malaria [40] and thus may be over diagnosed. Conversely, children enrolled in a large antimalarial trial, who were all considered to have malaria from a rapid diagnostic test, and were in the lowest tertile of plasma *Pf*HRP2, were found to have a low probability of malaria-attributable death and thus may have been misclassified [41]. Pre-planned subgroup analyses according to coma status, positive or negative status for malaria, presence or absence of severe anaemia (haemoglobin < g/dl vs ≥5g/dl), age, sex, base deficit (≥8 mmol/l vs. <8 mmol/l), lactate level (≥5 mmol/l vs. <5 mmol/l), and date of randomisation (before or after the protocol amendment) all supported the primary endpoint results with increased mortality in the bolus arms.

1.2.2 KEMRI Kilifi admissions dataset.

The Kenya Medical Research Institute (KEMRI) -Wellcome Trust Kilifi data are collected from all routine admissions into a district hospital on the coast of Kenya. The hospital admits 4400 paediatric patients and 3400 adult patients per year from a mainly rural population of approximately 240,000 [42] [43]. The data are also linked to demographic and surveillance information. The hospital has a 54 bed general paediatric ward and a 9 bed high dependency unit (referred to as the KEMRI ward). There is also a 20 cot newborn unit and occupancy for the units is between 150 and 200% with an average of 80-90 inpatients [44]. Research has

been carried out in the hospital and particularly the paediatric ward for many years and links into the research institute which also has a clinical trials unit. This has shown, for example, that there has been a decline in malaria paediatric admissions of up to 63% over 8 years from 1999-2007 and there have been changes in general in the disease burden of the area over the timespan of the dataset [45, 46].

Clinical data were collected by research clinical officers and doctors on standardised forms which also served as the clinical notes [47]. Specific training was also given for the recognition of some clinical signs [48].

The data from the Kilifi admissions dataset used for analyses in this thesis were collected from 2002 to 2012.

1.3 Aims of the thesis:

The overall aim of this thesis is to use statistical methods to delineate mortality risks and identify ways to improve management of the acutely sick African child. This will be achieved through the following objectives:

- Considering how to identify the most critically ill child and prioritise them within triage systems that have been introduced to hospitals across Africa. This will be done by identifying predictors of mortality, using multivariable Cox regression with fractional polynomials to build a prognostic model and create a risk score. The risk score will then be validated on external data and its performance compared to the performance of other risk scores in the literature applied to the FEAST trial data and the external data.

- Examining if a physiological measure that explained any of the (harmful) effect of boluses could be identified through estimating the proportion of treatment effect explained. Also examining and modelling the associations between continuous measures and mortality, both at baseline and during admissions to find the most appropriate functional form to calculate the proportion of treatment effect explained and understand if the effect of the boluses changed depending on the level of any continuous measure.

- Deconstructing mortality risk using flexible parametric models to directly estimate the changing baseline hazard over time. This will help discriminate between whether boluses immediately increased the mortality risk compared to the no bolus arm or whether it took a longer time from randomisation for the hazard to return to a baseline rate in the intervention arms compared to the no bolus arm.

2 What prognostic indicators for death could be used in African hospitals to identify children at greatest risk?

2.1 Introduction and objectives

The aim of this chapter is to identify what prognostic indicators could be used in African hospitals to identify children at greatest risk. There are many measures of clinical severity that can be recorded on admission to hospital and appropriate statistical methods will be used to examine the effect of these measures and how to combine them into a useful clinical bedside score for use in African hospitals. The primary objectives are to build a prognostic model using data from the FEAST trial, create a useful clinical bedside score from the prognostic model and validate the score on external data from the Kilifi admissions dataset. Secondary objectives are to validate other published scores using the FEAST trial data and to consider the impact of clinical measures that involve tests or additional equipment that may not always be available in African hospitals.

2.2 Overview of prognostic research

Prognosis in medicine is the term used when predicting the course of an illness, the outcome of an illness or predicting the health of individuals during a certain period of time.

Prognostic factors are the variables or characteristics that inform the prognosis that clinicians make regarding their patient's health and are also used to estimate the patient's risk of a particular outcome (whether that is recovery from a disease or recurrence of the disease or death). But rarely is one prognostic factor used in isolation to make predictions, as given the many factors that impact health and disease progression, it is likely to provide an inadequate estimate. Thus possible prognostic factors need to be evaluated in the presence of each other and so studies examining prognostic factors should be of multivariable approach and design. Combinations of prognostic factors can then be used to estimate outcome probabilities.

Tools to estimate outcome probabilities are commonly called prognostic models, prediction models, prediction rules or risk scores [49]. Although prognostic models evaluate population level risk (and often group patients into high or low risk strata) and predictive models estimate

individual risk (a given patient's probability of the event of interest) the literature uses these terms interchangeably to refer to both levels of risk. In this thesis I plan to, generally, use the term prognostic to cover all models. The aim of a good prognostic model and score is for the model to rarely fail to predict an event that will occur and also rarely predict an event when it will not occur [50]. This will lead to clinicians feeling confident in the score, rather than feeling that they would be able to make the same or a better prediction using their experience or knowledge only, and will lead to the score being used in clinical practice. In doing this the model is also successfully stratifying patients into high and low risk groups to inform decisions, and is able to give a summary of the type of patient being admitted to that particular hospital.

Prognostic factors or models can be used to: enable a balance of high and low risk patients enrolling into each arm in a clinical trial; enable economic analyses to take into account the risk level of patients in a study, or enable comparison of the performance of different hospitals to each other adjusting for their patient mix. One example of this is the Clinical Risk Index for Babies that was developed to compare neonatal intensive units in the UK [51]; another example are risk scores such as APACHE III [52] which enables calculation of expected hospital death rates to compare to actual death rates (the standardised mortality ratio) [53]. Predictive factors can also be used in sampling methods to estimate new disease within the population, for example HIV incidence.

Prognostic models are also used for patient care and determining treatment levels, to help with triage in emergency hospital departments (for example, there are a variety of scores to help assess prognosis in trauma cases [54]), to help inform end-of-life decisions or clinical decisions and to aid communication with carers and patients about prognosis or treatment [55]. For example, the APACHE III score mentioned above enables calculation of the standardised mortality ratio to compare hospitals by stratifying patients at admission, but is also used within a predictive equation to provide risk estimates for mortality for individual patients. Another example is the quick Sequential Organ Failure Assessment (qSOFA) score in adults, developed from the SOFA score to identify patients with suspected infection and who are at risk of sepsis [56]. The qSOFA can be used for risk stratification as patients with qSOFA ≥2 have a 3 to 14 fold higher mortality than those with <2 and is easy to calculate at the bedside (as it uses only hypotension (≤100mmHg), tachypnea (≥22/min) and altered mentation). [56]

A good prognostic score will have a clear aim that is described in the original published work and whenever used subsequently in validation studies. This will enable clinicians and

researchers to use and validate the score in the appropriate manner taking the aims and objectives of developing the score into account [57].

2.2.1 Prognostic models

2.2.1.1 Developing the model

To develop a prognostic model the most unbiased data is likely to come from a well-designed prospective cohort study [58], as data needs to be collected and patients need to be managed in way that is representative and generalizable, and for defined information on predictors and outcomes to be collected. Other observational studies such as case-control or cross-sectional studies can be used to investigate prognostic factors but they are weaker and can be more prone to bias than a cohort study, and retrospective cohort studies rely on previously collected data and so may not have measured all necessary factors or these may not have been recorded in the same way over time.

A randomised controlled trial is a good alternative study design to evaluate prognostic factors as the patients should have had the same treatment/management experience in all aspects apart from the intervention treatment, whereas for cohort studies, even if they had the same treatment available, the management may be different across different groups. If the randomised intervention has been ineffective then the randomised groups can be combined; otherwise the randomised groups can still be combined but the effect of the intervention will need to be included as a separate predictor in the multivariable model and care must be taken to check for possible interactions between randomised treatment and prognostic factors. One disadvantage of using clinical trial data to develop a prognostic model is that for some trials the eligibility criteria to the trial may have been very strict which could reduce the generalisability of the model. Another disadvantage is that in a clinical trial, often the clinical management is also stipulated in the protocol as well as the intervention and so may be different to that which the patient may have received in general clinical care. Although this may be the same for both arms it may be different to clinical practice outside of the trial, which may limit generalisability.

2.2.1.2 Building the model

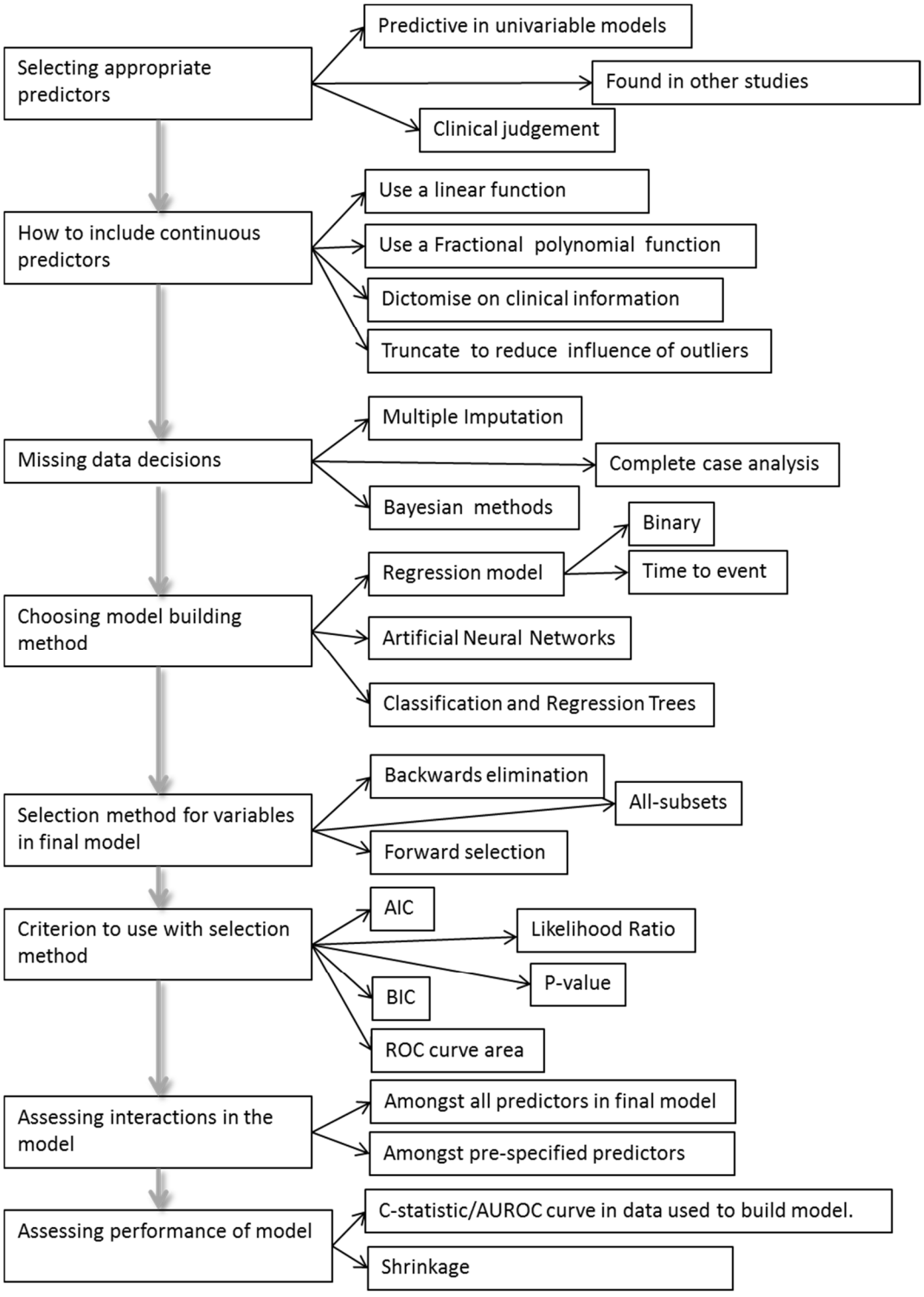
Multivariable analysis is the preferred method to evaluate prognostic factors to adjust for confounding bias. Factors should not be considered in isolation, although studies are published that still consider single rather than multiple predictors [59]. Choosing candidate predictors that are clinically relevant and easily measured in the situation under question is the first step in building a prognostic model, and this is then followed by formally selecting the variables to include in the model. This can be done with stepwise selection (either forwards or backwards elimination) with regression models or other methods such as artificial neural networks [60, 61] or Classification and regression tree modelling (CART) [62].

Artificial neural networks are used rarely due to clinicians wanting and needing to understand the model structure to be able to make sense of the predictions [50, 63]. Further, in a comparison with logistic regression for a heart procedure outcome, artificial neural networks did not perform any better than traditional modelling techniques [64, 65]. Artificial neural networks also have the disadvantage of needing complete datasets to analyse, i.e they cannot be extended to include incomplete observations [66]. Classification and regression tree modelling is a technique which repeatedly considers what pairwise split in any potential variable best explains variation in the outcome. It aims to create meaningful prognostic subgroups, and has been used successfully in some applications [67, 68]. Classification trees have the advantage of yielding a graphical display of results which is straightforward to understand and apply [69], but has the disadvantage of not being able to deal with many (particularly continuous) prognostic factors as too many nodes of the tree and thus subsets of the patients could be created.

The most commonly used method to build the model is backwards elimination with a regression model, either a Cox proportional hazards model for time to the outcome, or logistic regression with clinically relevant binary outcomes such as death, occurrence or remission of disease within a clinically relevant timeframe, or the presence or absence of a tumour. This reduces the number of candidate predictors to those that can form a model that best predicts the desired outcome. There are different options for the backwards elimination stepwise process including optimising the Akaike information criterion (AIC) [70], using p-values from Wald tests, or using the Bayesian information criterion (BIC) [71]. Prior to the model building, formal data reduction techniques can also be used to reduce the number of candidate variables to consider, including principal components, variable clustering, and deriving clinical summary indices. After the model has been built, then any selection bias that has arisen with weak predictors can be identified with shrinkage, or with bootstrap resampling and cross-validation [72].

The main choices that researchers make in building a prognostic model are summarised in the decision flow diagram in Figure 2.2.1 below.

Figure 2.2.1: Decision flow diagram for prognostic model building.



2.2.1.3 Validating the model

Once a prognostic model has been built then it needs to be validated. This validation preferably should be done on other data, known as external validation, but can also be done on a different part of the same dataset (internal validation or cross validation). Internal validation is a helpful process but can be over-optimistic as the two datasets (development and validation) are very similar and there is no information about the performance of the model outside the setting where it was created. External validation quantifies the performance of the model on a new set of patients, enabling researchers to evaluate the accuracy and generalisability of the model as set out in Justice A et al [55]. The main components of accuracy are calibration and discrimination. Calibration is looking at the extent of bias (predicted vs actual mortality (or event of interest)) and discrimination is measuring the model's ability to separate patients with different responses [72].

The most frequently used statistic to measure calibration is the Hosmer-Lemeshow test which compares the observed number of events with the number of events predicted from the model within categories on the basis of predicted risk [73, 74]. Discrimination for binary outcomes (used by the majority of prognostic models) is measured with the c-statistic or the equivalent 'area under the receiver operating curve (AUROC)'. The c-statistic is the proportion of all pairs of patients, one with the outcome and one without, where the patient with the outcome has the higher predicted probability of the outcome. The AUROC is the area under the curve produced by plotting the sensitivity against (1-specificity) when the linear predictor is dichotomised at many cut-offs to categorise high or low risk. The c-index as defined by Harrell et al [72] is also used for measuring discrimination (and is equivalent to the c-statistic for binary outcomes) but extends to survival models [75]. The generalisability of a model is looking at its reproducibility (whether the model is accurate in a similar set of patients) and transportability (whether the model is accurate in a different but related population (or where methods of data collection are different)). Internal validation methods cannot substitute for external validation of a model [76], and external validation of a prognostic model gives a much higher level of evidence and is thus a necessary step in moving from a research model to a useful tool for clinicians [77].

2.3 Prognostic factors for mortality in paediatric intensive care

2.3.1 Literature review

There were 1067 articles that were returned from a literature search of MEDLINE in February 2012 with a search term made up of “Child” or “Infant” or “Pedatrics/Paedatrics” in MeSH terms, and “death” or “mortality” in MeSH terms and “prognostic” or “predictor” in all fields within English language journals, excluding all those related to cancer (via the search term “neoplasm” in all fields). The search was designed to highlight the prognostic factors for mortality in children especially those in intensive care or critically ill with non-cancerous diseases. The search brought up a variety of literature including: literature on mortality following organ donation, transfusion, surgery, or trauma; literature on pregnancy outcomes and neonates; and literature on social behaviour and economics along with other topics. From this wide variety of topics, journal articles were retained on prognostic factors for mortality where children have not had trauma, surgery, or chronic illness, on modelling prognostic factors for mortality and also severity of illness, and on scoring systems, predictive risk scores and risk stratification in paediatrics. Literature was rejected from their title or abstract for the reasons in Table 2.3.1.

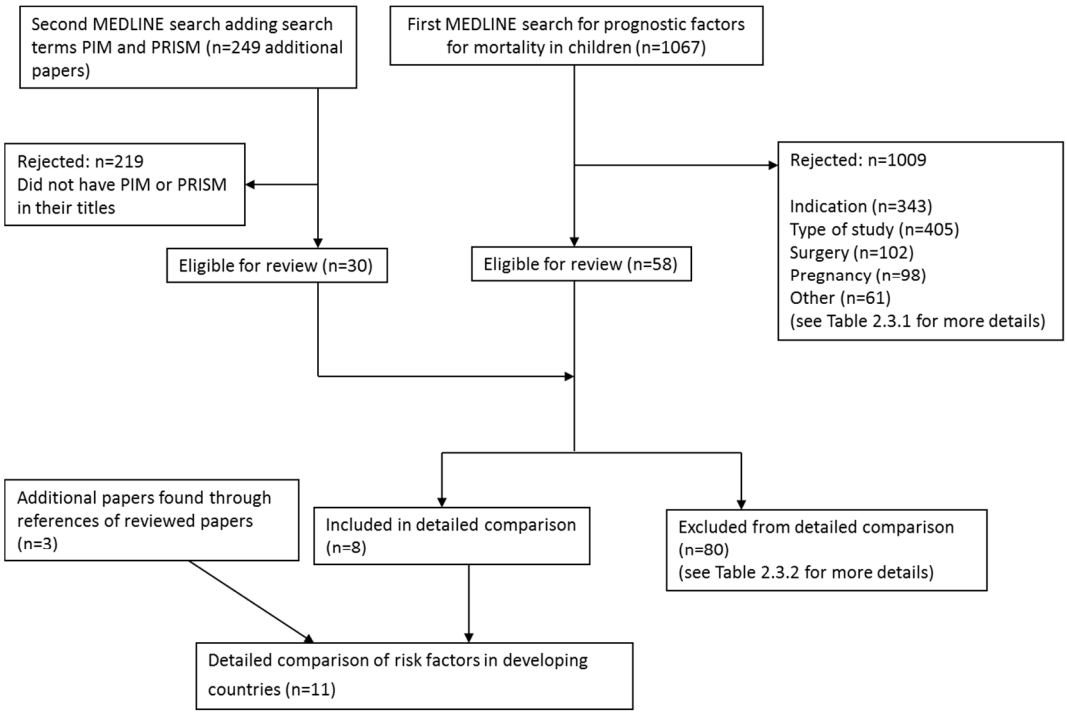
Table 2.3.1: Reasons for rejection of papers that appeared in first MEDLINE search.

Reason for rejection	**Number of papers (%)**
Organ disorders/conditions (including heart disease, kidney and liver disease, renal failure, heart failure)	167 (17%)
Fetal, neonatal or perinatal studies	128 (13%)
Surgery procedures	102 (10%)
Pregnancy outcome	98 (10%)
Transplant or transfusion medicine	92 (9%)
Trauma or Injury (including burns, poisoning, blunt trauma, broken bones, head injuries)	84 (8%)
Chronic diseases and cancer/stroke (including HIV, epilepsy, cystic fibrosis, stroke, cerebral palsy, diabetes and cancer)	56 (6%)
High Income setting	63 (6%)
Congenital disorders (spina bifada, cleft palate, congenital abnormalities)	36 (4%)
Environmental studies	38 (4%)
Adult studies	24 (2%)
Other types of studies (including animal studies, case reviews, trial protocols, forensics, genetics)	60 (6%)
Other*	61 (6%)

* including: non-mortality outcome, biomarkers, immunisation, ocular lesions, skin disease, thymus size, diphtheria, mitochondrial disease, tissue infection.

The search was then modified to include the search term 'pediatric risk of mortality' or 'PRISM' and 'pediatric index of mortality' or 'PIM' as these are two widely used and known paediatric risk scores and due to some papers on PRISM or PIM not including child mortality in their MeSH terms they had not come up in the previous search. This brought up an additional 249 papers to the original search. Their titles were reviewed and all those with PIM or PRISM (or the full name of either score) in their titles were kept (n=30) and all others rejected. Thus evaluations of these risk scores were included in this second search. These two searches are summarised in Figure 2.3.1.

Figure 2.3.1: Flow diagram describing literature search



There were 58 papers from the initial literature search that were kept for full review. They came under four main headings: paediatric risk of mortality scores and their evaluation (any country) (n=9); general methods (path analysis, neural networks, building prognostic scores) (n=12); risk factors for mortality in low-income countries especially Africa (n=21); risk factors of particular interest (i.e base excess, lactate) (n=5); and pneumonia, sepsis or meningococcal disease specific scores (n=11).

To enable a more detailed comparison with a clinical risk score developed using the FEAST data, I focused on papers that described predictors of mortality (using multivariable analyses) for critically ill children in low-income countries. Eighty papers were not included in the detailed comparison and the reasons for their exclusion are detailed below (Table 2.3.2). However, they were thoroughly reviewed to examine the application of their methods and a summary of that review is presented in Section 2.3.2.

Table 2.3.2: Reasons for non-inclusion in detailed comparison

Reason for non-inclusion	Number of papers (%)
Disease/Condition specific score (taking into account aetiology of disease/condition)	23 (29%)
Evaluation of pre-existing risk score	21 (26%)
No multivariable analysis for model building	12 (15%)
Methodology only	8 (10%)
Non-mortality outcome	4 (5%)
Other	12 (15%)

Other reasons are: systematic review (1), commentary (2), risk factors not measured at baseline (3), survey (1), subjective risk predictions (1), full text unobtainable (2), risk factors from mother not child (1)

2.3.2 Review of other literature

The literature that came up in the search but which was not included in the detailed comparison was reviewed and the summary of the review follows the categories presented in Table 2.3.2.

Disease specific score

Papers reviewed in this category included multivariable analyses of predictors of mortality but were specific to the disease or condition under examination, in that they included factors that were only relevant to that disease/condition. Several papers covered meningococcal septic shock or septicaemia and there are several scores for use in this area, including the Glasgow Meningococcal Septicaemia Prognostic Score [78], and the score from the Barcelona Meningococcal Disease Surveillance Group [79]. Barquet *et al* created their score using data from Spain by using logistic regression and selecting variables that were independently predictive to estimate a linear predictor and then calculated integer points based on the coefficients from the linear predictor which were summed for the score. The integer points were estimated from the linear predictor by using the smallest coefficient to divide all other coefficients by that number and rounding to the nearest integer. Scores for primary myelofibrosis were developed with stepwise Cox regression models [80].

Evaluation of pre-existing risk scores

The majority of the papers that were not included due to the analyses presented only evaluating a pre-existing risk score were examining and comparing the Pediatric Risk of Mortality (PRISM) scores and Pediatric Index of Mortality (PIM) scores. A summary of the findings from these papers is found in the overview given in Section 2.4.2. The other papers in this category not evaluating PRISM or PIM were examining the Acute Physiology and Chronic Health Evaluation (APACHE) score [81] or the Rapid Acute Physiology Score [82].

No multivariable analysis for model building process

The papers that came up in the literature search without multivariable analyses covered a range of disease areas and topics including commentaries on ethics [83] and case studies. The disease areas included meningitis, anaemia, hyperglycaemia, and cardiac surgery, and most papers considered individual predictors such as serum albumin, serum lactate, bacterial pathogens, or organ failure. These papers also identified possible risk factors using other methods such as ranking variables by their univariable area under the ROC curves or creating indices by counting the number of risk factors present [84]. One paper considered clinical factors outside of the hospital and examined the discriminative ability of nutritional and socio-demographic factors individually for mortality in rural Bangladesh [85].

Methodology only

The methodology found during the literature review varied with its application. Two papers modelled the whole birth weight distribution of neonates especially taking into consideration very low birth weight infants, and extremely low birth weight infants who lie at the edge of the distribution [86, 87]. Firstly, the distribution of birth weight was modelled with a 2 or 4 component normal mixture model which was chosen over other models such as the contaminated normal model. This was then combined with logistic regression for mortality and the odds of mortality at fixed values of the birth weight distribution were estimated. In a similar population a graphical tool was developed from a logistic regression model to help predict survival from four key variables available at birth [88]. The prematurity risk evaluation measure (PREM) was graphically presented using isosurv graphs (gestation-and-sex adjusted birth weight centiles plotted against gestation, with curves joining up points of equal survival probability, calculated by solving the logistic regression equation linking the predicted odds of survival to gestation (and other factors), over a series of z-score values) to predict survival on the basis of gestation time in days and birth weight or base deficit in the umbilical cord.

Confidence intervals as estimates of clinical certainty (i.e from the clinician) were compared to confidence intervals around estimates of survival made with risk scores, in particular the PRISM III score to look at difference in interpretation in one paper [89]. Methods to combine information from predictions from clinicians and predictions from risk scores were also evaluated.

Ruttimann et al extended the PRISM score to predict daily the probability of a patient dying in the next 24 hours [90]. Logistic regression was used to estimate the probability of dying the next day, from a weighted sum of all previous PRISM scores. A stepwise process found the admission score and most recent score were sufficient in the model. The linear predictor found was then named the Dynamic Objective Risk Assessment (DORA) and time spent with this value above a certain point could also be evaluated to predict mortality [91].

Most risk scores were developed for a binary outcome – most focused on mortality, predicting whether the patient was dead or alive at some pre-specified time point. However, it can be very useful to be able to predict more outcomes than two and, for example, include information on neurological status. Ruttimann et al [92] used stepwise polychotomous logistic regression analysis for three outcomes – functional, compromised or dead. The performance was then measured in a 3x3 classification table as there were no general measures of predictive performance for more than two outcomes. A lot of clinical information was used for the predictor including diagnostic factors from four different systems and six aetiologies, operative status, and the PRISM score as baseline function status. The authors felt it was feasible and useful to have a score for more than two outcomes.

Multivariable logistic regression models built with stepwise procedures as above are the most commonly used method to develop risk scores but artificial neural networks (ANN) have been used to predict mortality in a neonatal intensive care unit. The advantage is they do not rely on pre-defined models but the disadvantage is that they are unable to work with incomplete data [66]. By using a single imputation of 'normal' (categorised as zero) in place of missing values for binary variables and using the mean of the distribution for missing values in continuous variables, the ANN method was able to classify neonates with reasonable performance. Although the paper describing the application of ANN methods to neonates did not mention interactions between predictors nor non-linearity of predictors, interactions are examined by ANN implicitly. But non-linearity cannot be examined using this method unless specific non-linear terms are explicitly provided as predictors to the ANN.

In the reviewed papers multivariable logistic regression models were rarely tested for non-linearity and interactions in the sets of predictors chosen.

Non-mortality outcome

Although the MEDLINE search terms included mortality or death, some papers examining other outcomes were identified. Two of these papers investigated whether the PRISM score could predict other outcomes: one considered nosocomial infections, where they found that the score on admission may be useful to identify those at increased risk of infection [93]; and the other examined resource utilization as an outcome [94] and found PRISM could stratify patients for resource usage but underestimated mortality in paediatric trauma. One paper considered predictors for pneumonia in inpatient and outpatient children in Brazil [95], and the other looked at children in cardiac arrest as inpatients in Taiwan to predict primary and secondary outcomes following Cardio-pulmonary Resuscitation (CPR), including survival to hospital discharge, return of circulation, neurological status at discharge and post discharge survival up to 1 year [96].

Other

Several articles described the use of risk scores in clinical practice. There is particular interest in comparing PRISM and PIM and also considering under what conditions they are best used [97]. For example, Garcia et al concluded that PIM2 is an easy and efficient prognostic tool for low-income countries [98] compared to PRISM III, as PIM2 used fewer variables, was freely available, and only used information at admission compared to using all collected data within a 24 hour window. There were two meta-analyses looking at predictors of mortality, one for mortality in untreated HIV-infected children [99] and the other using clinical trial data evaluating neonatal vitamin A supplementation to reduce mortality and morbidity (which also reported predictors of mortality) [100]. Other papers considered risk factors that were not measured at admission to the hospital, such as whether the child lives in a female-headed household [101], or used updated measures of organ function and treatment measures once admitted [102].

2.3.3 Detailed comparison of selected papers

There were 11 papers that described predictors of mortality for critically ill children in low-income settings, with 7 based on the African subcontinent, 3 in south or central America, and 1 in Papua New Guinea. Three of these papers were identified only through references and had not come up in the MEDLINE search (as outlined in section 2.3.1 above) as they did not include the terms 'death' or 'mortality' in their keywords or MeSH terms ([103, 104]) and one did not include 'prognosis'([105]). The MeSH terms that they did include are listed in Table 2.3.3 below. Newton et al included 'predictive value of test' which may link to mortality but it is not clear, and von Seidlein included severity of illness to reflect the mortality risk score that was developed.

Table 2.3.3: Table of MeSH terms for three papers identified through references rather than the search strategy.

MeSH terms for papers identified through references		
Newton *et al* [104]	Maitland *et al* [105]	von Seidlein *et al* [103]
Acid-Base Equilibrium/physiology	Acidosis/drug therapy	Africa
Acidosis/metabolism	Acidosis/etiology	Antimalarials/administration & dosage
Adolescent	Blood Transfusion/methods	Artemisinins/administration & dosage
Child	Child	Child
Child, Preschool	Child, Preschool	Child, Preschool
Female	Female	Female
Ghana	Fluid Therapy/methods	Humans
Humans	Humans	Infant
Infant	Hypovolemia/complications	Injections, Intravenous
Kenya	Hypovolemia/drug therapy	Malaria, Falciparum/diagnosis
Lactic Acid/blood	Infant	Malaria, Falciparum/drug therapy
Linear Models	Infant, Newborn	Malaria, Falciparum/mortality
Malaria,	Kenya/epidemiology	Malaria,

Falciparum/blood Malaria, Falciparum/diagnosis Malaria, Falciparum/mortality Malaria, Falciparum/physiopathology Malawi Male Predictive Value of Tests Prognosis Sensitivity and Specificity	Malaria, Falciparum/complications Malaria, Falciparum/mortality Male Retrospective Studies Survival Rate	Falciparum/pathology Male Prognosis Quinine/administration & dosage Severity of Illness Index Treatment Outcome

The 11 papers fully reviewed are listed in the table below.

Table 2.3.4: Papers selected for detailed comparison

First Author	Title	Year published	Country	Number of children in analyses
Allen, S J [106]	Severe malaria in children in Papua New Guinea	1996	Papua New Guinea	489
Berkley, J [47]	Prognostic indicators of early and late death in children admitted to a district hospital in Kenya: cohort study	2003	Kenya	8091
de Leon, A [107]	Simplified PRISM III score and outcome in the pediatric intensive care unit	2005	Mexico	170
Evans, J [62]	Capillary refill time as an independent prognostic indicator in severe and complicated malaria.	2006	Ghana	2446

Maitland, K [105]	Severe *P. falciparum* malaria in Kenyan children: evidence for hypovolaemia	2003	Kenya	515
Marsh, K [108]	Indicators of Life-threatening Malaria in African Children	1995	Kenya	1844
Newton, CR [104]	The prognostic value of measures of acid/base balance in pediatric falciparum malaria, compared with other clinical and laboratory parameters	2005	Malawi, Ghana and Kenya	14605
Planche, T [109]	A prospective comparison of malaria with other severe diseases in African children: prognosis and optimization of management	2003	Ghana	1492
Roine, I [110]	Influence of admission findings on death and neurological outcome from childhood bacterial meningitis	2008	South America	654
von Seidlein, L [103]	Predicting the Clinical Outcome of Severe Falciparum Malaria in African Children: Findings From a Large Randomized Trial	2012	8 African countries	5426
Werneck, G [111]	Prognostic factors for death from visceral leishmaniasis in Teresina, Brazil	2003	Brazil	90

Each of the above studies was reviewed in close detail and described below. They cover a large time span – from 1995 to 2013. The majority of populations that were studied were children with malaria, although the paper by Berkley *et al* has no restriction on disease, Roine *et al* focussed on those with bacterial meningitis, and Wernack *et al* on leishmaniasis. One of the studies - von Seidlein *et al* - is a secondary analysis of data collected for a clinical trial (the AQUAMAT trial (Artesunate vs Quinine in the treatment of severe falciparum malaria in African Children)). The other studies are either designed to focus on prognostic factors for mortality or are secondary analyses of data collected for large studies (such as Evans *et al* and Newton *et al* where the data was collected for the Severe Malaria in African Children (SMAC) research network). The prognostic factors measured in each study and what factors were found to be associated with mortality in multivariable analyses are summarised in Table 2.3.4 at the end of this section.

Allen SJ et al, Severe Malaria in children in Papua New Guinea, Q J MED 1996

489 children with severe falciparum malaria admitted to Madang Hospital in Papua New Guinea were included in the analysis for this paper. 22% also had severe anaemia, 16% were in a coma and malaria was hyperendemic in the area. The study was a cohort study restricted to children living in the province for >12 months at entry upon admission to hospital with malaria (based on WHO criteria of severe and complicated malaria) and were managed according to national guidelines. The criteria for WHO severe and complicated malaria are split into defining and supporting criteria and are: defining criteria - coma, severe malarial anaemia, pulmonary oedema, hypoglycaemia, circulatory collapse, renal failure, spontaneous bleeding, repeated convulsions, acidosis, haemoglobinurea; supporting criteria - impaired consciousness, jaundice, prostration, hyperpyrexia, hyperparasitaemia. Severe malarial anaemia was defined as haemoglobin <5g/dl and parasitaemia ≥ 10000/µl, level of consciousness was defined by the Blantyre Coma Score (BCS), with coma defined as BCS≤2, and shock was defined as a systolic blood pressure < 50 mmHg with cold clammy skin. The BCS is made up of three components – eye movement, motor response and verbal response; each is scored from 0-2 (or 0-1 for eye movement) with 0 being no response, and the components are summed to give the BCS [112]. There was no intensive care (thus no mechanical ventilation) or increased monitoring available for the children included in the study (children were reviewed twice a day and as required at other times).

The paper looked at associations between different severe manifestations of malaria (for example the association between severe anaemia and cerebral malaria) with tabulations and

chi-squared tests. Predictors of death were analysed using a forward stepwise logistic regression model. The paper closely describes the manifestations and consequences of severe malaria including neurological effects and metabolic complications. Overall mortality was low with only 17/489 dying (3.5%; 95% CI 1.9-5.1%). 16 died within the first 24 hours (12 in first 12 hours, 4 in second 12 hours) and 1 at 5 days. All severe manifestations of malaria (cerebral malaria, acidosis, hyperlactataemia) were found in univariable analysis to be significantly associated with mortality apart from anaemia (it is unclear whether the univariable analyses are adjusted for age). Laboratory measures (apart from haemoglobin) were found to have non-linear relationships with the odds of mortality and thus were log-transformed (using log base 2) when included as continuous variables in modelling. High levels of lactate had the strongest association. The final model (with only 15 events, 2 events excluded due to missing data) contained lactate, bicarbonate, creatinine with odds ratios per 2-fold increase and haemoglobin as a linear function. Having adjusted for these variables, coma and glucose did not add further prognostic information and the authors indicated that acidosis was the strongest predictor of mortality. There was less heavy parasitaemias than observed in Kenyan settings (≥500,000 in 1.7% of children compared to 8.9% in Kenya) and the authors considered that the difference between settings could be an explanation for coma not being identified as a risk factor. However, low mortality rates have been shown in children in similar areas in the South West Pacific, where malaria is hyperendemic [113]. Further, as only 489 children were included, the power to find independent associations in this study by Allen *et al* is low, and they also have a possible spurious finding within the multivariable model as higher haemoglobin was associated with mortality.

Berkley J et al, Prognostic indicators of early and late death in children admitted to a district hospital in Kenya: cohort study, BMJ 2003

Berkley *et al* considers not just predictors of inpatient mortality in children but predictors for immediate (<4 hours), early (4-48 hours) or late (after 48 hours) deaths in admitted children regardless of diagnosis. The study took place in a rural district hospital in Kilifi, Kenya and was a cohort study using an admissions database. They used data from 8091 children greater than 90 days old admitted over a period of 2 years with 436 (5%) deaths to develop scores, and then data from 4082 children enrolled in a subsequent period for validation of the scores. Their aim was to develop a score to use in clinical practice, audit and research settings in sub-Saharan Africa. To select variables predictive for each type of death, the authors used positive and negative likelihood ratios (LR) as calculated in a diagnostic test study. A positive LR is defined as (sensitivity/(1-specificity)), a negative LR as ((1-sensitivity)/specificity). These ratios were then adjusted for potential confounding from other variables in an analysis following the

methods of Spiegelhalter and Knill Jones [114] which is based on decision support systems. Variables were removed that had an adjusted likelihood ratio of either ≥0.67 or ≤1.5 as these had the least independent predictive value. The score was then prospectively validated using area under receiver operating curve analyses. The univariable methods used in the paper are intuitive for clinicians and closely linked with sensitivity and specificity. These methods also have the advantage of being able to include variables in the score when their presence is beneficial (i.e when present they reduce the score total), but the method to adjust the predictors for other variables and to select variables is unusual and not as clear as other more frequently used methods such as regression.

There were 60 immediate deaths, 193 early and 183 late deaths in the development dataset. General clinical predictors not specific for any aetiology of a particular disease were compared by calculating crude likelihood ratios (as described above) for each type of death. The predictors were length of history of illness, history of fever, history of diarrhoea, history of cough, history of seizures, jaundice, severe anaemia, wasting, kwashiorkor, temperature, raised respiratory rate, (subcostal) indrawing, deep breathing, and neurological status. The multivariable models that were built for each type of death were overlapping in terms of variables that they include but each model added or excluded a few different ones. Although this is of interest in comparing risk between different types of deaths, this could lead to confusion if the score was put into practice and it is less intuitive than one set of risk factors for mortality in general. A score was constructed by assigning points approximating to the natural logarithm of the adjusted likelihood ratio for each indicator. Validation of the scores created gave area under the ROC equal to 0.93 (95% CI 0.92-0.94) for immediate, 0.82 (0.80-0.83) for early and 0.82 (0.81-0.84) for late deaths. These showed very good sensitivity and specificity for the scores but this might be expected as the validation data was from the same setting and from a period immediately after the development data.

De Leon A et al, Simplified PRISM III score and outcome in the pediatric intensive care unit, Pediatric International 2005

This small study enrolled 170 children with complete clinical or laboratory test data to a cohort within a Mexican paediatric intensive care unit. It is unclear how many were screened but not enrolled and the age range of the children was from 1 month to 16 years which is very wide encompassing infants in post-natal period through to early adulthood. The Pediatric Risk of Mortality (PRISM) score was calculated at the time of admission and within the first 8 hours. Multivariable logistic regression analysis was carried out using the backwards elimination

procedure by considering the 17 variables that contribute to the PRISM score as independent variables. The major admitting diagnoses were in the postoperative period from major surgery, septic shock and severe head trauma. There were 42 (25%) deaths but it is not clear in what time period they occurred and there are conflicting statements on length of hospital stay for those enrolled. There were also not enough deaths to reliably assess 17 covariates in a model. Only 4 of the 17 measured variables were found to be independently significant and they were abnormal pupillary reflexes, acidosis, blood urea nitrate (BUN) and white blood cell count. The number of children in the categories for acidosis or abnormal pupillary reflexes was not presented and the confidence interval for abnormal pupillary reflexes was very wide (OR 9.9 (95% CI 3.5-28.4). This may be due to the small sample size and the number of variables considered in the model (n=17) but the authors still felt that the PRISM score in their setting could be reduced to just 4 measures. The authors did not present a model using the PRISM score on their data and instead they considered the specificity and sensitivity of certain cut-offs of the PRISM score. A cut-off of 13 was found to have the best sensitivity (0.71) and specificity (0.64) - this is only just showing fair discriminatory performance and may not be high enough to be considered clinically useful, which could also be due to the small sample size.

Evans J et al, Capillary refill time as an independent prognostic indicator in severe and complicated malaria. Journal of Pediatrics 2006

This was a prospective study that enrolled 2466 children with severe and complicated malaria (defined as being positive for malaria parasitaemia and with haemoglobin <5g/dl, or lactate >5mmol/L or BCS<3) admitted to a tertiary referral hospital in Ghana as part of a study looking at genetic factors influencing malaria. Logistic regression was used to build multivariable models using stepwise process (it is not clear whether this was forwards or backwards elimination) with a cut-off of p=0.2. Classification and regression tree (CART) modelling was independently used to identify the most prognostic factors for subgroups. Then the sensitivity and specificity of the independent prognostic factors the model identified, alone and in combination, were calculated from the same data. There were 172 (7%) deaths but it is unclear as to whether these were inpatient deaths only. The multivariable analysis identified coma, prolonged capillary refill time (CRT) (>2s), respiratory distress and acidosis to be independently associated with mortality and the decision tree used all these variables apart from acidosis. Using all four signs they were able to predict 90.7% of the deaths in the cohort (82.5% without using acidosis) and prolonged capillary refill time had the highest specificity

but low sensitivity. The authors concluded that prolonged CRT may be an indicator of a distinct malaria complication and discussed whether it reflected the presence of hypovolemia and suggested that it should be added to criteria defining severe and complicated malaria. The cohort of children was pre-selected to already have coma, severe anaemia or hyperlactataemia and so it is only in addition to these risk factors that prolonged CRT can be included. Also, the decision tree was only evaluated on the data used to create the model.

Maitland K et al, Severe P. falciparum malaria in Kenyan children: evidence for hypovolaemia, QJM 2003

This is a retrospective review of the admission records of children with malaria admitted to a paediatric high-dependency unit of a hospital in Kilifi, Kenya. This is the same hospital that Berkley et al used for their analyses and is a subset of those patients. The analysis included 515 children with malaria (slide positive test, and at least one of the following: coma or prostration (reported as a single entity as cerebral malaria ie either prostration or BCS ≤2), seizures, respiratory distress, circulatory collapse, anaemia, or jaundice). The overall mortality rate was 12.8% (66/515). Laboratory variables were categorised from clinical factors defined by Paedatric Advanced Life Support (PALS) guidelines. Logistic regression was used for both univariable and multivariable models. In addition to the models, a clinical index for shock was derived from guidelines and other information and was evaluated using the ROC curve. The score created gave each of the following one point each (apart from hypotension (systolic blood pressure <80 mmHg for children ≥1 year and systolic blood pressure <70 mmHg for children <1 year) for which two points was given): hypoxia (<90%), tachypnoea, delayed or prolonged capillary refilling time ≥3 seconds), hypothermia (<36), acidosis (and/or deep breathing), elevated creatinine or dehydration. The selection process for the multivariable model is not outlined in the paper but the variables found to be significantly associated with a fatal outcome in the model were oxygen saturation <90%, deep breathing, raised creatinine (>80 μmol/l) and hypoglycaemia (<2.5 mmol/l). There was some overlap between the multivariable model and the shock score, and the mortality rate increased as the shock score increased. However, the shock score focused on hypotension (present in 66/507 (13%) children, with 26% mortality) which was not found to be a risk factor after adjusting for other factors in the multivariable analysis. Also, all the variables included in the model were dichotomised at clinically appropriate cut-points and none were assessed continuously. Thus, information or associations could have been lost in the process. The score was also dichotomised at less than vs greater than or equal to 2 to assess sensitivity and specificity. The

study focused on assessing the impact of volume resuscitation on outcomes as the authors felt that they had evidence that hypovolaemia is likely to be a major factor in the aetiology of acidosis. This meant they did not extend the multivariable model or clinical shock score any further.

Marsh K et al, Indicators of Life-threatening Malaria in African Children, NEJM 1995

1866 children whose primary diagnosis was malaria (defined as peripheral parasitaemia and no other detectable clause for clinical presentation) were enrolled into a study based at Kilifi district hospital, Kenya. The enrolled children had a median age of 26 months. 86/1866 children died, but deaths during the admission process (n=18) and deaths not from malaria (n=4) were excluded from all analyses giving an overall case fatality rate of 3.5% (64/1844). The children were assessed using the WHO criteria for severe malaria (see definition above in the detailed review of Allen et al, QJ Med 1996), but replacing pulmonary oedema with a definition of respiratory distress and severe respiratory distress. The authors used reasonable clinical cut-offs for categorising variables but did not evaluate them continuously, and used backwards stepwise logistic regression starting with all variables that had $p<0.1$ in univariable analyses to find a 'minimal-effects' model. The variables entered into the multivariable-model building process were coma, anaemia, respiratory distress, convulsions, impaired consciousness (defined as a depressed level of consciousness but can localise a painful stimulus), jaundice, hyperpyrexia, hyperparasitaemia and prostration. Major predictors were found to be impaired consciousness, jaundice and respiratory distress and hypoglycaemia (though the model was built on a reduced number of observations (n=673) for this last variable). They compared these predictors and any combination of them to the WHO criteria, and a prognostic index created from the four major risk factors found in their model performed better, as the index predicted 92% of observed deaths compared to 80% of observed deaths predicted by the WHO criteria. The paper reported other analyses including looking at, in turn, malaria with impaired consciousness, malaria with anaemia, and malaria with respiratory distress but these were all very descriptive with no regression models built.

Newton C et al, The Prognostic Value of Measures of Acid/Base Balance in Pediatric Falciparum Malaria, Compared with Other Clinical and Laboratory Parameters, CID 2005

This study enrolled 14,605 children at 3 sites across Africa (Kilifi in Kenya, Blantyre in Malawi, Kumasi in Ghana) through the Severe Malaria in African Children network over a period of 2 years. The sites each contributed around 5000 children and some differences in severity of disease and mortality were found between sites. The study used the point at which sensitivity

and specificity curves crossed to find optimal cut-offs for lactate and base excess and to examine the association between these variables and mortality. A multivariable logistic regression model was built for each site, first focussing on clinical and demographic variables initially and then adding affordable laboratory parameters on 10,559 children with complete information. Base excess and lactate were added at the end to assess their value in addition to variables already in the model. Model selection was based on the all-possible subsets regression procedure to maximise the adjusted-R^2. This is where all possible models with one, two, three or more predictors (given the set of candidate predictors) are fitted and the best fitting model according to the adjusted-R^2 is chosen. The variables that remained in the final model were weight-for-age score (whether WHO or CDC standards used to calculate this are not reported), deep breathing, BCS ≤2, unable to sit, hypoglycaemia, lactate and base excess (within sites other variables were included: vomiting in Blantyre, irregular breathing and seizures in Kumasi, and irregular breathing, seizures and indrawing in Kilifi but addition to model did not increase c-index by >0.01). The authors concluded that base excess and lactate did not appreciably add to the predictive ability of a model based on clinical features and a blood glucose test, but with the caveat that there were some issues with the base excess samples not being tested immediately and becoming alkalotic (due to evaporation of CO_2). This is a well-powered study even though they built separate models for each site, as each model included between 60 to 183 deaths. However, the c-index is not very sensitive to additions of new variables and so is not the best method to compare predictive ability between models.

Planche T et al, A prospective comparison of malaria with other severe diseases in African children: prognosis and optimization of management, CID 2003

Planche *et al* evaluated prognostic factors in severely ill children admitted to a teaching hospital in Ghana (this is the same hospital as Evans J *et al* used for their study) comparing factors found in those with and those without severe malaria. They enrolled 1654 children into the cohort study and a total of 1492 children had evaluable outcomes. The authors used logistic regression and area under the ROC (AUROC) as summary measure of the predictive value of each variable individually. The multivariable model was built using a forward selection procedure using the AUROC by adding variables to the logistic model when there was >5% increase to the AUROC based on the linear predictor. Separate models were built for malaria positive and malaria negative children. Overall case-fatality rate was 8.5% (127/1492) with 6% malaria vs 11.3% without malaria (diagnoses included meningitis, pneumonia and sickle cell disease). In those with malaria the predictive variables were low BCS, hyperlactatemia, and low Body Mass Index (BMI); in those without malaria it was low BCS, hyperlactatemia, respiratory distress, and low haematocrit (severe anaemia). The malaria model had an overall

AUROC of 0.84 which would be expected as it is calculated on the data the model was built with. The model had BCS and BMI fitted as linear and lactate as a dichotomous variable with the cut-off of ≥5 or <5 based on the clinical definition of hyperlactatemia. In the non-malaria model haematocrit was fitted as a linear function. The authors discussed their models and felt that respiratory distress may have identified those most at risk in the non-malaria group, as children with pneumonia were included in this group. This paper contrasted with others as it concluded that only two predictive factors were needed for those with severe malaria (hyperlactatemia and coma) whereas other studies often found that respiratory distress was also very predictive of mortality in children with malaria. But they had another result that was not explained further, namely higher BMI was independently associated with increased mortality which may be confounded by a different clinical sign. Overall, however, this was a well powered study that carefully examined risk factors in both malaria and non-malaria children.

Roine I et al, Influence of Admission Findings on Death and Neurological Outcome from Childhood Bacterial Meningitis, CID 2008

This was a retrospective review of 654 children with bacterial meningitis presenting to hospital across South America, with a case fatality rate of 13% (86/654). The study investigated overall mortality as well as other outcomes, defined as mortality or severe neurological sequelae, and mortality or any neurological sequelae, and was focusing on simple criteria which would be easy to ascertain in low-income countries. The authors did not report what percentage of children admitted to the hospitals had bacterial meningitis and all had CSF measured, which is often rare in low-income settings. They used a multivariable logistic regression and built a model for mortality keeping independent variables where p≤0.05 but with a reduced number of observations (n=332), as only complete case analyses were used and some variables had missing data. The method for building the model was not clear and it may be that they included all variables and reported only those for which p≤0.05. The model found that conscious level, capillary refill time >3s or a CSF protein concentration >250 g/dL increased the risk of death independently of each other. Conscious level was the variable with prognostic importance for all three outcomes and it was not the aetiology of the disease that was most predictive.

von Seidlein L et al, Predicting the Clinical Outcome of Severe Falciparum Malaria in African Children: Findings from a Large Randomised Trial, CID 2012

This study analysed data from a randomised controlled trial comparing artesunate with quinine for the treatment of severe falciparum malaria (AQUAMAT) in 9 African countries (The

Gambia, Mozambique, Nigeria, Rwanda, Kenya, Democratic Republic of the Congo (DRC), Tanzania, Ghana and Uganda). They recruited 5426 children in the trial with the final model fitted on 4089 children (402 (10%) deaths). It is unclear exactly why it is such a reduced number although part of the missing data stems from continuous variables measured using an i-STAT cartridge, as witnessed in the FEAST trial. The authors evaluated 20 different predictors with multivariable logistic regression (keeping those variables with $p<0.01$ when all variables were included in the model), adjusted for randomisation arm and stratified by site, and also considered fractional polynomials for continuous variables (glucose, log parasitaemia, and respiratory rate). ROC curves were used to evaluate the predictive ability of the final model within the same dataset. They concluded that the predictors that remain in the model (and are thus independently associated with increased risk of death) were acidosis defined with base excess, coma and/or convulsions (cerebral manifestations of malaria), an elevated Blood Urea Nitrogen (BUN) or signs of chronic illness on admission (any one of the following: lymphadenopathy, malnutrition, candidiasis, severe visible wasting and desquamation). At clinically defined cut-offs each predictor individually had high specificity but lower sensitivity (from 28-76%). The authors also created a score where the presence of each of the following variables contributed one point each to the score: base excess <-8mmol/L, BUN≥20mg/dL, combined coma score<3, chronic illness and convulsions. The score was not validated within this study, and the rationale behind allocating one point per variable to create the score was not clear.

Werneck G et al, Prognostic factors for death from visceral leishmaniasis in Teresina, Brazil, Infection 2003

This was a very small case control study set in a hospital in Brazil in patients with confirmed visceral leishmaniasis. Cases were those that died during treatment (n=12) and controls were a random sample of those alive (n=78). The authors used logistic regression in univariable analyses to identify important factors (although it is unclear what significance level was used to define importance), and then went on to use backwards elimination to select variables for the multivariable model with a cut-off of $p<0.1$ which may not be appropriate given the number of events included. They then built a prognostic score by dividing the regression coefficients by the lowest one and rounded to the nearest integer. The predictive performance of the model was then evaluated using AUROC curves but on the same data that the model was developed with. The multivariable analysis identified four variables independently predicting death: diarrhoea, jaundice, fever and haematocrit<20%. The odds ratios calculated by the multivariable model had very low precision (and very wide confidence intervals) due to the very small number of cases in each group and in general the very low power of this study

indicated that the findings may be due to chance. The prognostic score described in the results assigned 1 for each variable, giving a score from 0 – 4 as coefficients are all between 2.06-2.76. The authors acknowledged the very limited statistical power of their study but felt it may be useful to generate suggestions for interventions. There has been no validation of the score in data other than that it was generated with.

Table 2.3.5 below summarises which prognostic factors were measured within each study and also which were found to be significantly associated with mortality in multivariable analyses. The most common factor that was found to provide independent information, after adjustment for other factors, was conscious level, with those in a coma at highest risk of mortality and this held for the majority of studies that included it in models (7/10 studies). In the Maitland *et al* QJM 2003 paper impaired consciousness or cerebral malaria was defined as either prostration or Blantyre Coma Scale ≤2, which was present in 78% of the study population with overall mortality 12% which may give an explanation as to why coma was not identified by the study as an independent risk factor [115]. Other factors prognostic for mortality that were found frequently (in 3 or 4 studies) are raised temperature, raised respiratory rate or respiratory distress, jaundice, low haemoglobin, and deep breathing. Factors that were not measured in many studies (3 or less), but when measured were often found to have a strong association with mortality in multivariable analyses, were raised creatinine, prolonged capillary refill time and low weight-for-age z-score. As expected in these settings, most of the prognostic factors were measured clinically at the bedside with few laboratory values reported or included in models.

Table 2.3.5: Prognostic factors measured in each study included in the detailed comparison

A shaded box is a risk factor evaluated in analyses, **x** is a variable found to be independently predictive of mortality in multivariable models.

	Independently predictive in any papers? (excluding FEAST).	Allen	Berkley J			de Leon	Evans	Maitland	Marsh	Newton	Planche	Roine	von Seidlein	Werneck	FEAST
			Immediate deaths	*Early deaths*	*Late deaths*										
Age	**0/9**														
Base Excess	**2/5**						**x**						**x**		
Bicarbonate	**1/2**	**x**													
Blood pressure	**0/6**														
Bulging Fontenelle or stiff neck															
BMI	**1/1**										**x**[a]				
Capillary Refill time	**2/3**						**x**					**x**			
Conscious level	**7/10**		**x**	**x**	**x**		**x**		**x**	**x**	**x**	**x**	**x**		
Convulsions	**2/5**											**x**	**x**		
Creatinine	**2/2**	**x**						**x**							
CSF	**1/2**											**x**			

	Independently predictive in any papers?*	Allen	Berkley (immediate) deaths	Berkley (early deaths)	Berkley (late deaths)	de Leon	Evans	Maitland	Marsh	Newton	Planche	Roine,	von Seidlein	Werneck	FEAST
Decreased skin turgor	**0/1**														
Deep breathing	**3/4**		**x**					**x**		**x**					
Ethnicity	**0/1**														
Gender	**0/8**														
Glucose	**2/9**							**x**	**x**						
Haemoglobin	**3/10**	**x**	**x**								**x**[b]			**x**	
Heart rate	**0/7**														
Height	**0/1**														
History of cough	**0/2**														
History of diarrhoea	**1/2**													**x**	
history of seizures	**0/2**														
HIV positive	**0/0**														
Indrawing	**1/2**		**x**	**x**											
Jaundice	**3/4**		**x**	**x**					**x**					**x**	

	Independently predictive in any papers?*	Allen	Berkley (immediate) deaths	Berkley (early deaths)	Berkley (late deaths)	de Leon	Evans	Maitland	Marsh	Newton	Planche	Roine,	von Seidlein	Werneck	FEAST
Kwashiorkor	**1/1**			**x**	**x**										
Lactate	**2/4**	**x**									**x**				
Liver size >2cm below costal margin	**0/1**														
Malaria parasitaemia	**0/6**														
MUAC	**0/0**														
Other lab values	**2/5**					**x (pH, Blood Urea, WBC)**							**x (Blood Urea)**		
Oxygen saturation	**1/1**							**X**							
Residence	**0/1**														
Respiratory	**3/7**						**x**		**x**		**x**[b]				
Severe pallor	**0/0**														
Sunken eyes	**0/1**														
Temperature	**2/9**		**x**		**x**									**x**	

	Independently predictive in any papers?*	Allen	Berkley (immediate) deaths	Berkley (early deaths)	Berkley (late deaths)	de Leon	Evans	Maitland	Marsh	Newton	Planche	Roine,	von Seidlein	Werneck	FEAST
Temperature gradient															
Unequal pupils	**1/1**					**x**									
Vomiting	**0/1**														
Wasting	**1/2**			**x**	**x**										
Weight	**2/3**							**x (for age)**		**x (for age)**					

[a] The authors indicate this finding to a lesser extent than the other risk factors in their model.

[b] This was for malaria slide negative children only.

* excluding FEAST

2.4 Reviewing risk scores found in literature using the FEAST clinical trial data

2.4.1 Available variables

There were 48 clinical variables collected on children admitted to hospitals in the FEAST trial, recorded at randomisation (focusing on triage measures such as blood pressure, heart rate, axillary temperature, conscious level) or shortly after the administration of randomised treatment started (with focus on clinical history and full clinical examination). The children were critically ill and thus informed assent was used alongside informed consent enabling randomisation and treatment initiation to happen very quickly after admission. The clinical variables were well-completed and on average only had <2% missing data apart from height, head circumference and mid upper arm circumference (MUAC) which had 9%, 8% and 7% missing data respectively. Some clinical variables were also available from different time points after randomisation.

There were 22 laboratory variables collected on the children which varied by site and by availability of results from a handheld blood analyser (i-STAT, made by Abbott laboratories). The laboratory information comes from three sources: bedside tests (oxygen saturation, malaria rapid diagnostic test, haemoglobin from Haemocue test, glucose and lactate (from handheld machines)); blood tests (white blood cell count, platelets, blood culture, malaria slide); and i-STAT tests (sodium, potassium, base excess, pH, BUN and others). The i-STAT devices were used at admission and 24 hours but some sites found them difficult to use routinely throughout the trial and so there is 33-43% missing data at baseline and 70-79% missing data at 24 hours for these test results. Malaria slides were well completed (1% missing) but white blood cell counts and platelets were done on very few children (73% missing data). Blood culture was available in 1085/3170 (34%) children.

Some clinical signs can be subjective and studies have shown inter-observer variation in measurement of various signs in children on admission to hospital [48, 116]. This would be expected to cause dilution bias of any effect, but would be something seen in practice anyway. Thus if there is variation in measurement between observers then the results of analyses should be even more applicable to clinical situations.

2.4.2 Validating appropriate models already developed

The performance of models designed to predict mortality in paediatric intensive care units (PICUs) (PRISM and PIM) was first assessed on the control arm data from FEAST (n=1044). Models from the 11 papers fully reviewed were then assessed along with the PRISM and PIM scores as they were well known scores and had been validated in a variety of settings.

2.4.2.1 PRISM and PIM scores overview

The PRISM (Pediatric Risk of Mortality) score was developed by Pollack *et al* in 1988 from the Physiologic Stability Index (PSI) to assess the severity of illness in children in PICUs in the USA [117]. It simplified the PSI by reducing the number of variables to be measured from 34 to 14 and restricting the many ranges that were present in the PSI. They used a development dataset of 1415 children with 116 deaths (8%), and validation dataset of 1227 children and 105 deaths (9%). The authors simplified the PSI by first removing all variables in the PSI that were not associated with mortality by a chi-squared test ($P>0.25$), and then all the factors that were present in <1% of the population, which is appropriate when the development dataset is large; however, there is a risk that information on variables that occur rarely, but could be very predictive, is lost. They performed stepwise logistic regression (not reporting whether they used backwards elimination or forward selection), keeping factors with $p\leq0.3$ and then pooled pre-assigned ranges (from PSI) with adjacent ranges if in logistic regression models the coefficients for adjacent ranges were not different from each other ($p>0.15$). This helped to reduce the number of individual categories of variables included in the model.

As well as internal validation where the score was validated on data from separate ICUs within the same database, the score has been externally validated in various situations and in other countries to varying degrees of success. PRISM over-estimated the severity of illness in children in Sheffield Hospital, UK [118], and it did not discriminate mortality well in a PICU in South Africa, indicating the score may not be population independent [119], but was found to be valid in The Netherlands [120] and France [121].

The researchers behind the PRISM score then looked to re-evaluate the score in the light of developments in interventions and monitoring in PICUs over 1988-1995, and in 1996 they published the PRISM-III score [122]. The score was split into PRISM-III-12 using only measurements from the first 12 hours, and PRISM-III-24 where measurements in the first 24

hours are used. They then refined this further to create PRISM-III-APS (Acute Physiology Score) which was felt to be more sensitive to change in physiological variables [123]. PRISM-III has been validated in various settings. In Italy it was found to under-estimate mortality and so had unsatisfactory calibration [124] and over-predicted death in Australia and New Zealand [125], but it performed well in The Netherlands [97], Hong Kong [126] and India [127]. PRISM-III-APS has not been externally validated or discussed as widely in the literature (from a citation search carried out in February 2018).

The Paediatric Index of Mortality (PIM) score was first developed in Australia and published in 1996 as a response to PRISM which the authors felt was very complicated to calculate (with 14 variables to measure) with the further disadvantage that it uses the worst value for each variable over the first 24 hours so can only be applied retrospectively [128]. The PIM score is calculated using baseline values and was developed using standard methods very similar to those of PRISM, excluding those variables not associated with death in a univariable analysis and then performing backwards stepwise regression. The PIM score was internally validated in data from Australia where it was developed and also from a UK PICU, and in general felt that it was most useful in identifying groups of patients rather than making decisions about the management of an individual patient. It has also been validated and compared to PRISM in other settings [97, 126, 129, 130]. The authors revised PIM in 2002 to create PIM2 which replaced the factor 'specific diagnosis' with two factors - 'high-risk diagnosis' and 'low-risk diagnosis'; and recovery from surgery and bypass were separated in the new score [131]. The authors of PIM2 also internally validated their score on a wider range of PICUs.

All these scores focus on predicting mortality for the PICU and estimating Standardised Mortality Ratios and so aim to adjust for case-mix when comparing units rather than being used to inform individual patient's management [132].

2.4.2.2 PRISM score applied to FEAST data

The PRISM score was the first score published following the PSI and has been subsequently modified since to give the PRISM III score which is the most up to date version of the score[117, 122]. The authors of the PRISM and PRISM III scores did not include the coefficients for the scores in their publications, but the PRISM score since its original publication has had its coefficients made publicly available in a separate publication and so its performance can be fully assessed in the FEAST data [133]. The value allocated to each variable in the PRISM score is outlined in Figure 2.4.1 below and was published along with the regression coefficients in a paper evaluating the score.

Figure 2.4.1: PRISM score calculation taken from Balakrishnan *et al*, Arch Dis Child 1992 [133]

Variable	*Age restrictions and ranges*		*Scores*
Systolic blood pressure (mm Hg)	*Infants*	*Children*	
	130–160	150–200	2
	55–65	65–75	2
	>160	>200	6
	40–54	50–64	6
	<40	<50	7
Diastolic blood pressure (mm Hg)	*All ages*		
	>110		6
Heart rate (beats/min)	*Infants*	*Children*	
	>160	>150	4
	<90	<80	4
Respiratory rate (breaths/min)	*Infants*	*Children*	
	61–90	51–70	1
	>90	>70	5
	Apnoea	Apnoea	5
Arterial oxygen tension:fractional inspired oxygen ratio*	*All ages*		
	200–300		2
	<200		3
Arterial carbon dioxide tension (kPa)†	*All ages*		
	6·80–8·66		1
	>8·66		5
Glasgow coma score‡	*All ages*		
	<8		6
Pupillary reactions	*All ages*		
	Unequal or dilated		4
	Fixed and dilated		10
Prothrombin time:partial thromboplastin time ratio	*All ages*		
	>1·5×control		2
Total bilirubin (μmol/l)	*>1 Month*		
	>60		6
Potassium (mmol/l)	*All ages*		
	3·0–3·5		1
	6·5–7·5		1
	<3·0		5
	>7·5		5
Calcium (mmol/l)	*All ages*		
	1·75–2·00		2
	3·00–3·74		2
	<1·75		6
	>3·74		6
Glucose (mmol/l)	*All ages*		
	2·2–3·3		4
	13·9–22·2		4
	<2·2		8
	>22·2		8
Bicarbonate (mmol/l)§	*All ages*		
	<16		3
	>32		3

*Cannot be assessed in patients with intracardiac shunts or chronic respiratory insufficiency. Requires arterial blood sampling.
†May be assessed with capillary blood gases.
‡Assessed only if there is known or suspected central nervous system dysfunction. Cannot be assessed in patients during iatrogenic sedation, paralysis, anaesthesia, etc. Scores <8 correspond to deep stupor or coma.
§Use measured values.

Not all variables were available in the FEAST dataset (as outlined in Table 2.4.1 below), and there were differing amounts of missing data in variables that were recorded, and so two versions of this score were assessed; one including potassium, glucose and bicarbonate which were only available for two thirds of the cohort, and the other without these variables.

Table 2.4.1: The availability of variables used to calculate PRISM in the FEAST control arm data.

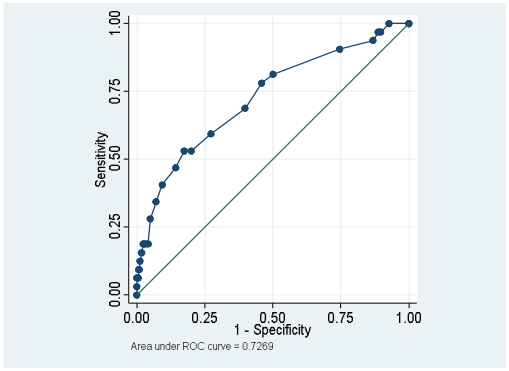
Number of variables	8	5
Number of control group children with all variables recorded (of total 1044)	612	1015
Variables:		
Systolic blood pressure	x	x
Diastolic blood pressure		
Heart rate	x	x
Respiratory rate	x	x
Arterial Oxygen Tension		
Arterial Carbon Dioxide Tension		
Glasgow coma score	x	x
Pupillary reactions	x	x
Prothrombin time		
Total bilirubin		
Potassium	x	
Calcium		
Glucose	x	
Bicarbonate	x	

The PRISM score is calculated using all available information for the above variables in the first 24 hours and taking the value that gives the highest score in that time period (although pre-terminal measurements should not be used). To calculate the score in FEAST data all values taken within 1 hour prior to death were excluded and any children that died within one hour of admission were excluded. The variables that were measured more than once in the initial 24 hour period were systolic blood pressure, heart rate, respiratory rate, consciousness (mapped to the Glasgow coma score) and pupillary reactions, and the worst of the values that was not pre-terminal was used. The PRISM authors do not give details about when the children in their study died and whether any died within 24 hours of admission to the PICUs in the study. This suggests that they were able to use complete data for the first 24 hours on all patients and assess mortality after this point though it is not clear from the published work. The PRISM score in the FEAST data was calculated for mortality after 1 hour from admission up to 48 hours to enable AUROC curves to be estimated and these are presented in Figure 2.4.2. The

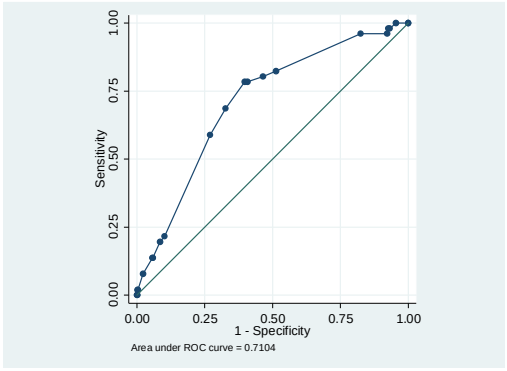
curves gave AUROC estimates as 0.73 (0.63-0.82) for the score with 8 variables and 0.71 (0.64-0.78) for the curve with 5 variables.

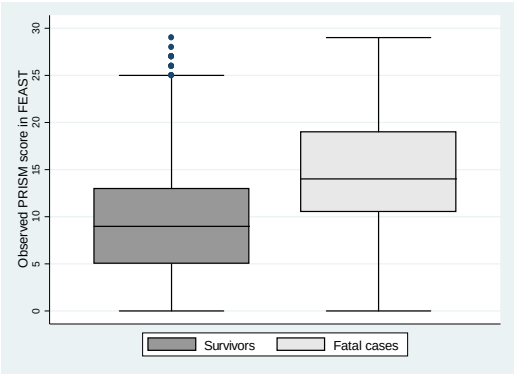
Figure 2.4.2: AUROC plots for mortality risk estimated using PRISM score and box plots comparing the distribution of PRISM score in survivors compared to fatal cases.

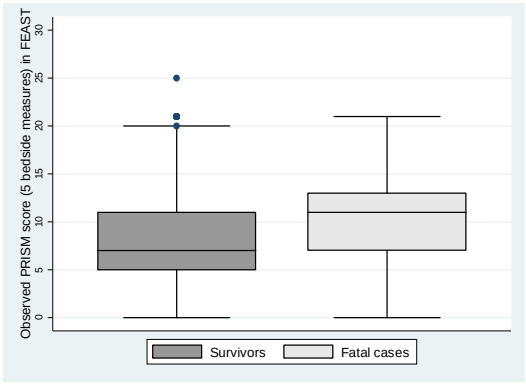
a) AUROC analysis for PRISM score with 8/14 variables



b) AUROC analysis for PRISM score with 5/14 variables







Balakrishnan *et al* published the original equation to use the PRISM score to calculate the probability of death for each patient. The equation with the regression coefficients is as follows:

$$Probability\ of\ death = \frac{\exp(r)}{1 + \exp(r)}$$

where $r = 0.207xPRISM - 0.005x\ age\ (in\ months) - 0.433\ x\ operative\ status - 4.782$

(As none of the children in our cohort were post-operative the variable *operative status* is set to 0).

This equation was used to calculate the number of predicted deaths by each quintile of risk (based on observed PRISM scores in FEAST) by summing the probabilities over the

observations in that quintile. This was then compared with observed deaths. Table 2.4.2 below presents this information for the PRISM score calculated with 8 variables (thus including glucose, potassium and bicarbonate).

Table 2.4.2: Observed and expected deaths in the control arm FEAST data by quintiles of mortality risk estimated with PRISM

Quintiles	N	Probability	Observed Deaths	Predicted Deaths	Scaled Predicted Deaths*
First	121	0.004-0.021	1	2	1
Second	123	0.021-0.031	4	3	2
Third	123	0.032-0.059	2	6	3
Fourth	120	0.06-0.117	9	10	5
Fifth	125	0.117-0.758	12	30	17
Total	612	0.004-0.758	28	51	28

*normalised to total observed deaths.

The PRISM score predicted 51 deaths within those children where information was available to calculate the score, based on 8 of the 14 variables; for comparison, the predicted deaths have been scaled in the table to match the total of 28 observed deaths. The Hosmer-Lemeshow goodness of fit test for calibration gives p=0.07 indicating that it is not particularly well calibrated for the FEAST dataset.

As the score was calculated both with 8 variables on a small proportion of the dataset and 5 variables on a much larger proportion of the dataset, the quintiles of risk have also been calculated for the score calculated with only 5 variables. The Hosmer-Lemeshow goodness of fit test gives a p-value=0.41 for the bedside score showing better calibration of the score compared to when 8 variables of the score were used. The total number of deaths predicted by the score with 5 variables was 49 which was scaled to the 51 deaths observed in these 1015 children in Table 2.4.3.

Table 2.4.3: Observed and expected deaths by quintiles of mortality risk estimated with the PRISM score calculated using 5 bedside measures.

Quintiles	N	Probability	Observed Deaths	Predicted Deaths	Scaled Predicted Deaths*
First	203	0.004-0.018	2	2	2
Second	193	0.018-0.022	3	4	4
Third	213	0.022-0.043	8	5	5
Fourth	200	0.043-0.067	16	13	14
Fifth	206	0.068-0.576	22	25	26
Total	1015	0.004-0.576	51	49	51

*normalised to total observed deaths.

Table 2.4.3 shows that with 5 variables the PRISM score is reasonable at predicting an appropriate number of deaths per mortality risk quintile and is an improvement compared to using the score with 8 variables. This may be due to the limited number of children that the 8 variable score could be evaluated on, and that 2 of the 3 additional measurements used in this score come from the i-STAT cartridge where there is more potential variability in their ascertainment.

2.4.2.3 PRISM III score applied to FEAST data

The PRISM-III score is based on the PRISM score above and is calculated in a similar way, though there are some differences in the variables included [122]. The score uses 17 variables (10/14 from PRISM (variables excluded are diastolic blood pressure, calcium, bicarbonate, bilirubin) and the following variables added: temperature, creatinine, platelets, white blood cells, urea, acidosis, and pH) with scores assigned to severe high or low values, with the cut-offs defined by clinical judgement and where there were two or more categories these were combined if the regression coefficients of the categories were within the standard errors (Figure 2.4.3). The ranges for systolic blood pressure, heart rate, urea, prothrombin time, creatinine were also age dependent. It similarly uses the most severe of the measures recorded in the first 24 hours or prior to death if within 24 hours. As above, all values taken within 1 hour prior to death will not be used for calculating the score and children that died within 1 hour of admission are excluded.

Figure 2.4.3: PRISM III-24 scoring sourced from Tan *et al* , Ann Acad Med Singapore 1998 [134].

APPENDIX II [PRISM III - 24]

Variable	Age restrictions				Score appointed	Score given
Systolic blood pressure (mmHg)	Neonate	Infant	Child	Adolescent		
	40-55	45-65	55-75	65-85	3	
	<40	<45	<55	<65	7	
Temperature	All ages <33°C or >40°C				3	
Mental status	All ages: stupor or coma (GCS <8)				5	
Heart rate	Neonate	Infant	Child	Adolescent		
	215-225	215-225	185-205	145-155	3	
	>225	>225	>205	>155	4	
Pupillary reflexes	All ages = One pupil fixed, pupil >3mm				7	
	All ages = Both fixed, pupil >3mm				11	
Acidosis (pH) or total CO_2 (mmol/L)	All ages = pH 7.0-7.28 or total CO_2 5-16.9				2	
	All ages = pH <7.0 or total CO_2 <5				6	
pH	All ages = 7.48-7.55				2	
	All ages >7.55				3	
PCO_2 (mmHg)	All ages = 50.0-75.0				1	
	All ages >75.0				3	
Total CO_2 (mmol/L)	All ages >34.0				4	
Arterial PaO_2 (mmHg)	All ages = 42.0-49.9				3	
	All ages <42.0				6	
Glucose	All ages >11.0 mmol/L				2	
Potassium	All ages >6.9 mmol/L				3	
Creatinine (µmol/L)	Neonate	Infant	Child	Adolescent	2	
	>75	>80	>80	>115		
Urea (mmol/L)	Neonate	All other ages			3	
	>4.3	>5.4				
White blood cells	All ages < 3000 cells/mm³				4	
Prothrombin time (PT) or partial thromboplastin time (PTT)	Neonate	All other ages			3	
	PT >22.0 sec or PTT >85.0 sec	PT >22.0 sec or PTT >57.0 sec				
Platelets (cells/mm³)	All ages = 100,000 to 200,000				2	
	All ages = 50,000 to 99,999				4	
	<50,000				5	
Total PRISM III-24 score						

(NB: Neonate= <1 months; infant=≥1 month – 12months; child ≥12 months – 144 months; adolescent > 144months)

In the FEAST data there were 12/17 variables available for 224/1044 (20%) of the control arm data, 10/17 variables available for 623/1044 (60%), and 5/17 variables for 1015/1044 (98%). This is due to some variables involving haematology (such as platelets and white blood cells) that had limited availability, and others using information from the i-STAT cartridge which was also not recorded for every child. If all variables in the score were available then the range of the score is from 0 to 74. Within the FEAST data the possible range for the PRISM III score created from 12 variables was 0 to 56, from 10 variables it was 0 to 47, and 0 to 30 for the score created with 5 variables. The actual ranges seen in the FEAST control arm data were 0 to

19 for the score with 12 or 10 variables and 0 to 15 where only 5 variables were used. But not all values within this range were found in the control arm data. By not using the full score we may have a situation where there may be significant loss of sensitivity as a predictor of mortality [135], and also that the remaining variables may have the wrong weighting or values assigned to them within the score. The variables available at each level are summarised in Table 2.4.4 below.

Table 2.4.4: Variables used to calculate PRISM III that are available in the FEAST control arm data.

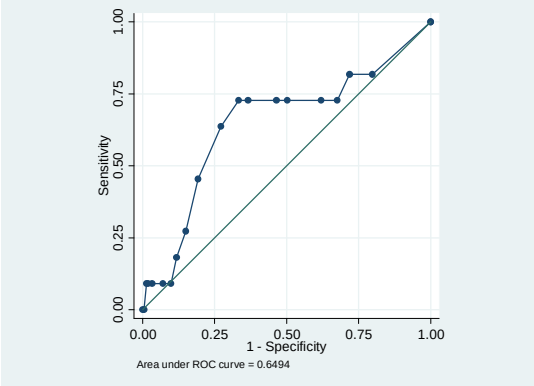
Number of variables recorded	12	10	5
Number of control group children with all variables recorded of total N=1044	224	620	1015
Variables:			
Systolic blood pressure	X	x	x
Temperature	X	x	x
Mental Status	X	x	x
Heart rate	X	x	x
Pupillary Reflexes	x	x	x
Acidosis or total CO_2	x	x	
pH	x	x	
PCO_2	x	x	
Total CO_2			
Arterial PaO_2			
Glucose	x	x	
Potassium	x	x	
Creatinine			
Urea			
White blood cells	x		
Prothrombin time			
Platelets	x		

The fit of the score can be assessed initially by estimating the ROC curve and AUROC using non-parametric methods on data from the control arm of the FEAST trial only. These give AUROC of 0.65 (0.59-0.72) for the score with 12 variables, 0.66 (0.55-0.77) for the score with

10 variables and 0.66 (0.59-0.74) for those score with 5 variables. These were calculated on data from within the first 24 hours of admission but excluding measurements that were recorded within 1 hour of death.

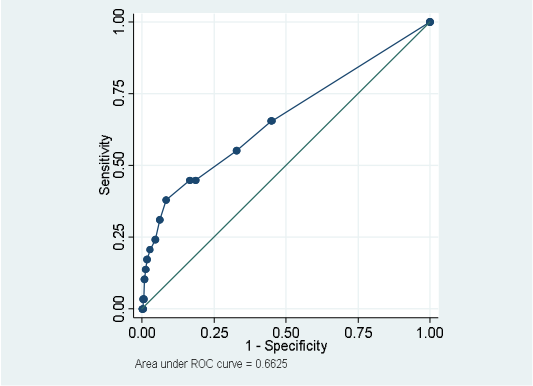
Figure 2.4.4: AUROC plots for mortality risk estimated using PRISM III score at baseline.

a) PRISM III score with 12/17 at baseline (N=224).



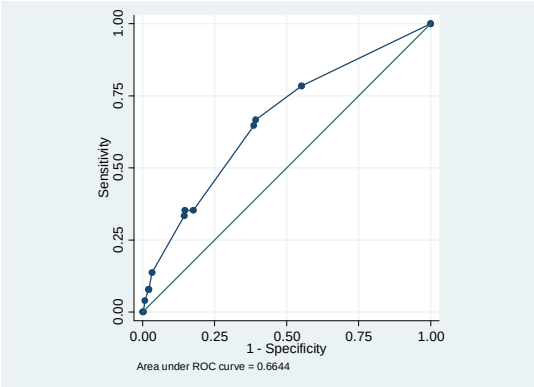
Range from 0-19, not all values taken.

b) PRISM III score with 10/17 variables at baseline (N=620)



Range from 0-19, not all values taken.

c) PRISM III score with 5/17 variables at baseline (n=1015)



Range from 0-15, not all values taken.

In similar settings to the FEAST trial, bedside observation measurements may not be as actively recorded as they were during the trial so a score using 'from admission only' information was calculated and included children that died in the first hour (14 children in the control arm).

This gives values of the AUROC curve of 0.65 (0.43-0.86) for twelve variables, 0.71 (0.61-0.81) for ten variables and 0.66 (0.59-0.72) for five variables. These values are very similar to those that are estimated including observations up to 24 hours.

In considering whether to exclude deaths within the first hour from the dataset, the papers that implemented PRISM or PRISMIII in low-income settings were reviewed but none gave details regarding using scores from pre-terminal children on admission.

Thus, in comparison with the PRISM score which gave AUROC values of 0.73 (0.64-0.82) and 0.71 (0.64-0.78), the PRISM-III score with or without updated observations within the first 24 hours appears inferior.

2.4.2.4 PIM II score applied to FEAST data

The Paediatric Index of Mortality II is presented in Figure 2.4.5 [131]. The authors used a logistic regression model to build the model and 13% of children included in the model were <1 month of age on admission.

Figure 2.4.5: PIMII scoring chart taken from Slater A *et al*, Int Crit Care Med 2003[131].

1. Systolic blood pressure, mmHg (unknown=120)[1]
2. Pupillary reactions to bright light (>3 mm and both fixed=1, other or unknown=0)[2]
3. PaO_2, mmHg (unknown=0)FIO_2 at the time of PaO_2 if oxygen via ETT or headbox (unknown=0)
4. Base excess in arterial or capillary blood, mmol/l (unknown=0)
5. Mechanical ventilation at any time during the first hour in ICU (no=0, yes=1)[3]
6. Elective admission to ICU (no=0, yes=1)[4]
7. Recovery from surgery or a procedure is the main reason for ICU admission (no=0, yes=1)[5]
8. Admitted following cardiac bypass (no=0, yes=1)[6]
9. High risk diagnosis. Record the number in brackets. If in doubt record 0.
   - [0] None
   - [1] Cardiac arrest preceding ICU admission[7]
   - [2] Severe combined immune deficiency
   - [3] Leukaemia or lymphoma after first induction
   - [4] Spontaneous cerebral haemorrhage[8]
   - [5] Cardiomyopathy or myocarditis
   - [6] Hypoplastic left heart syndrome[9]
   - [7] HIV infection
   - [8] Liver failure is the main reason for ICU admission[10]
   - [9] Neuro-degenerative disorder[11]
10. Low risk diagnosis. Record the number in brackets. If in doubt record 0.
    - [0] None
    - [1] Asthma is the main reason for ICU admission
    - [2] Bronchiolitis is the main reason for ICU admission[12]
    - [3] Croup is the main reason for ICU admission
    - [4] Obstructive sleep apnoea is the main reason for ICU admission[13]
    - [5] Diabetic keto-acidosis is the main reason for ICU admission

$$PIM2 = \{0.01395 * (1)\} + \{3.0791 * (2)\} + \{0.2888 * (3)\} + \{0.104 * [absolute\ (4)]\} + \{1.3352 * (5)\} - \{0.9282 * (6)\} - \{1.0244 * (7)\} + \{0.7507 * (8)\} + \{1.6829 * (9)\} - \{1.5770 * (10)\} - 4.8841$$

$$\text{Probability of Death} = \frac{e^{(\text{PIM2})}}{1 + e^{(\text{PIM2})}}$$

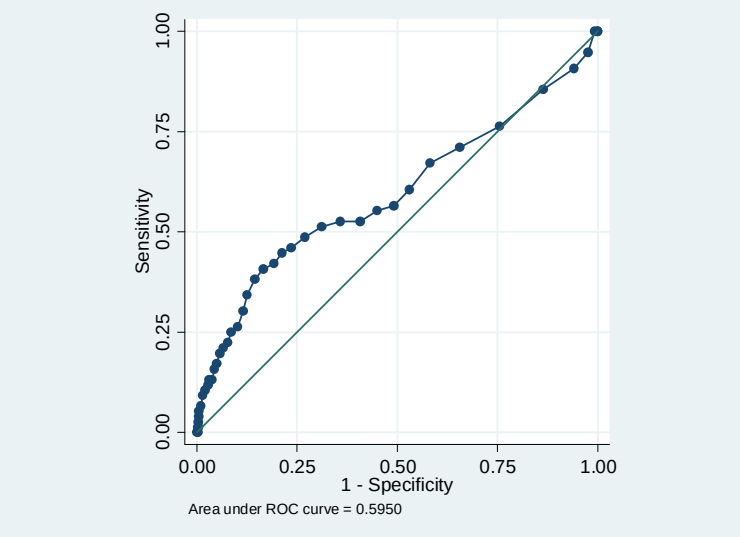
Many of the variables in the PIM II risk score are not applicable to the sub-Saharan African setting and so were not able to be used when applying the risk score to the control arm of the FEAST data. Variables from Figure 2.4.5 and their availability are as follows: (1) Blood pressure information was available on 1034/1044 (99%) children and the value was set at 120 as recommended for those that had missing values; (2) The only variable regarding eye pupils in the FEAST dataset was a binary variable recording whether they were equal or unequal (and a very small proportion were unequal 12/3153 (0.4%)) and as it is not known whether the equal pupils were '>3mm and both fixed' the value for the second variable is fixed at 0 = "other or unknown"; (3) oxygen was not recorded in a way that can be included in the score and so this value is set to 0; (4) base excess was available in the FEAST data for 676/1044 (65%) and the missing values were set to 0 as recommended above for main analysis. However, 0 is not the centre of the base excess distribution, meaning that this assumes these children were abnormal. Therefore these missing values were also set to the median value of -8 (estimated from the distribution at baseline in the control arm of FEAST) in a second analysis. The PIM validation dataset, had a base excess distribution of mean -5.4mmol/l (sd 9.2) and mean -1.6mmol/l (sd 5.2) for those that died and lived respectively, compared with mean -13.2mmol/l (sd 8.9) and mean -8.1mmol/l (sd 6.6) for died and lived respectively in the FEAST data. No mechanical ventilation was available in FEAST settings and it was very unlikely there were any elective admissions to the paediatric wards or that the admission was for recovery from surgery or a procedure, or that they were admitted following a cardiac bypass, so variables (5) to (8) were set to 0. The diagnoses for parts 9 and 10 of the score are very region specific and none were recorded in the FEAST dataset apart from HIV results. So the low risk diagnosis is always set to zero and the high risk diagnosis is 1 if HIV positive or 0 otherwise.

Thus the PIMII score used in the FEAST data reads as follows:

$$PIM2 = \{0.01395 * [absolute(sbp - 120)]\} + \{0.104 * [absolute\,(base\ excess)]\} + \{1.6829 * [1\ if\ HIV\ positive, 0\ otherwise]\} - 4.8841$$

The PIM2 score calculated on the control arm data and rounded to the nearest 1 decimal place between 0 and 4.9 gave a AUROC value of 0.60 (0.52-0.67) which is presented in Figure 2.4.6.

Figure 2.4.6: AUROC plot estimating mortality risk using the PIMII score.



The probability of mortality can be calculated using the PIM2 score and the equation as above. Slater *et al* in the PIM II paper split the data into 10 groups based on centiles of mortality risk, and a similar table for the FEAST control arm data is presented below. The actual number of predicted deaths was calculated by summing the probabilities over the observations within each centile. In total, this led to 31 deaths predicted by the PIMII score. But to enable better comparison with the observed deaths, the total was normalised to 76 deaths to give relative predicted deaths per group, as presented in Table 2.4.5. The Hosmer-Lemeshow goodness of fit test with 10 groups gave p=0.02 showing that the score does not calibrate well to the FEAST data.

Table 2.4.5: Observed and expected deaths by centiles of mortality risk estimated with PIM II

Centile	N	Probability from PIMII	Observed Deaths	Predicted Deaths	Relative Predicted Deaths*
First	93	0.008 - 0.01	8	1	2
Second	105	0.01 - 0.011	8	1	2
Third	113	0.011 - 0.012	4	1	2
Fourth	102	0.012 - 0.014	5	1	2
Fifth	109	0.014 - 0.017	8	2	5
Sixth	104	0.017 - 0.022	3	2	5
Seventh	103	0.022 - 0.027	1	3	7
Eighth	105	0.027 - 0.039	7	3	7
Ninth	105	0.039 - 0.064	13	5	12
Tenth	105	0.065 - 0.494	19	13	32
Total	1044	0.008 - 0.494	76	31	76

*normalised to total observed deaths.

When the base excess values were set to -8, the AUROC for the PIMII score was 0.68 (0.61-0.74), showing a better discriminatory ability.

Large groups were then created to give wider probability bands; keeping the unknown base excess values set to -8, the following estimates were observed using the PIM II score (Table 2.4.6). The Hosmer-Lemeshow test gave p=0.48 indicating that calibration improved when the value used in place of missing base excess was -8 instead of zero for 5 groups.

Table 2.4.6: Observed and expected deaths by quintiles of mortality risk estimated with PIM II

Quintile	N	Probability	Observed Deaths	Predicted Deaths	Relative Predicted Deaths*
First	205	0.008 -0.018	4	3	6
Second	209	0.019 -0.023	12	4	8
Third	205	0.024 -0.027	10	5	10
Fourth	214	0.028 -0.039	18	7	14
Fifth	211	0.040 -0.494	32	19	38
Total	1044	0.005-0.285	76	37	76

*normalised to total observed deaths.

The two tables above describing the calibration of the PIM II score showed that it was better at assigning low probabilities to children who survived when the missing values of base excess were set to -8 instead of zero. Both tables, though, had lower groups with very similar numbers of observed deaths, despite the increasing observed probability of dying, and then a jump to high numbers in the higher groups (i.e groups 8 and 9 in Table 2.4.5 or group 4 in Table 2.4.6).

2.4.2.5 Summary for PRISM, PRISM III, and PIM II

The PRISM score had a somewhat better ability to predict FEAST mortality with both the full data and reduced dataset compared to the PRISM III score. One factor that may explain this is that the PRISM score takes into account low and high heart rate and in the FEAST data increased mortality risk is present for those with bradycardia and tachycardia. PRISM III only gives a score for those with very high heart rates thus not reflecting the increased risk of those with bradycardia in the FEAST dataset. There are also other variables that PRISM III did not use from the original PRISM score that are available in the FEAST dataset (respiratory rate, glucose and bicarbonate) which may have enabled PRISM III to perform better in this population had they been included. Both PRISM and PRISM III may have performed better had more of their variables been recorded and hence included in the analyses.

PRISM discriminated most well when 8/14 variables were used but was better at predicting the appropriate number of deaths with 5/14 variables. The PRISM III score discriminated best with 10/17 variables giving an AUROC of 0.69 with the score calculated on both more or less variables giving lower AUROC values.

The PIMII score's poor calibration and low discriminatory ability may be due to the limited number of variables available in the FEAST dataset for the score. This meant that values of 0 were used for many factors, as recommended, and so it was assumed the child did not satisfy that criteria (i.e low risk diagnoses, pupil reactions, mechanical ventilation) or their value was in the normal clinical range of the distribution of continuous variable (i.e PaO_2), as the authors had centred the continuous variables in the model. This was the case for 6 of the measures that make up the score and so these measures were the same value for all children thus affecting the score's ability to discriminate between survivors and those who died. The calibration and discriminatory ability of PIMII improved in the FEAST population when a value from the centre of the distribution of base excess in that population was used. This shows

there is some impact of centring variables in the model around the values in the development or validation dataset for PIMII and demonstrates one aspect of the difference between the populations.

There has not been much discussion of calculating the PRISM or PIM scores with missing values in the evaluations of these scores that were found during the literature review. Even in countries such as Senegal, the full PRISM score was evaluated on children presenting to a paediatric intensive care unit [136]. In Mexico the study authors simply did a complete-case analysis and excluded those who did not complete all the clinical or laboratory variables [107]. The PIM scores gives guidance for some of the variables when they are unknown, but the PRISM and PRISM-III scores do not. For evaluating the PRISM and PRISM-III scores in FEAST, only children with complete data have been used but others in the dataset may only have one or two measurements missing and it is a disadvantage that there is no mechanism of including them in the score without imputing data.

The use of the most severe of the measures recorded in the first 24 hours in the PRISM and PRISM-III scores has been noted as a disadvantage of the score by the PIM authors as it can only be applied retrospectively after the first 24 hours in practice [132]. But the application of these scores to the FEAST setting also raises the difficulty that the most severe of several measurements has more variability than a single measurement being an extreme, and some of the variables in the score are from laboratory tests or the i-STAT tests which may be more prone to error.

2.4.2.6 Berkley et al's score applied to FEAST data

Berkley *et al* in their paper published in the BMJ in 2003 devised and validated a prognostic score for children admitted to hospital in Kenya [47]. The study has been described in detail in the literature review above (Section 2.3.3).

The deaths were split into immediate, early and late where immediate was within 4 hours of admission, early deaths were deaths from 4 to 48 hours from admission and late deaths were ones that took place >48 hours after admission but still whilst in hospital. Models were built to predict each type of death separately. Table 2.4.7 below summarises the points given to each indictor for immediate, early and late deaths.

Table 2.4.7: Berkley's scores for immediate, early and late deaths summarised

	Immediate		Early		Late	
Indicator	Present	Absent	Present	Absent	Present	Absent
Severe anaemia (haemoglobin <5g/dl)	1	0	-	-	-	-
Jaundice	1	0	1	0	-	-
Indrawing	1	-1	1	0	-	-
Deep breathing	1	0	-	-	-	-
Prostrated with seizures	1	0	2	0	-1	0
Prostrated without seizures	3	0	2	0	-	-
Impaired consciousness with seizures	2	0	2	0	1	0
Impaired consciousness without seizures	3	0	3	0	1	0
Axillary temperature <36°C	1	0	-	-	1	0
Axillary temperature >39°C	-1	0	-	-	-1	0
Wasting	-	-	1	0	1	0
Kwashiorkor	-	-	1	0	1	0
History of condition > 7 days	-	-	-	-	1	0
Constant	2		0		2	
Maximum possible	10		7		6	
Minimum possible	0		0		0	

The table above shows that some indicators are common across all scores but others, such as severe anaemia and jaundice, only appear in the score for immediate deaths. Wasting and kwashiorkor are only important in the early and late death scores, and history greater than 7 days is only present in the late death score. Also, the table indicates that different numbers of points are given to the presence or absence of the same indicators within the different scores. For example, a score of -1 is given for a very high temperature indicating that a child is at lower risk of immediate and late mortality compared to those with a normal temperature, probably due to those with fever being treated effectively for malaria, however a score of 1 is given for hypothermia. There is variation between scores where points awarded to the presence or absence of an indicator are different (thus, being prostrated with seizures is an apparent risk factor for immediate and early deaths but then has a protective effect for those that have survived for 48 hours, presumably reflecting the fact that these patients have mostly already died in the intermediate/early periods and those prostrated with seizures who survive to the

late period are intrinsically a less sick group). It may be argued that although it was useful to have these different definitions of death to separate risk of immediate, early and late deaths, it is difficult to make sense of the different scores associated with the same factor clinically. Information on the number of patients that died with each score was also presented in the paper and is replicated below in Table 2.4.8.

Table 2.4.8: Distribution of prognostic scores and outcome in the validation dataset of 4802 children admitted to hospital in Kenya, as presented in Berkley *et al.*

Score	Immediate		Early		Late	
	Number admitted	**Number (%) died**	**Number admitted**	**Number (%) died**	**Number admitted**	**Number (%) died**
0	698	0	2779	11 (0.4)	15	0
1	2196	1 (0.1)	1101	16 (1.5)	1081	2 (0.2)
2	657	1 (0.2)	668	23 (3.4)	2310	18 (0.8)
3	612	5 (0.8)	183	23 (13)	869	26 (3.0)
4	217	4 (1.8)	55	10 (18)	347	38 (11)
5	72	3 (4.2)	15	5 (33)	170	22 (13)
6	34	3 (8.8)	1	0	10	2 (20)
7	31	4 (13)	0	0	0	0
8	12	3 (25)	—	—	—	—
9	2	1 (50)	—	—	—	—
10	1	1 (100)	—	—	—	—

These scores were used by the authors to calculate the AUROC in a validation dataset created from a subsequent time period of the hospital admissions data (n=4802, 222 (4%) deaths) giving an AUROC of 0.93 (95% CI 0.92-0.94) for immediate, 0.82 (95% CI 0.80-0.83) for early, and 0.82 (95% CI 0.81-0.84) for late deaths.

I assessed the performance of these scores using the children enrolled into the control arm of the FEAST clinical trial (n=1044). The values of wasting and kwashiorkor were set to 0 as children with these conditions were excluded from the trial. History of >7 days for the child's condition was not explicitly recorded but there is information on whether the child had a history of fever >14 days and so this has been used in place of this variable. Some children

have missing information on the indicators in the score, in particular haemoglobin measures at baseline, and so they are presented in Table 2.4.9 below under 'missing score'.

Table 2.4.9: Validating Berkley's scores using information from the control arm of FEAST

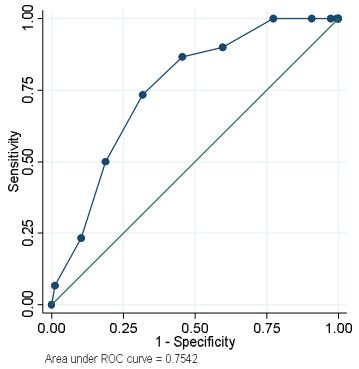
	Immediate		Early		Late	
Score	Number	Deaths	Number	Deaths	Number	Deaths
0	0	0	28	0 (0%)	21	0 (0%)
1	5	0 (0%)	196	4 (2%)	253	4 (2%)
2	23	0 (0%)	273	1 (0.3%)	626	8 (1%)
3	66	0 (0%)	310	21 (7%)	111	2 (2%)
4	139	0 (0%)	208	13 (6%)	8	1 (11%)
5	227	4 (2%)	20	4 (20%)	-	-
6	130	2 (2%)	-	-	-	-
7	118	5 (4%)	-	-	-	-
8	154	9 (6%)	-	-	-	-
9	127	8 (6%)	-	-	-	-
10	18	2 (11%)	-	-	-	-
missing score	37	3 (8%)	9	0 (0%)	10	0 (0%)
Total	1044	33 (3%)	1044	43 (4%)	1044	15 (1%)

The tables above show that according to the scores the FEAST trial population control group was in general more critically ill than the validation population in Berkley *et al* (admissions to a general paediatric ward at a district hospital in Kenya between June 2000 and July 2001) as they had generally higher scores. For the immediate deaths in the FEAST control arm, the median score was 6 whereas it was 2 in Berkley's validation dataset. The median score was 3 for early deaths, compared to 0 in Berkley's validation dataset, but then they were similar for the late deaths (median score 2 for both). There was a greater difference between the datasets in the distribution of scores for the late deaths, as Berkley's had 0-6 whereas FEAST had 0-4, with a higher proportion of zeros. The difference in the distributions may have been due to not having the full 'history' variable available in FEAST and using a proxy for this, and the similarity in median score for late deaths may be due to the fact that the deaths are conditional on surviving to 48 hours and so any difference between the FEAST trial population and the population in the paper is minimised as all have had 48 hours' worth of treatment. Non-parametric analyses of the area under the ROC curve using the children in the control arm of FEAST is presented below for each of the scores, and give AUROC of 0.75 (95% CI 0.68-0.83)

for immediate deaths, 0.70 (95% CI 0.63-0.77) for early deaths and 0.55 (95% CI 0.40-0.69) for late deaths. The Hosmer-Lemeshow goodness of fit test for comparing observed and predicted mortality in groups defined by each score value (1 to 10 for immediate deaths, 0 to 5 for early, 0 to 4 for late) showed good calibration for immediate (p=0.64) and late deaths (p=0.35) but poor calibration for early deaths (p=0.02).

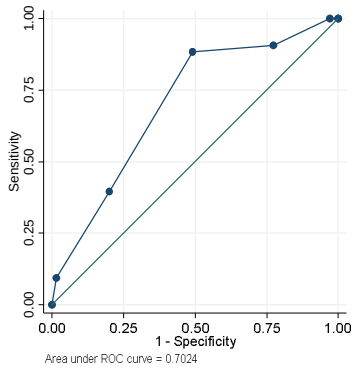
Figure 2.4.7: AUROC analyses for Berkley's mortality risk score evaluated on the FEAST data.

a) AUROC analysis for immediate deaths.



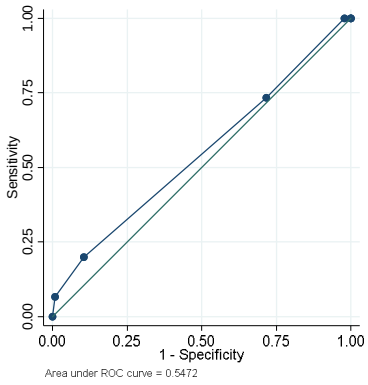
NB: Score values are from 1-10.

b) AUROC analysis for early deaths.



NB: Score values are from 0-5

c) AUROC analysis for late deaths.



NB: Score values are from 0-4

The validation of the scores shows that the immediate death score still performed reasonably well in the population recruited to the FEAST trial, but that the early and late death scores did not transfer across to this population, with the late death score no better than chance in predicting mortality, and the early score had poor calibration.

2.4.2.7 Von Seidlein's (AQUAMAT) score applied to FEAST data

Von Seidlein *et al* examined prognostic factors in a cohort of 5426 children enrolled into an anti-malarial trial in 9 African countries [103]. From multivariable analyses including 21 variables they found 5 independently significant variables and used these to create a score. The presence of each variable scored 1 point and their score went from 0 to 5. The variables were base excess <-8mmol/L, Blood Urea Nitrate (BUN) ≥20mg/dl, a combined coma score <3, chronic disease and convulsions. The mortality rate for each value of the score in their cohort is presented in Figure 2.4.8 below.

Figure 2.4.8: Table taken from von Seidlein *et al* describing mortality by level of their score.

Table 5. Score Based on 5 Independently Significant Variables: Base Excess (<−8 mmol/L), Blood Urea Nitrogen (≥ 20 mg/dL), Combined Coma Score <3, Chronic Disease, and Convulsions.[a]

Score	Survived	Died (%)	Total
0	1056	9 (1)	1065
1	1339	75 (5)	1414
2	923	118 (11)	1041
3	311	131 (30)	442
4	58	66 (53)	124
5	1	2 (67)	3
Total	3688	402 (10)	4089

[a] Presence of each variable provides 1 point to the score.

The chronic diseases that were included in the score were as lymphadenopathy, malnutrition, candidiasis, severe visible wasting and desquamation combined as an indicator. Most of these chronic diseases were not recorded in the FEAST data and if a child had visible severe wasting or malnutrition they were excluded from the trial. There were some children with a mid-upper arm circumference of <11.5 cm which is a definition of malnutrition and so that has been used in place of the chronic disease variable. Convulsions was defined as a history of convulsions of 30 minutes or longer or ≥2 convulsions in previous 24 hours reported by the caregiver. Information on base excess and blood urea nitrogen is taken from the i-STAT cartridge and so is only available on approximately 66% of the children in the FEAST dataset.

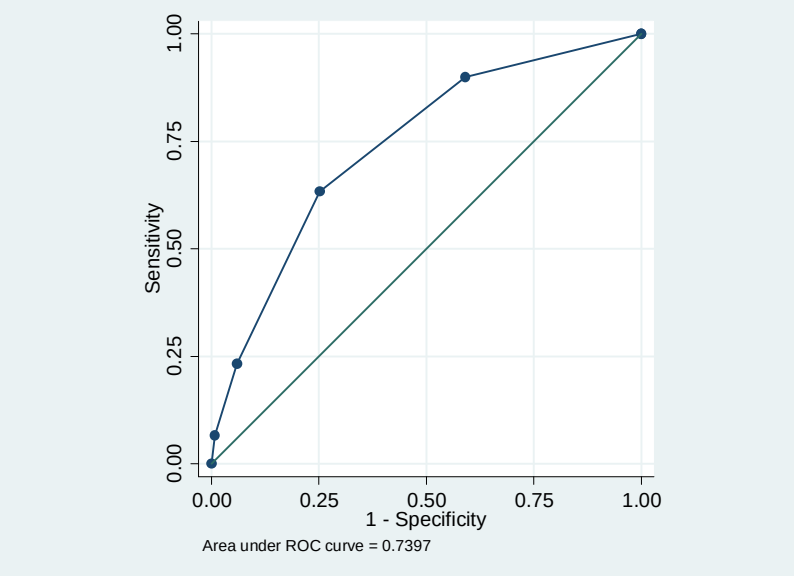
Table 2.4.10: Validating the AQUAMAT score using information from the control arm of FEAST.

Score	Survived	Died (%)	Total
0	253	3 (1%)	256
1	209	8 (4%)	217
2	119	12 (9%)	131
3	32	5 (13%)	37
4	5	2 (29%)	7
Total	618	30 (5%)	648

The information in the table above indicates that the mortality in each score follows the mortality that von Seidlein *et al* found in the AQUAMAT trial cohort. There were no children with all 5 risk factors to give a score of 5 in our cohort and there was only <0.1% in the AQUAMAT cohort, so given the reduced numbers in FEAST it might be expected not to find any children fitting that criteria in our data. Also there was a lower mortality in the group that had information available to calculate the score (5%) compared to the mortality in the AQUAMAT cohort (10%) and also compared to the mortality in the control arm of FEAST estimated from complete data (7%). The mortality for those that had missing information was 46/396 (12%).

Non-parametric analyses of the AUROC curve using the children in the control arm of FEAST is presented below giving a value of 0.74 (0.65-0.83). The Hosmer-Lemeshow goodness of fit test for comparing observed and predicted mortality in groups defined by each score value (0 to 4) showed good calibration for the score (p=0.84).

Figure 2.4.9: AUROC analyses for the AQUAMAT risk score evaluated on the FEAST data.



The children enrolled into the AQUAMAT trial all had malaria and so the score was also evaluated on the 360 children in the FEAST control arm data that had all the information needed for the score and had malaria at admission. The AUROC was 0.80 (0.68-0.93) which is an improvement but is calculated on small numbers.

2.4.2.8 Summary of Berkley and AQUAMAT scores

Both the AQUAMAT and Berkley's scores were developed in similar settings to the FEAST trial although Berkley's score was calculated on children coming into the general admissions wards and the AQUAMAT score was developed on children with malaria as part of a clinical trial. Berkley's score has been included in a review of clinical prediction rules in practice [77] and was allocated a level of evidence which indicated a narrow validation of the prediction rule but that further validation was needed in varied settings. The levels of evidence have been defined by the Evidence Based Medicine Working Group and described under a set of articles published in JAMA called 'User's Guide to the Medical Literature' [137]). One validation study of Berkley's scores in febrile children presenting to hospital in Uganda was found from a citation search (in November 2017) [138]. The authors combined Berkley's three scores into one (including jaundice, subcostal indrawing, prostration with/without seizures, altered consciousness with/without seizures and wasting) and found good discrimination and calibration with an AUROC of 0.90 (0.88-0.91) and Hosmer-Lemeshow p-value of 0.22. There is no further work validating the AQUAMAT score in other settings as yet.

2.4.3 Overall summary of published scores validated in FEAST control arm data

The summary table below presents each score, the AUROC calculated on FEAST control arm data and which variables were used to create the score. The table shows that there were several scores that had fair discriminatory ability with the FEAST control arm data. PRISM with 8 variables had an AUROC of 0.73, Berkley's score for immediate deaths had an AUROC of 0.74 and the AQUAMAT score of 0.74. Impaired consciousness is a very strong predictor of mortality and appears in all three scores. Signs of acidosis appear in Berkley's and the AQUAMAT scores as: deep breathing; low base excess; and seizures/convulsions. PRISM uses more laboratory values and was developed in high-income settings whereas the AQUAMAT and Berkley's scores were developed in very similar resource-limited settings to the FEAST

trial. PIM II and Berkley's score for early or late deaths do not discriminate well in the data. PIM II uses several variables that were not available in the FEAST dataset and so the score was just reduced to base excess and systolic blood pressure which may explain its poor performance. The kwashiorkor and wasting variables in Berkley's late death score are not available in FEAST as this was an exclusion criteria and the length of the illness prior to admission was also not recorded, which may affect the performance of this score. Also, the late deaths used in developing the score were inpatient deaths and could be a longer time after admission compared to the FEAST data, where follow up for deaths was censored at 28 days and included deaths outside of the hospital.

Table 2.4.11: Summary table of performance of all scores assessed with the FEAST control arm data.

Score	Restrictions	Variables included	AUROC (95% CI)
PRISM - 8/14 variables	Uses all information from first 24 hours- excluding measurements taken within one hour of death, (glucose, potassium and bicarbonate only at baseline). Only ranges for infant and child were applicable to FEAST data.	bicarbonate, glucose, heart rate, mental status, potassium, pupillary reactions, respiratory rate, SBP.	0.73 (0.64-0.82)
PRISM - 5/14 variables	Uses all information from first 24 hours- excluding measurements taken within one hour of death. Only ranges for infant and child were applicable to FEAST data.	heart rate, mental status, pupillary reactions, respiratory rate, SBP.	0.71 (0.64-0.78)
PRISM III - 12/17 variables	Uses information from first 24 hours - excluding measurements taken within one hour of death.	axillary temperature, acidosis, glucose, heart rate, mental status, pco2, pH, platelets, potassium, pupillary reflexes, SBP, white blood cells	0.65 (0.59-0.72)
PRISM III - 10/17 variables	Uses information from first 24 hours - excluding measurements taken within one hour of death.	axillary temperature, acidosis, glucose, heart rate, mental status, pco2, pH, potassium, pupillary reflexes, SBP	0.65 (0.46-0.84)
PRISM III - 5/17 variables	Uses information from first 24 hours - excluding measurements taken within one hour of death.	axillary temperature, heart rate, mental status, pupillary reflexes, SBP	0.66 (0.59-0.74)
PRISM III -	Only using baseline information,	axillary temperature,	0.66 (0.59-0.72)

Score	Restrictions	Variables included	AUROC (95% CI)
baseline only (5 variables)	and including measurements on pre-terminal children (death within 2 hours of admission)	heart rate, mental status, pupillary reflexes, SBP	
PIM II	Only using baseline information, and including measurements on pre-terminal children (death within 2 hours of admission)	base excess, SBP	0.60 (0.52-0.67)
Berkley - immediate deaths	All baseline information included - only for deaths <4 hours from admission	axillary temperature, deep breathing, impaired consciousness with or without seizures, indrawing, jaundice, prostrate with or without seizures, severe anaemia	0.74 (0.67-0.81)
Berkley - early deaths	All baseline information included - only for deaths 4-48 hours after admission	impaired consciousness with or without seizures, indrawing, jaundice, kwashiorkor, prostrate with or without seizures, wasting	0.49 (0.62-0.77)
Berkley - late deaths	All baseline information included - only for deaths >48 hours from admission	axillary temperature, history of illness, impaired consciousness with or without seizures, kwashiorkor, prostrated with seizures, wasting	0.51 (0.39-0.65)
AQUAMAT	Children with severe wasting or malnutrition were excluded from trial but MUAC <11.5cm was used in place of chronic disease variable. Base excess and BUN only measured on 60% of children.	base excess, BUN, chronic disease, combined coma score, convulsions,	0.74 (0.65-0.83)

2.5 Building a clinical prognostic model

2.5.1 Methods

An approach outlined in Royston *et al*'s paper on Prognostic research [139] was followed to develop a clinical prognostic model for children admitted to an intensive care ward in a low income setting. They outline six areas in which key decisions are made during a prognostic model building process and these areas have decisions as outlined in the diagram in section 1.1 above. The six key areas are:

- Selecting clinically relevant candidate predictors
- Evaluating the quality of the data and the amount of missing data
- Data handling decisions
- The strategy used to select important variables in the final model
- Modelling continuous variables
- Selecting measures of model performance

The methods for each area are outlined in the sections below.

2.5.1.1 Selecting clinically relevant candidate predictors

The first decision is to select clinically relevant candidate predictors to be considered for inclusion into the model.

The maximum number of coefficients for a prognostic model with which it is advised to start the model building process is approximately m/10, where m is the number of events in the sample on which the model is initially fitted (i.e the training sample) [72, 140]. This gives the maximum number of candidate variables that can be considered. Using less than this maximum number will help ensure that there is (at least theoretically) an event for every way of combining the variables in the model selection process. This issue is also explored in Peduzzi *et al* [141] as an 'events per variable' ratio and they find using simulation studies that 10 events per variable is the suitable boundary to avoid problems with the modelling process.

Predictors already found in the literature are very suitable as candidate predictors, as well as variables that are reliably measured and variables for which there is a reasonable number of

observations falling in each category of the variable. If candidate predictors are not reliably measured they may have large amounts of measurement error which would dilute their prognostic information [139]. Harrell *et al* [72] indicate that multiple comparison problems can occur if many univariable models are fitted using the outcome, and so recommends not using the outcome in any decision making regarding candidate predictors.

Candidate predictors can also be identified by thinking about elements of a good model - generalisability, transportability and practical usefulness [142], and focusing on what settings the model will be useful in and what variables are able to be measured there.

2.5.1.2 Evaluating the quality of data and making decisions about missing data

For many clinical bedside variables there is approximately 1% of missing data. This is a little higher in some variables that measure growth such as height/length (9%), MUAC (6%) and head circumference (7%), but importantly in these variables it is also higher in those that died (for example 164/315 (52%) in height, 126/315 (40%) in MUAC and 132/315 (42%) for head circumference). Thus, many events may be excluded from models if simple univariable and multivariable analysis was done. As part of a sensitivity analysis, univariable models for all the potential risk factors will be fitted and these will have greater numbers of observations compared to the multivariable models where some children will have been excluded due to missing data in other variables. If the missingness is non-ignorable then the estimates from multivariable models will be biased and, due to the fewer observations in all variables, the variance is increased [72]. If missing is at random the variance will increase due to fewer observations, but the methods of including observations can ensure that bias does not occur by using variables that predict the missingness. These assumptions are tested by considering if missingness can be explained by other variables in the database (missing at random) or if it cannot be explained (missing not at random or non-ignorable missingness).

One way to include observations with missing data is through multiple imputation. This method can be used for all of the clinical bedside variables as well as laboratory values which have a greater percentage of missing data (from 3% to 45%). The multiple imputation model will take into account the planned analyses and include the outcome variable (mortality by 48 hours) as binary, and as the analysis model will be Cox regression then time from randomisation to censoring or death will also be included (this can be included as log(time) to avoid influence of high values, but as the censoring for the analysis is 48 hours then time will

not be transformed). Multiple imputation has been found to be a robust way to include more observations in models but is reliant on the appropriate model being used for imputation and the correct assumptions about the data being made.

2.5.1.3 Data handling decisions

The way that the variables are used or handled in the model also needs to be considered. Variables that are continuous should not be dichotomised or categorised in the model building process [139] as otherwise key predictive information may be lost. Fractional polynomials (implemented in Stata through the mfp command) are a way to find the appropriate function, as simply fitting a linear function for continuous variables makes strong assumptions that may not hold. The selection level (or significance level) for choosing a non-linear function over a linear function (and choosing between fractional polynomial models of different degrees) is usually set at an alpha of 0.05 in models compared through the mfp command. Categories that are already present for variables may need to be reduced or merged to make them more generalisable or to ensure a reasonable number of events within each category.

2.5.1.4 Selecting the important variables in the final model – which strategy?

Once the number of candidate predictors has been determined then the strategy for identifying which of them are important is chosen. The model that would avoid over-fitting and selection bias is the full model with all the predictors included. This, though, is usually impractical, because many covariates will not have important effects and their inclusion will increase variability in other effect estimators and so a more parsimonious model is desired [139]. The backwards elimination approach is preferable to forward selection and uses a selected significance level or information criterion (AIC, BIC) to include or exclude predictors having started with the full model [143]. The selected significance level was 0.05 in the following analyses in the FEAST data as there was a desire to select as few variables as possible for simplicity. Other options were 0.1 or 0.157 which for a single predictor equates to the Akaike Information Criterion. The larger the significance level for the backwards elimination process the greater the chance of unimportant variables being included, but this may ensure that all confounding biases that may be present (even if not significant at the threshold chosen) are appropriately adjusted for. The exit significance level can be chosen in line with

the aims of the study, with low values of the significance level (i.e 0.01) useful for identifying only strongly predictive variables, and high values, for example 0.20, for selecting confounders in an epidemiological study [142].

Using a selection process based on significance level is known to produce selection bias and optimism, but where there are selected predictors with very small p-values (i.e <0.001) then they are much less prone to these issues. Weak predictors with p-values near the significance level (of 0.05) are more affected. Bootstrap resampling can help identify which are strong and weak predictors amongst the variables included in the model by looking at the number of times the variable is included in the model out of all the replications [144].

I concluded that the model for these analyses should be built with backwards elimination with a significance level of 0.05 as I wanted a small number of important predictors for individual level prognosis. The dataset is also large enough to identify important predictors at this level, whereas in smaller datasets a larger significance level may be needed to ensure all important predictors are included in the model [145].

2.5.1.5 Assessing a model's performance

The performance of a model can be assessed both internally (on the same data) or externally on different data. Two key aspects of this are calibration and discrimination. The Hosmer-Lemeshow test compares the observed number of events with that predicted from the model within categories on the basis of predicted risk and so is used to assess calibration [73]. But this test can be oversensitive in very large samples (and thus produces significant p-values regardless of calibration) and has limited power to detect poor calibration [59]. As observed and predicted probabilities should be very close for the data on which the model was built, it is most often used for external validation of the model. A calibration plot should also be presented comparing these probabilities to assess performance. It plots the observed proportions of events against predicted probabilities for groups based on risk. The predicted probabilities for calibration come from fitting a logistic regression model with the score as the explanatory variable and calculating the probability of death directly using the linear predictor from the model. The area under the receiver operating curve (AUROC) (or equivalent c-statistic) is used to examine discrimination and plots sensitivity against 1-specificity. In general, if the AUROC is equal to 0.9 or more this is considered excellent discriminatory power, 0.8 –

0.89 is considered good and 0.70-0.79 is considered fair [146], and potentially if the discriminatory power is 0.75 or more it would be considered clinically useful [126].

2.5.1.6 Moving from a prognostic regression model to a bedside score

Prognostic regression model coefficients are often published and used to validate the scores on other datasets. To use the coefficients in everyday clinical practice, computers or online calculators are needed so that information can be inputted, the linear predictor calculated for that specific information and the probability of an event then estimated. For example, the Framingham study risk functions have been used for predicting cardiovascular disease events by different time points (2 years, 10 years, 30 years) for many years and have been turned into an interactive calculator (www.framinghamheartstudy.org) [147]. Risk calculators have also been developed in other areas such as the Antiretroviral Therapy Cohort Collaboration (ART-CC) calculator that estimates progression rates to AIDS or death after starting Highly Active Antiretroviral Therapy (HAART) (www.art-cohort-collaboration.org) [148].

Thus, there are two stages to building a score, the first is to develop the most appropriate prognostic model and the second is to use that to inform a score. The methods to turn a linear predictor into a score have not been widely discussed but each reviewed paper from the literature search was examined to identify the method used.

The majority of the papers built a multivariable logistic regression model. Two used backwards elimination, two used forward selection, five did not specify how the model was developed (although one was indicated to be stepwise), one used positive and negative likelihood ratios and one used improvement in the C-statistic. Any integerised scores, if applicable, were then developed in a variety of ways including: CART modelling; assigning points that were approximately equivalent to log of the adjusted likelihood ratio; dividing all regression coefficients by the lowest one and rounding to the nearest integer; and considering the sensitivity and specificity of the presence or absence of the independent predictors (when all were binary).

I decided that to reduce the loss of information that will happen when moving from a linear predictor to an integer bedside score, dividing each coefficient by the lowest coefficient within the linear predictor and rounding to the nearest integer would be most appropriate. This method has been used to develop a bedside score for meningococcal disease [79]. Categories were created for continuous variables based on the previously modelled fractional polynomial

functions to enable the bedside score to be practical and for integers to be linked with clinical measurements.

2.5.1.7 Assessing value of additional predictors

In comparing different models for discriminative ability, the c-index may not be the most useful tool, because the rank method by which the c-index is calculated does not differentiate the distance between prediction and outcome in the pairs it is evaluating [72]. A recently developed method is reclassification based on stratification of individuals into clinical categories on the basis of risk [149]. The new variables are then assessed on their ability to more accurately reclassify individuals into higher or lower risk strata (which are pre-defined). The statistics that are used for this new method are the net reclassification improvement (NRI) statistic, the integrated discrimination improvement (IDI) and discrimination slope (also known as Yates' slope). The NRI is the net increase minus decrease in risk categories out of the number of patients with the outcome (cases), minus the net increase minus decrease out of the number of patients without the outcome (controls). This value is then compared against the null hypothesis of no change (0%) to give a p-value [149]. The discrimination slope is defined as the difference between mean predicted probabilities of the outcome for those with the outcome and the mean for those without the outcome, and the IDI is an alternative method to the NRI and is calculated as the difference in discrimination slopes between models [73]. As with the NRI an asymptotic test has been calculated for the null hypothesis of the IDI=0 to give a p-value [149]. These measures (NRI, IDI) have been shown to be useful to better compare the addition of new variable(s) or biomarker(s) to existing models.

The disadvantages of the NRI and IDI is that they are less easily interpretable in disease areas where there are not useful clinical categories of risk that can be pre-specified (especially for the NRI) [149] and the NRI and IDI do not use information regarding the underlying probability of the event [150]. The IDI does not, though, require predefined risk thresholds as described in Moons K *et al*, Heart 2012 [76] "The IDI is looking at the estimated improvement in the average sensitivity of the basic model with addition of the new predictor minus the estimated decrease in the mean specificity, summarised over all possible risk thresholds."

Candidate predictors not consistently available across low-income settings could add value to the prognostic model when they are able to be measured and so models were built including these variables to examine their prognostic influence over and above the easily measured

bedside clinical variables already available. The NRI measure and IDI were calculated using logistic regression to see if the new model was better able to classify children into the appropriate risk strata.

2.5.2 Results

2.5.2.1 Selecting appropriate candidate predictors

Following Harrell's guidance the FEAST dataset has a total of 315 deaths within 48 hours (297 deaths in A and 18 in B), or 363 in total within 28 days (345 deaths in A and 18 in B) so a reasonable number of variables to include in the model would be 25-30 variables.

There are 48 clinical variables available in the FEAST dataset (clinical variables are defined as not relying on the result of a test and being measured quickly at the bedside). If a variable is highly correlated to another variable or is not informative about the clinical state on admission (such as height) then it can be removed from further consideration. It is also important to consider the availability of the variable in the validation dataset and in these settings in general. A list of all 48 clinical variables is presented in Table 2.5.1, with discussions regarding variables that needed further consideration presented beneath the table.

Table 2.5.1: Description of candidate predictors for the regression model.

Variable	Candidate for model	Reason if not using
Age (months)	**X**	
Axillary Temperature	**X**	
Blantyre Coma Scale		See reason below
Bulging fontenelle		See reason below
Capillary refill time	**X**	
Capillary refill time >2s		Dichotomised from continuous variable
Cold hands and/or feet		See reason below
Conscious level	**X**	
Cough	**X**	
Crackles	**X**	
Decreased skin turgor	**X**	
Deep breathing	**X**	
Diarrhoea		See reason below
Difficulty in breathing	**X**	
Epilepsy		Potentially applicable to very few children and can be a subjective diagnosis.
Ethnicity		Too study specific and many categories
Fits greater than 30 minutes in this illness	**X**	
Fits in this illness	**X**	
Fitting currently		See reason below
Fitting/convulsions at admission	**X**	
Haemoglobinurea		Not a widely known clinical measure and very rare at some sites.
Head circumference		Not informative regarding clinical state
Heart rate	**X**	
Height		Not informative regarding clinical state
History of fever	**X**	
History of fever of more than 14 days		Highly correlated with history of fever
Indrawing	**X**	
Jaundice	**X**	
Liver size >2cm below costal margin	**X**	

Variable	Candidate for model	Reason if not using
MUAC	**X**	
Neck stiffness		See reason below.
Neck stiffness or bulging fontenelle (sign of meningitis)	**X**	
Pupil symmetry		12/3153 had this clinical feature so may not be informative and not available in validation dataset.
Received a blood transfusion in this illness		Describing prior treatment rather than clinical state.
Respiratory distress	**X**	
Respiratory rate	**X**	
Severe pallor	**X**	
Severe tachycardia		Dichotomised from heart rate
Sex	**X**	
Site		Study specific
Sunken eyes	**X**	
Systolic blood pressure		To be evaluated following model building as not available in validation dataset.
Temperature gradient	**X**	
Vomiting	**X**	
Was child able to sit unsupported before current illness		Age – dependent variable
Was child able to walk before current illness		Age – dependent variable
Weak pulse	**X**	
Weight	**X**	

Blantyre Coma Score (BCS) or Conscious level

Conscious level was measured in triage before randomisation in FEAST and was one of three options: alert, prostrate (inability to sit unsupported or to breastfeed if <9 months) or coma (inability to localise a painful stimulus). Blantyre Coma Score was assessed potentially up to an hour after randomisation but it may be a more comparable score to other sites/settings. This was investigated using the reviewed papers and scores. This is summarised in Table 2.5.2 below.

Table 2.5.2: Definitions of impaired consciousness in reviewed papers.

First Author	**Prostration:**	**Coma:**
Maitland K (FEAST) [33]	*Inability to sit unsupported or to breastfeed if <9 months*	*Inability to localise a painful stimulus.*
Evans J [62]	age dependent inability of the child to suck, sit, stand or walk	BCS≤2
Planche T [109]	BCS≤4	BCS≤2
Allen SJ [106]		BCS≤2
Berkley J [47]	Inability to sit unassisted (>1 year); inability to drink or breast feed (<1 year).	(Impaired consciousness) Inability to localise a painful stimulus (>8 months); non-directed eye movements (<8 months).
Marsh K [108]	Patient is not in a coma but cannot sit	Patient is unable to localise a painful stimulus (>8 months) or no motor response after painful stimuli (<8 months)
Roine I [110]	Glasgow Coma Score in four levels (≤6, 7-9, 10-12, 13-15); 3 is deep coma and 15 is alert and awake.	
Newton C [104]	Inability to sit	BCS≤2 (although not defined precisely as use all of BCS scale).
Von Seidlein [103]	defined as the inability to sit unsupported (for children over 6 months of age) or the inability to drink or breast-feed in younger children.	BCS≤2 for children <2 years and GCS≤9 for older children
Pollack M [117]	Glasgow Coma Score <8 (stupor or coma).	

From the summary table above it seems that prostration and coma are very similarly defined in studies and that the admission definition could be used as a more practical assessment than calculating the Blantyre coma score. Kilifi admissions has BCS recorded and also a text field called conscious level which has 'unconscious, normal, prostrate, lethargic and agitated' for 99.93% in dataset (26 records have other text recorded). There is also a yes/no field for 'prostrate or unconscious'. There is a WHO definition for lethargic but it is not as serious as prostration, and conscious level and prostration can be very subjective [48]. When BCS and the admission definition of conscious level used in FEAST were compared using the FEAST data, the BCS indicated more children with coma and more classified as alert with fewer in the prostration category. For example, prostration according to the admission assessment was found in 61% and alert in 24%, whereas using BCS this was 38% for prostration and 41% alert. Coma was 15% in admission definition compared to 21% with BCS. Taking all these factors into consideration, I decided to use the admission definition of consciousness in FEAST in the model building process as it was recorded immediately on admission (whereas BCS was assessed after randomisation and potentially after the beginning of treatment). In the validation dataset, BCS was recoded as in the table above to reflect alert, prostrate and coma, as the Kilifi assessments (BCS and text field) were made at the same time, and the text field includes categories not used in the FEAST admission definition,

Bulging fontenelle or neck stiffness or both

As bulging fontenelle is applicable only to children ≤18 months then it would be useful to combine this with neck stiffness as they are describing a similar feature of meningitis. After discussions with a senior clinician it was felt they could be combined to be one variable.

'Cold hands or feet' or temperature gradient

Positive temperature gradient was defined by a notable temperature change from cold (dorsum of foot) to warm (knee) when running the back of the hand from the toe to the knee. The Kilifi admissions dataset does not have 'cold hands or feet' so I decided to use temperature gradient in preference to 'cold hands or feet'.

Diarrhoea

Diarrhoea was found in one other study to be a prognostic factor in mortality in children with leishmaniasis, but children with more than 3 watery stools in last 24 hours were excluded from FEAST, so I decided that it was not appropriate to use in building a model (and only 0.5% children in FEAST had diarrhoea).

Difficulty breathing or respiratory distress or deep breathing or indrawing

Respiratory distress yes/no was recorded at triage before treatment but difficulty breathing, deep breathing and indrawing were recorded up to one hour afterwards. Indrawing was found to be a significant predictor of mortality in Berkley's paper and deep breathing was found to be a significant predictor in Roine's paper. I therefore felt that both should be kept for the model. Difficulty breathing was not closely correlated with either respiratory distress or indrawing, but had a slightly larger correlation with respiratory distress. Thus all variables were included as candidates in the model.

Fitting currently, fits at admission, or fits in this illness.

In FEAST, fits at admission records whether the child was fitting when they were screened and being randomised. Fitting currently was evaluated when a full clinical examination was carried out and so for baseline purposes fits at admission would be a better variable to consider. They were the same for 3078/3146 (98%) observations.

If fits at admission is considered to describe the current situation at baseline, then fits in this illness could be considered to describe the history. 738/2680 (27%) had fits in their illness but did not have fits at admission so this does not correlate closely, and so both fits in this illness and fits at admission were included as candidate predictors.

Respiratory rate or respiratory distress

As a continuous variable, respiratory rate has more information than respiratory distress and the WHO definition of 'fast breathing' can be calculated from respiratory rate. But 'respiratory distress' was defined simply in FEAST as 'increased work of breathing' so may not be measuring the same thing. Evans *et al* found respiratory distress defined as "irregular or deep, acidotic breathing" to be a significant predictor of mortality. Planche *et al* also found it significant in malaria slide-negative children using the definition of "presence of any of the following conditions: alar flaring, chest recession (intercostal or subcostal), the use of the accessory muscles, or abnormally deep breathing." I therefore included both as candidates taking into account the fact that respiratory distress is a very practical simple bedside variable.

Weak pulse (radial) volume

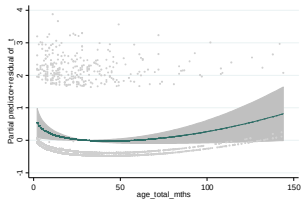
This was a dichotomised variable in the FEAST dataset. Although related to heart rate it appeared to be measuring something different as the distribution of heart rate split by weak pulse yes/no was very similar. Kilifi admissions dataset had three options for pulse volume – normal, bounding or weak, so bounding and normal categories together were joined for use in the validation dataset.

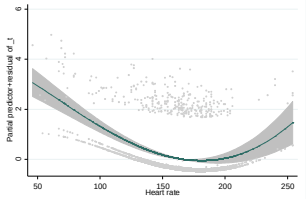
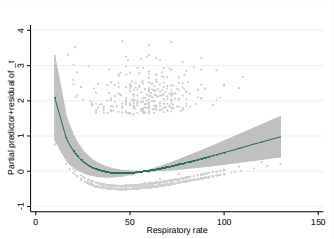
Weight or weight-for-age

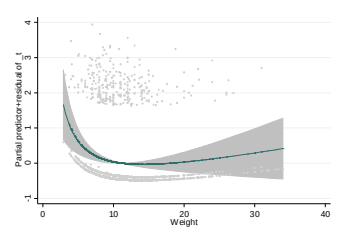
As there was a wide range of ages included in the study, then weight-for-age may be a more meaningful parameter than weight as it gives an indication of possible malnutrition or wasting. Z score charts are available in some hospitals so in theory weight-for-age could be estimated at the bedside though it may be very approximate (and not always easily available) and would take time to calculate. Weight-for-height is also a good indicator of wasting or malnutrition but Mogeni *et al* [151] found that it is not necessary to measure both weight for height and MUAC. Thus weight was kept in the model as more intuitive and easy to measure than weight-for-age and MUAC was used as a measure of those with malnutrition/wasting in the model.

All candidate predictors were evaluated in models estimating just their association with mortality adjusted for fluid arm (using the full trial data) and the results are presented in the table below. Candidate predictors were not discarded if they did not have an association with mortality at this stage [142]. The continuous variables were modelled with fractional polynomials – if only a linear effect is shown below, then there was no evidence (at p=0.05 level) for non-linear effects.

Table 2.5.3: Univariable analyses in Cox Regression models of all candidate predictors.

		Mortality	**Hazard Ratio (HR) (95% CI)**	p-value for HR
Age (months)[a]:				
√(age/100)			1.15 (1.03-1.27)	0.01
(age/100)²			1.52 (0.99-2.33)	0.02
Axillary Temperature (°C)			0.66 (0.61-0.71)	<0.001
Capillary refill time (seconds)			1.91 (1.71-2.14)	<0.001
Conscious level	Alert	23/742 (3%)		
	Prostrate	166/1949 (9%)	2.77 (1.79-4.29)	<0.001
	Coma	126/476 (26%)	9.47 (6.07-14.78)	<0.001
Cough	No	87/891 (10%)		
	Yes	227/2274 (10%)	1.03 (0.80-1.32)	0.82
Crackles	No	195/2463 (8%)		
	Yes	116/699 (17%)	2.18 (1.73-2.74)	<0.001
Decreased skin turgor	No	266/2972 (9%)		
	Yes	46/192 (24%)	2.91 (2.12-3.98)	<0.001
Deep breathing	No	47/1122 (4%)		
	Yes	265/2041 (13%)	3.23 (2.37-4.41)	<0.001
Difficulty in breathing	No	54/857 (6%)		
	Yes	260/2308 (11%)	1.84 (1.38-2.47)	<0.001
Fits >30mins in this illness	No	269/2903 (9%)		
	Yes	41/246 (17%)	1.83 (1.32-2.54)	<0.001
Fits in this illness	No	175/1973 (9%)		
	Yes	137/1187 (12%)	1.29 (1.03-1.62)	0.03
Fitting/convulsions at admission	No	257/2689 (10%)		
	Yes	56/463 (12%)	1.27 (0.95-1.69)	0.11

		Mortality	Hazard Ratio (HR) (95% CI)	p-value for HR
Heart rate[b]: (heartrate/100)² (heartrate/100)²xln(heartrate/100)			0.07 (0.04-0.11) 12.39 (6.67-23.02)	 <0.001 <0.001
History of fever	No Yes	2/16 (13%) 313/3152 (10%)	0.80 (0.20-3.23)	0.76
Indrawing	No Yes	71/1014 (7%) 241/2150 (11%)	1.65 (1.26-2.15)	<0.001
Jaundice	No Yes	172/2149 (8%) 142/1014 (14%)	1.82 (1.46-2.27)	<0.001
Liver size >2cm below costal margin	No Yes	150/1886 (7%) 162/1124 (14%)	2.02 (1.62-2.53)	<0.001
MUAC (cm)			0.86 (0.78-0.94)	0.001
Neck stiffness or bulging fontenelle	No Yes	297/3083 (9%) 14/80 (18%)	1.86 (1.09-3.18)	0.02
Respiratory distress	No Yes	25/547 (5%) 289/2610 (11%)	2.53 (1.68-3.80)	<0.001
Respiratory rate[c]: ln(resprate/100) (ln(resprate/100))²			4.38 (1.91-10.07) 2.55 (1.59-4.09)	0.001 <0.001
Severe pallor	None Eyes or tongue All	91/1550 (6%) 61/501 (12%) 161/1109 (15%)	 2.12 (1.54-2.94) 2.62 (2.03-3.39)	 <0.001 <0.001
Sex	Male Female	167/1705 (10%) 148/1465 (10%)	1.03 (0.83-1.29)	0.79
Sunken eyes	No Yes	292/3067 (10%) 19/96 (20%)	2.11 (1.33-3.36)	0.002
Temperature Gradient	No Yes	72/1287 (6%) 243/1883 (13%)	2.41 (1.85-3.13)	<0.001

		Mortality	**Hazard Ratio (HR) (95% CI)**	p-value for HR
Vomiting	No	131/1536 (9%)		
	Yes	183/1628 (11%)	1.34 (1.07-1.68)	0.01
Weak pulse	No	154/2490 (6%)		
	Yes	161/680 (24%)	4.29 (3.44-5.36)	<0.001
Weight (kg)[d]:				
weight/10			2.17 (1.30-3.61	0.002
1/(weight/10)			1.47 (0.91-2.39)	0.003

[a] This gives a L-shaped overall function with a small decline to 30 months and then a steady increase.

[b] This gives an u-shaped overall function with a clear long decline from 50 – 150 beats per min and then an increase from 150 bpm.

[c] This gives an u-shaped overall function with a very steep decline from 20-60 breaths per min, then a gradual but clear increase after this point.

[d] This gives an u-shaped overall function with a steep decline from 2-11 kg and then a gradual increase from this point.

Candidate predictors were also examined to check for any interactions with the treatment arms. The variables presented below had a p-value<0.1, following a likelihood ratio test to examine interactions between the fluid allocation and exposure variable. The HR per fluid arm category was presented to examine the evidence for an interaction further.

Table 2.5.4: Univariable analyses of potential interactions between candidate predictors and randomisation arm.

	HR (95% CI)	p-value for HR	p-value for interaction
Respiratory distress			
yes vs no			0.07
- Albumin	1.68 (0.94-2.98)	0.08	
- Saline	5.33 (2.18-13.07)	<0.001	
- Control	2.13 (0.98-4.63)	0.06	
Deep breathing			
yes vs no			0.02
- Albumin	1.99 (1.28-3.09)	0.002	
- Saline	5.19 (2.86-9.43)	<0.001	
- Control	4.07 (2.09-7.92)	<0.001	
Bulging fontenelle or stiff neck			
yes vs no			0.03
- Albumin	0.35 (0.05-2.52)	0.3	
- Saline	2.76 (1.34-5.65)	0.006	
- Control	2.81 (1.13-6.96)	0.03	

The hazard ratios show that having respiratory distress or deep breathing still increases the hazard for mortality but sometimes more for one particular arm than the other arms. Bulging fontenelle or stiff neck appears protective for those in the albumin arm but there was a wide confidence interval and a small number of children in this group with events – mortality for those with bulging fontenelle or a stiff neck was 1/24 (4%) for albumin, 8/29 (28%) for saline and 5/27 (16%) in the control group. Mortality for those without bulging fontenelle or a stiff neck was very similar to the overall mortality distribution. The interactions were all ≥0.02 and the majority of effect estimates were in the same direction and so I decided to use an overall estimate for candidate predictors in the models.

2.5.2.2 Data handling decisions

Bulging fontenelle has been combined with stiff neck as they represented a similar physiological sign and bulging fontenelle was age-dependent so had a lot of missing data; combining it with stiff neck gave a new variable with missingness at <1%.

Severe pallor was reduced from four categories to three categories by combining palmar with tongue or conjunctiva to give a middle category for 'some pallor' which compares with categories 'none' or 'all'. This was to make this variable more similar to the validation dataset which does not specify where the pallor is.

Correlations between continuous candidate predictors were also examined in a matrix plot (Figure 2.5.1). This identified 4 outliers in temperature, MUAC and respiratory rate when plotted with age and weight. These were truncated at the maximum of the distribution for the six month interval around their age. The matrix plot was then redrawn and showed that, as expected, weight, age and MUAC were highly correlated (with spearman correlations of 0.89 for age and weight, 0.50 for MUAC and age, and 0.67 for MUAC and weight), and heart rate had weak correlation with respiratory rate (0.42) and temperature (0.41).

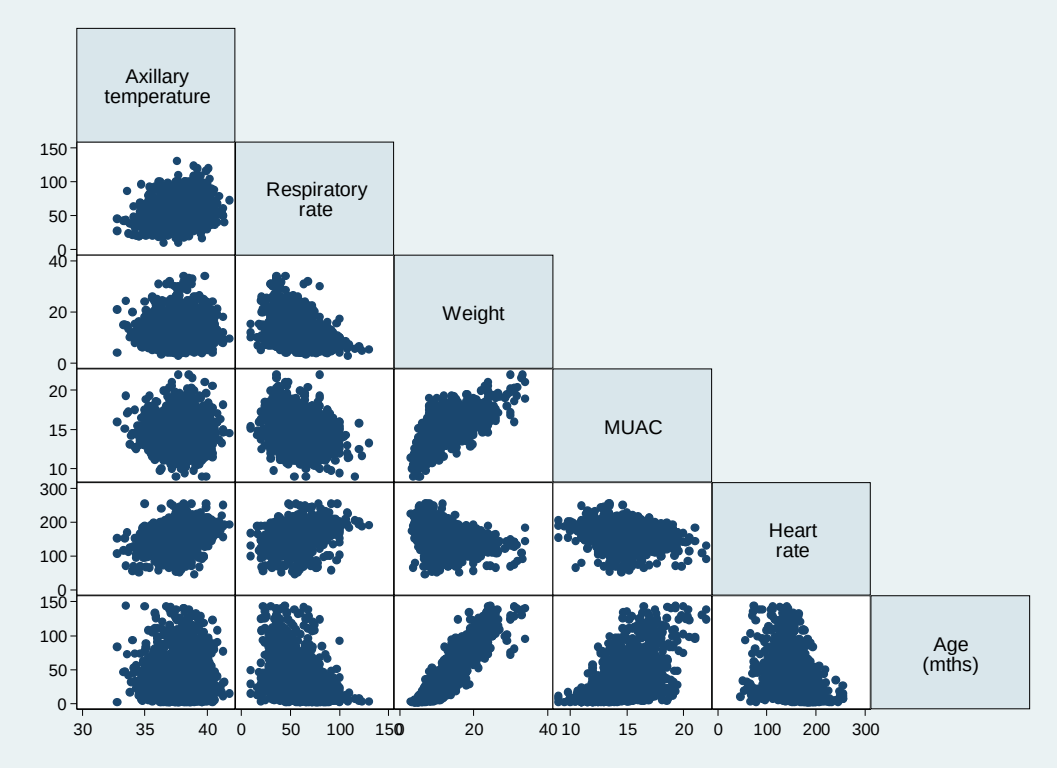
The distributions of the continuous variables were also examined and summarised. The table shows that the distributions were approximately normal with no strong skewedness or kurtosis.

Table 2.5.5: Distributions of continuous candidate predictors.

Variable	Mean	SD	Min	Max	Skewedness	kurtosis	missing (%)
Axillary Temperature (°C)	38.07	1.27	30.2	41.8	-0.52	3.50	<1%
Respiratory rate (breaths per min)	57.73	15.23	10	140	0.36	3.70	<1%
Age (months)	29.86	23.77	2	144	1.77	6.81	0%
Weight (kg)	11.16	4.01	2.9	34	1.53	7.05	<1%
MUAC (cm)	14.74	1.60	9	26	0.34	4.77	6%
Heart rate (beats per min)	166.03	26.15	37	254	-0.93	5.31	<1%

Due to the amount of missing data, and that missing data occurring disproportionately in those that died, MUAC was not included in the initial model building process (MUAC was missing in 126/315 (40%) children who died before 48 hours).

Figure 2.5.1: Matrix plot of continuous candidate predictors.



Fractional polynomials were used to model the association between continuous variables and mortality in both univariable and multivariable models. Consideration of the strong correlation between weight, age and MUAC was taken into account when evaluating these variables in the model. Although useful to examine the functions that were selected in univariable models for continuous variables, the most appropriate function is best found as part of the model building process in the full multivariable model because this fully adjusts for confounding.

2.5.2.3 Selecting a final regression model

All the candidate variables were included in a multivariable Cox regression model with the outcome of death by 48 hours. Time was measured in hours and observations were censored at time of death, time of absconding or 48 hours from randomisation, whichever was the earliest. There were 16 children who were censored alive before 48 hours (0.5%) and 315 deaths before 48 hours in the FEAST trial dataset. MUAC was also initially included in the model but as this reduced the number of events for the model (from 315 events to 189) disproportionately to the amount of missing data in that variable (6% missing overall), I

decided that it would be better explored with multiple imputation after the main clinical model has been developed and it was therefore not included in complete case analyses for a clinical model.

The backwards elimination process started with the 25 clinical variables described above, and 3056 observations with observed values for all of them (96% of total observations) of which there were 296 deaths (10%)). Missingness was spread across the candidate predictors but was never more than 17 observations within one variable (<1%). The model was always adjusted for randomisation arm ie this variable was not made available for selection (although remained statistically significant). The Royston and Altman model-selection algorithm was used which increases the power to detect non-linear relationships [152]. For a given degree *d* of fractional polynomial (FP), all FP*d* models are fitted by maximum likelihood and the power vector maximising the likelihood is selected as the best-fitting (for the discrete powers considered). The steps in the algorithm are: Firstly, FP2 functions are evaluated to find the best function as described above, then the best FP2 function in the model is tested against the best FP1 function (found using the process described above) at an appropriate significance level called alpha (0.05 is used here). If it is significant then the best FP2 function is kept, if not then the best FP1 function is tested against a straight line (or linear relationship) at the alpha significance level. If this is significant then the straight line is tested against omitting the variable from the model at the model selection significance level. If this is significant then the function is a straight line, otherwise the variable is dropped.

The variables that were found to be independently associated with mortality in the first model, with the model and alpha significance level both set at 0.05, were temperature, weight, heart rate, capillary refill time, weak pulse, conscious level, respiratory distress, deep breathing, crackles and severe pallor. Age was not kept in the final model, but as it was an *a-priori* confounder a model was re-fitted where age (in months) was kept in the model and not included in the selection process. This gave coefficients that suggested opposite effects of weight and age, probably as they are so highly correlated. Thus, although age and weight were key variables, their co-linearity gave contradictory model results, suggesting over-fitting when both were included, indicating that only one should be retained. The relationship between age in months and weight was examined further by fitting univariable models to find the best fitting fractional polynomial model without adjustment for other variables. These were found to be linear functions with increasing weight or age associated with lower mortality.

The residuals of models with just age and weight, and then each one with all candidate predictors included, were examined. Weight was strongly associated with mortality at the low

end of the distribution (babies and toddlers) but other predictors were stronger when weight was above approximately 10kg and this association remained in the models with all candidate predictors (except age). The fractional polynomial chosen was an FP1 function of $1/(\text{weight})^2$.

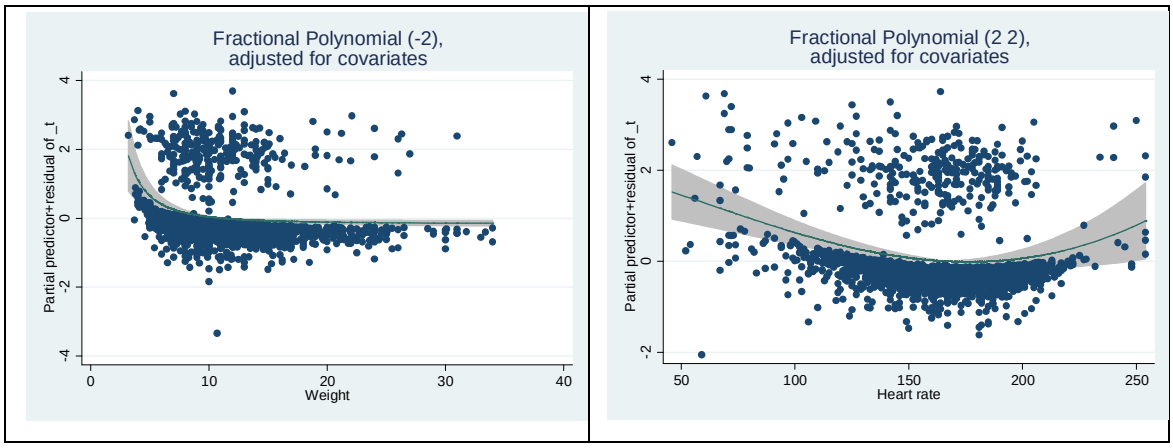
Weight-for-age was also considered as a solution to the difficulties of age and weight being highly correlated. The weight-for-age z-score can be estimated at the bedside by examining charts which are available in most hospitals as it is a key measurement of malnutrition. It was calculated for the dataset using the WHO Child Growth Standards 2006 (and 2007 extension for 5-19 year olds) [153, 154]. But when weight-for-age was evaluated in models including the candidate predictors it was not retained during the stepwise process and did not have a strong independent association with mortality once adjusted for other factors. Thus, as weight had a stronger association with mortality, especially for those with weight<10kg, the FP1 function for weight was kept in the model and age was not kept.

In models with the other candidate predictors, adjusted for randomisation, heart rate had an association with mortality as an FP2 function (quadratic plus quadratic multiplied by the log function) as shown in Figure 2.5.2. This function may have been influenced by outliers, which can have large impact on the fractional polynomial function chosen in the model. Thus the influence of values in the 1st and 99th percentile of heart rate, as well as weight, axillary temperature, and respiratory rate were closely examined through histograms and by plotting the data and the fit of the model. Truncation is one method to reduce influence of outliers whilst retaining the observations in the model and can be used to reduce their impact and for heart rate especially it was important to consider this. Thus the model was also fit with heart rate truncated at the 1-99th percentiles of the distribution and restricting the highest FP power to 2. The values <76 (to the minimum of 46) were changed to 76 and values >216 (up to the maximum of 254) were changed to 216 (n=61 changed altogether). This gave a linear function for heart rate. It is important to note that values above the 99th or below the 1st percentile are possible in this population (as bradycardia is defined as <80 beats per minute, and severe tachycardia for the youngest age group (<12 months) is >180 beats per minute) and so care needs to be taken when truncating setting such values to missing. Piecewise linear functions were also considered for heart rate and weight in the models to help with simplifying the functions for ease of use in hospitals and to be intuitive for clinicians to understand what the fractional polynomial effects mean.

Taking all these factors together, I decided not to truncate heart rate and so it was retained in the model as a FP2 function of $(\text{heart rate}/100)^2 + (\text{heart rate}/100)^2 \log(\text{heart rate}/100)$ as including the information from the extreme values was important in this population, despite

the fact these values may be exerting undue influence. Weight was retained as described above and axillary temperature and capillary refill time were modelled as linear functions (as chosen by the FP model selection procedure). These functions were plotted against the data points and are presented in Figure 2.5.2.

Figure 2.5.2: Fitted fractional polynomial functions for weight and heart rate plotted against the data points**.**



The elimination process is outlined in Table 2.5.6, indicating which variables were entered into the model for selection and which remained and, for the continuous variables, which function was most appropriate.

Table 2.5.6: Backwards elimination stepwise process using fractional polynomials for continuous functions.

Variable	**Status**	**Powers**	**Transformation**
Axillary temperature	in	1	
Respiratory rate	out		
Gender	out		
Age (months)	out		
Weight (kg)	in	-2	$1/(\text{weight}/10)^2$
Heart rate	in	2 2	$(\text{heart rate}/100)^2+$ $(\text{heart rate}/100)^2\log(\text{heart rate}/100)$
Capillary refill time (s)	in	1	
Temperature gradient	out		
Weak pulse	in	1	
Conscious level	in	1	
Respiratory distress	in	1	
Cough	out		
History of fever	out		
Vomiting	out		
Jaundice	out		
Fits	out		
Fits for longer than 30 minutes	out		
Fits at admission	Out		
Indrawing	Out		
Deep Breathing	in	1	
Crackles	in	1	
Liver size >2cm below costal margin	Out		
Skin turgor	Out		
Bulging fontenelle or stiff neck	Out		
Severe pallor	in	1	

Note: none of the variables were truncated.

The model that was selected from the candidate predictors using 3096 complete cases for all variables was refitted to the complete case data with just the selected variables, and is presented in Table 2.5.7 below with the hazard ratio, coefficient and 95% confidence interval for the coefficient. The model had 3121 observations (98% of total sample size) and 306 deaths.

Table 2.5.7: Final clinical model coefficients for linear predictor.

Variable	**Hazard Ratio***	**Coefficient***	**95% Confidence Interval for coefficient**	**Wald test p-value**
Axillary temperature (per 1°C increase)	0.85	-0.16	(-0.25, -0.07)	<0.001
$1/(weight/10)^2$	1.129	0.18	(0.07, 0.28)	0.001
$(Heart\ rate/100)^2$	0.25	-1.39	(-1.97, -0.81)	<0.001
$(Heart\ rate/100)^2 ln(heart\ rate/100)$	3.64	1.29	(0.69, 1.90)	<0.001
Capillary refill time (s)	1.22	0.20	(0.07, 0.34)	0.004
Weak pulse	1.93	0.66	(0.41, 0.92)	<0.001
Conscious level	2.23	0.80	(0.59, 1.01)	<0.001
Respiratory distress	1.81	0.59	(0.12, 1.06)	0.013
Deep breathing	1.49	0.40	(0.04,0.76)	0.029
Crackles	1.87	0.63	(0.38, 0.87)	<0.001
Severe pallor	1.62	0.48	(0.20, 0.76)	0.001

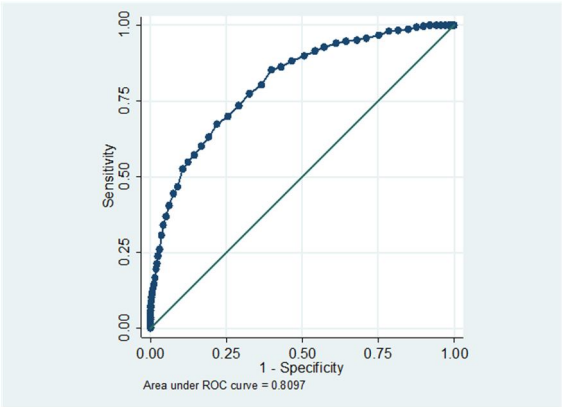
* Adjusted for fluid arm and all other variables in the model.

The linear predictor was calculated for each child and used to investigate the discriminatory ability of the model. It was rounded to the nearest 0.1 (to enable the AUROC to be calculated more easily) and non-parametric AUROC was calculated and plotted in Figure 2.5.3. The AUROC was 0.81 (0.79-0.84).

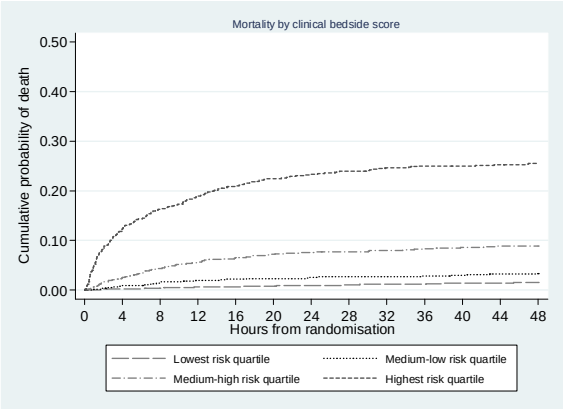
The observed mortality risk was also examined by splitting the data into quartiles of this linear predictor. Figure 2.5.3 shows that the group with the highest predicted risk has a much higher cumulative probability of death (26% at 48 hours) compared to the other groups and thus has been separated out reasonably well. The medium-high risk quartile has a cumulative probability of death at 48 hours of 9% and is also separated out from the lowest risk (1% at 48 hours) and medium-low risk groups (3% at 48 hours).

Figure 2.5.3: a) Area under ROC curve plot and b) cumulative probability of death estimated from the linear predictor.

a)



b)



2.5.2.4 Sensitivity analyses around final clinical regression model

As there was a very small percentage of missing data in the clinical model (49/3170 (2%), 9 deaths), multiple imputation with chained equations was used to impute data and ensure that the coefficient estimates did not differ widely when compared to the complete case analysis. Categorical ordered variables were imputed with ordered logistic regression (there were no categorical unordered variables), binary variables with logistic regression, skewed variables or those transformed for the analysis model were imputed in their transformations with predictive mean matching drawing from the nearest 10 neighbours, and other continuous variables were imputed with linear regression [155]. Cox Regression was used to recreate the final model adjusting the coefficients and standard errors by Rubin's rules [156]. The coefficients and 95% confidence intervals differed by less than 0.01 for their values in the model confirming the validity of the complete case analysis.

Other sensitivity analyses included examining the impact of censoring on the model building process as I used survival data and Cox regression models in comparison to logistic regression models which are commonly found in prognostic model building. The multivariable logistic regression model found using backwards elimination with a selection level of 0.05 included all of the same variables as the Cox regression model apart from weight. But when this was added back into the final model as $1/(\text{weight})^2$, the Wald test p-value was 0.014. The selection of the same variables was as expected given the very small amount of censoring in this dataset.

The survival analysis examined mortality over 2 days (from randomisation to 48 hours) and was measured in hours, as death by 48 hours was the endpoint of the trial. Not all datasets may have such accurate data and so logistic regression with the outcome of mortality within 2 calendar days of admission was examined to make the outcome more comparable with other datasets. For the FEAST data this made little difference to the coefficients and only added 11 deaths to the 315 that had already occurred by 48 hours.

It is also of interest to understand the predictive ability of the overall model in different sites across the trial. The AUROC for the linear predictor is presented per site in Table 2.5.8 below. There was, as expected, some variability in the AUROC estimate between 0.77 and 0.90 but these were not correlated with higher mortality rate or size of site.

Table 2.5.8: Area under ROC curve values for different trial sites estimated from the linear predictor.

Site	Mortality	AUROC for linear predictor
Kilifi	21/215 (10%)	0.81 (0.71-0.91)
Mulago	60/732 (8%)	0.78 (0.72-0.84)
Soroti	58/623 (9%)	0.77 (0.70-0.84)
Lacor	37/230 (16%)	0.90 (0.84-0.96)
Mbale	111/1226 (9%)	0.84 (0.81-0.88)
Teule	19/95 (20%)	0.83 (0.74-0.91)

2.5.2.5 Moving from a regression model to a bedside score

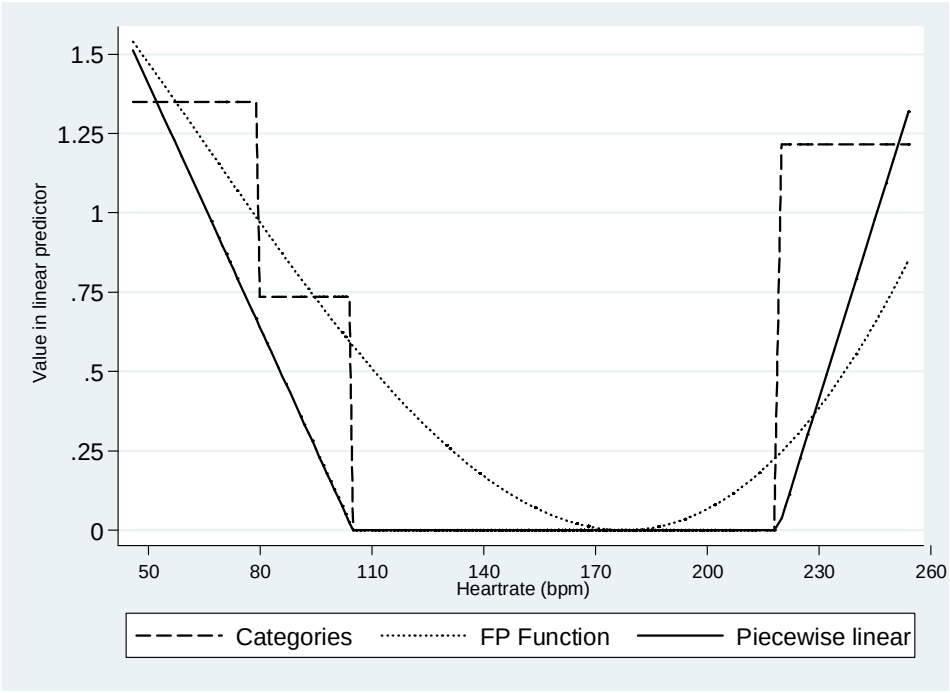
To move from a prognostic model to a clinical score I used the method of dividing the coefficients in the linear predictor by the lowest coefficient and so each coefficient needed to link to a subset of children in the dataset. This was straightforward for categorised variables but the continuous variables needed to be categorised.

Piecewise linear functions were fitted for weight and heart rate and the profile log likelihood was maximised to find the two knots for heart rate and one knot for weight. The log likelihood was maximised at heart rate of 105 and 220 beats per minute and a weight of 10kg.

As bradycardia is defined as <80 beats per minute this cut-off was used to create one category for heart rate, the next was a category from 80 up to the knot of 105 beats per minute, a

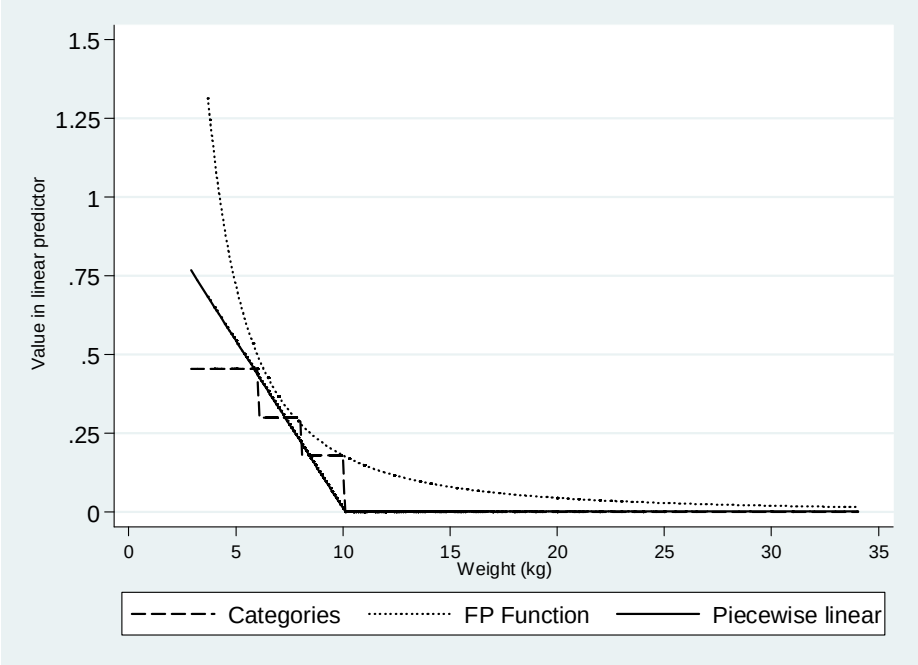
category for normal heart rate and then a category for very high heart rate as identified above. So heart rate categories were <80, 80-105, >105 - <220, ≥220 beats per minute. See graph in Figure 2.5.4 below.

Figure 2.5.4: Linear predictors from different functions for heart rate



As the knot for weight as a linear piecewise function was at 10kg then the base category was created to include all children with a weight >10kg. It was felt there would be too many categories if all integers under 10kg were considered separately and so 2kg increments were examined. The numbers in each 2kg category become very few when weight ≤6kg (n=153) and so this was created as one category. Then the other categories were 6.1 – 8kg and 8.1 – 10kg. See graph in Figure 2.5.5 below.

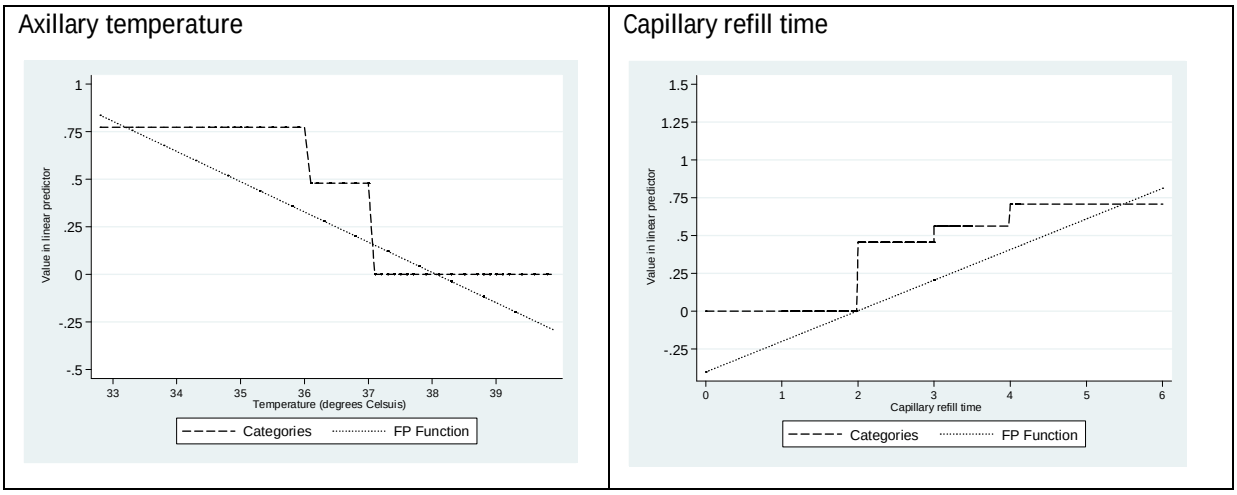
Figure 2.5.5: Linear predictors from different functions for weight



The variable axillary temperature was modelled as linear and so known clinical cut-offs were considered when categorising this variable for a bedside score. Hypothermia is usually defined as a temperature <36.0°C and this was used for the first category, and then an increase by one degree to 37.0 for the second category. When more categories were created beyond this cut-off in one degree groups the Wald test p-values were >0.1 and the coefficients would be too small when points are allocated for the score, and so one category for all those >37.0°C was created (Figure 2.5.6).

The variable capillary refill time was also modelled as linear. The lowest group was 1 second which forms the baseline, and then 2 seconds and 3 seconds were considered as separate categories, and then as there were very small numbers at 4 and 5 seconds these were combined together to give the categories 1 second , 2 seconds, 3 seconds, 4 or more seconds (Figure 2.5.6).

Figure 2.5.6: Linear predictor functions for axillary temperature and capillary refill time



These newly categorised continuous variables were fitted in a Cox Regression model along with the variables that were already binary or categorical and coefficients were estimated (this model is referred to as model A). The coefficient for weight category 8.1-10kg was the smallest in absolute magnitude in the model at 0.179 and so all other coefficients were divided by this value and rounded to the nearest integer to produce the score. The score is presented in Table 2.5.9.

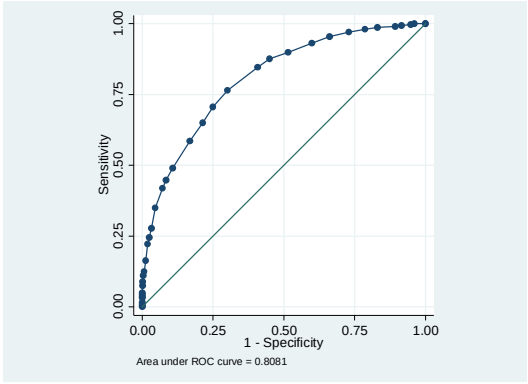
Table 2.5.9: Clinical bedside score

Variable	Coefficient/0.179	Score given if present	Maximum possible score
Weight ≤6.0kg	2.53	3	3
Weight 6.1-8kg	1.66	2	
Weight 8.1-10kg	1	1	
Axillary temperature 36.1-37.0	2.68	3	4
Axillary temperature ≤36.0	4.32	4	
Heart rate <80	7.54	8	8
Heart rate ≥80- <105bpm	4.13	4	
Heart rate ≥220 bpm	6.76	7	
Capillary refill time 2s	2.56	3	4
Capillary refill time 3s	3.13	3	
Capillary refill time 4 or more seconds	3.96	4	
Conscious level – prostrate	4.41	4	9
Conscious level – coma	8.8	9	
Deep breathing	2.26	2	2
Respiratory distress	3.17	3	3
Severe pallor	2.46	2	2
Crackles/lung crepitations	3.37	3	3
Weak pulse	3.74	4	4
Total			42

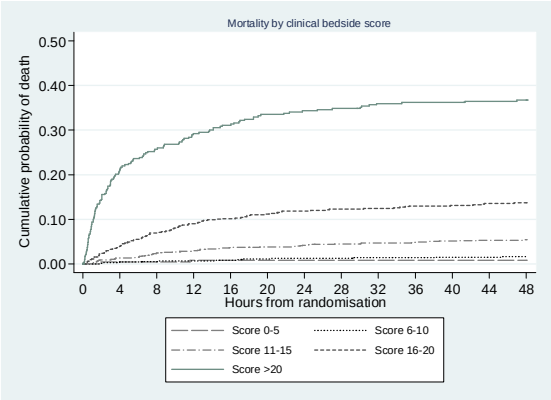
The integer score was then evaluated on the FEAST data to examine its discriminative ability and values of the score ranged from 3 to 39. It gave an overall AUROC of 0.81 (95% CI 0.78-0.83) as seen in Figure 2.5.7. This shows that there was very little information lost when the integer score is used rather than the linear predictor.

Figure 2.5.7: a) Area under ROC curve plot for clinical bedside score and b) cumulative probability of death over time from randomisation.

a)



b)



The increased mortality risk from having a high score can be calculated using the predicted survival probability at 48 hours from Kaplan Meier. For those with a score of 5 the estimated risk is 1.0%. A score of 5 would be a child who is alert with a weight above 6kg and a heart rate in the normal range and only one or two of the other prognostic factors, such as deep breathing, but not severe pallor nor weak pulse. For those with a score of 15 their 48h mortality risk is 7.7%. An example of a child with a score of 15 is one with a high temperature but normal weight and heart rate, a slightly longer capillary refill time of 3 seconds, and who is prostrate with severe pallor, respiratory distress and deep breathing. For those with a score of 30, the estimated 48h mortality risk is 36.0% and this would be a child with most of the high scoring prognostic factors such as a hypothermia and bradycardia, severe pallor, long capillary refill time, who is prostrate or in a coma and has respiratory distress, deep breathing and a weak pulse.

The prognostic score was then modified further to ensure an easy-to-use straightforward scale from 1-10 by reducing some categorisations (for example combining the two temperature categories to ≤37°C, and dropping one weight category (8-10kg), and combining capillary refill time categories to create one category representing 2 or more seconds). Score categories that were allocated a score >3 in the clinical bedside score were assigned 2 in the revised scale and those ≤3 in the clinical bedside score to 1 in the revised scale. The variables were also assessed with Net Reclassification Index (NRI) and weight and deep breathing added the least predictive value and so were dropped from the score (Table 2.5.10). It was also titled the FEAST Paediatric Emergency Triage (FEAST PET) score.

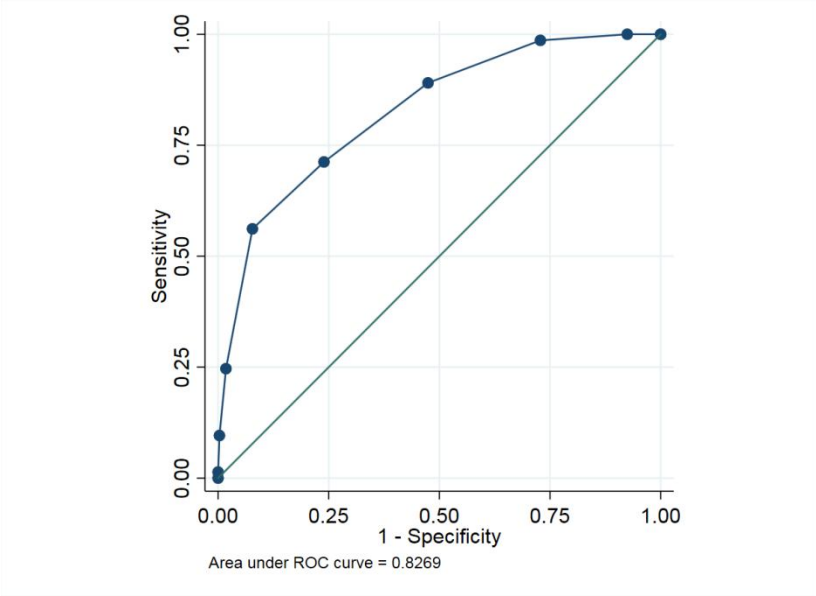
Table 2.5.10: The FEAST PET score

Factor	Coefficient (95% CI) from multivariable model*	p-value from multivariable model	Score value given if present	Maximum possible value for PET score
Axillary temperature: ≤37°C	0.63 (0.38-0.87)	<0.001	1	**1**
Heart rate: <80bpm (bradycardia)	1.34 (0.92-1.77)	<0.001	2	**2**
≥80- <105bpm	0.70 (0.11-1.30)	0.02	1	
≥220 bpm (severe tachycardia)	1.34 (0.92-1.77)	<0.001	2	
Capillary refill time: 2 or more seconds	0.53 (0.21-0.85)	0.001	1	**1**
Conscious level: prostrate	0.68 (0.23-1.13)	0.003	1	**2**
coma	1.53 (1.06-2.00)	<0.001	2	
Respiratory distress	0.55 (0.07-1.02)	0.02	1	**1**
Lung crepitations	0.60 (0.36-0.85)	<0.001	1	**1**
Severe pallor	0.49 (0.22-0.76)	<0.001	1	**1**
Weak pulse	0.73 (0.48-0.97)	<0.001	1	**1**
Weight: <6kg	0.41 (-0.05-0.88)	0.08	-	**-**
6-8kg	0.21 (-0.03-0.45)	0.09		
Deep breathing	0.42 (0.06-0.77)	0.02	-	**-**
Total				**10**

*Coefficient from linear predictor of multivariable cox regression model on complete cases adjusted for randomisation arm which was not included in the score.

The FEAST PET score was then internally validated on the control arm data giving an AUROC of 0.83 (95% CI 0.77-0.87) (Figure 2.5.8).

Figure 2.5.8: Area under ROC curve plot for FEAST PET score on FEAST control arm data.



2.5.3 Evaluating the addition of other candidate predictors – a 'clinical plus' score

As the score does not have as high as might be expected AUROC in the development dataset it was felt that other variables could be considered as candidate predictors and they might improve the model. These other variables were also measured, with the result being available 'at the bedside' rather than having to be sent to a local laboratory, but may have had larger amounts of missing data compared to those used in the clinical score meaning a complete case analysis was potentially biased.

Multiple imputation was therefore used to evaluate these variables in the model. The candidate variables included those measured using the i-STAT machine: sodium (mmol/L), potassium (mmol/L), chloride (mmol/L), pH, BUN (mg/dL), base excess (mmol/L), PCO_2 (mmHg) and TCO_2 (mmol/L). It also included variables not measured using the i-STAT machine: haemoglobin (g/dL), malaria test result (positive or negative), oxygen saturation (%), mid-upper arm circumference (MUAC) (cm), systolic blood pressure (mmHg), blood sugar (glucose) (mmol/l), lactate (mmol/L), and HIV result (positive or negative).

Multiple imputation with the clinical variables and the additional laboratory variables described above was carried out to create 25 datasets. Heart rate was imputed as an FP2 function as this was the function that was found in complete case analyses with very few missing values. Weight was also used in the imputation model as an FP1 function from complete case analyses, and other variables were included as in the sensitivity analyses in

section 2.5.2.4. There was more missing data in general in the laboratory variables with greater proportions of missing data for those that died compared to those that lived in some variables (MUAC, sbp, and oxygen saturation). Thus choosing appropriate functions from complete case analyses would be more prone to bias for these variables and so in the imputation models all the continuous laboratory and additional variables were not transformed and were imputed using predictive mean matching (PMM). As PMM was used, fractional polynomials were restricted to FP1 functions in the multivariable fractional polynomial model building process, as methods to adapt the mfp command in imputed data to higher FP powers have not been developed yet. Imputed and observed values were compared visually to check ranges.

As the AUROC measure is not as useful to compare models directly, the Net Reclassification Index (NRI) was used to evaluate the additional variables in the model. Logistic regression was used to calculate the NRI in each dataset and 4 risk categories were created with arbitrary cut-offs at 5%, 10%, and 15% for 48 hour mortality. The range and mean of the NRI across the 25 imputed datasets was used to assess whether the additional laboratory variables could be usefully added to the clinical bedside variables already included in the FEAST PET score. The range of calculated NRI measures across multiple imputations was presented, because methods to formally combine estimates of NRI and IDI in multiply-imputed data have not yet been described in the literature. Backwards elimination (with an exit threshold mean p=0.05 calculated from all imputed datasets), including all laboratory markers, was then used to identify those with the largest NRIs (Table 2.5.11).

Table 2.5.11: Net Reclassification Index ranges across 25 imputed datasets for candidate laboratory markers when added individually and in combination to the clinical model.

Univariable analyses (added individually to clinical model)			
	NRI range	Two-sided p-value range	Mean p-value
Lactate	20.4-23.1%	<0.001	<0.001
TCO_2 (mmol/L)	18.2-23.0%	<0.001	<0.001
pH	13.8-19.7%	<0.001	<0.001
BUN	9.9-16.4%	<0.001	<0.001
Base excess	18.3-23.4%	<0.001	<0.001
Potassium	6.3-11.8%	<0.001-0.03	0.003
HIV positive	2.4-6.0%	0.004-0.2	0.03
Glucose	2.8-5.3%	0.015-0.2	0.08
Oxygen Saturation	1.1-5.3%	0.001-0.4	0.08
Malaria positive**	2.3-5.4%	0.02-0.3	0.1
Systolic Blood Pressure	2.0-3.4%	0.02-0.2	0.1
Haemoglobin	1.1-3.3%	0.04-0.6	0.2
Chloride	1.2-7.6%	0.002-0.6	0.2
PCO_2	-0.9-4%	0.01-1.0	0.3
Sodium	-1.1-2.3%	0.07-1.0	0.6
Factors identified through backwards elimination process, included multivariably*.			
Lactate	10.6-16.7%	<0.001	<0.001
BUN	3.1-8.2%	<0.001-0.11	0.02
pH	2.9-9.0%	<0.001-0.22	0.03
Combined effect of Lactate, BUN and pH	24.6-28.9%	<0.001	<0.001

* NRI's calculated from one multivariable model considering each factor separately, and then adding all three together to the clinical model to estimate the NRI's for a combined effect.

** As defined in the FEAST Statistical Analysis Plan

Using backwards elimination including clinical factors and all laboratory markers, lactate, BUN and pH added independent information to the FEAST PET score, as shown in Table 2.5.11

above. These variables were then categorised using appropriate clinical cut-offs and added to the clinical score. This extended the range to 0-14 and created the FEAST Paediatric Emergency Triage and Laboratory score (FEAST PETaL) (Table 2.5.12).

Table 2.5.12: FEAST Paediatric Emergency Triage (PET) score and the FEAST Paediatric Emergency Triage and Laboratory (PETaL) score.

Factor	Coefficient (95% CI) from multivariable model*	Score value given if present	Maximum possible value for PET score	Maximum possible value for PETaL score
Axillary temperature: ≤37°C	0.63 (0.38-0.87)	1	**1**	1
Heart rate: <80bpm (bradycardia)	1.34 (0.92-1.77)	2		2
≥80- <105bpm	0.70 (0.11-1.30)	1	**2**	
≥220 bpm (severe tachycardia)	1.34 (0.92-1.77)	2		
Capillary refill time: 2 or more seconds	0.53 (0.21-0.85)	1	**1**	1
Conscious level: prostrate	0.68 (0.23-1.13)	1	**2**	2
coma	1.53 (1.06-2.00)	2		
Respiratory distress	0.55 (0.07-1.02)	1	**1**	1
Lung crepitations	0.60 (0.36-0.85)	1	**1**	1
Severe pallor	0.49 (0.22-0.76)	1	**1**	1
Weak pulse	0.73 (0.48-0.97)	1	**1**	1
Weight: <6kg	0.41 (-0.05-0.88)	-	**-**	
6-8kg	0.21 (-0.03-0.45)			
Deep breathing	0.42 (0.06-0.77)	-	**-**	
Total			**10**	
Additional laboratory values to be added if measured				
Lactate >5mmol/l	1.12 (0.79-1.46)	2		2
pH<7.2	0.97 (0.69-1.25)	1		1
BUN > 20 mg/dl	0.58 (0.26-0.90)	1		1
Total (laboratory score)				**14**

*Coefficient from linear predictor of multivariable cox regression model on complete cases. First section includes clinical factors only. Second section (laboratory values) adjusted for all clinical factors. Multivariable model also adjusted for randomisation arm which was not included in the score.

Rubin's rules were used to combine AUROCs from the multiply imputed datasets to validate the PETaL score in the FEAST control arm data [145] and gave 0.86 (95% CI 0.82-0.90).

2.5.4 Conclusions

An AUROC value of 0.81 indicates that the clinical bedside score has good discriminatory ability but this could be expected for the data that the score was developed from. As the model should be optimistic for the data that it was built with, the AUROC value might even have been expected to be higher. There was not a large loss of information when moving from using the linear predictor to using an integer score for particular categories, nor when reduced further to create the FEAST PET score, as the AUROC values were very similar. This suggests intergerisation and category reduction is not the cause of the slightly poorer than expected performance.

The discriminative ability of the clinical bedside score varied a little by site involved in the trial, with the AUROC ranging from 0.77 (Soroti, Uganda) to 0.90 (Lacor, Uganda). However, the variation was not obviously related to size of the site nor mortality rate and so no clear pattern emerged from this comparison. There could be variability in the underlying causes that brought children to hospital when comparing sites; for example, in Uganda there is a much higher incidence of haemoglobinurea (dark or red coloured urine) which may impact mortality in Soroti and Mbale but not at other sites and was thus not included in the model because it was less relevant elsewhere. However, these two sites have different values of AUROC themselves. Other underlying diseases may have more impact on the discriminative ability of the score and were simply not predicted well by the clinical observations.

Nevertheless, the model covers a variety of presentation syndromes and takes into account neurological disorders on admission with the inclusion of conscious level, respiratory presentation with the inclusion of deep breathing, crackles and respiratory distress, and severe shock or acidosis on admission with the inclusion of weak pulse, capillary refill time, and severe pallor[157]. This is reassuring as children included in the trial had a variety of diagnoses and the model would be expected to reflect this situation.

The FEAST PET score improves on the clinical bedside score by reducing the maximum of the score to 10, reducing each score value to either 1 or 2 and reducing the number of factors that need to be measured. This should make it more appealing and thus more useful to clinical staff working in triage but also for staff working in research if it is used for stratification or as part of

entry criteria for a trial or research study. It retained a good discriminatory ability when validated on the control arm data with an AUROC value of 0.83, similar to that of the clinical bedside score.

The three laboratory measures (lactate, BUN, and pH) that were added to the PET score based on their NRI gave a slightly higher AUROC of 0.86 in the derivation dataset but this was not significantly different from the PET score, thus showing that clinical models could be sufficient for a good score. It also highlights that improving prediction does not always improve ability to discriminate children at high and low risk.

2.6 Validating the scores in the Kilifi dataset

The Kilifi admissions dataset was identified as an appropriate validation dataset for the prognostic model and score as the hospital was a trial site (with any children enrolled in the trial excluded from the validation dataset) and admits children in shock with diagnoses that were seen in the trial (along with others with different diagnoses and not in shock). The admissions dataset was split into two datasets for validation, general admissions and the KEMRI ward. The FEAST trial actually took place in the KEMRI ward which is a high dependency paediatric unit and so admissions to this unit were used as an initial validation dataset. The full admissions dataset includes other more general admissions and was restricted to the time span March 2011 - December 2012 (selected to be after the trial took place to avoid overlap) to provide a second validation dataset. Table 2.6.1 below compares admission characteristics between the three datasets.

Table 2.6.1: Admission characteristics for children in the different datasets.

	FEAST (2009-2011)	KEMRI (2002-2012)	Kilifi admissions (2011-2012)
Variable **(in bold if included in clinical model)**			
Number with score	3121	5791	4976
Age (months) - median (IQR)	24 (13-38)	28 (12-55)	24 (10-53)
Gender (Female, %)	1443 (46%)	2609 (45%)	2097 (42%)
Weight (kg) - median (IQR)	**10 (9-13)**	**10 (7-14)**	**10 (7-14)**
Conscious level - prostrate	**1918 (61%)**	**1379 (24%)**	**1012 (20%)**
- coma	**463 (15%)**	**1780 (31%)**	**192 (4%)**
Temperature (°C) - median (IQR)	**38.2 (37.3-39)**	**37.6 (36.8-38.6)**	**37.6 (36.8-38.5)**
History of fever (%)	3106 (99%)	4422 (76%)	3492 (70%)
Heart rate (beats per min) – median (IQR)	**169 (153-183)**	**149 (127-169)**	**144 (124-163)**
Weak pulse (%)	**660 (21%)**	**739 (13%)**	**91 (2%)**
Capillary refill time (s) – median (IQR)	**2 (1-3)**	**2 (1-2)**	**1 (1-2)**
Temperature gradient (%)	1847 (59%)	1278 (22%)	238 (5%)
Respiratory rate (breaths per min) - median (IQR)	58 (48-67)	41 (32-55)	38 (32-50)
Respiratory distress (%)	**2582 (83%)**	**2013 (35%)**	**1355 (27%)**
Deep breathing (%)	**2019 (65%)**	**1663 (29%)**	**358 (7%)**
Indrawing (%)	2127 (68%)	1490 (26%)	1174 (24%)
Crackles (%)	**692 (22%)**	**785 (14%)**	**624 (13%)**
Cough (%)	2253 (72%)	2134 (37%)	2050 (41%)
Severe pallor (%)	**1586 (51%)**	**2343 (40%)**	**1282 (26%)**
Convulsions (%)	452 (15%)	2398 (41%)	1003 (20%)
Decreased skin turgor (%)	187 (6%)	700 (12%)	323 (7%)
Vomiting (%)	1601 (51%)	1901 (33%)	1530 (31%)

Bold = variables in prognostic model

The children in the KEMRI high dependency unit (HDU) ward were a similar weight, age, and gender as those in the FEAST trial. There were more children in the KEMRI ward data in a coma but fewer were prostrate, more had convulsions (as coma and complex convulsions were admission criteria for HDU admission) but fewer had severe pallor, crackles, respiratory distress or deep breathing. The Kilifi general paediatric admission data show that these children were not as critically ill, as their median temperature was lower, their heart rate was lower, and there were low numbers of children in a coma or prostrate on admission.

2.6.1 Preparing the Kilifi dataset and assumptions made

To avoid overlap between the children enrolled into the FEAST trial from the Kilifi site used in the developmental data set (N=215) and the children admitted to the KEMRI ward and recorded in the Kilifi admissions dataset, these have been identified via their hospital number and removed from the validation dataset.

To enable calculation of the score on the data from Kilifi, some assumptions needed to be made regarding variables and the way they had been recorded. Conscious level was recorded both as one of five levels (unconscious, prostrate, lethargic, agitated and normal) as well as being recorded using the BCS. The BCS was used for validation as the categories lethargic and agitated were not easy to fit into FEAST definitions. As shown previously in Table 2.5.2, a BCS≤2 was used as a definition of coma and BCS≤4 was used as a definition of prostrate based on the literature [108].

Time of death was rarely recorded in the Kilifi dataset although the date of death and date of admission were well recorded. Death within 48 hours was defined as within two calendar days from admission to the hospital which is likely to slightly over-estimate the number of deaths within the hour defined time frame. Children that absconded before this time were considered alive in these analyses. For validation of Berkley's score, immediate deaths were defined as those that occurred on the same day as admission to hospital. Berkley's early death score was calculated on mortality by 2 calendar days but not the same day as admission. Late death was defined as strictly greater than 2 days after admission. Immediate deaths were not included in the early death analysis, immediate and early deaths were not included in the late death analysis as in the original publication [47].

For the validation dataset of just the KEMRI ward, the date of admission to the hospital was used as the baseline, as treatment will have started at that point prior to admission to the

KEMRI ward. Of those admitted to the ward, 72% were admitted on the same day as they were admitted to hospital, 14% were admitted on the next day and the median time from admission to hospital to the KEMRI ward if not the same day was 2 days (IQR 1-5 days). The dataset included all those admitted to the KEMRI ward from 1st Jan 2002 – 31st Dec 2012 not enrolled in the FEAST trial.

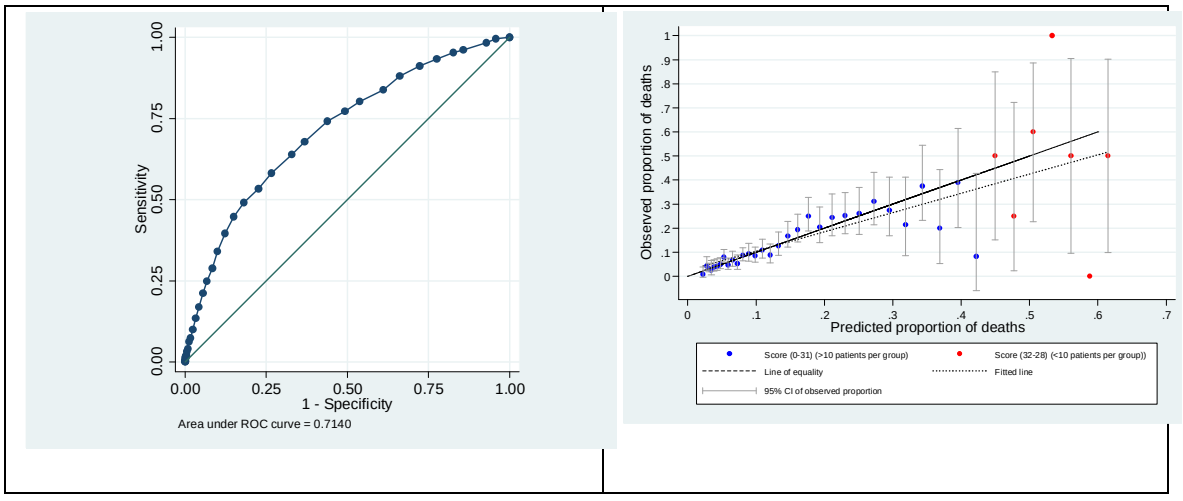
In the validation of the score on the wider admissions dataset, data from those that were admitted to the KEMRI was included (1019/4976 (20%)). This validation dataset used data from 1st March 2011 to 31st December 2012 as this was the most recent data and thus the most complete, and excludes the period when the FEAST trial was enrolling. Respiratory distress was not recorded in the admissions dataset and so 'difficulty breathing' was used in its place which is synonymous to respiratory distress.

2.6.2 Results of validation of clinical score

2.6.2.1 Validation on KEMRI ward dataset

The clinical bedside score used on the KEMRI ward validation dataset gave an AUROC of 0.71 (95% CI 0.69-0.74) assessed on 5791 children admitted to the KEMRI ward with 543 deaths (9%). This was presented in Figure 2.6.1 below.

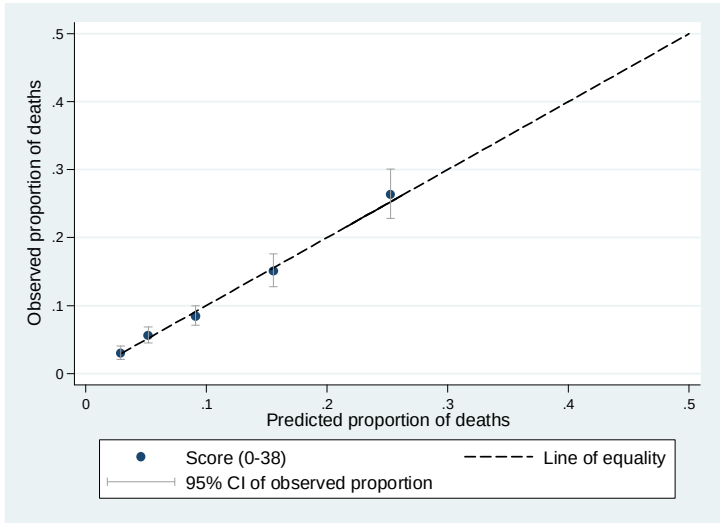
Figure 2.6.1: AUROC and calibration plot for the external validation of the clinical score in the KEMRI ward.



The discriminatory ability of the score was only fair, with AUROC of 0.71. The Hosmer-Lemeshow test was also carried out on a logistic regression model with the same variables as the Cox regression model; with p=0.49 it showed reasonable calibration of the model in the

KEMRI ward data. The calibration plot shows good calibration up to a score value of approximately 20 but for scores >20 the data was very sparse and observed probabilities per unit score value varied between 0 and 100%, with wide 95% confidence intervals due to the small numbers in each category (in the plot all scores above 32 (in red) have less than 10 patients). The score was divided into 5 equally spaced categories (0-5, 6-10, 11-15, 16-20, 21+) to consider the calibration further and address the issue of small numbers within each category. This categorisation gave better calibration as presented in the figure below. As the KEMRI ward dataset covers a long period of time, to make this more comparable to the other datasets, the AUROC was also calculated just for patients admitted from the 1st March 2011 onwards and showed a slightly better discriminatory ability at 0.77 (95% CI 0.73-0.82) (from 1019 patients) compared to those admitted prior to this date with an AUROC of 0.70 (95% CI 0.68-0.73) (from 4772 patients).

Figure 2.6.2: Calibration plot using five categories of the clinical score in the KEMRI ward.

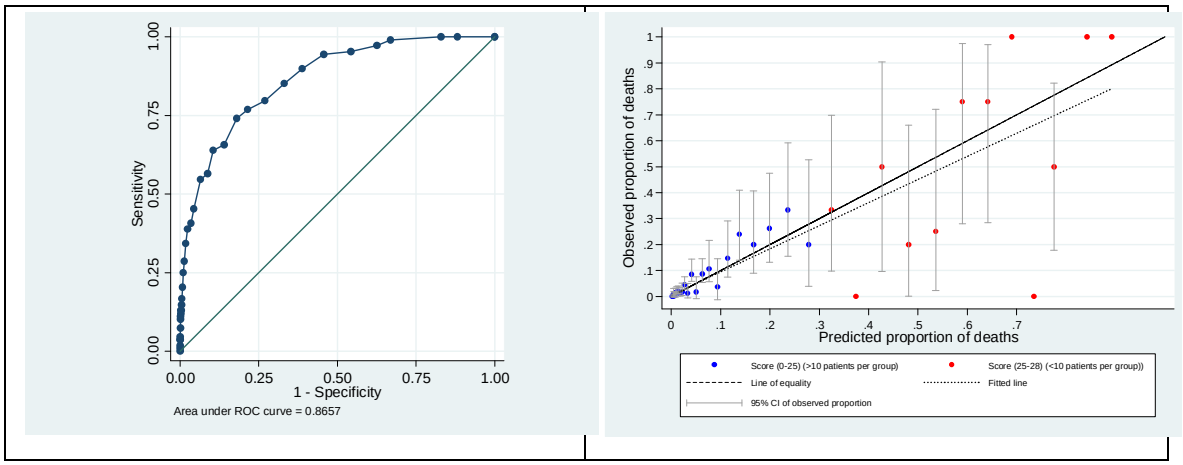


Score	0-5	6-10	11-15	16-20	>20
Observed deaths (%)	36/1179 (3%)	88/1485 (6%)	129/1527 (8%)	133/881 (15%)	157/596 (26%)

2.6.2.2 Validation on Kilifi admissions dataset (2011-2012)

The clinical bedside score was also validated on the admissions dataset between March 2011 and December 2012. This gave an AUROC of 0.87 (95% CI 0.83-0.90) assessed on 4976 children with 108 (2%) deaths (Figure 2.6.3 below). The Hosmer-Lemeshow test for goodness-of-fit gave p=0.19 showing that the calibration was reasonable but there was a similar pattern in the calibration plot compared to the KEMRI ward data where the proportion of deaths at high scores varied considerably, predominantly due to low numbers.

Figure 2.6.3: AUROC and calibration plot for the external validation of the clinical score in the Kilifi admissions dataset.



The AUROC plot showed that very low values of the score discriminated survivors well with very low death risk. For example, for scores between 0 and 6 the maximum observed mortality was 0.9%. High scores above 30 were also discriminatory for those at high risk though there were less than 16 children with these scores.

2.6.2.3 Validation of the FEAST PET score

The FEAST PET score was also validated on the Kilifi data and gave very similar discrimination to the full clinical score, with an AUROC of 0.77 (95% CI 0.72-0.82) and Hosmer-Lemeshow p-value of p=0.30 indicating a good fit in the KEMRI ward (admitted from 2011-2012) (1,053 children, 98 (9%) deaths) and an AUROC of 0.86 (95%CI 0.82-0.89) with Hosmer-Lemeshow p-value of p=0.51 for the general admissions dataset (5,098 children, 117 (2%) deaths) (Figure 2.6.4). The calibration plots also show good calibration apart from when the proportion of deaths at high scores (≥7) varies (Figure 2.6.5), as there were low numbers of children in these categories. The numbers included in this analysis are slightly higher than that for the clinical score as there were fewer variables in the score and those with missing weight or deep breathing could thus be included.

Figure 2.6.4: Receiver operating characteristic curves for the FEAST PET score in A) the Kilifi KEMRI ward and B) the Kilifi general admissions dataset.

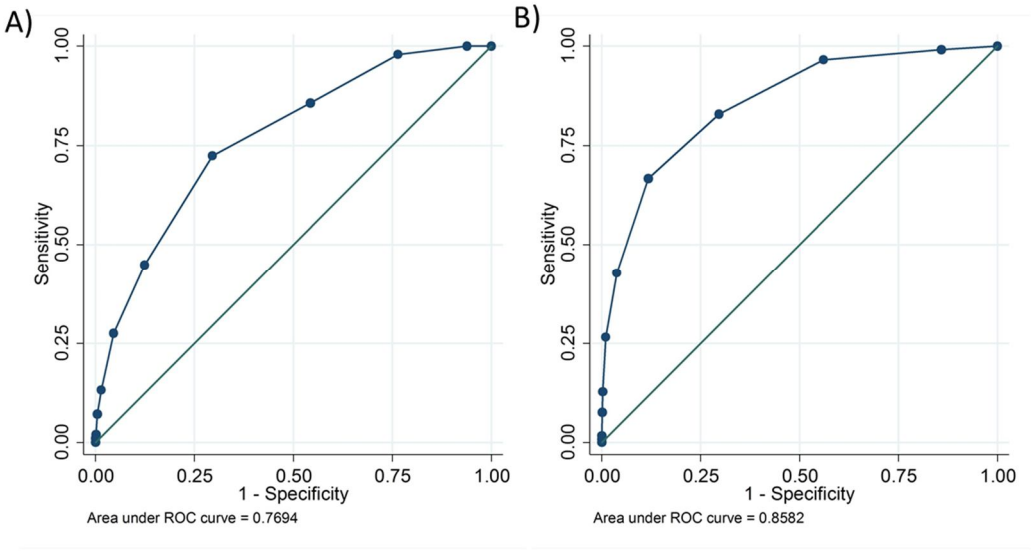
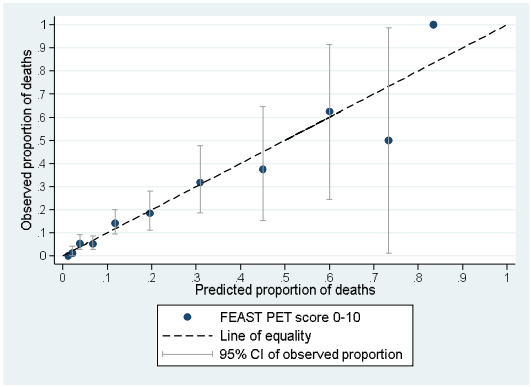


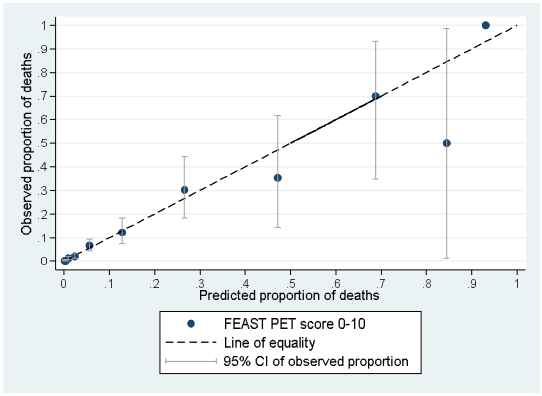
Figure 2.6.5: Calibration plots for the FEAST PET score in A) the Kilifi KEMRI ward and B) the Kilifi general admissions dataset

A)



Score	0	1	2	3	4	5	6	7	8	9	10
Deaths	0/59 (0%)	2/168 (1.2%)	12/224 (5.4%)	13/250 (5.2%)	27/191 (14.1%)	17/92 (18.5%)	14/44 (31.8%)	6/16 (37.5%)	5/8 (62.5%)	1/2 (50%)	1/1 (100%)

B)



Score	0	1	2	3	4	5	6	7	8	9	10
Deaths	1/706 (0.1%)	3/1486 (0.2%)	16/1332 (1.2%)	19/911 (3.1%)	28/423 (6.6%)	19/156 (12.2%)	16/53 (30.2%)	6/17 (35.3%)	7/10 (70%)	1/2 (50%)	1/1 (100%)

The FEAST PETaL score could not be validated on the Kilifi admissions data as the laboratory measures included in that score (lactate, BUN, and pH) were not recorded in the validation datasets.

2.6.3 Validation of other scores in the Kilifi datasets

The performance of the FEAST PET score for mortality by 48 hours was compared to other scores identified through the literature search (Figure 2.6.6) by testing the AUROC of each score in each validation dataset directly against the AUROC scores for FEAST PET for equality (Table 2.6.2). In the FEAST dataset there was no evidence of a difference between the AUROC for the FEAST PET score versus the AQUAMAT score overall (p=0.19) and in children with malaria only (p=0.65); however, the FEAST PET score was significantly better than PRISM III (p=0.02), and Berkley's scores for immediate (p=0.002) and early death (p=0.04). In the Kilifi validation datasets (KEMRI/general) there was no evidence of a difference between the AUROC for the FEAST PET score versus Berkley's scores for immediate (p=0.34/0.82) and early (p=0.63/0.47) death, and the FEAST PET score was significantly better than PRISM III (p=0.003/<0.001) and the AQUMAT scores (p<0.001/<0.001).

Figure 2.6.6: Discriminatory ability of different scores when applied to FEAST and Kilifi admissions datasets.

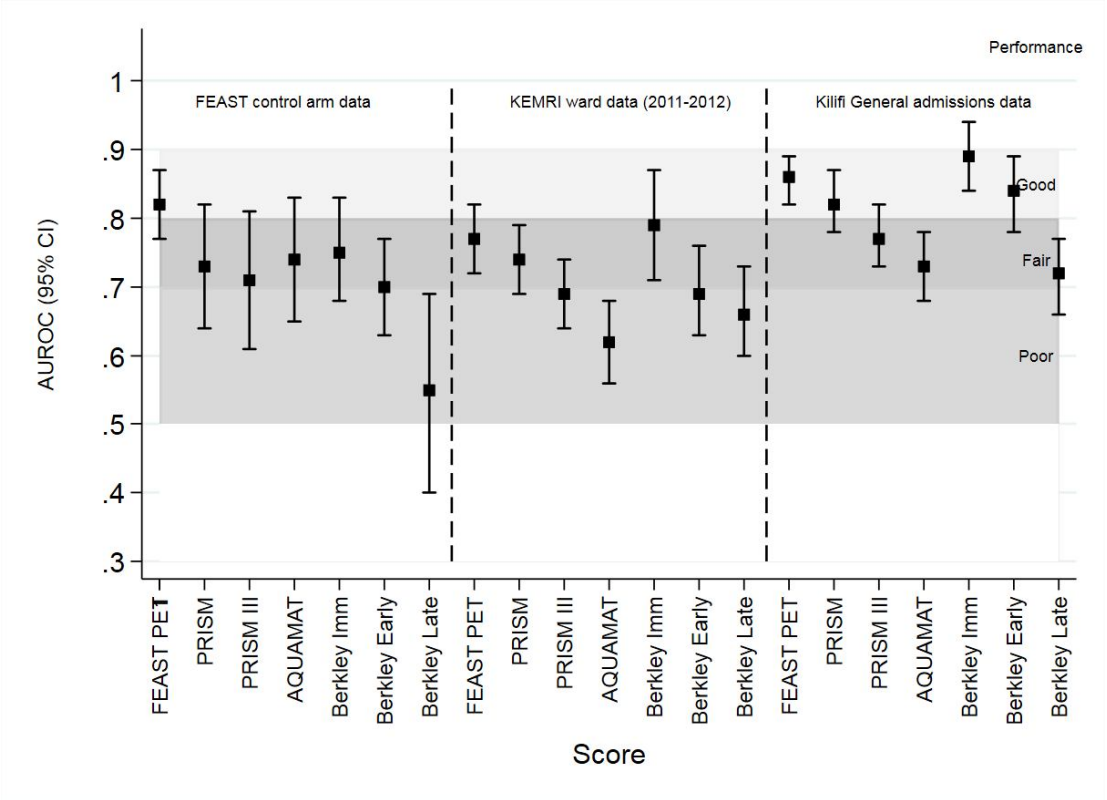


Table 2.6.2: Discriminatory ability of different scores when applied to data from FEAST and Kilifi.

		FEAST (data from the control arm only) (n=1044)			Kilifi KEMRI Ward (n=1058)			Kilifi General admissions (n=5107)		
Score	Variables included	Number with score (% died)	AUROC (95% CI)	Hosmer-Lemeshow test	Number with score (% died)	AUROC (95% CI)	Hosmer-Lemeshow test	Number with score (% died)	AUROC (95% CI)	Hosmer-Lemeshow test
FEAST PET score	**Axillary temperature, heart rate, capillary refill time, conscious level, deep breathing, respiratory distress, lung crepitations, weak pulse**	**1024 (7%)**	**0.82 (0.77-0.87)**	**p=0.56**	**1053 (9%)**	**0.77 (0.72-0.82)**	**p=0.30**	**5098 (2%)**	**0.86 (0.82-0.89)**	**P=0.50**
PRISM[a]	Heart rate, respiratory rate, conscious level, systolic blood pressure, potassium, glucose, pupillary reflexes, bicarbonate	620 (6%)	0.74 (0.65-0.83)	p=0.07	1054 (9%)	0.74 (0.69-0.79)	p=0.001	5095 (2%)	0.82 (0.78-0.87)	P=0.001
PRISM III[b]	Heart rate, temperature, conscious level, systolic blood pressure, glucose, potassium, PCO2, pH, acidosis, pupillary reflexes	627 (6%)	0.71 (0.61-0.81)	p=0.26	1056 (9%)	0.69 (0.64-0.74)	p=0.10	5099 (2%)	0.77 (0.73-0.82)	P=0.01
AQUAMAT score[c] (overall)	Conscious level, chronic disease, convulsions, blood urea nitrogen (BUN), and base excess (BE).	648 (5%)	0.74 (0.65-0.83)	p=0.84	1011 (9%)	0.62 (0.56-0.68)	p=0.79	4964 (2%)	0.73 (0.68-0.78)	P=0.04

		FEAST (data from the control arm only) (n=1044)			Kilifi KEMRI Ward (n=1058)			Kilifi General admissions (n=5107)		
Score	Variables included	Number with score (% died)	AUROC (95% CI)	Hosmer-Lemeshow test	Score	Variables included	Number with score (% died)	AUROC (95% CI)	Hosmer-Lemeshow test	Score
AQUAMAT score [c] (malaria positive only)	Conscious level, chronic disease, convulsions, blood urea nitrogen (BUN), and base excess (BE).	360 (3%)	0.80 (0.68-0.93)	p=0.65	355 (6%)	0.54 (0.42-0.66)	p=0.83	781 (3%)	0.60 (0.49-0.72)	p=0.41
Berkley Immediate death score[c]	Anaemia, Jaundice, Indrawing, deep breathing, conscious level, convulsions/seizures, temperature	1007 (3%)	0.75 (0.68-0.83)	p=0.64	680 (4%)	0.79 (0.71-0.87)	p=0.47	3504 (1%)	0.89 (0.84-0.94)	p=0.15
Berkley Early death score	Jaundice, indrawing, conscious level, convulsions/seizures, wasting, kwashiorkor	1003 (4%)	0.70 (0.63-0.77)	p=0.02	1016 (9%)	0.69 (0.63-0.76)	p=0.76	5071 (2%)	0.84 (0.78-0.89)	p=0.08
Berkley Late death score	History > 7 days, conscious level, convulsions/seizures, temperature, wasting, kwashiorkor.	959 (1%)	0.55 (0.40-0.69)	p=0.35	664 (10%)	0.66 (0.60-0.73)	p=0.34	3472 (2%)	0.72 (0.66-0.77)	p=0.08

[a] Variables in the score but not measured: diastolic blood pressure, arterial oxygen tension, arterial carbon dioxide tension, pupillary reactions, prothrombin time: partial thromboplastin time ratio, total bilirubin, calcium. Underlined variables were available in the FEAST dataset but not in the Kilifi datasets. For consistency with other scores all deaths were included in analyses.

[b] Variables in the score but not measured: pupillary reflexes, pH, total CO_2, PCO_2, arterial PaO_3, creatinine, urea, white blood cells, prothrombin time, and platelets. Underlined variables were available in the FEAST dataset but not in the Kilifi datasets. For consistency with other scores all deaths were included in analyses.

[c] Underlined variables were available in the FEAST dataset but not in the Kilifi datasets.

2.6.4 Conclusions

The clinical bedside score only showed fair discriminatory performance in the KEMRI ward dataset which was seen to be the most similar setting to the FEAST trial with an AUROC of 0.71. When this was explored further by calendar time then the score performed better on data collected in more recent years (2011 onwards (AUROC 0.77)) which is a more comparable time period to the general Kilifi admissions data set. This suggests that the score may be clinically useful but it is still only fair performance for a risk score. Overall, the calibration was good for low values of the score but was variable for high scores, and was better when the clinical bedside score was grouped into categories.

The clinical bedside score did show good discriminatory performance in the general admissions dataset for 2011 and 2012, with an AUROC of 0.87. This validation dataset had a much lower 48h mortality of 2% compared to 9% in the KEMRI ward and 10% in the FEAST dataset, and children in this validation dataset were more heterogeneous than those admitted to the KEMRI ward.

Reducing the clinical bedside score to the FEAST PET score with 8 variables with a maximum score of 10 gave a more practical triage score and the discriminatory performance was very similar. The FEAST PET score gave an AUROC of 0.86 in the general admissions dataset and 0.77 in the KEMRI ward dataset from 2011 onwards. This was also shown to be similar to or better than the other risk scores validated in the FEAST and Kilifi datasets.

2.7 Discussion

This analysis of the FEAST trial data and the Kilifi admissions data has identified the prognostic indicators that identify children at greatest risk of mortality on arrival to African hospitals, and has enabled me to develop and externally validate a bedside clinical risk score. In addition, I have validated published risk scores that had been developed in a variety of settings, both in the FEAST data and the Kilifi datasets, and evaluated if other measures that are not so easily recorded at the bedside or need special equipment can be usefully added to a clinical score.

The analysis identified axillary temperature, conscious level, capillary refill time, lung crackles, deep breathing, respiratory distress, pallor, heart rate, weak pulse, and weight as prognostic

indicators that best identified children at greatest risk of mortality by 48 hours. Different levels of these clinical signs were then weighted to give a score that could be calculated by a clinician at the bedside. The score was externally validated and gave good discriminative ability in the Kilifi admissions dataset (0.87) and fair discriminative ability in the KEMRI ward data (0.77) using data from the same time period. A disadvantage of this score was that it had a scale of 0-42, and so this was reduced to 0-10 and named the FEAST PET score. The PET score kept 8/10 of the original clinical bedside score variables and dropped weight and deep breathing as they added the least predictive ability to the model. The FEAST PET score showed good discriminatory ability in the Kilifi admissions dataset (AUROC 0.86) and the KEMRI ward data using data from the same period (0.77). Oxygen saturation has been found a predictor of mortality in other studies [158, 159], but it was not a significant predictor in multivariable models in our dataset. The limited predictive ability of hypoxaemia was found despite the significance of lung crepitations in the PET score, which supports WHO recommendations of the value of this sign to reinforce diagnosis of pneumonia in children with severe breathing difficulties. Although crepitations could be considered a subjective sign, a sensitivity analysis showed it was important to include in the PET score, as excluding it worsened the discriminatory ability of the score (AUROC without crepitations 0.80 (0.75-0.86); p=0.04 in the control arm data).

The advantages of the FEAST PET score are that it is straightforward to use, built on a large amount of good quality data, and externally validated on a suitable dataset. The clinical signs identified are commonly known, easily and quickly measured at admission, and are prognostic of mortality by one clear time point (48 hours). This satisfies Moons *et al*'s [160] comment that "the application of prognostic models requires unambiguous definitions of predictors and outcomes, and reproducible measures using methods in clinical practice." The score is also not specific to defined disease aetiologies and includes large subgroups of children with sepsis and severe malaria, which is advantageous in these settings as it is often difficult to diagnose the underlying condition on admission. The data for the bedside score came from a clinical trial, which is a good study design for building prognostic models, and the trial was set up both within hospitals experienced with research and hospitals that had not conducted clinic trials before. Therefore, although generalisability of prognostic models built within clinical trial data can be limited, the pragmatic nature of the FEAST trial gives reassurance that the settings of the trial were not far from standard hospital care. The number of events in the dataset was the appropriate amount for the number of candidate predictors considered in the model building process, and the decisions made in the process followed published recommendations [139].

The FEAST PETaL score had a slightly increased AUROC in the control arm of the FEAST data with an AUROC of 0.86 (95% CI 0.82-0.90) although not significantly different from that of the PET score; it was not able to be validated in the Kilifi datasets. This shows that in settings where lactate, BUN and pH are recorded, that it would be useful to include them but that the FEAST PET score is sufficient for good discrimination where these tests are not possible.

Previously published scores, such as Berkley *et al*'s [47] and the AQUAMAT score [103] that are built from data in similar settings to the FEAST trial, have not been widely externally validated and have disadvantages that may have led to them not being implemented in clinical practice. Berkley *et al*'s score evaluated prognostic variables for death at three different time points (immediate (<4 hours after admission), early (4-48 hours after admission), and late (>48 hours after admission)) and created three scores. Clinical signs could be positive in one score and negative in another score which could be confusing and only the immediate score showed fair discriminative ability and calibration in the FEAST data (AUROC 0.74), and better discriminatory ability in the Kilifi datasets (0.79 in KEMRI ward (2011 onwards), 0.89 in general admissions). Although this score has a published validation on a different dataset to the development dataset, the two datasets were from the same hospital. Berkley *et al*'s immediate, early and late scores were also combined into one score (using jaundice, subcostal indrawing, prostration with/without seizures, altered consciousness with/without seizures and wasting) for validation in children with fever presenting to a hospital in Uganda [138]. It showed good discrimination in children with malaria (AUROC 0.92) and with a non-malaria febrile illness (AUROC 0.86). Wasting and kwashiorkor were included in Berkley *et al*'s scores but the NRI calculated on the multiple imputed data did not show MUAC to be a variable that improved discriminative ability of the score. This could be as there were not many children with MUAC <11.5cm included in the dataset as malnutrition was an exclusion criteria for the trial.

The AQUAMAT score was developed only in children with malaria, and uses information from i-STAT cartridge tests that are not commonly available outside of clinical trials in low-income settings. The score performed well in the FEAST control arm data malaria subgroup but poorly in the Kilifi datasets, possibly because the two measures from the i-STAT cartridge were not routinely available. The score also included a measure of chronic disease which was not clearly defined and would be decided by the clinician, and each measure within the score was given the same weight; for example, being in a coma was given the same weight as having convulsions. Conscious level was found to be a prognostic factor in all the validated risk scores (Berkley *et al*, AQUAMAT, PRISM and PRISM III) and most scores gave more weight to being prostrate or in a coma compared to other risk factors.

The FEAST PET score developed through these analyses has only a fair discriminatory power in the KEMRI ward validation dataset which was from the most similar setting to the FEAST trial. This may limit its uptake and use in clinical practice, as clinicians need to feel that a score is able to identify children at high risk alongside (or be an improvement on) their own ability to make the appropriate diagnosis. However, it is important to note that some of the children in the KEMRI Ward had already spent time on the general wards before deteriorating, therefore admission criteria may not be as helpful in identifying those at high risk within this group. The FEAST PET score included crackles which is often considered to have inter-observer variability, but hypoxia, which is a more reproducible sign of poor lung function was not found to be important to include. The FEAST PET score included clinical signs that cover different presentation syndromes and reflected well the population of children presenting to hospital in these settings. But the diverse nature of presentations could be a reason why the score was not able to perform better, as there were too many critical clinical signs (i.e several within each presentation) instead of a few key signs for one type of presentation that would be applicable to more children. This shows that it may be difficult to create a very good discriminatory score using bedside clinical measures in this population, with diverse presentations and a variety of underlying conditions. Despite this the FEAST PET score does discriminate well in the general admissions dataset, suggesting this would be the appropriate setting to explore implementing it in clinical practice. Improved triage has been shown to reduce mortality rates in these settings [10] and the FEAST PET score could help the introduction, or an improvement, of triage systems, encourage better examination of clinical signs by staff, and ensure consistent comparisons between patients by clinical staff compared to clinical opinion. This could facilitate the rapid prioritisation of care for, or closer monitoring of, the sickest children and thus improve outcomes.

3 Exploring mechanisms of action from bolus administration in the FEAST trial

3.1 Aims and Objectives

There has been much discussion following the FEAST trial results of mechanisms by which the bolus intervention was potentially doing harm to the children in the trial. One way to examine this is by considering if there was a physiological measure that explained any of the (harmful) effect of the bolus. Some of the potential physiological measures have been examined in Chapter 2 in order to build a prognostic model for mortality. Those analyses were restricted to clinical features at baseline in order to be generalisable to paediatric triage in emergency rooms and for cross-site severity comparisons in sub-Saharan Africa. They led to a prognostic score, the FEAST PET score, and an extended score which included laboratory parameters (FEAST PETAL score) [161]. In addition to the baseline assessments, FEAST children were systematically assessed (largely clinical assessments) at regular time intervals after randomisation, and so their relationship with mortality and the effect of the bolus can be examined over time. Furthermore, other bedside measures, expanding on those in Chapter 2, can be considered using the whole trial dataset to explore if there were any physiological measures that explained all or part of the bolus effect.

In this chapter physiological measures over time will be considered as potential surrogate outcomes or endpoints and, by analysing them in this way, more information may be found on how the bolus intervention influenced the primary outcome. Also, as exploratory analyses, the associations between the continuous measures and mortality, both at baseline and during admission, can be examined and modelled using the most appropriate functional form. Once the appropriate functional form has been identified, tests for heterogeneity of the effect of boluses across the values of the continuous measures can also be carried out. As above, my previous work has modelled a selection of these continuous measures, and other studies have examined some of these measures with cut-offs or binary categories, as described in Chapter 2, but this work will be able to use the true values of the measure, rather than reducing them to categories [162].

Surrogate outcomes are usually considered for trials of a long duration (over many years) or rare clinical outcomes [163], whereas the primary outcome of FEAST was mortality at 48 hours, with an overall mortality of 9.5%. The size of this trial was large (3170 children) and this sample size was driven by the binary endpoint and the expected modest relative difference

between the arms of 33% (as reported in the protocol). Nevertheless, if there was a continuous physiological measure that could be used as a surrogate endpoint then smaller future trials, especially in populations with much lower predicted mortality, which have been proposed but not yet designed, could nevertheless be adequately powered to detect a difference in such a surrogate endpoint. These studies could then be used to prioritise other interventions to take forward into larger trials in these populations.

3.2 Statistical Background

All the physiological measures that are associated with the true endpoint and are affected by the intervention, and are available post-baseline, can be considered as possible 'surrogates' i.e intermediate indicators in a disease progression process. Prentice *et al* define a surrogate variable to be one that captures the whole relationship between the treatment and the true endpoint, so that the test of null hypothesis of no treatment difference in the surrogate endpoint is also a valid test of the corresponding null hypothesis based on the true endpoint [163].

Prentice *et al* capture the notion of a potential surrogate response variable S(t) for the outcome H (a time-to-failure endpoint) being able to capture the dependence of H on treatment R as follows:

$$\lambda_H\{t; S(t), R\} = \lambda_H\{t; S(t)\} \quad (1)$$

where λ_H is the outcome failure hazard.

For a binary outcome (H), where S is the surrogate and binary treatment is R, this is [164]:

$$P(H=1|S, R)=P(H=1|S) \quad (2)$$

There is strong evidence against this criterion holding when a model for H adjusting for S still has a significant treatment effect, and also when there is an interaction between S and R [164].

Freedman *et al* acknowledge that the phrase "data were consistent with equation 2" i.e there was no interaction and that there was no evidence of a treatment effect after adjustment for the surrogate marker, is a weak statement for surrogacy. They also note that there would not be high precision in the validation of a surrogate where the unadjusted treatment effect is 2 standard errors or less away from the null. A more useful approach of estimating the proportion of the treatment effect explained (PTE) by the surrogate was proposed. The approach compares the estimate of treatment effect in an unadjusted model to the estimate of treatment effect in a model adjusted for the putative surrogate [164].

$$\text{PTE (\%)} = [1 - (\text{coefficient of adjusted estimate})/(\text{coefficient of unadjusted estimate})]\text{x}100 \quad (3)$$

This estimate of PTE is extended to survival analysis by Lin *et al* [165]. There are two models where R is a binary indicator for intervention vs control and S(t) is the time varying potential surrogate covariate measured at time t.

$$\lambda_1(t|R) = \lambda_{10}(t)\exp(\beta R) \quad (4)$$

$$\lambda_2(t|R,S) = \lambda_{20}(t)\exp[\gamma_0 R + \gamma_1 S(t)] \quad (5)$$

where $\lambda_{10}(t)$ and $\lambda_{20}(t)$ are unspecified baseline hazard functions and β, γ_0 and γ_1 are unknown regression coefficients, where β = net log hazard ratio for treatment effect, γ_0 = log hazard ratio for treatment effect adjusted for $S(t)$, and γ_1 = log hazard ratio for the effect of $S(t)$. Then the PTE can be described as:

$$\hat{p}_1 = 1 - \frac{\hat{\gamma}_0}{\hat{\beta}} \quad (6)$$

or

$$\hat{p}_2 = 1 - \frac{1 - \exp(\hat{\gamma}_0)}{1 - \exp(\hat{\beta})} \quad (7)$$

These are the two conventional estimates of the PTE used in the statistical literature with the second ($\hat{p}_2$) being a direct measure of the relative difference in risk reduction between unadjusted and adjusted analyses [166] as $(1 - \exp(\hat{\beta}))$ = estimated net relative reduction in risk for intervention vs control, and $(1 - \exp(\hat{\gamma}_0))$ = estimated adjusted relative reduction in risk for intervention vs control.

For a measure to be considered a suitable surrogate marker then I would expect to see a high PTE, with a perfect surrogate giving a PTE of 100%. A restriction of the PTE is that, following on from Freedman's definition, it is not a meaningful statistic when there is an interaction between R and S.

Also, the definition of PTE from Lin *et al* requires two models to hold true simultaneously, one giving the unadjusted estimate and one giving the adjusted estimate (equations (4) and (5)) [165]. This has been criticised in the literature as only one model can be the true model to describe the data, and this also presents statistical difficulties when calculating appropriate confidence intervals. It has also been criticised for not being a true proportion as the estimate can be outside of the range of [0,1] and for the authors not providing a cut-off value for a definition of an valid surrogate marker [167].

An additional caveat is also that PTE cannot, in general, have a causal interpretation. Although the association between the intervention and the true outcome, and the association between the intervention and the biomarker, do have a causal interpretation in clinical trials because of the randomisation, the association between the biomarker, and the outcome is not protected

by randomisation. For the PTE to have a causal interpretation it is necessary that there are no confounders between the biomarker and the outcome that are unaccounted for in equations (4) and (5). Even if all biomarker-outcome confounders were measured, the adjustment for these confounders may not be straightforward unless one can assume that the confounders are not affected by the intervention.

Various alternative ways to calculate the PTE and also alternative statistics describing the effectiveness of a surrogate marker have been proposed over the last 20 years. The key alternatives to the Lin *et al* PTE are summarised in Table 3.2.1 below.

Table 3.2.1: Summary of key papers proposing statistical methods to calculate how much of a treatment effect is explained by a surrogate marker.

Authors	Year	Journal	Number of citations*	Abbreviated name	Brief Description	Advantages	Disadvantages
Lin DY, Fleming TR and DeGruttola V [165]	1997	Statistics in Medicine	218	PTE (LFD method)	Based on proportional reduction in regression coefficient for treatment indicator due to inclusion of surrogate marker. Equations (4) and (5).	Described in Cox model, well known and intuitive.	Not a true proportion, and estimate and CI can fall outside range of [0,1]. Assumes two models hold true.
Cowles, MK [168]	2002	Statistics in Medicine	14		Based on a Bayesian approach with a Markov chain Monte Carlo based method for estimating the Bayesian posterior distribution of PTE conditioning on the truth of a single model.	Uses only one model.	Not well known, and Bayesian methods not as widely understood by clinicians.
Wang Y and Taylor JMG [169]	2002	Biometrics	39	PE	It is a quantity that looks to capture what the effect of the treatment would be if the values of the surrogate marker in the treatment group were distributed as those in the control group [170].	There are fewer assumptions about the data which allows greater flexibility in modelling.	Estimate and CI can fall outside range of [0,1].

Authors	Year	Journal	Number of citations*	Abbreviated name	Brief Description	Advantages	Disadvantages
Chen C, Wang H and Snapinn SM [166]	**2003**	**Statistics in Medicine**	**22**	**PTE (Chen method)**	**Based on well-known definition of proportion of treatment effect explained, but uses one model to estimate both unadjusted and adjusted coefficients of treatment effect.**	**Very intuitive and uses Lin definition of PTE, easy to understand and to calculate confidence intervals.**	**Estimate and CI can fall outside range of [0,1].**
Qu Y and Case M [171]	2007	Biometrics	14	PIG	Generalisation of path analysis to generalised linear models or time to event analyses using three hierarchical regression models.	Always bounded by [0,1].	Can be seen as combining individual level effects with trial-level effects [172].
Huang J and Huang B [173]	2010	Statistics in Biopharmaceutical Research	4		Based on a counterfactuals approach.	Interpretation of measure can be illustrated with Ordinal Dominance curves.	Logistic or probit models only.
Kobayashi F and Kuroki M [167]	2014	Statistics in Medicine	2	PCS	Based on the same principles as Wang and Taylor's PE measure and looks to decompose the treatment effect into parts captured and not-captured by the proposed measure.	Considers impact of dependence between treatment and surrogate outcome, has suitable cut-offs from inflection points.	Not well known, uses Half-range mode method with bootstrap for Cis.

*Web of Science (June 2017)

The method I have decided to use is that of Chen *et al* as it is the most straightforward and intuitive method with a measure well-recognised by clinicians [166]. It also overcomes one of the arguments against the Lin *et al* PTE measure, as Chen *et al* approximate the unadjusted and adjusted estimate of effect from one model. The method uses the model that includes the surrogate marker of interest and recovers the unadjusted estimate using the covariance matrix of the coefficients of the model. The unadjusted estimate and adjusted estimate are then combined in the same way as Lin *et al*'s definition above to give PTE (equations (6) and (7)).

Chen *et al* start with the assumption that equation (5) is the true model and that the asymptomatic distribution of $(\hat{\gamma}_0, \hat{\gamma}_1)$ is that of a bivariate normal vector with a covariance matrix estimated to be

$$\hat{V}(\hat{\gamma}_0, \hat{\gamma}_1) = \begin{pmatrix} \hat{\sigma}_0^2 & \hat{\rho}_{01}\hat{\sigma}_0\hat{\sigma}_1 \\ \hat{\rho}_{01}\hat{\sigma}_0\hat{\sigma}_1 & \hat{\sigma}_1^2 \end{pmatrix} \quad (8)$$

Motivated by linear models they then rewrite equation (5) as

$$\lambda(t|R,S) = \lambda_{10}(t)\exp[\gamma_0 R + c\gamma_1 R - c\gamma_1 R + \gamma_1 S(t)] \quad (9)$$

which can also be written as

$$\lambda(t|R,S) = \lambda_{10}(t)\exp[(\gamma_0 + c\gamma_1)R + \gamma_1(S(t) - cR)] \quad (10)$$

where c is a parameter for estimating the unadjusted treatment effect though $\gamma_0 + c\gamma_1$. The authors then estimate c through information separation between R and $S(t) - cR$ by forcing $cov(\hat{\gamma}_0 + c\hat{\gamma}_1, \hat{\gamma}_1) = 0$. Thus c is estimated to be

$$\hat{c} = -\hat{\rho}_{01}\hat{\sigma}_0/\hat{\sigma}_1 \quad (11)$$

Then as $\hat{\gamma}_0 + c\hat{\gamma}_1$ and $\hat{\gamma}_1$ are assumed to be uncorrelated, the information in R with respect to clinical outcomes would be largely separated from that in $S(t) - cR$ which leads to the overall treatment effect before covariate adjustment being largely retained in $\hat{\gamma}_0 + c\hat{\gamma}_1$.

This leads to the following estimates of PTE:

$$\hat{p}_1 = 1 - \frac{\hat{\gamma}_0}{\hat{\gamma}_0 + \hat{c}\hat{\gamma}_1} \quad (12)$$

And

$$\hat{p}_2 = 1 - \frac{1 - \exp(\hat{\gamma}_0)}{1 - \exp(\hat{\gamma}_0 + \hat{c}\hat{\gamma}_1)} \quad (13)$$

The confidence intervals for this measure can then be calculated using the Monte Carlo bootstrapping technique using a large number of independent samples from the exact (asymptotically normal) joint distribution of $(\hat{\gamma}_0, \hat{\gamma}_1)$.

This method of calculating the PTE can be used to estimate the PTE by the time-updated absolute values adjusted for the baseline value or by the change from baseline also adjusted for the baseline value (when $t_0 = 0$).

These two models are:

$$\lambda_1(t|R,U) = \lambda_{10}(t)\exp[\gamma_0 R + \beta f(U(t_0)) + \mu g(U(t))] \quad (14)$$

$$\lambda_2(t|R,U) = \lambda_{20}(t)\exp[\gamma_1 R + \beta^* h(U(t_0)) + \theta k(U(t) - U(t_0))] \quad (15)$$

where $\lambda_{10}(t)$ and $\lambda_{20}(t)$ are unspecified baseline hazard functions, $\gamma_0, \gamma_1, \beta, \mu, \beta^*$ and θ are unknown regression coefficients, $f(\ \ ), g(\ \ ), h(\ \ )$, and $k(\ \ )$ are fractional polynomial (fp) functions, R is a binary indicator of intervention vs control, and $U(t)$ is the time varying covariate measured at time t. Where all the fp functions are linear then the two models are equivalent and $\beta = \beta^* - \theta$ and $\theta = \mu$. The first model (14) is then simpler in demonstrating the impact of the baseline measure as it does not need to be calculated from two separate coefficients. But if the fp functions are not linear then the two models may not be equivalent, the relationship between the coefficients as above may not hold and one model representation may have a better fit than the other. This may then affect the PTE estimate, as if one model has a better fit then it may be more accurately estimating the proportion of the effect explained by the surrogate marker.

A paper by Li, Qu and Kulkarni compared using the change or actual value of a biomarker in calculating Proportion of Information Gain (PIG) as a measure of PTE [174]. They compared three models of which two were not adjusted for the baseline value: Model A – just the time-updated value, B – just the change from baseline and C – the change from baseline with adjustment for the baseline value. They concluded that C was the most appropriate model but also that using the time-updated value and adjusting for the baseline was as valid due to the equivalence shown above when the same function is used in each model.

3.3 Methods

3.3.1 Data

Respiratory rate, heart rate, systolic blood pressure, oxygen saturation, glucose and temperature were recorded at baseline, 1 hour, 4 hours, 8 hours, 24 hours and 48 hours. Haemoglobin (Hemocue, Angleholm) and lactate (Lactate Pro® (Arkray KDK, Kyoto, Japan) were measured by hand-held analysers at baseline, 8 hours and 24 hours. BUN, base excess, PH, sodium, chloride and potassium were measured with a handheld blood analyzer and cartridge (i-STAT, Abbott Laboratories) at baseline and 24 hours. As only 30% of children have i-STAT measures at both time points and most deaths occurred before 24 hours I decided to only use baseline values for the i-STAT measures in analyses in this chapter. The clinical measures and when they were measured are described in the table below.

Table 3.3.1: Clinical measures and assessment time points included in models.

Clinical measure	Baseline	1 hour	4 hours	8 hours	24 hours	48 hours
Respiratory rate	X	X	X	X	X	X
Heart rate	X	X	X	X	X	X
Oxygen saturation	X	X	X	X	X	X
Systolic blood pressure (SBP)	X	X	X	X	X	X
Glucose	X	X	X	X	X	X
Temperature	X	X	X	X	X	X
Haemoglobin	X			X	X	
Lactate	X			X	X	
BUN*	X					
Base excess*	X					
Sodium*	X					
PH*	X					
Chloride*	X					
Potassium*	X					

* Denotes measure from i-STAT cartridge

All measures were truncated at the 1st and 99th percentile to ensure that outliers did not have undue influence on the selection of an appropriate fractional polynomial.

There was a small amount of missing data at time points after baseline and the most straightforward way to maximise the available information was to carry forward values to the next time point whilst the child remained alive, thus imputing missing values after baseline with a single 'hard' imputation. The number of values for each measure that were carried forward was very low (Table 3.3.2).

Table 3.3.2: Number of observations carried forward in time-updated analyses.

Measure	Number carried forward (% of all values)
Respiratory rate	116/15071 (0.8%)
Temperature	86/15076 (0.6%)
Oxygen saturation	133/14925 (0.9%)
Heart rate	83/15084 (0.6%)
SBP	146/14999 (1.0%)
Glucose	270/14460 (1.8%)
Haemoglobin*	208/8765 (2.4%)
Lactate*	180/8614 (2.1%)

*There were fewer time points in these measures.

3.3.2 Statistical modelling

I estimated the effect of continuous baseline measures as well as continuous time-updated values adjusted for baseline measures on mortality through 48 hours. Cox proportional hazards regression was used to estimate hazard ratios for mortality up to 48 hours, with fractional polynomials for continuous measures where it was found to be appropriate. In Chapter 2 fractional polynomials were used to model continuous measures that had complete data (respiratory rate, heart rate, temperature, age and weight) as part of a prognostic model. But other baseline continuous measures were not modelled with fractional polynomials in that chapter, either as they had a higher percentage of missing data (oxygen saturation (4% missing), SBP (2% missing)) or were not as easily measured at baseline (haemoglobin (3% missing), lactate (5% missing) and glucose (6% missing)). They were instead analysed to see by how much they improved the clinical prognostic score using the Net Reclassification Index. In this chapter, using data from all arms, baseline analyses were performed using complete cases,

and time-updated analyses were performed on baseline complete cases data with missing post-baseline data whilst a child was alive imputed with the closest prior observed value (Table 3.3.2). A sensitivity analysis used multiple imputation to explore the impact of missing baseline measures on baseline models.

The multivariable fractional polynomial (mfp) process in Stata was used to choose the best function for the measures (with an alpha of 0.05) [175]. It was restricted to choose only powers from -2 to 2 as introducing a cubic term can sometimes lead to overfitting to values at the extremes. The fit of the fractional polynomial models and functions chosen were examined using residual plots. Then the functional form of the measure was plotted using the predicted hazard ratio relative to a child in the no bolus arm, with the median baseline value as reference.

Time-updated models estimated the treatment effect adjusted for each baseline measure and its time-updated values, acknowledging the caveat noted in the statistical background section above regarding using change or absolute value. Even if there was a better fit, using change from baseline, for one variable in one or more groups there may not be for another variable or group. To ensure consistency in modelling across the different variables and groups I used absolute values adjusted for baseline values for all analyses with time-updated models.

All models were tested for interactions between the time-updated value and the bolus effect in the time-updated models, and baseline measure and bolus effect in the baseline models using the likelihood ratio test. Any evidence ($p \leq 0.05$) of interactions was investigated further and the hazard ratio over the values of the time-updated or baseline measure was plotted.

The estimates of the hazard ratios from the time-updated models were used to calculate the proportion of treatment effect explained (PTE) using the Chen *et al* method (equation (12)[166]). These were then plotted to identify which markers explained the most of the treatment effect. The unadjusted hazard ratios for each measure were also plotted along with their corresponding adjusted hazard ratios to show the impact of the adjustment for each variable. Confidence intervals for the PTEs were calculated by bootstrapping 10,000 samples from a joint normal distribution with means equal to the coefficients from the adjusted Cox regression model and the covariance matrix from the same model. The calculated PTEs were also compared to a proportion of treatment effect measure estimated using two models (one unadjusted and the other adjusted for the time-updated and baseline measure) which I will refer to as the LFD (Lin, Fleming and DeGruttola) method (equation (4)).

A multivariable Cox model was used to estimate PTE, including all continuous time-updated variables and adjusted for their baseline values (rather than considering each measure one at a time). As this model used information from all time points, lactate and haemoglobin values from baseline were carried forward to 1 and 4 hours. The PTE from the multivariable model was then calculated in subgroups defined by malaria status and baseline base excess, and by oxygen saturation, as these were found to have had an interaction between baseline values and receipt of bolus.

Models were also fitted for four pre-defined subgroups of interest: children with malaria, children without malaria (who were likely to have conditions such as sepsis), children who were 2 years old or younger, and children over 2 years old. Malaria was defined following the FEAST Statistical Analysis Plan (SAP) definition as, in (hierarchical) order of preference:

positive for malaria parasitaemia on a quality-controlled double read microscopy slide; or if this result was unavailable then

positive result on the Plasmodium falciparum lactate dehydrogenase Rapid Diagnostic Test (RDT) (Optimal, Diamed, Cressier, Switzerland) which detects current or recently cleared (~ 48 hours) infections; or if this was unavailable then

positive result on any microscopy slide for that patient at admission.

There was a similar distribution of malaria positive children in both age groups (Table 3.3.3). The reason for considering these different subgroups is that underlying mechanisms may be very different, and the trial may have been underpowered to detect this through heterogeneity tests.

Table 3.3.3: The distribution of malaria and non-malaria by age subgroup.

	Age ≤ 2 years n (column %) *(row %)*	**Age > 2 years** n (column %) *(row %)*	**Total** n (column %) *(row %)*
Non malaria	703 (45%) *(53%)*	627 (40%) *(47%)*	1330 (43%) *(100%)*
Malaria	860 (55%) *(48%)*	935 (60%) *(52%)*	1795 (57%) *(100%)*
Total	1563 (100%) *(50%)*	1562 (100%) *(50%)*	3125 (100%) *(100%)*

Analyses also used types/causes of death called terminal clinical events (TCEs) as defined by the Endpoint Review Committee (ERC) [157]. TCEs were defined as:

1) *Cardiogenic/cardiovascular collapse:* signs of shock at the point of death - severe tachycardia or bradycardia plus one of prolonged capillary refill time >2 seconds, cold peripheries or low SBP (undefined), or where hypoxia was present but circulatory failure was deemed to be the primary problem.
2) *Respiratory:* Ongoing or development of hypoxia (PaO2 <90%) with chest signs (crepitations or indrawing); or the attending clinician had assigned the primary cause of death as pneumonia and/or possible pulmonary oedema.
3) *Neurological*: Possible raised intracranial pressure (high SBP or relative bradycardia) or severely reduced conscious level (Blantyre Coma Score $\leq$2), focal neurological signs, abnormal pupil response to light or posturing at the point of death.
4) *Unknown/Other:* Death was unwitnessed or it was an unknown or other cause of death.

Children could have one or more of these TCEs but the predominant TCE was assigned as cardiogenic for cardiogenic and neurological TCEs, and neurological for neurological and respiratory TCEs (largely terminal lung aspiration in a comatose child).

3.3.3 Sensitivity analyses

3.3.3.1 Missing data

The measures at baseline from the i-STAT cartridge were missing for up to 38% of children (Table 3.3.4). The reasons for this missingness were mostly due to the technical difficulties in performing the tests, including air bubbles introduced when loading the cartridge, failure of the machine to operate at high temperatures and, finally, short expiry dates leading to 'stock outs' during international shipment of next batches and thus unavailability of the cartridges at some sites for some time periods. Additionally, a further reason for missingness may have also been connected with the blood samples used in the cartridge, as if the blood sample was very haemolysed (for example if the child had high parasitaemia) then it would be difficult to get results. Missingness was associated with malaria status (chi^2 test $p<0.001$), with greater missingness in those with malaria.

Multiple imputation was used to explore the impact of missing baseline values on the selection of the appropriate fractional polynomial modelling risk of mortality at different levels of the baseline measure. The data was assumed to be missing at random (MAR) and multivariable imputation by chained equations was used to produce 25 complete datasets. Data for incomplete baseline variables were imputed using four methods depending on the distribution of the variable with missing data (Table 3.3.4): logistic regression, ordered logistic regression, normal regression, and predictive mean matching (PMM) (using the closest 10 individuals). Skewed continuous variables were log transformed towards normality but where the transformation was not fully satisfactory predictive mean matching was also used. To enable interaction tests to be performed the imputation was done with two separate models – one for children in the bolus arms and one for those in the no bolus arm. Also, as an interaction was found between base excess and bolus in the malaria group only in the complete case analysis, an interaction between malaria status and base excess was included in the imputation model. Variables used in the models are presented in Table 3.3.4 below.

Table 3.3.4: Baseline variables included in the multiple imputation models

Baseline variable	% missing	Log transformed?	Method
Systolic Blood Pressure	2%	Yes	PMM
Haemoglobin	3%	No	PMM
Oxygen Saturation	4%	Yes	PMM
Lactate	5%	Yes	PMM
Glucose	6%	Yes	PMM
Sodium	33%	Yes	PMM
Potassium	35%	Yes	PMM
Chloride	34%	Yes	PMM
PH	34%	Yes	PMM
BUN	38%	Yes	PMM
Base excess	34%	Yes	PMM
MUAC	6%	Yes	Normal regression
Heart rate	<1%	No	Normal regression
Respiratory rate	1%	No	Normal regression
Temperature	<1%	No	Normal regression
Weight	<1%	Yes	Normal regression
Fits at admission	1%	No	Logistic regression
Respiratory distress	<1%	No	Logistic regression
Malaria	1%	No	Logistic regression
Capillary refill time	<1%	No	Ordered logistic regression
Conscious level	<1%	No	Ordered logistic regression
Bolus/No bolus arm	Complete		
Gender	Complete		
Age (months)	Complete		
Site	Complete		
FEAST A or B	Complete		
Survival time estimate (Nelson-Aalen)	Complete		
Death by 48 hrs (yes/no)	Complete		
Weak pulse (yes/no)	Complete		

Cox regression was used with Rubin's rules [156] to evaluate fractional polynomials for each continuous measure, testing between possible fractional polynomials with the 'stack' method [176]. 'Stacked data' treats all the imputed datasets as one dataset, with all observations weighted by $w_c = (1 - f_c)/M$ where f_c is the fraction of missing data for the c-th covariate and M is the number of imputed datasets [176]. The alternative method suggests using Wald tests based on Rubin's rules to test between possible fractional polynomials [177]. The functions selected were then compared with the complete case analyses by plotting predicted

mortality risk across a range of values for that measure. Where variables had <1% missing data the stacked method selected fractional polynomials that were the same as the functions selected in the complete case analysis, whereas the Wald method selected functions that when plotted were similar but differed in the tails of the distributions. Thus the stacked method was used.

3.3.3.2 Definition of malaria

Malaria was defined in the analysis plan for the FEAST trial using results from both the rapid diagnostic test and microscopy as above (page 154). Children who live in a malaria endemic areas frequently have asymptomatic parasitaemia, thus may be positive on one or both of these tests (incidental parasitaemia) when they present to hospital with severe illness even though the illness may not be caused by malaria. Distinguishing which children have 'true' severe malaria remains challenging at a clinical level since many of the clinical features for bacterial infections and severe malaria are very similar. Thus, from a management perspective, it is challenging to separate out children with severe malaria that will respond to anti-malarials or malaria-specific supportive therapy from children who have incidental malaria parasitaemia. The original analysis plan definition of malaria, in particular, may not distinguish the two clinically relevant groups and further refinements would be useful.

Since malaria parasitaemia may be incidental and measures only the circulating parasite load, a more accurate measure of current malaria clinical status is *P.falciparum* histidine-rich protein 2 (P*f*HRP2). P*f*HRP2 measures the total parasite burden (biomass) in a patient by counting mature parasites that have released proteins in the last 48 hours compared to parasites on a blood film that are from circulating parasites that have not yet sequestered in the vital organs [178]. Studies involving Asian adults [178, 179] and African children [41, 180] have demonstrated that plasma PfHRP2 concentration of > 1000 ng/mL, in contrast with the peripheral blood parasite density, was strongly correlated with disease severity, end-organ damage and mortality. A re-analysis of a trial comparing anti-malarial treatments in children (AQUAMAT [28]) demonstrated that only the children in the middle and upper tertiles of P*f*HRP2 had a substantial benefit from artesunate (over quinine) [41]. For children in the lowest tertile there was no difference in mortality between the two groups implying those with low levels of P*f*HRP2 had co-incidental parasitaemia with another cause of severe infection [181].

The FEAST trial stored samples from admission, which were batch-tested retrospectively for P*f*HRP2 protein and the results used to define a subgroup of severe malaria to carry out sensitivity analyses. The refinement to the clinical definition of severe malaria included any child having P*f*HRP2>1000ng/ml [41]. Furthermore, total parasite burden (P_{tot}) was calculated using the equation below [182]

$$P_{tot} = 7.3\ X\ PfHRP2g/L\ X\ (1 - Haemocrit)\ X\ body\ weight\ (kg) X\ 10^{13}$$

Any interactions with bolus and malaria previously found in baseline or time-updated models were then explored in children with severe malaria as defined above. The interactions were also investigated with a continuous measure of malaria - total parasite burden - rather than malaria as a binary variable.

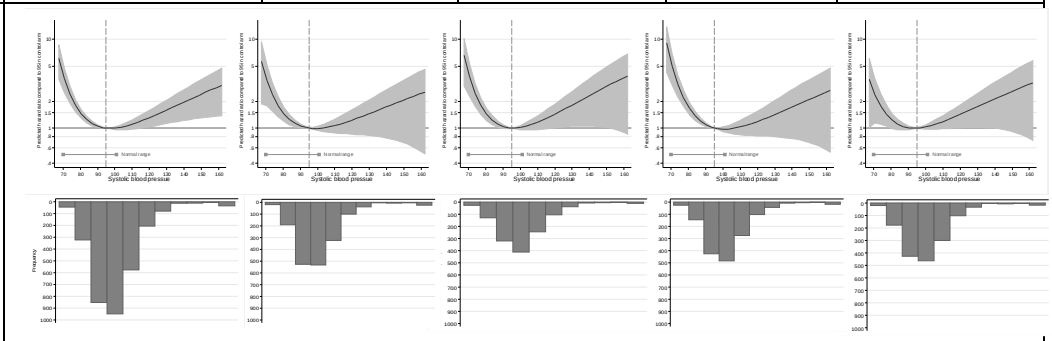
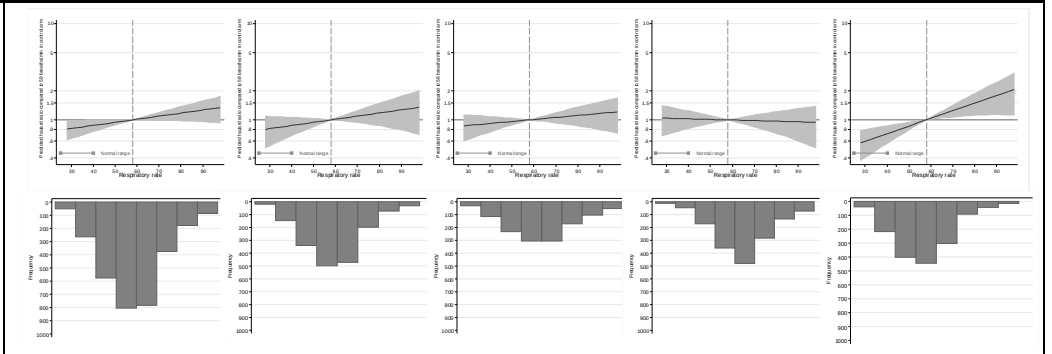
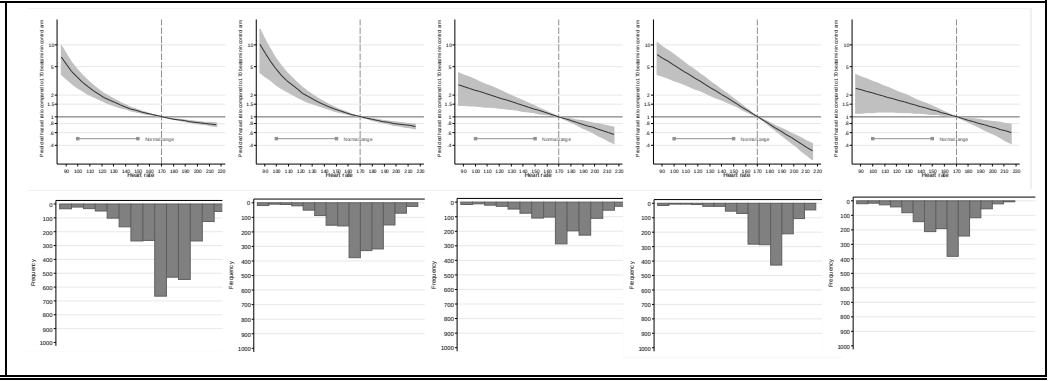
3.4 Results

3.4.1 Baseline Models

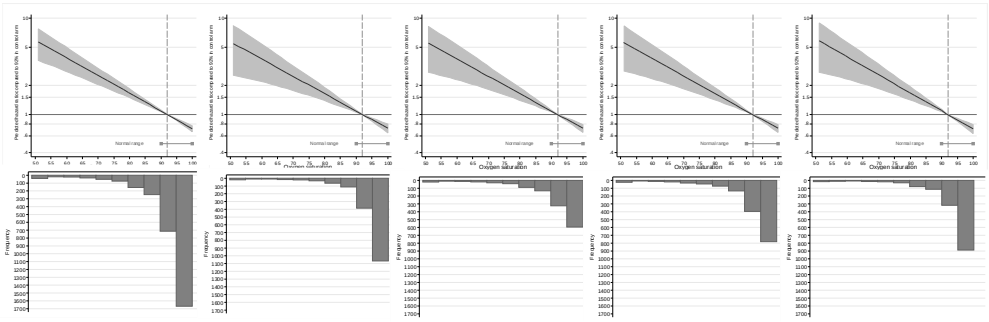
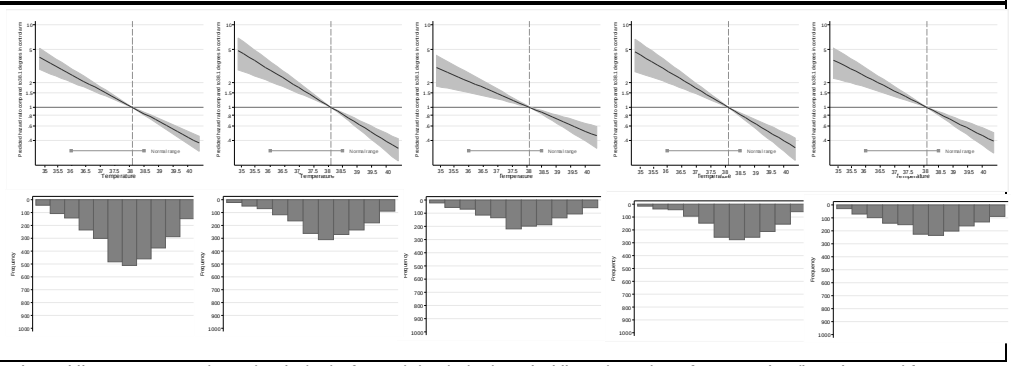
3.4.1.1 Bedside Vital signs

Cox regression models were first used to estimate mortality risk according to each baseline measure in complete cases, and hazard ratios for the effect of bolus vs no bolus adjusted for each baseline measure. These models, for exploratory analysis, included all children and randomised group as a single binary factor (bolus vs no bolus) but were not adjusted for other baseline measures. Figure 3.4.1, Figure 3.4.2 and Figure 3.4.3 below present the shape of the functions selected to best reflect the association between each baseline measure and mortality risk, for each baseline measure within each group only, adjusted for bolus vs no bolus. Malaria is defined using the SAP definition as above. The y-axis is the predicted hazard ratio relative to a child with the median reference value (e.g. 58 breaths per min for respiratory rate). The median reference value is plotted as a vertical dotted line and the clinical normal range for the measure is also indicated on the plot. The distribution of the measure is also described in histograms underneath and absolute numbers are provided in the appendix.

Figure 3.4.1: Associations between mortality and baseline bedside measures: SBP, respiratory rate, heart rate, oxygen saturation and temperature (rows) for all children, those with malaria, those without malaria, those ≤2 years old and those >2 years old (columns). Distribution of bedside measure by group in the histogram underneath each plot.

	All children	Malaria	Non malaria	Age ≤2 years	Age>2 years
Baseline SBP					
Respiratory rate					
Heart rate					

Notes: The grey horizontal lines represent a hazard ratio (HR) of 1, and the dashed vertical lines show the reference value (i.e value used for centering each distribution).

	All children	Malaria	Non malaria	Age ≤2 years	Age>2 years
Oxygen saturation					
Temperature					

Notes: The grey horizontal lines represent a hazard ratio (HR) of 1, and the dashed vertical lines show the reference value (i.e value used for centering each distribution).

The mfp process selected very similar functions across groups for each measure at baseline, with monotonic functions in all measures apart from SBP. The plots indicated that there was high mortality risk at both high and low levels of SBP in all groups although the confidence intervals were very wide for the high mortality risk at high SBP due to the low numbers of children with SBP>120 (n=128 (4%)).

Children with SBP>120 were more likely to be in coma (26% compared to 14% for those with SBP<120) but were not more prevalent in those with malaria (4.5% had SBP>120 in the non-malaria group versus 3.9% in the malaria group). Coma (inability to localise a painful stimulus) in the presence of malaria defines cerebral malaria and in non-malaria this may have been secondary to a bacterial or viral encephalopathy. Cushing's Triad (hypertension, relative bradycardia and slowing of the respiratory rate in children with coma) is a well-recognised but late phenomenon in central nervous system (CNS) infections indicative of compensatory intracerebral perfusion pressure in meningitis, complicated by raised intracranial pressure (ICP). Cushing's Triad is less well described in children with cerebral malaria as the major pathological event is cerebral hypertension and not raised ICP [183-185]. However, I found a similar mortality risk associated with high blood pressure in children with and without malaria.

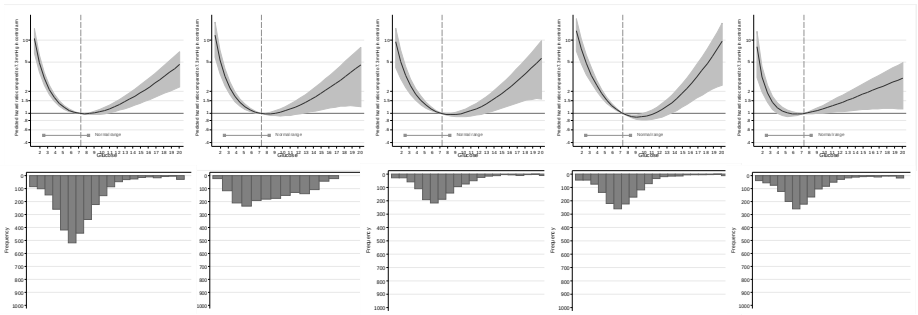
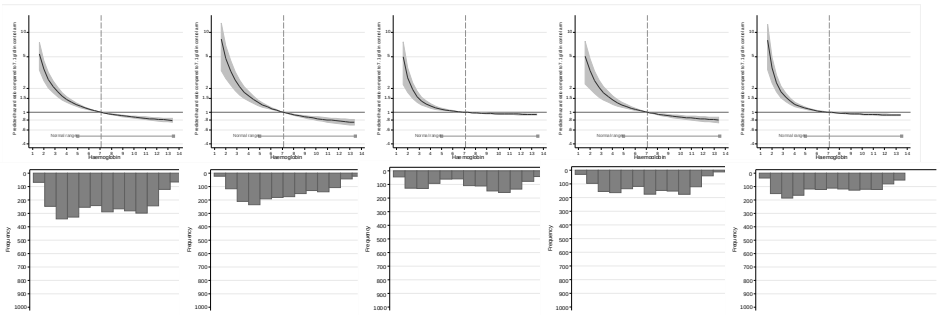
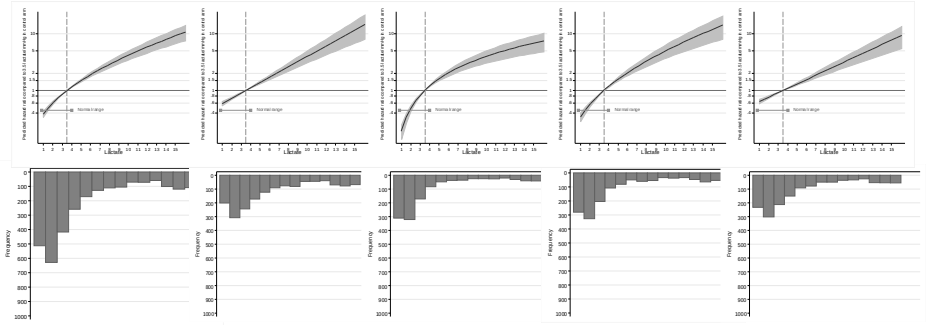
There was a high mortality risk associated with low baseline heart rate (relative bradycardia), also demonstrated previously in children in the no bolus arm in Chapter 2 prognostic score. However, the high mortality risk at baseline of severe tachycardia (very high heart rate) identified in Chapter 2 was not observed in the analyses in this chapter. One reason may be that heart rate was not truncated in Chapter 2 to capture the information needed for a risk score, whereas it was truncated at the 99th centile for these analyses, reducing the impact of outlying values.

Respiratory rate was modelled as a linear function in all groups, with a small increase in mortality risk as the respiratory rate increased in those with and without malaria. In older children the mortality risk increased faster with increasing respiratory rate, especially when compared to children aged ≤2 years who had no significant change in mortality risk across their levels of respiratory rate (test for heterogeneity p=0.01). Respiratory rates in children in the FEAST trial were also much higher than the normal range (median of 58 breaths/min compared to a normal range of 20-40 breaths/min) reflecting the inclusion criteria of respiratory distress (increased work of breathing) on admission. As expected, oxygen saturation had increased mortality risk at low levels of hypoxaemia (<90%) and a linear association for all groups.

Mortality risk decreased linearly as temperature increased. A high temperature at baseline was protective and children with low temperature (hypothermia) had a much higher mortality risk. This has been shown before in other studies and beneficial effects of fever have been discussed in the literature [47, 186]. Also, hypothermia can be seen as maladaptive and is found in children in a late stage of decompensation. Of the 189 children with hypothermia (<36°C) on admission, 54 (29%) died, with 25 (46%) of the 54 deaths occurring within 4 hours of randomisation. In contrast, there were 113 (3.5%) children with a very high temperature (>40°C) of which 5 (4.4%) children died (only 1 before 4 hours).

3.4.1.2 Laboratory indices

Figure 3.4.2: Associations between mortality and baseline glucose, haemoglobin, and lactate (rows) for all children, those with malaria, those without malaria, those ≤2 years old and those >2 years old (columns). Distribution of each measure by group in the histogram underneath each plot.

	All children	Malaria	Non-Malaria	Age ≤ 2 years	Age >2 years
Glucose					
Haemoglobin					
Lactate					

Notes: The grey horizontal lines represent a HR of 1, and the dashed vertical lines show the reference value (i.e value used for centering each distribution).

The best functional form for associations with mortality was similar across the groups for glucose, haemoglobin and lactate. Only glucose had a non-monotonic relationship with risk.

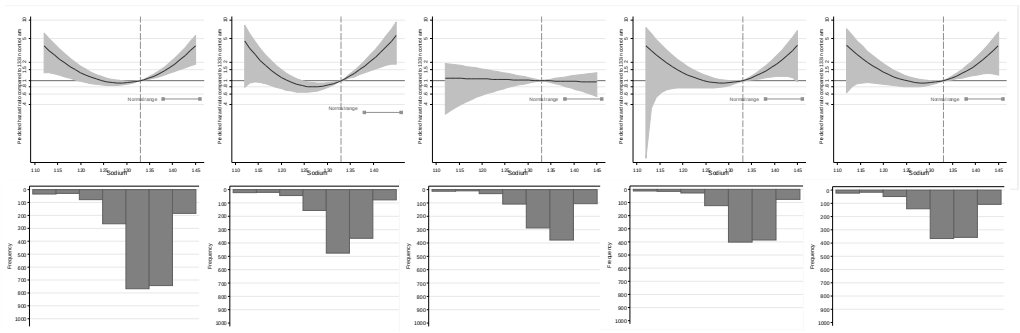
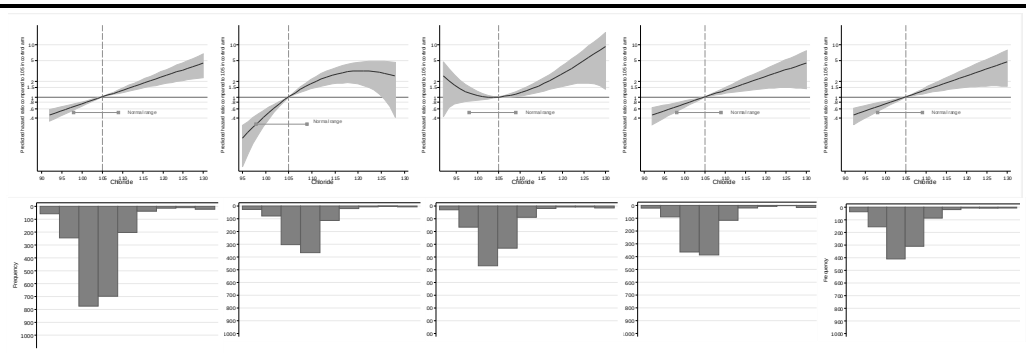
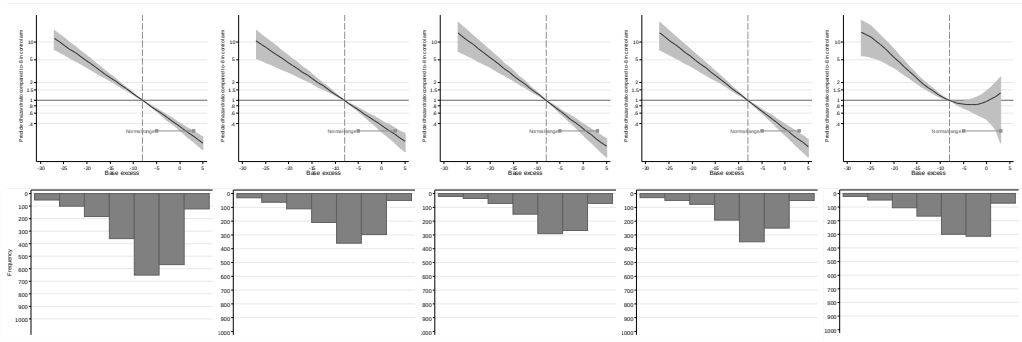
The fp(2) function for glucose demonstrated that mortality risk was increased for those with very low glucose or very high glucose, which is biologically plausible and well recognised in critical care where metabolic derangement is common. The NICE-SUGAR trial in adults

compared tight glycaemia control (maintaining glucose levels between 4.5-6mmol/L) against a conventional threshold for glycaemic control (set at maintaining glucose levels <10mmol/L) to prevent hyperglycaemia, with severe hypoglycaemia as an adverse event. Hypoglycaemia occurred more frequently in the arm that had intensive or 'tight' glycaemia control and thus conventional control was recommended in adults [187]. A post-hoc analysis of that trial also indicated that hypoglycaemia was associated with increased mortality [188]. A similar trial in children also showed an increase in hypoglycaemic episodes with strict glycaemic control, and increased mortality associated with hypoglycaemia in a subgroup of children undergoing cardiac surgery [189]. Further, a study in Tanzania in children with febrile illness showed increased odds of mortality for children with glucose <2.5 mmol/L and in those with glucose levels of 2.5-5mmol/L, but did not report on hyperglycaemic events [190].

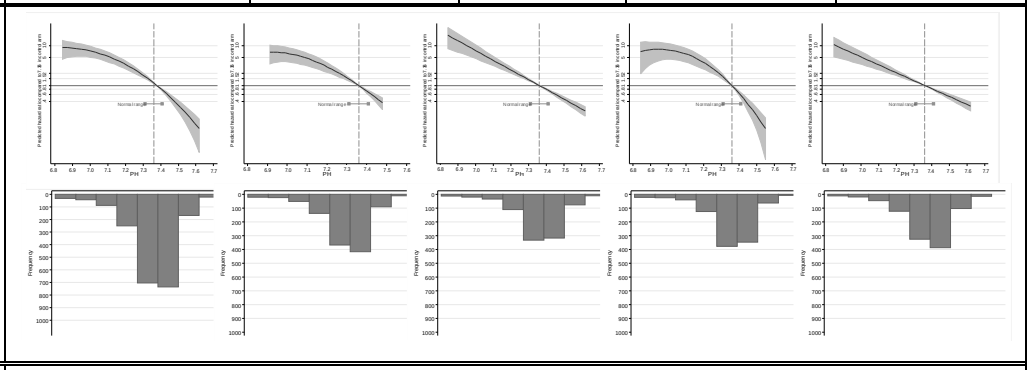
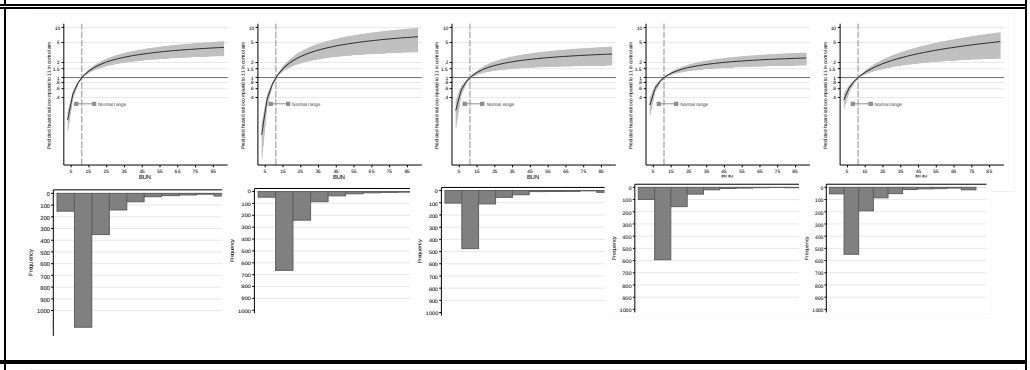
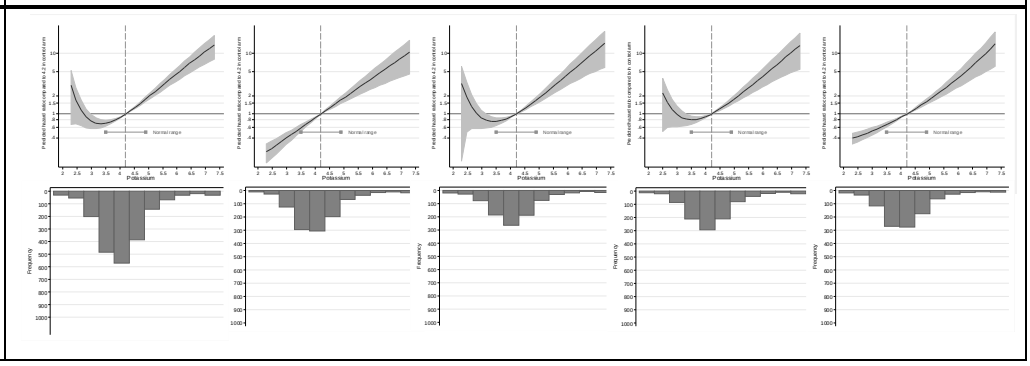
Lactate has not been modelled with fractional polynomials in previous chapters and here showed high mortality risk across all groups at high levels of lactate, with a monotonic increasing function as lactate increased. The distribution of mortality risk across lactate levels between the non-malaria and malaria groups may be expected to be different but in fact was very similar, with the only small difference being that the mortality risk tended to plateau for non-malaria children with lactate over 11 mmol/L. Median lactates were lower in the non-malaria group (median 3.0 (IQR 1.3, 6.3) compared to 4.4 (IQR 2.7, 8.8) in malaria). Lactic acidosis in severe malaria is multifactorial, with the high frequency of severe anaemia, shock and microcirculatory compromise evidenced by the correlation between lactate concentration and circulating and total parasite biomass [191], and with direct demonstration of impaired perfusion by orthogonal polarization spectral (OPS) imaging.

Risk of mortality was very high at very low haemoglobin and then decreased as haemoglobin increased, plateauing as haemoglobin increased beyond 8g/dL. There were 989/3054 children in FEAST A with severe anaemia (haemoglobin <5g/dL) of whom 94% received a transfusion. Of those with haemoglobin <5g/dl who did receive a transfusion by 8 hours, 38/892 (4%) died (after receiving a transfusion), compared to 48/97 (50%) children who died out of those who did not receive a transfusion. Thus it was likely to be those that did not receive a transfusion that contributed most to the high risk at very low haemoglobin.

Figure 3.4.3: Associations between mortality and baseline i-STAT measures: BUN, base excess, sodium, PH, chloride and potassium (rows) for all children, those with malaria, those without malaria, those ≤2 years old and those >2 years old (columns). Distribution of baseline i-STAT measure by group in the histogram underneath each plot.

	All children	Malaria	Non-malaria	Age ≤2 years	Age >2 years
Sodium					
Chloride					
Base excess					

Notes: The grey horizontal lines represent a HR of 1, and the dashed vertical lines show the reference value (i.e value used for centering each distribution).

	All children	Malaria	Non-malaria	Age ≤2 years	Age >2 years
PH					
BUN					
Potassium					

As for the previous measures, the mfp process generally selected very similar functions reflecting associations between baseline BUN, base excess, sodium, chloride, potassium, PH and mortality risk across the different groups. However, there were some exceptions, notably sodium and chloride in the non-malaria group, potassium in the malaria group and age>2 years group, and base excess in the age>2 years group. Some of these minor differences were due to mfp selecting only the 'best' model, with other models very close in terms of goodness of fit. For example, the AIC for the linear model for sodium in the non-malaria group shown in Figure 3.4.3 was 1101, but if a fp(0 0) function was used, similar to the malaria group, then the AIC was 1100. Although the best fp functions in the malaria, age ≤2 years and overall groups were different - fp(0 0), fp(0.5 0.5) and fp(2 2) respectively - they all gave a very similar functional form for the non-linear association between baseline sodium and mortality risk. If the other fp(2) functions were used to model sodium in the non-malaria group then they gave AICs of

1098 (fp(2 2)) and 1100 (fp(0.5 0.5)). This shows that the mortality risk plausibly varied in a similar way in the non-malaria group compared to the other groups, with increased risk at both high and low sodium, and any of the models would be suitable.

In contrast, applying the same process to look at the association between baseline chloride and mortality risk in the non-malaria group showed that the AIC for the model shown in Figure 3.4.3 with fp(-2 -2) was 1247, whereas a linear function (as in overall and each age group) gave an AIC of 1257. As the difference between the AICs was >3.84 (chi-squared (1)) this suggests that mortality risk may have genuinely varied in a different way with chloride in the non-malaria group compared to the other groups. However, the confidence intervals for low chloride in the non-malaria group were wide, indicating the mortality risk could have plateaued rather than increased, and the number of children with chloride<95 in the non-malaria group was small (n=27).

This situation was similar for base excess: in those aged >2 years the selected function (fp(2 2)) indicated that the mortality risk plateaued above -5 rather than continuing to decrease as in the other subgroups. The AIC of the fp(2 2) model shown in Figure 3.4.3 was 1197 compared to 1205 from a model assuming a linear association between base excess and mortality risk, indicating evidence that there was a different risk profile for those in the older age group. There was a linear association between mortality risk and base excess in the other groups. Base excess <-8mmol/L was used as a cut-off value for high risk in the AQUAMAT score [103] which in the FEAST trial data would put 50% of children at high risk as -8mmol/L was the median value.

Mortality risk decreased with increased pH as expected, although this was not a linear association in all groups. Overall, and for children ≤2 years, the high mortality risk plateaued at very low levels of PH. Models with a linear relationship between mortality risk and pH levels were fitted for children >2 years and non-malaria. The AICs from these models compared to the models with a plateauing of the mortality risk showed >3.84 difference in the AICs. This indicated that there was evidence for the risk of mortality plateauing when pH<7.0 in these groups rather than continuing to increase as pH decreased under 7.0. The number of children with pH<7.0 (where the mortality risk plateaus in Figure 3.4.3 was 70/2082 (3%) and is at a level where many other measures would be deranged as well). A pH<7.2 would be considered severely acidaemic and there were 204/2082 (10%) children overall under this level.

The mortality risk increased sharply over the normal range of BUN (5-18mg/dl) before plateauing around 20 which is considered a cut-off for abnormally high values. The cut-off of 20 in BUN was also used as a prognostic value in the AQUAMAT score [103]. However, this

analysis suggests that there was substantial variation in mortality risk as BUN decreased below this value.

Overall, and in the non-malaria and age ≤2 years groups, the mortality risk was increased at both low and high levels of potassium. However, 95% CI were wide at low levels and the best fp function chosen was also consistent with plateauing risk below 3.5 mmol/L. For malaria and children aged >2 years a linear function reflecting a monotonic increase in mortality risk was selected. For both these groups the AICs of the linear functions were within 5 of the AIC of an fp(-1 -1) model where mortality risk would increase at both low and high levels; thus there was only relatively weak support for non-linearity. Hyperkalaemia at admission is known to complicate malaria, and high levels have been shown to be associated with higher mortality [105], whereas low potassium usually occurs after admission following treatment for acidosis [192]. In this population overall there were 86 (4%) children with potassium <3 mmol/L on admission.

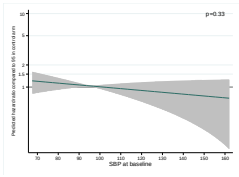
3.4.1.3 Baseline measures combined

These were exploratory analyses with models including just the measure of interest and adjusting for the randomised group of bolus vs no bolus in order to investigate what would be the best functional form for further modelling and testing for interactions. However, many of these measures would be impacted by levels of the other measures, and the high mortality risk for one particular group of children may not be independent of the levels of other baseline measures; thus, interpretation of these exploratory analyses requires caution. To further investigate this a Cox regression model was used to estimate mortality risk adjusted for all of the continuous baseline measures described above in children with all measures complete at baseline (n=1576). Thus, estimated mortality risk at each level of baseline measure was adjusted for all other baseline measures. This showed that some of the non-linear relationships between mortality risk and baseline measures in univariable models did not remain when adjusted for other baseline measures (for example, SBP, glucose, sodium, pH, BUN and potassium) (Figure 3.4.4). When adjusting for all the other baseline measures, I also found that mortality risk did not vary significantly over levels of SBP (p=0.33), respiratory rate (p=0.70), temperature (p=0.65), glucose (p=0.81), base excess (p=0.31), sodium (p=0.09), potassium (p=0.20), and pH (p=0.10). Associations with lactate (p<0.001), oxygen saturation (p=0.005), heartrate (p=0.006), haemoglobin (p<0.001), chloride (p<0.001) and BUN (p=0.06) remained in generally the same direction as unadjusted models, with the exception of haemoglobin, with some attenuation of effects at extremes for some factors.

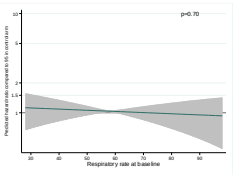
Interactions were not included in this model for comparisons with Figure 3.4.1, Figure 3.4.2 and Figure 3.4.3. In the FEAST data an interaction has been identified between baseline haemoglobin and lactate when both were included as continuous variables in a multivariable model for mortality [161]. This showed that rather than mortality risk uniformly increasing with increasing lactate and decreasing with decreasing haemoglobin, the higher mortality risk associated with higher lactate values (>7mmol/l) was restricted to those with high haemoglobin (>6 g/dl) [161]. As the interaction was not included in the adjusted model above, this may explain why the mortality risk was highest at high levels of haemoglobin. Subgroups were not explored due to the large number of measures included in the model and the reduced number of children with all measures recorded at baseline.

Figure 3.4.4: Associations between baseline measures and mortality risk estimated from a cox regression model adjusted for all other measures in the figure and randomised group (bolus vs no bolus).

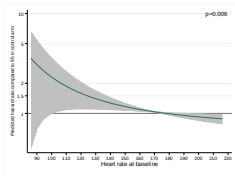
SBP



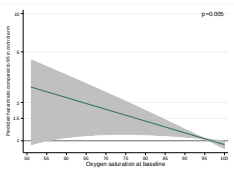
Respiratory rate



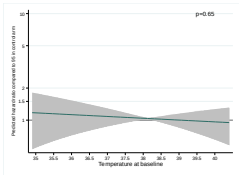
Heart rate



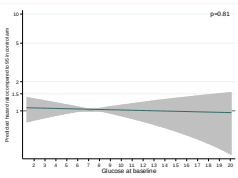
Oxygen saturation



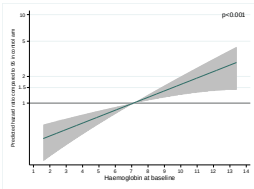
Temperature



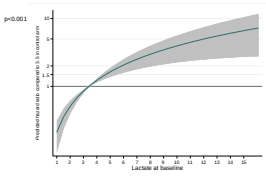
Glucose



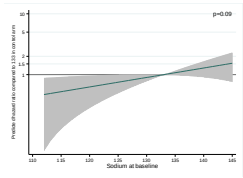
Haemoglobin



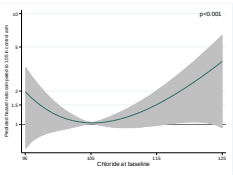
Lactate



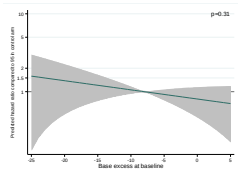
Sodium



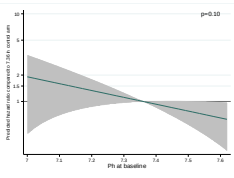
Chloride



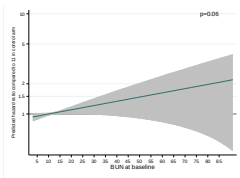
Base excess



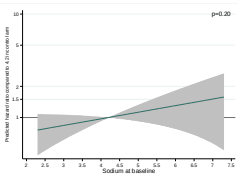
pH



BUN



Potassium



3.4.1.4 Interactions

Each model was investigated to identify whether the impact of randomised group (bolus vs no bolus) varied according to the level of the baseline measure (an interaction between the baseline measures and bolus) using likelihood ratio tests (Table 3.4.1). Three measures across three groups were found to have interactions with bolus ($p \leq 0.05$): three interactions were found between oxygen saturation and bolus - one in a model with all children ($p=0.02$), one in the malaria group ($p=0.04$) and one in the aged ≤2 years group (0.04); one interaction between baseline base excess and bolus ($p=0.02$), and one between chloride and bolus ($p=0.05$) both in the malaria group. This was out of 70 interaction tests in total for baseline models, giving a rate of 5/70 (7%) which was only slightly higher than would be expected by chance. As a significance level of 0.05 or 1/20 was used, there was a chance that 1 out of 20 tests (or 3 tests out of 60) would be a false positive.

Table 3.4.1: Results of heterogeneity tests between each baseline measure and impact of bolus vs no bolus.

Measure	Interaction p-values from each model				
	Overall	Malaria	Non-malaria	Age ≤2 years	Age >2 years
Systolic Blood Pressure	0.91	0.86	0.42	0.47	0.50
Respiratory rate	0.28	0.76	0.22	0.92	0.12
Heart rate	0.87	0.97	0.33	0.79	0.09
Oxygen saturation	**0.02**	**0.04**	0.15	**0.04**	0.19
Glucose	0.82	0.95	0.51	0.31	0.83
Temperature	0.72	0.10	0.30	0.41	0.93
Haemoglobin	0.58	0.67	0.92	0.97	0.61
Lactate	0.92	0.70	0.50	0.45	0.74
Sodium	0.44	0.12	0.47	0.70	0.99
Chloride	0.76	**0.05**	0.67	0.83	0.93
Base Excess	0.11	**0.02**	0.85	0.08	0.95
pH	0.96	0.89	0.18	0.82	0.55
BUN	0.44	0.76	0.62	0.07	0.85
Potassium	0.63	0.24	0.97	0.59	0.19

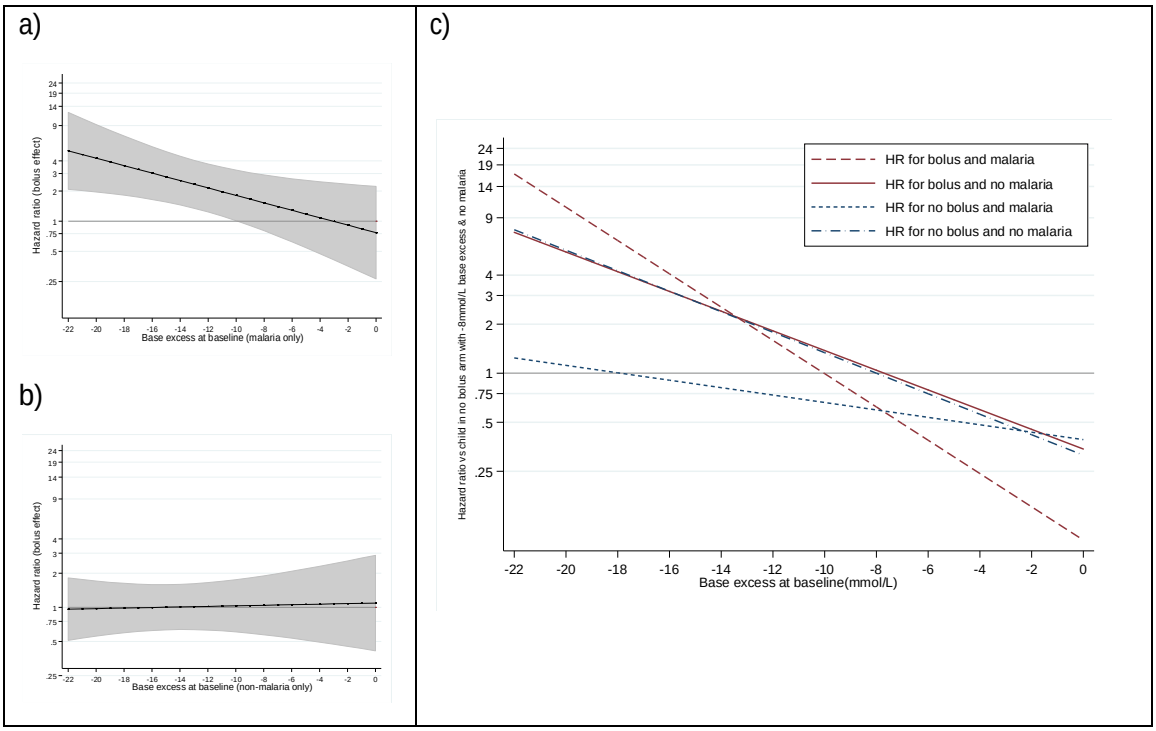
NB: Significant p-values are presented in bold font.

3.4.1.4.1 Interactions with baseline base excess

In children with malaria, there was some evidence that the effect of baseline base excess on mortality risk varied between bolus and no bolus groups or, equivalently, that the effect of bolus vs no bolus varied according to baseline base excess (p=0.02). There was borderline evidence for heterogeneity within the overall model (p=0.11) and children ≤2 years old (p=0.08) and no evidence within the non-malaria (p=0.85) model or for age>2 years old (p=0.95).

Those with malaria with very low base excess <-8 (acidosis) had a much greater detrimental effect of bolus compared to those with base excess >0 and malaria (Figure 3.4.5a). The impact of bolus on those with no malaria was similar across the levels of base excess (Figure 3.4.5b). Overall, absolute mortality risk was higher for low values of base excess (Figure 3.4.3), but those with malaria in the bolus arms had an even higher risk at low values (Figure 3.4.5c). Given the differences between malaria and non-malaria subgroups, a three-way interaction for bolus, malaria and base excess was examined with the likelihood ratio test including all children (p=0.01). This suggested that the impact of the bolus on those in the malaria group compared to the non-malaria group was different depending on the level of base excess the child had at baseline (Figure 3.4.5c). Compared to a child in the no bolus arm with no malaria and base excess of -8mmol/L, those with no malaria all have a similar distribution of absolute risk across base excess levels regardless of whether they had received a bolus or not; whereas, for children in the malaria group, those who had a bolus had an increased risk of mortality at extreme levels of base excess Figure 3.4.5c.

Figure 3.4.5a) Estimated hazard ratio comparing bolus vs no bolus at different levels of base excess in children with malaria (n=1139), b) Estimated hazard ratio comparing bolus vs no bolus at different levels of base excess in children with no malaria (n=925) and c) estimated hazard ratios relative to a child with base excess of -8 mmol/L in the no bolus arm without malaria from a model with a three-way interaction.



NB: The two lines that overlap in Figure 3.4.5c describe mortality risk versus a child with no malaria and base excess of -8 in bolus (solid) vs no bolus (dashed). The fact that they overlap shows there is no evidence in the non-malaria group that bolus affected mortality differently according to baseline excess (Figure 3.4.5b)

To further understand the base excess interaction with malaria and bolus, the TCEs as defined by the ERC (described in more detail on page 155) were examined in the malaria and non-malaria groups and the base excess was categorised at -8mmol/L (median value in FEAST, threshold used in AQUAMAT) to create a binary variable. The majority of children with malaria and acidosis that died in the bolus group had a cardiogenic TCE (38/65 (59%)) and the majority of TCEs in the corresponding no bolus group were neurological (6/11 (55%)) (Table 3.4.2; exact p-value for all TCEs=0.15). The deaths of children in the non-malaria group with acidosis were distributed across the TCEs similarly between the bolus and no bolus groups (Table 3.4.2; exact p-value for all TCEs = 0.97).

Table 3.4.2: Effect of bolus on mortality risk and TCEs (causes of death) by base excess level and malaria status at baseline.

	Base excess <-8mmol/L (**acidosis**)		
	Malaria	Non-malaria	Total
Bolus (% died)	**65/369 (18%)**	**53/267 (20%)**	118/637 (19%)
Cardiogenic deaths	38/65 (59%)	22/53 (42%)	60/118 (59%)
Neurological deaths	17/65 (26%)	14/53 (26%)	31/118 (26%)
Respiratory deaths	7/65 (11%)	14/53 (26%)	21/118 (18%)
Other deaths	3/65 (5%)	3/53 (6%)	6/118 (5%)
No Bolus (% died)	**11/177 (6%)**	**19/105 (18%)**	31/283 (11%)
Cardiogenic deaths	3/11 (27%)	8/19 (42%)	11/30 (37%)
Neurological deaths	6/11 (55%)	6/19 (32%)	12/30 (40%)
Respiratory deaths	2/11 (18%)	4/19 (21%)	6/30 (20%)
Other deaths	0/11 (0%)	1/19 (5%)	1/30 (3%)
Hazard ratio (HR) for bolus	**3.01 (1.59, 5.71)**	**1.09 (0.64, 1.83)**	1.74 (1.18, 2.59)
Exact p-value comparing type of deaths between bolus and no bolus	p=0.15	p=0.97	
	Base excess ≥ -8 mmol/L		
	Malaria	Non-malaria	Total
Bolus (% died)	**15/397 (4%)**	**10/356 (3%)**	25/755 (3%)
Cardiogenic deaths	7/15 (47%)	2/10 (20%)	9/25 (36%)
Neurological deaths	6/15 (40%)	3/10 (30%)	9/25 (36%)
Respiratory deaths	2/15 (13%)	4/10 (40%)	6/25 (24%)
Other deaths	0/15 (0%)	1/10 (10%)	1/25 (4%)
No Bolus (% died)	**5/196 (3%)**	**7/197 (4%)**	12/393 (3%)
Cardiogenic deaths	1/5 (20%)	0/7 (0%)	1/12 (8%)
Neurological deaths	3/5 (60%)	2/7 (29%)	5/12 (42%)
Respiratory deaths	1/5 (20%)	4/7 (57%)	5/12 (42%)
Other deaths	0/5 (0%)	1/7 (14%)	1/12 (8%)
HR for bolus	**1.48 (0.54, 4.09)**	**0.79 (0.30, 2.09)**	1.09 (0.56, 2.17)
Exact p-value comparing type of deaths between bolus and no bolus	p=0.65	p=0.89	

	Observations that have non-missing base excess [a] (interaction between bolus and malaria p=0.009)		
	Malaria	Non-malaria	Total
HR for bolus	**2.57 (1.50, 4.40)**	**1.00 (0.64, 1.59)**	1.58 (1.13, 2.23)
	All observations (including those with missing base excess)[b] (interaction between bolus and malaria p=0.72)		
	Malaria	Non-malaria	Total
HR for bolus	**1.62 (1.10-2.37)**	**1.46 (1.02, 2.12)**	1.47 (1.13, 1.91)

[a] Model adjusted for continuous base excess. [b] Model not including base excess.

Base excess had 34% missing values and the estimates for the effect of bolus in the malaria and non-malaria groups in children with base excess recorded were slightly different from the complete data estimates: HR 1.62 (95%CI (1.10, 2.37)) in children with malaria and HR 1.46 (95% CI (1.02, 2.12)) in children with non-malaria (Table 3.4.2). The three-way interaction with malaria and continuous base excess was confirmed in multiple imputation analyses (which correctly incorporated the interaction into the imputation model) (p=0.001) and gave evidence that the interaction is not an artefact related to the missingness of base excess. The data was also split into the categories in Table 3.4.3 and by using Rubin's rules the HR and confidence intervals restricted by base excess category and malaria status were estimated (Table 3.4.3).

Table 3.4.3: Estimated effect of bolus on mortality by malaria and acidosis using multiple imputation analyses. Percentage died is estimated across imputation datasets.

	Base excess <-8 (acidosis)		
	Malaria	Non-malaria	Total
Bolus (% died)	**17%**	**24%**	20%
No Bolus (% died)	**10%**	**19%**	13%
	1.72 (1.12-2.65)	1.25 (0.46-3.39)	1.52 (1.13, 2.04); interaction p = 0.15
	Base excess ≥ -8		
Bolus (% died)	**3%**	**3%**	3%
No Bolus (% died)	**2%**	**3%**	3%
	1.25 (0.83-1.89)	1.06 (0.45-2.57)	1.12 (0.59, 2.14); interaction p=0.41

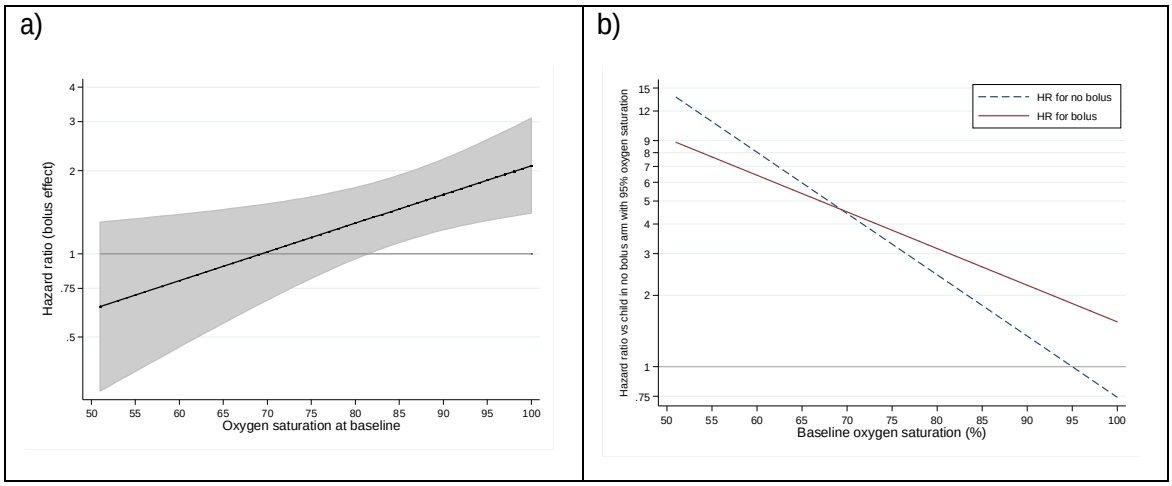
Table 3.4.3 showed that the multiple imputation analyses support the complete case analyses, but that the majority of the estimates of the bolus effect in Table 3.4.2 were further away from 1 compared to the imputation analyses.

3.4.1.4.2 Interaction with baseline oxygen saturation

In the whole trial population, there was some evidence that the effect of baseline oxygen saturation on mortality risk varied between bolus and no bolus groups or, equivalently, that the effect of bolus vs no bolus varied according to baseline oxygen saturation (p=0.02). There was also evidence that the effect of bolus vs no bolus varied according to oxygen saturation in those with malaria (p=0.04) and those age ≤2 years (p=0.04), with no evidence for those without malaria (p=0.15) or age>2 years (p=0.19). There was no evidence that the varying effect of bolus vs no bolus over oxygen saturation values was significantly different between the malaria and non-malaria groups (p=0.17 for a three-way interaction) or between the two age groups (p=0.18 for a three-way interaction). The varying of the bolus impact according to continuous baseline oxygen saturation in the whole trial population had been found before (Appendix Figure S2 of Maitland *et al*, BMC Med 2013) [157].

The hazard ratios were estimated from a model allowing the bolus effect to vary continuously across the values of oxygen saturation in all children (Figure 3.4.6a). Children with good lung function at baseline had a greater detrimental effect of the bolus, for example with a hazard ratio of 1.98 if they had 98% oxygen saturation. If they had very low oxygen saturation then there was no evidence for an effect of bolus increasing mortality, but their overall mortality risk was higher. This is shown in Figure 3.4.6b which estimates the hazard ratio relative to a child with 95% oxygen saturation in the no bolus arm. The mortality risk was much higher for both groups at very low oxygen saturation (<70%) with no significant difference between them (Figure 3.4.6a), but at higher oxygen saturation (>85%) overall mortality risk was lower, but significantly higher with bolus compared to no bolus.

Figure 3.4.6a) Estimated hazard ratio for bolus vs no bolus at different levels of baseline oxygen saturation. b) Estimated hazard ratio relative to a child in the no bolus arm with 95% oxygen saturation.



The median (IQR) of oxygen saturation was 95 (91-98) and a healthy/good level of oxygen saturation is generally considered to be 99% or 100%. The WHO definition of hypoxia is oxygen saturation <90% [193]. The HRs for bolus vs no bolus across this range, corresponding to Figure 3.4.6a, are presented in Table 3.4.4 below.

Table 3.4.4: Estimated hazard ratios for bolus vs no bolus at specific values of oxygen saturation, taken from Figure 3.4.6a) above.

Oxygen level	HR for bolus vs no bolus (95% CI)
98%	1.98 (1.37-2.87); p<0.001
95%	1.89 (1.34-2.68); p<0.001
90%	1.63 (1.22-2.20); p=0.001
85%	1.45 (1.10-1.92); p=0.009
80%	1.29 (0.96-1.74); p=0.09

The interaction was further explored by splitting oxygen saturation baseline into two categories (<90% and 90-100%) (Table 3.4.5). Oxygen saturation <90% (measured by a pulse oximeter) is the definition of hypoxaemia in children given by WHO [194]. This gave hazard ratios for each subgroup in the two categories showing the increased risk at high levels of oxygen saturation.

Table 3.4.5: Effect of bolus by dichotomised baseline oxygen saturation level in each group.

	Oxygen saturation <90%	Oxygen saturation 90-100%
Group	HR for bolus vs no bolus	HR for bolus vs no bolus
Overall	1.09 (0.72, 1.64)	1.91 (1.30, 2.81)
Malaria	1.10 (0.57, 2.12)	2.21 (1.29, 3.81)
Non-malaria	1.10 (0.65, 1.87)	1.83 (1.04, 3.24)
Age ≤ 2 years	0.97 (0.56, 1.69)	1.86 (1.07-3.22)
Age > 2 years	1.26 (0.69, 2.31)	1.95 (1.14, 3.35)

As children with ≥90% oxygen saturation at baseline (n=2384) was a large subgroup, TCEs were examined in them to further explore the interaction. Sub hazard ratios were calculated for each TCE (Table 3.4.6) taking into account the competing risks.

Table 3.4.6: Sub hazard ratios for bolus vs no bolus in those with baseline oxygen saturation≥90% taking into account the competing risks.

	Cumulative incidence of death at 48 hours			
TCE	Bolus	No bolus	Sub-hazard ratio	p-value
Cardiogenic	3.7%	1.9%	1.95 (1.11, 3.45)	0.02
Neurological	2.7%	2.5%	1.08 (0.63, 1.83)	0.78
Respiratory	1.6%	0.01%	12.65 (1.71, 93.3)*	0.01
Other/Unknown	1.1%	0.9%	1.29 (0.54, 3.10)	0.56

* There was only 1 respiratory death in the no bolus arm in children with oxygen saturation ≥90%, compared to 22 respiratory deaths in the bolus arm.

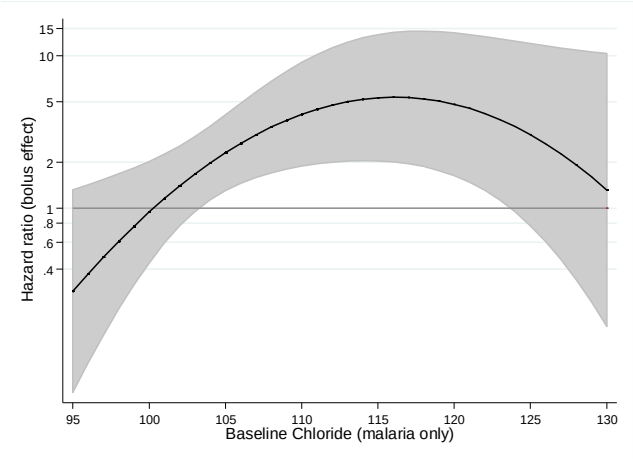
This shows that the major difference between the bolus and no bolus groups for children with ≥90% oxygen saturation at baseline was in the higher proportion of deaths adjudicated as having a cardiogenic or shock TCE, as well as in the much higher proportion of deaths adjudicated as having a respiratory TCE. There was some weak evidence for a difference in respiratory TCE between the bolus and no bolus group using all the children in the trial (p=0.09) [157], but the analysis above showed stronger evidence when restricting to those with ≥90% oxygen saturation at baseline.

In addition to analysing how children with ≥90% oxygen saturation died, the distribution of shock within these children was described by examining the four clinical characteristics that were part of the eligibility criteria for impaired perfusion at baseline – temperature gradient, capillary refill time >2s, severe tachycardia and a weak pulse. Children needed to have at least one of these characteristics, along with a fever or history of fever and respiratory distress and/or impaired consciousness, to be eligible for the trial. Children with ≥90% oxygen saturation had more severe tachycardia compared to children with <90% oxygen saturation (1745/2381 (73%) vs 423/651 (65%) chi-squared p<0.001). They were more likely to have only one feature of impaired perfusion (1197/2381 (50%) vs 288/651 (44%)) and less likely to have three features (309/2381 (13%) vs 124/651 (19%)) but were similarly likely to have two (783/2381 (33%) vs 210/651 (32%)) or four features (92/2381 (4%) vs 29/651 (4%); chi-squared over all four categories p<0.001).

3.4.1.4.3 Possible weak interaction with baseline chloride

There was weak evidence for a varying effect of the bolus across levels of baseline chloride (p=0.05) in those with malaria. There was no evidence of the bolus impact varying in the other groups (p=0.67, p=0.83, p=0.93 for non-malaria, age≤2 years and age>2 years respectively) nor overall (p=0.76). The evidence for a varying impact across levels of chloride disappeared when the model for children with malaria was adjusted for the base excess interaction (p=0.28). The increased risk, in the unadjusted model, associated with bolus was highest in children with malaria and baseline chloride of 115mmol/L, which was only slightly higher than the clinically normal range of 98-109mmol/L, rather than the increased risk being highest at very high levels of chloride (Figure 3.4.7). Thus it was not simply that children with very high levels of chloride at baseline had highest risk from the bolus, as those with chloride of 105-110 mmol/L had similar mortality risk from bolus to those with chloride >120mmol/L. The fact that the interaction disappeared in the presence of the base excess interaction (which remained significant), suggests that this did not describe a different mechanism to the base excess interaction. This may have been expected since chloride is a component of base excess.

Figure 3.4.7: Estimated hazard ratio for bolus vs no bolus at different levels of baseline chloride in children with malaria.

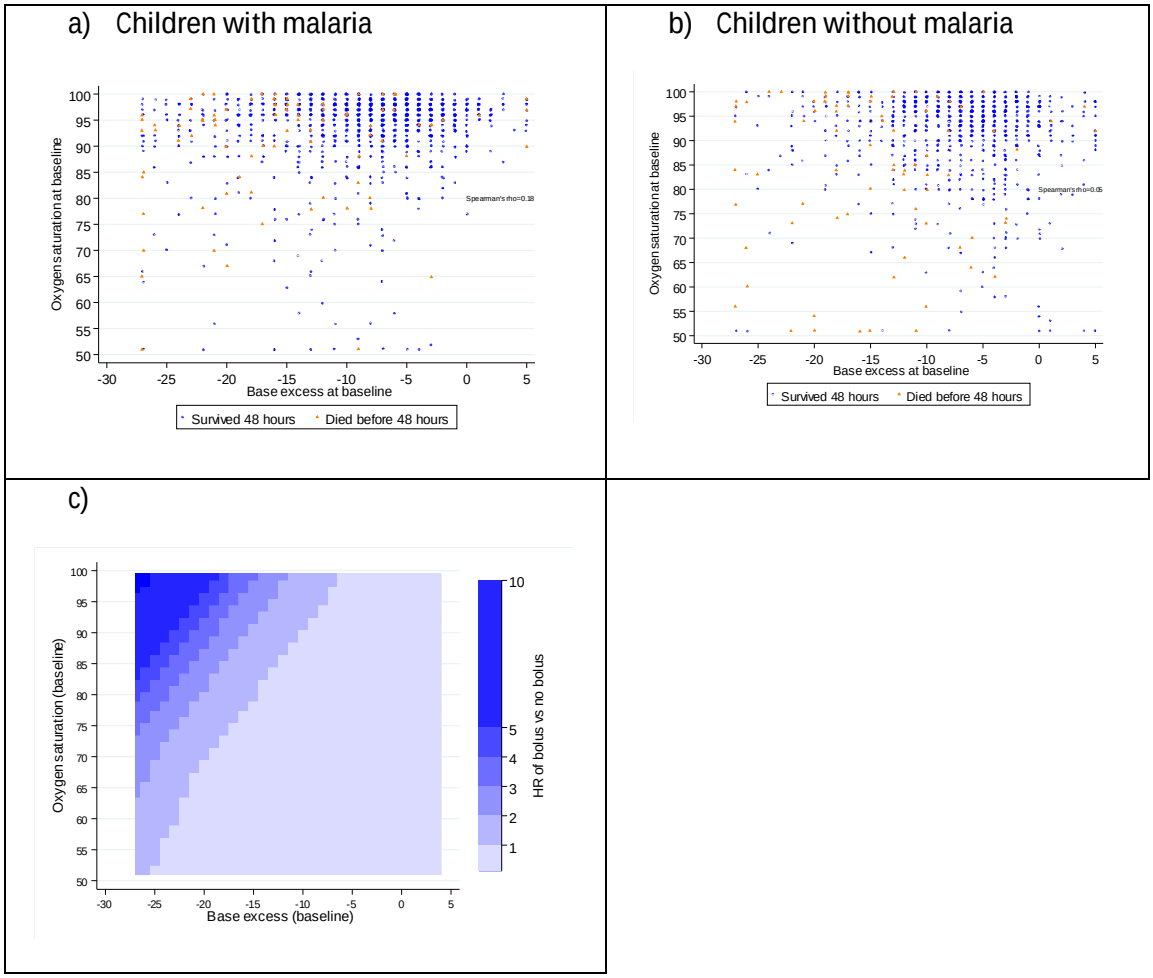


3.4.1.4.4 Combining interactions

The interaction between bolus and baseline oxygen saturation and interaction between bolus, malaria and baseline base excess were included in a model, with all children that had base excess and oxygen saturation measured (n=1995) to investigate the impact of one interaction on the other and whether both were occurring independently. Likelihood ratio tests showed that both interactions held in the presence of the other (p=0.007 for oxygen saturation and p=0.002 for base excess three-way interaction with malaria). This indicated that there were two groups that had a greater impact of the bolus, those with higher oxygen saturation where there was relative increased risk of a respiratory TCE (and a slightly higher relative risk of a cardiogenic TCE) in the bolus group compared to the no bolus group, and those with malaria and low base excess where risk of a cardiogenic TCE was increased. Oxygen saturation and base excess were weakly positively correlated in children with malaria (rho=0.19, p<0.001) but not in children with non-malaria (rho=0.05, p=0.18) (Figure 3.4.8a and b).

Figure 3.4.8c shows the impact of bolus across oxygen saturation levels and base excess levels for children with malaria. The different shades of blue on the contour plot indicated different levels of risk: the darkest represent highest risk (hazard ratio ≥5) and the lightest shade represent lowest risk (hazard ratio <1) with base excess at baseline on the x-axis and oxygen saturation at baseline on the y-axis. The highest risk of mortality from receiving bolus was in those with malaria and a high oxygen saturation level and low base excess.

Figure 3.4.8: Correlations between base excess and oxygen saturation at baseline in a) children with malaria and b) children without malaria, c) Contour plot of bolus vs no bolus effect on mortality risk across baseline base excess and oxygen saturation levels in children with malaria.



The bolus effect was also estimated for a binary categorisation of base excess and oxygen saturation (defined as above) in the malaria and non-malaria group to examine the bolus impact on children that had high oxygen saturation, low base excess and malaria (Table 3.4.7). Consistent with the contour plot, this showed the greatest impact of bolus was in those with malaria and oxygen saturation ≥90% and base excess<-8 (HR 7.06 (2.19-22.82). Of note, I did not identify any subgroup where bolus was significantly beneficial.

Table 3.4.7: Impact of bolus vs no bolus on mortality defined by base excess, oxygen saturation and malaria status.

	Base excess <-8 (acidosis)		
	Bolus (% died)	No bolus (% died)	HR for Bolus
Malaria	**65/369 (18%)**	**11/177 (6%)**	**3.01 (1.59, 5.71)**
oxygen saturation <90%	14/72 (19%)	6/36 (17%)	1.22 (0.47, 3.17)
oxygen saturation ≥90%	**41/282 (15%)**	**3/137 (2%)**	**7.06 (2.19-22.82)**
Non-malaria	53/267 (20%)	19/105 (18%)	1.09 (0.64, 1.83)
oxygen saturation <90%	25/82 (30%)	8/28 (29%)	1.05 (0.47, 2.33)
oxygen saturation ≥90%	21/165 (13%)	8/70 (11%)	1.11 (0.49, 2.51)
Total*	**118/637 (19%)**	**31/283 (11%)**	**1.74 (1.17, 2.59)**
	Base excess ≥ - 8[a]		
	Bolus (% died)	No bolus (% died)	HR for Bolus
Malaria	15/397 (4%)	5/196 (3%)	1.48 (0.54, 4.09)
oxygen saturation <90%	3/46 (7%)	1/17 (6%)	1.08 (0.11, 10.36)
oxygen saturation ≥90%	12/339 (4%)	4/174 (2%)	1.55 (0.50, 4.80)
Non-malaria	10/356 (3%)	7/197 (4%)	0.79 (0.30, 2.09)
oxygen saturation <90%	2/80 (3%)	6/57 (11%)	0.23 (0.05, 1.14)
oxygen saturation ≥90%	7/272 (3%)	1/138 (<1%)	3.59 (0.44, 29.23)
Total*	25/755 (3%)	12/393 (3%)	1.18 (0.55, 2.17)

*There were 4 children with missing malaria status and 70 with missing oxygen saturation in those with recorded base excess.

[a] There were few deaths in this group (37 in total) and so when split into subgroups by malaria and oxygen saturation (8 subgroups in total) there were very wide confidence intervals around the estimates and estimates of effect based on <10 deaths.

3.4.1.5 Missing data sensitivity analysis on baseline measures

Many of the continuous variables used in the analysis above had missing data, some with up to a third missing (Table 3.3.4, page 157). Multiple imputation (MI) was used to create 25 complete datasets which were then analysed to confirm the most appropriate function for each baseline continuous measure, which was then compared with the complete case analyses by plotting predicted values (Figure 3.4.9, Figure 3.4.10, and Figure 3.4.11).

Figure 3.4.9: Fitted functions for associations between SBP, respiratory rate, heart rate, oxygen saturation and mortality comparing between complete case analyses and multiple imputed data (MI) by group (overall, malaria, non-malaria, age ≤2 years and age > 2 years).

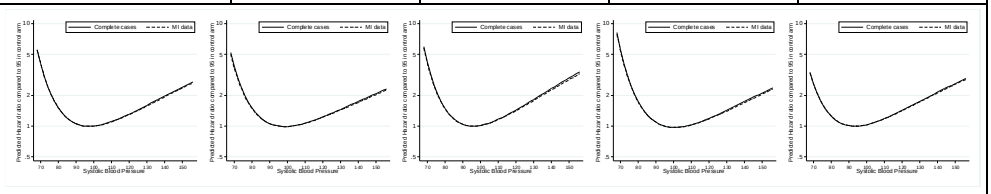
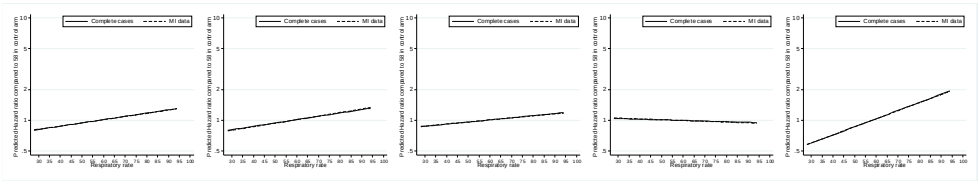
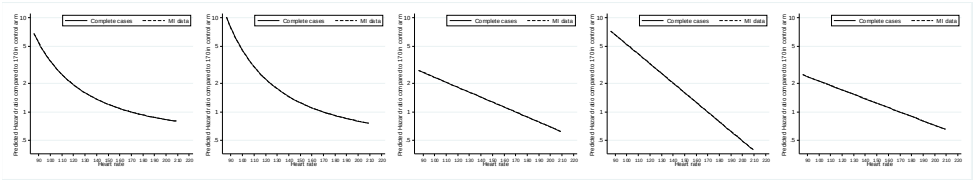
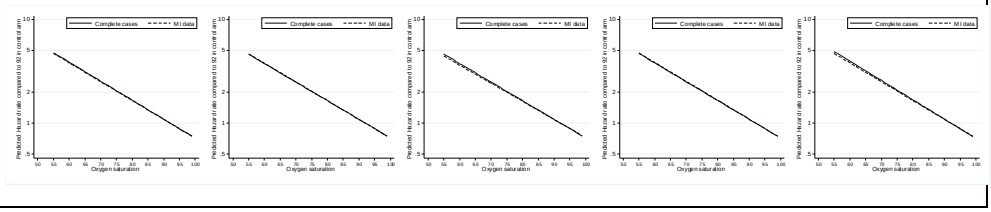
	Overall	Malaria	Non-malaria	Age≤2 years	Age > 2years
SBP *(2% missing)*					
Respiratory rate *(1% missing)*					
Heart rate *(<1% missing)*					
Oxygen saturation *(4% missing)*					

Figure 3.4.10: Fitted functions for associations between glucose, temperature, haemoglobin, lactate and mortality comparing between complete case analyses and multiple imputed data (MI) by group (overall, malaria, non-malaria, age ≤2 years and age > 2 years).

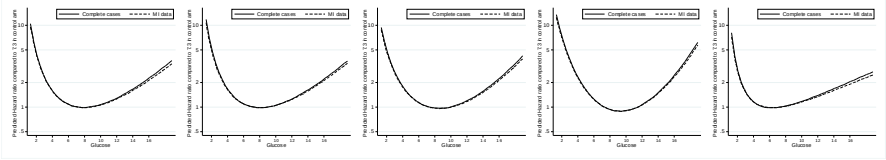
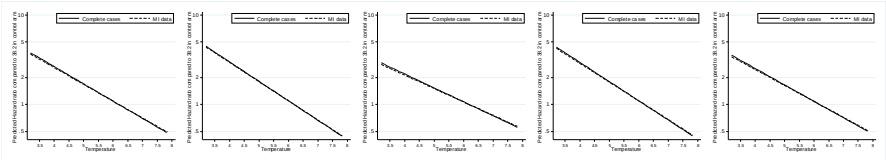
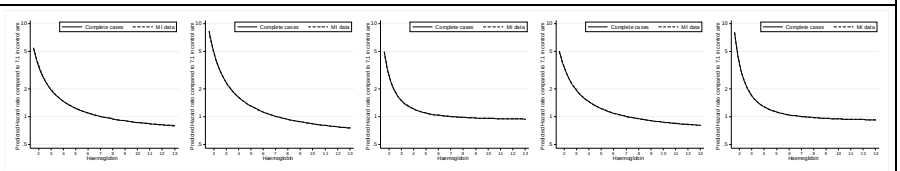
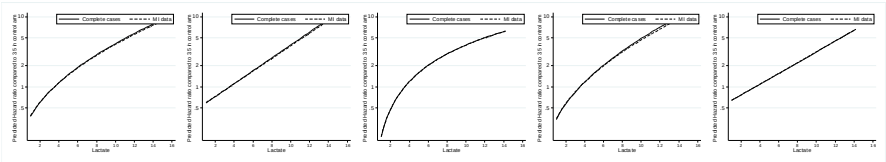
	Overall	Malaria	Non-malaria	Age ≤2 years	Age > 2 years
Glucose *(6% missing)*					
Temperature *(<1% missing)*					
Haemoglobin *(3% missing)*					
Lactate *(5% missing)*					

Figure 3.4.9 and Figure 3.4.10 show that the functions describing mortality risk over a range of values of baseline measures selected in the complete case analyses were supported by the multiple imputation analyses. This was as expected as there were low levels of missingness in these measures.

Figure 3.4.11: Fitted functions associations between sodium, chloride, base excess, pH, BUN, potassium and mortality comparing between complete case analyses and multiple imputed data (MI) by group (overall, malaria, non-malaria, age ≤2 years and age > 2 years).

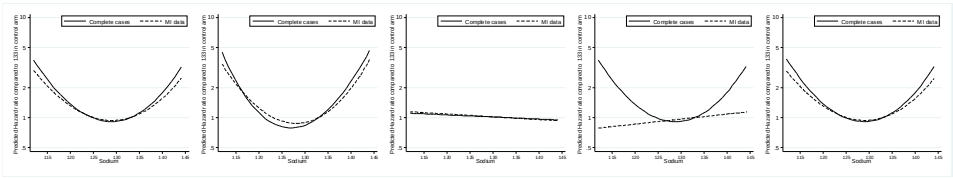
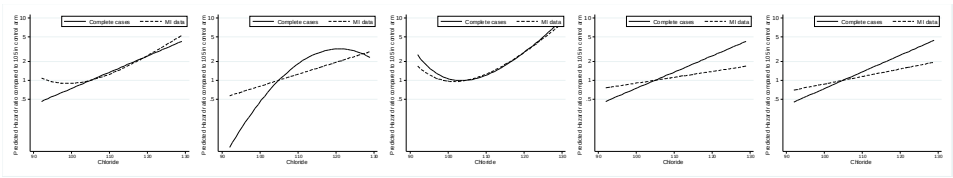
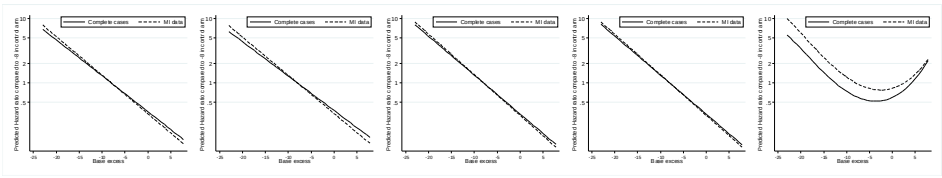
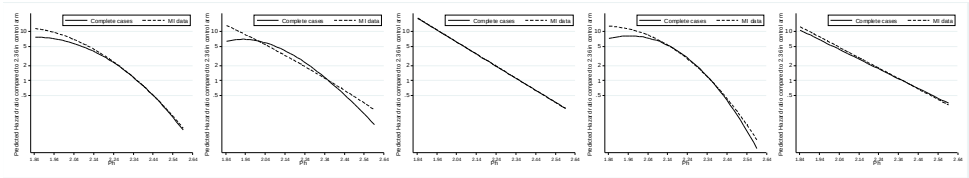
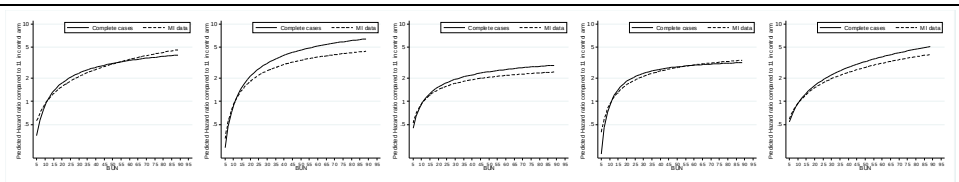
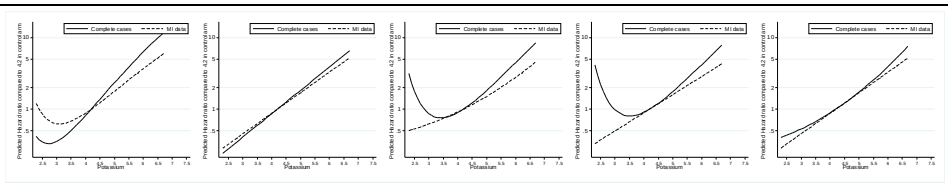
	Overall	Malaria	Non-malaria	Age ≤2 years	Age > 2years
Sodium *(33% missing)*					
Chloride *(34% missing)*					
Base excess *(34% missing)*					
pH *(34% missing)*					
BUN *(38% missing)*					
Potassium *(35% missing)*					

Figure 3.4.11 showed there were some differences between functions selected with the complete case analyses compared to analyses in the multiple imputed datasets, although the majority of functions chosen were the same or very similar.

The MI analysis for sodium confirmed a linear relationship between levels of sodium and mortality risk in the non-malaria group, which indicated that the increased mortality risk at low and high sodium seen overall and in both age groups may not be present for those without

malaria, although in complete case analyses the model with a linear function had an AIC within 3.84 of the AIC from models with FP2 functions for this group. A linear function was also selected by the MI analysis for those aged ≤2 years, although on further investigation this may also be due to the 'stacked data' model selection process. If Wald tests were used to select the fractional polynomial for sodium in MI data then it selected the same FP2 function as the complete case analysis.

The functions that the MI analyses selected to describe mortality risk over different levels of chloride were more consistent across the groups compared to the complete case analyses. The selected functions in MI analyses showed an increased risk of mortality at high chloride levels supporting the complete case analyses, and a suggestion of some increased risk at low levels of chloride, especially in the non-malaria group.

The functions selected for base excess, pH and BUN by the mfp analyses were similar to those selected in the complete cases. The complete case analysis for base excess in aged >2 years children had a small increase in mortality risk at levels of base excess >0mmol/L which was supported by the MI data, though every other group had a linear function showing protection from risk at those levels. Also, there were few children with a base excess >0mmol/L (n=95/2068) which was reflected in the confidence intervals widening in Figure 3.4.3. The MI analysis for potassium suggested that the increased risk found at low levels of potassium in complete case analyses in non-malaria and children ≤2 years may have been exaggerated by missing data, as the increase was not found in MI in those two groups. There was a small increase in mortality risk at low potassium in the overall group but the MI analyses in the other groups did not support this.

3.4.1.5.1 Interactions

The interactions between base excess, malaria and bolus and between oxygen saturation and bolus were tested in the MI data using a global Wald test that the coefficients of the interaction terms were different from zero and thus the interaction was also present in MI data. The test gave p=0.001 for the interaction between base excess, malaria and bolus, confirming the evidence from the complete case analysis that the mortality risk from bolus was much higher at very low base excess in those with malaria. The test also confirmed the oxygen saturation interaction that had previously been found in complete cases with p=0.02. As the bolus and no bolus groups were imputed separately then other interactions between baseline

measures and bolus could be tested. I planned to conduct these tests only for i-STAT cartridge measures which had weak evidence for an interaction in the complete case analyses ($p<0.1$ in Table 3.4.1, page 172). As other non-i-STAT measures had low levels of missingness, I felt that the complete case analysis was sufficient to identify any interactions between bolus and the baseline measure in those measures.

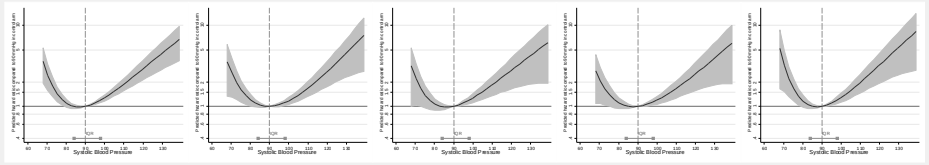
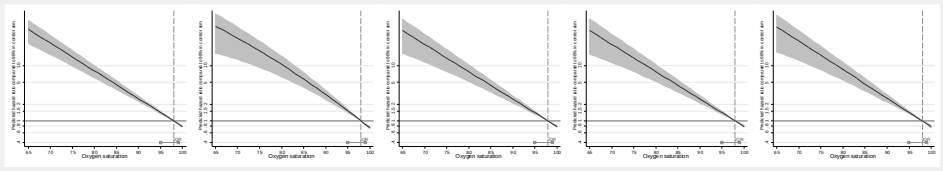
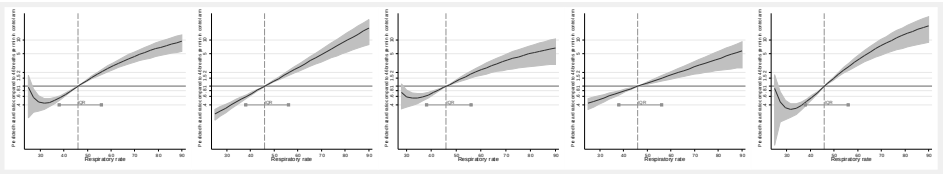
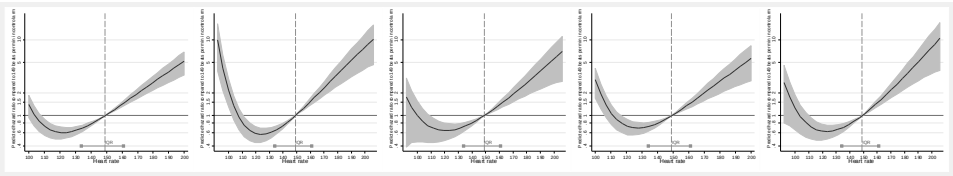
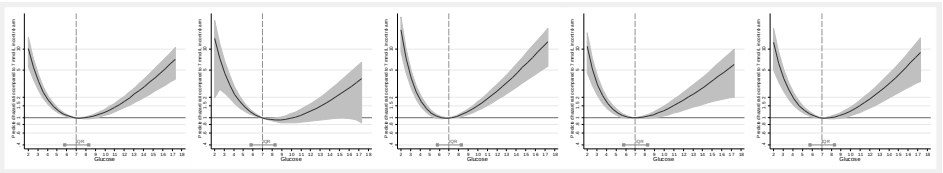
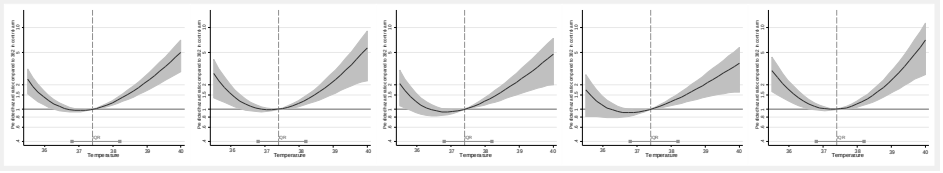
BUN and chloride satisfied the above criteria but the mfp command had identified that these measures did not have a linear association with mortality risk. In order to ensure that the analysis model was not mis-specified the transformed function needed to be imputed alongside the original variable [176]. I was not able to run the imputation model with the transformed variables for either BUN or chloride as the model would not converge. I had previously found that it is important to have the correct specification of the variables in the imputation model as I had tested the base excess malaria interaction in the MI data without the interaction in the imputation model and the global Wald test gave $p=0.62$. When the interaction was correctly included in the imputation model then the analysis model confirmed the interaction detected in the complete case analyses ($p<0.001$).

Overall, the majority of the MI analyses were consistent with the complete-case analyses for the functions chosen to describe each baseline measure's association with mortality. For potassium and chloride the MI analyses also demonstrated that the mortality risk at different levels of these measures may have been more similar across the groups than shown in complete case analyses. The analyses also confirmed the interactions between base excess and bolus in the malaria group and between oxygen saturation and bolus overall, found previously in compete-case analyses. However, the weak evidence for an interaction between chloride and bolus could not be examined in the MI data due to non-convergence of the imputation model.

3.4.2 Time-updated models

Cox time-updated models were used to estimate the association of time-updated values for each measure with mortality, adjusted for their value at baseline. These models show the mortality risk from the most recent value that was recorded for each child whilst also taking into account their baseline value. These models included all children and randomised group as a single binary factor comparing bolus vs no bolus but were not adjusted for any other measures. Missing data at or after the 1st post baseline time point was imputed by carrying forward the last non-missing value to time points whilst the child was still alive. This was done for 1-2% of data values (Table 3.3.2, page 152). The estimated unadjusted and adjusted hazard ratios were then combined to give the proportion of treatment effect explained for respiratory rate, glucose, temperature, heart rate, SBP, oxygen saturation, haemoglobin and lactate.

Figure 3.4.12: Association between mortality and time-updated systolic blood pressure, respiratory rate, heart rate, oxygen saturation and temperature (rows) for all children, those with malaria, those without malaria, those ≤2 years old and those >2 years old (columns).

	All children	Malaria	Non-malaria	Age ≤2 years old	Age>2 years old
SBP					
Oxygen saturation					
Respiratory rate					
Heart rate					
Glucose					
Temperature					

Notes: The grey horizontal lines represent a HR of 1, and the dashed vertical lines show the reference value (i.e value used for centering each distribution).

In general, the association between time-updated values and mortality risk was similar across the groups for each measure in Figure 3.4.12. Also, the relationship with mortality risk in the time-updated models selected for oxygen saturation, glucose and SBP were the same as those selected in baseline models (Figure 3.4.1). This showed that abnormal SBP (either high or low) and abnormal glucose (either high or low) after admission increased mortality risk, and low

oxygen saturation after admission increased mortality risk with risk decreasingly linearly as oxygen saturation increased.

Mortality risk increased linearly as respiratory rate increased for the malaria and age≤2 years groups. However, for the non-malaria, age>2 years groups and overall trial population, mortality risk plateaued at low values with then a small increase at very low levels of respiratory rate. The confidence intervals widened around the small increase in risk indicating larger uncertainty around the estimate of mortality risk at those levels, and the risk still remained lower than that of a child with the median value of 46 breaths per minute. These FP2 models (that described a non-linear relationship) were compared with models fitting respiratory rate as a linear relationship with mortality risk, and the AICs from the FP2 models (Figure 3.4.12) were significantly lower which suggests that mortality risk may have genuinely varied in a different way at very low levels of respiratory rate (<30 breaths per min) in these groups. However, high respiratory rate after admission consistently across the groups had much higher mortality risk than very low respiratory rate, and the amount that mortality risk may have varied at low levels of respiratory rate was small. Respiratory rates are also known to vary by age in healthy children and thus the small difference at low levels of respiratory rate in the shape of the association between respiratory rate and mortality risk in the age≤2 years group compared to age>2 years group may be due to the difference in ages. Respiratory rate after admission was weakly inversely correlated with oxygen saturation after admission (spearman's rho=-0.10 ($p<0.001$)). Children had lower mean oxygen saturation if they had a respiratory rate>70 breaths per min compared to those with ≤70 breaths per min although within a normal range (96% vs 97% median oxygen saturation respectively at 1hr, Wilcoxon test $p<0.001$; 97% vs 98% at 4hrs, $p<0.001$).

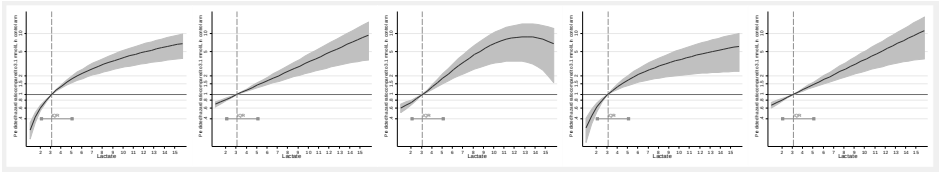
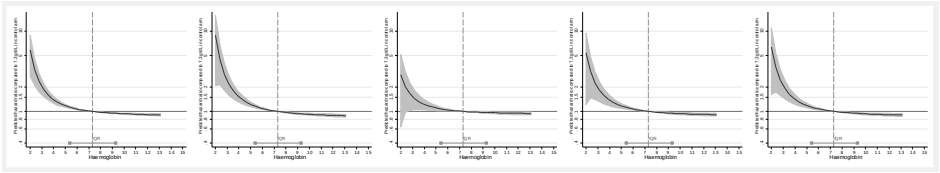
The association between heart rate after admission and mortality was best described with a U-shaped function (with increased mortality risk at high and low values of heart rate) which was consistent with the U-shape for the mortality risk by heart rate at baseline identified in Chapter 2 of this thesis. This indicated that the highest mortality risk was in those with high or low heart rate after admission. Thus if a child still had an abnormal heart rate (either high or low) or their heart rate became abnormal after admission then their mortality risk increased.

The relationship between glucose and mortality risk was non-monotonic, similar to the relationship between heart rate and mortality risk in time-updated models, with an increase in mortality risk at both high and low values across all groups and overall. Having hypoglycaemia at admission has been shown to increase risk of mortality [105] and this analysis confirmed that if the child had low glucose levels (<3mmol/l) or recurrent hypoglycaemia (due to

impaired gluconeogenesis) after admission the mortality risk also increased. This indicated that either they did not receive the treatment they should have got for hypoglycaemia (either identified at baseline or subsequently), or despite treatment their glucose remained low or they had rebound hypoglycaemia, all of which plausibly would increase mortality risk. There were 187 (6%) children with glucose <3mmol/L at baseline, with 246 (8%) children that received treatment for hypoglycaemia during admission. There were more children with hypoglycaemia in the malaria group (131/1687 (8%)) vs the non-malaria group (56/1275 (4%); chi-squared p<0.001) at baseline, but no evidence for a difference in the number of children with hypoglycaemia between the malaria and non-malaria groups after baseline (90/1705 (5%) vs 57/1277 (5%) ever had hypoglycaemia post-baseline; p=0.31). Although there are concerns that quinine can cause hypoglycaemia in adults [195], these findings support evidence that this is not a concern in children [196]. The analysis also showed that mortality risk was high in children with high levels of glucose after admission.

The mfp process selected a FP2 function for temperature leading to an increased mortality risk for those at very high or very low temperatures after admission, whereas there was a linear association with mortality risk from baseline measures. The increase in mortality risk at high levels of temperature which was not detected in the baseline models (Figure 3.4.1) may have been expected as these are post-admission measures. For example, if antimalarial or antibiotic treatment was immediately effective then those with high fever, especially with malaria, should have had their temperature reduced and so a continued high or increased temperature after admission may indicate the first line treatments (or dosages) may not have been effective in treating the underlying cause, thus putting the child at an increased risk of mortality. Quinine was the standard treatment for malaria in FEAST, which was conducted before the change in policy to artesunate (following the results of the AQUAMAT trial showing superiority of artesunate over quinine [28]). However, quinine remains an effective antimalarial if dosed correctly (since there were no reports of quinine resistance in Africa at the time of the FEAST trial).

Figure 3.4.13: Association between mortality and time-updated lactate and haemoglobin (rows) for all children, those with malaria, those without malaria, those ≤2 years old and those >2 years old (columns).

	All children	Malaria	Non-malaria	Age ≤2 years old	Age>2 years old
Lactate					
Haemoglobin					

Notes: The grey horizontal lines represent a HR of 1, and the dashed vertical lines show the reference value (i.e value used for centering each distribution).

Haemoglobin and lactate were only measured at 8 hours and 24 hours compared to the other measures that were measured at 1, 4, 8 and 24 hours. The mortality risk across time-updated values of haemoglobin and lactate was similar to that for baseline values showing increased risk at low levels of haemoglobin as expected and increased risk at high levels of lactate. The associations between different levels of haemoglobin and mortality risk, and lactate and mortality risk were also all similar across the different groups.

3.4.2.1 Examining association between randomised treatment and time-updated measures

For a time-updated measure to be considered as explaining any of the treatment effect or to be a surrogate it also needs to have an association with treatment; for example, the measure should change differently over time between the two treatment arms. To assess this, generalised estimating equations (GEE) with an independent correlation structure were used to analyse whether the mean absolute value of the measures described in this chapter varied between the treatment arms across the time points (1, 4, 8, 24, and 48 hours for SBP, oxygen saturation, respiratory rate, heart rate, glucose and temperature; 8 and 24 hours for haemoglobin and lactate) adjusted for the baseline value. SBP, glucose, temperature, oxygen saturation and lactate were log transformed in the models due to non-normal distributions. Global tests of difference were performed comparing the bolus and no bolus arms (Figure 3.4.14). These exploratory analyses did not account for mortality, but overall mortality by 48

hours was 9.5% and the absolute difference between arms was small (although significant) at 3.3%; thus, there was likely to be low impact on these analyses of the missing data due to deaths.

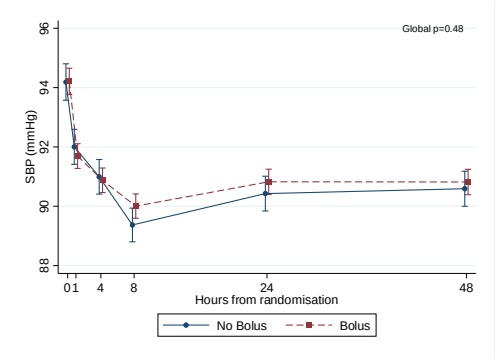
There was evidence for associations between bolus treatment and respiratory rate, glucose, heart rate, and haemoglobin over time (Figure 3.4.14). Those in the bolus arm had a higher mean respiratory rate at each time point compared to the no bolus arm (although the difference was <1 breath per min), and higher mean heart rate until 4 hours (a difference of 2 beats per min at 1 hour) but after this point were similar to the no bolus arm.

Children in the bolus arm had lower haemoglobin at each time point compared to the no bolus arm. The lower haemoglobin may have been due to blood transfusions for those with haemoglobin <5g/dl in the bolus arm being withheld until after the bolus had finished, but the difference between the arms remained when the model was restricted to children ≥5g/dl (where fewer children received an immediate transfusion) [197]. Some of the difference may also have been due to a dilution effect as haemoglobin was measured as g/dL and the children in the bolus arm had more fluid in their body by 8 hours. There was no difference in the number of children with hb<5g/dl who got transfused by 8 hours between bolus and no bolus arms (89% vs 91%; p=0.44). By 8 hours there were 34/357 (5%) with hb <5g/dl that died before they could have a transfusion in the bolus arms, and 14/332 (4%) that died before they could have a transfusion in the no bolus arm.

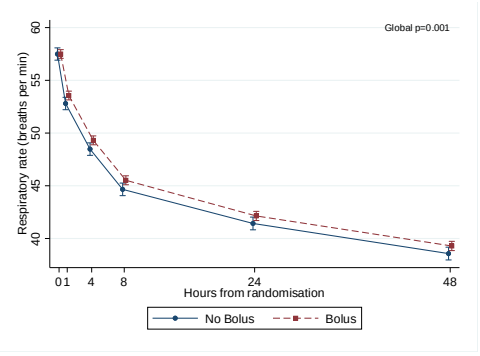
There was a dip in glucose seen in the bolus arm at 1 hour which could be due to metabolic stress; alternatively, the maintenance fluids containing glucose in the bolus arms only started at 1 hour (post bolus). Those in the no bolus arm, who received maintenance fluids from admission, had an increase in glucose post admission which returned to similar levels to the bolus arm at 8 hours. This increase in the no bolus arm compared to the bolus arm between baseline and 8 hours remained even when all children with hypoglycaemia (defined as <3mmol/L and who would have received a glucose correction according to the trial protocol) were removed from analyses. There were more children in the bolus arm that developed hypoglycaemia after admission at 1 and 4 hours (glucose <3mmol/L after a glucose level ≥3 mmol/L at admission) compared to the no bolus arm but the absolute numbers were low overall (38/1865 (2%) vs 9/931 (1%), exact p=0.04 at 1 hour; 20/1821 (1%) vs 3/921 (0.3%), exact p=0.04 at 4 hrs). After 8 hours the percentage of children developing hypoglycaemia was the same between arms, and there was no evidence the decrease and then increase over the first four hours seen in the bolus arms was more common in those that died.

Figure 3.4.14: GEE models for each measure over time split by whether they received bolus or not. P-values from a global Wald test

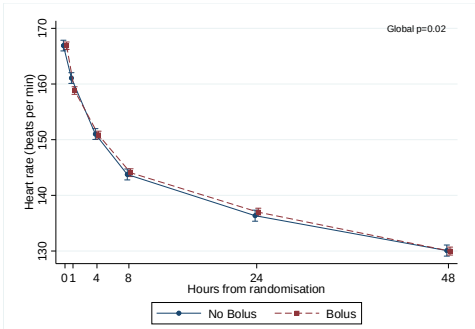
Systolic blood pressure over time



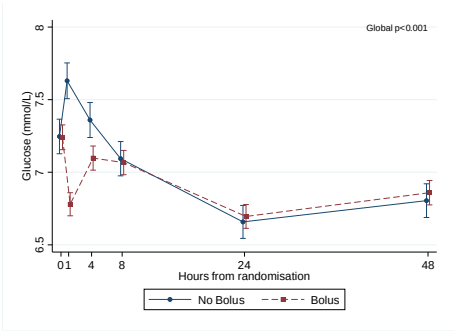
Respiratory rate over time



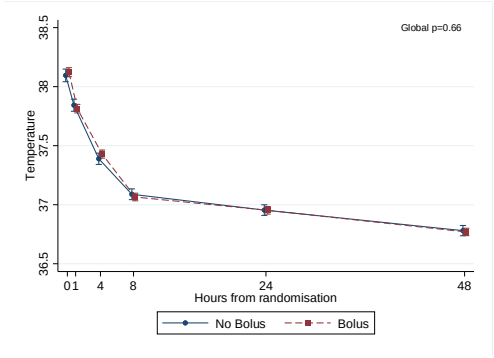
Heart rate over time



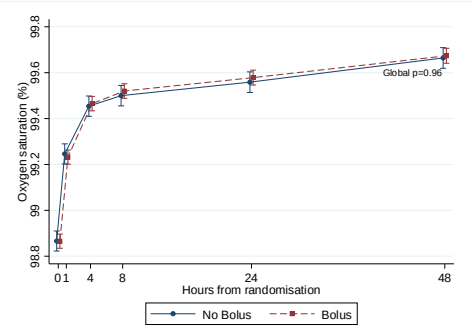
Glucose over time



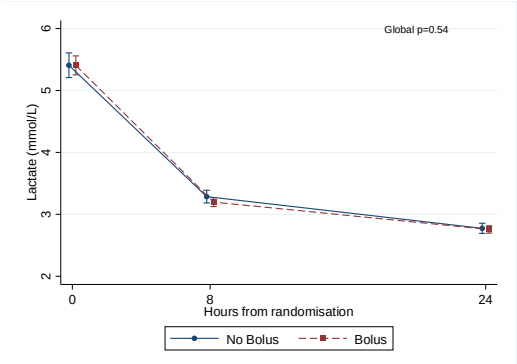
Temperature over time



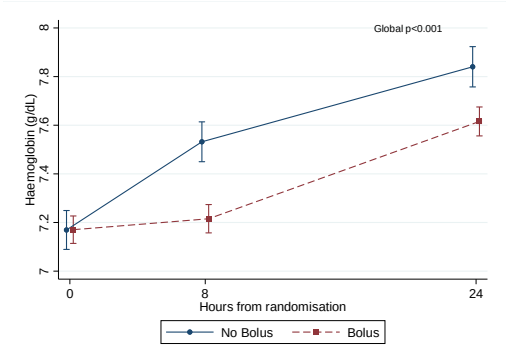
Oxygen saturation



Lactate over time



Haemoglobin over time



3.4.2.2 Comparing unadjusted and adjusted HRs, and estimating PTE

The estimated adjusted hazard ratio for bolus vs no bolus from time-updated models, including the time updated value and baseline value for each measure and subgroup, and the unadjusted hazard ratio for bolus vs no bolus for each subgroup (a model with just bolus vs no bolus restricted to children within each subgroup with the measure recorded) were then plotted together in Figure 3.4.15 below, and the PTE was calculated using the Chen method (Figure 3.4.16).

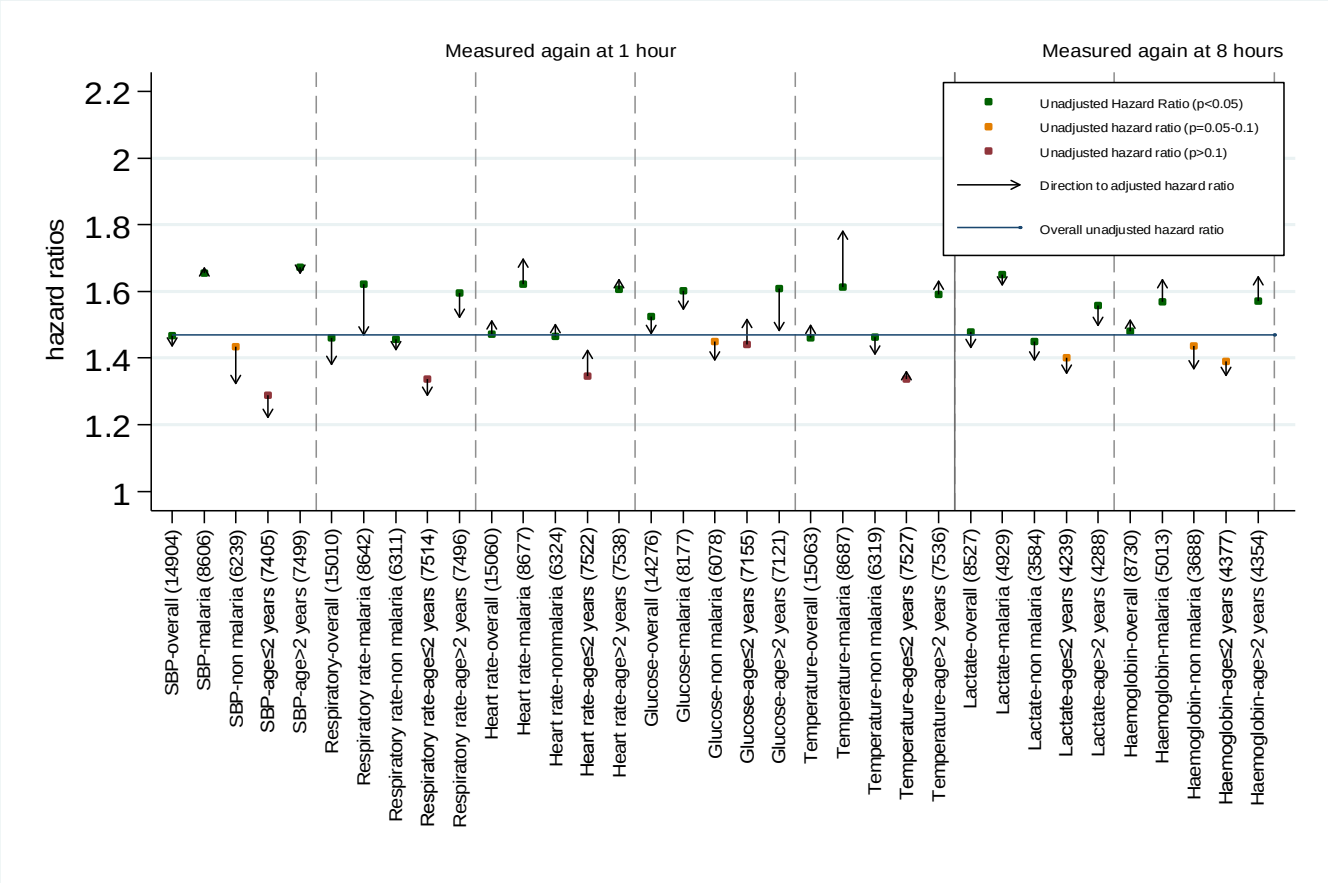
Figure 3.4.15 shows which measures, either overall or within a subgroup, when included in a model with bolus vs no bolus, moved the HR towards the null (which would have been an indication the measure may have been explaining some of the treatment effect). The square symbols represent the unadjusted hazard ratio for bolus vs no bolus with the population indicated in brackets on the x-axis. The population sizes varied reflecting how much missing baseline data there was for each variable. The point estimates were approximately the same level in each subgroup across the plot and were plotted in three different colours depending on the strength of the bolus effect on mortality as indicated by the results of the Wald test. The estimate of the hazard ratio was plotted in green if the Wald test $p<0.05$ from the unadjusted model, in orange if $p=0.05 – 0.1$, and in red if $p>0.1$. The solid line represents the main trial result and the arrows point to the estimated hazard ratio for bolus vs no bolus after adjusting for baseline and time-updated values using the functional form from Figure 3.4.12 and Figure 3.4.13. The first five measures (SBP, respiratory rate, heart rate, glucose and temperature) were recorded at baseline, 1, 4, 8, 24 and 48 hours from randomisation whereas the last two measures (lactate and haemoglobin) were only recorded at baseline, 8 and 24 hours from randomisation. The symbols and arrows match with the PTE plotted in Figure 3.4.15, for example, adjusting for respiratory rate moved the hazard ratios across all the groups towards the null and the corresponding PTEs were positive (which showed that a small amount of the bolus effect may have been explained by the measure); but adjusting for heart rate moved all the hazard ratios away from the null and the PTEs were negative (which showed none of the bolus effect was explained by the time-updated measure and adjusting for it gave a stronger effect of bolus).

The proportion of treatment effect (PTE) explained by time-updated values in each covariate measured over the first 48 hours of randomisation and the 95% CI were calculated using the

Chen method (Figure 3.4.16). For comparison, the PTE calculated using the unadjusted hazard ratio from a separate model (a time-updated model with bolus vs no bolus only as used in the LFD PTE method) was also plotted (as an x) showing that the methods gave close estimates (30/35 were within the 95% CI for the Chen method) but for the majority of measures (24/35 (69%)) the Chen method estimate was further away from the null (Figure 3.4.16). All the PTEs with confidence intervals that have upper and lower bounds outside of the interval [-100, 100] were from models where the unadjusted bolus effect was non-significant ($p \geq 0.05$ as indicated in Figure 3.4.15). These very wide confidence intervals, as well as the very large negative PTE estimate for glucose in the age ≤2 years group, may have been due to the fact there was no strong effect of bolus in these groups.

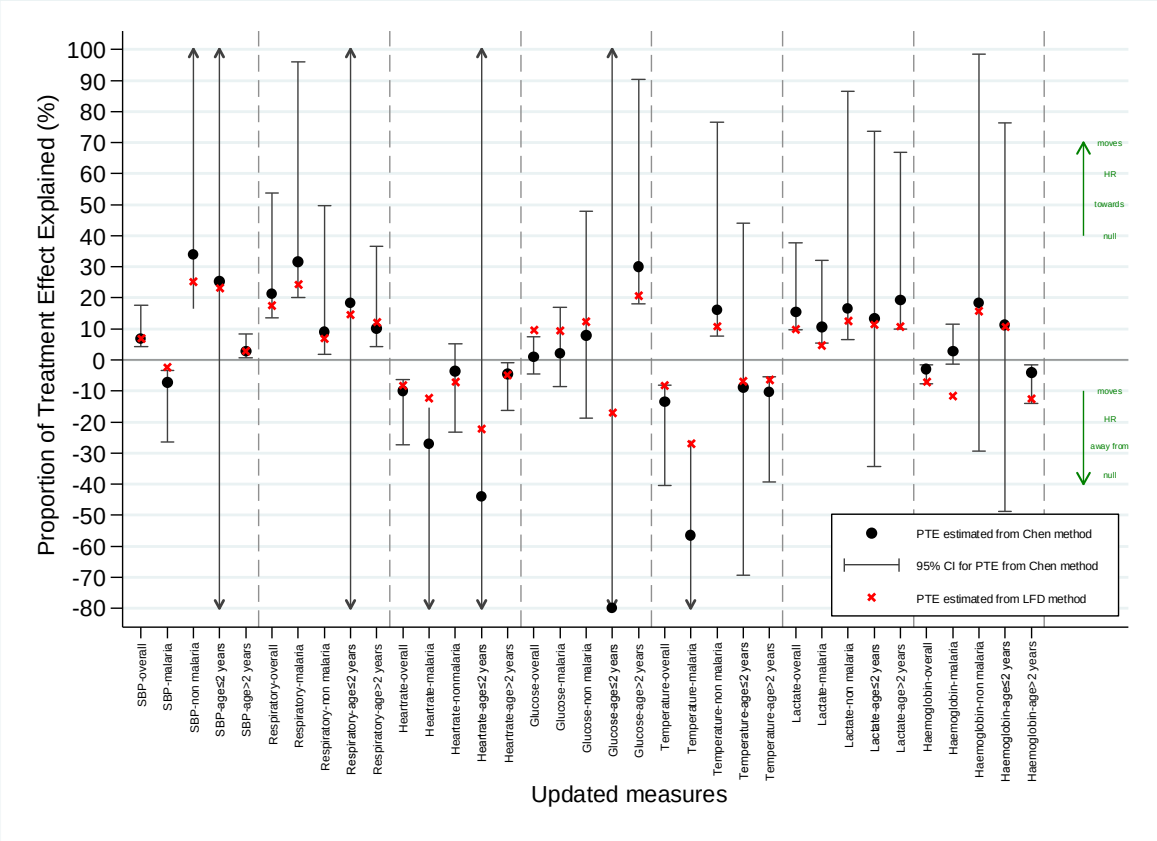
Oxygen saturation has not been included in Figure 3.4.15 or Figure 3.4.16 as the effect of bolus varied by baseline oxygen saturation, and appropriately adjusting for that interaction with baseline would lead to separate PTE estimates per level of baseline oxygen saturation. There was also evidence for an interaction between bolus and oxygen saturation after admission in children with malaria; thus, estimates of overall hazard ratios and a PTE were not appropriate. This is discussed further below.

Figure 3.4.15: Change in bolus vs no bolus hazard ratio once adjusted for the baseline and time-updated covariate.



Footnote: Overall there were 3141 children available for analysis, 1795 in malaria positive group, 1330 in non-malaria group, 1571 aged ≤2 years and 1570 aged > 2 years. They had up to 5 observations each for this analysis (0, 1, 4, 8, 24 hours (8 and 24 hours only for haemoglobin and lactate)). Each group for each covariate gave a slightly different unadjusted HR as there are a different number of observations in each model (depending on the size of the group (malaria, non-malaria, aged ≤2 or >2 years) and how much missing baseline information there is). The overall unadjusted HR plotted was 1.47.

Figure 3.4.16: Proportion of treatment effect explained by measures over time from randomisation.



None of the covariates in Figure 3.4.16 seem to explain the treatment effect; the largest PTE was 35% (from SBP in the non-malaria group) which was still small. For a measure to be considered a surrogate, Freedman *et al* [164] suggest the estimate would need to be around 75% or higher and the lower bound of the confidence interval would be expected to be above 50%. Some of the covariates, such as heart rate and temperature, even moved the hazard ratio away from the null giving negative values for PTE. Haemoglobin had very little impact on the treatment effect, and lactate, respiratory rate and glucose had PTE<33% in all groups. Lactate and respiratory rate gave consistent positive PTE values across all groups (all >10%); some also had confidence intervals excluding zero and so, of all the measures, these were the most likely to have explained a small amount of the treatment effect. But when analysing the association between bolus and lactate in a model with lactate as an outcome, I found that the bolus intervention had no impact on the mean lactate level at each time point (Figure 3.4.14). Lactate may still explain a small amount of the treatment effect but this may not have been seen in the GEE analyses, as the first post-randomisation time point was at 8 hours and so any impact of the bolus on lactate may have resolved by that time point. It may also have been that the impact of bolus on lactate was only in small groups of children at high levels of lactate. However, a quantile regression analysis of lactate at 8 hours, adjusted for lactate at baseline, showed no difference in the 90th percentile level between the bolus and no bolus group (p=0.79). A similar result was found for the 90th percentile level using quantile analysis for lactate at 24 hours (p=0.54). Absolute change from baseline in lactate at 8 and 12 hours was described as a useful surrogate for the effect of anti-malarial treatment on mortality in adults with severe malaria, in a reanalysis of the AQ study in Vietnam, with PTEs of 73% and 77% [198, 199]. However, serial lactate measurements were not available to calculate PTEs from a similar dataset in children with severe malaria (AQUAMAT study) [199]. Quantile regression of lactate at 8 hours adjusting for baseline lactate restricted to those in the malaria group only also showed no impact of bolus on the 90th percentile of lactate in this subgroup (p=0.68).

Respiratory rate differed significantly between treatment arms over time (Figure 3.4.14) and may have explained a small amount of the treatment effect but the PTE estimates within the groups and overall were all <33%.

3.4.2.3 Interactions

Each model was investigated to identify whether the effect of bolus vs no bolus on mortality risk varied over different values of the time-updated measures adjusted for the baseline measures (Table 3.4.8). Out of 40 tests, five gave p<0.1 and one gave p<0.05 giving a rate of 1/40 (3%) for significant interactions.

Table 3.4.8: Tests for heterogeneity between time-updated measures and bolus effect.

	Interaction p-values from each model				
	Overall	Malaria	Non-malaria	Age ≤2 years	Age > 2 years
Heart rate	0.38	0.81	0.39	0.85	0.49
Respiratory rate	0.14	0.58	0.12	0.53	0.25
Glucose	0.46	0.33	0.47	0.33	0.17
Temperature	0.94	0.60	0.61	0.54	0.91
Systolic Blood Pressure	0.41	0.50	0.58	0.23	0.09
Haemoglobin	0.95	0.76	0.73	0.88	0.83
Lactate	0.06	0.17	0.08	0.85	0.06
Oxygen saturation*	0.49	**0.02**	0.28	0.39	0.92

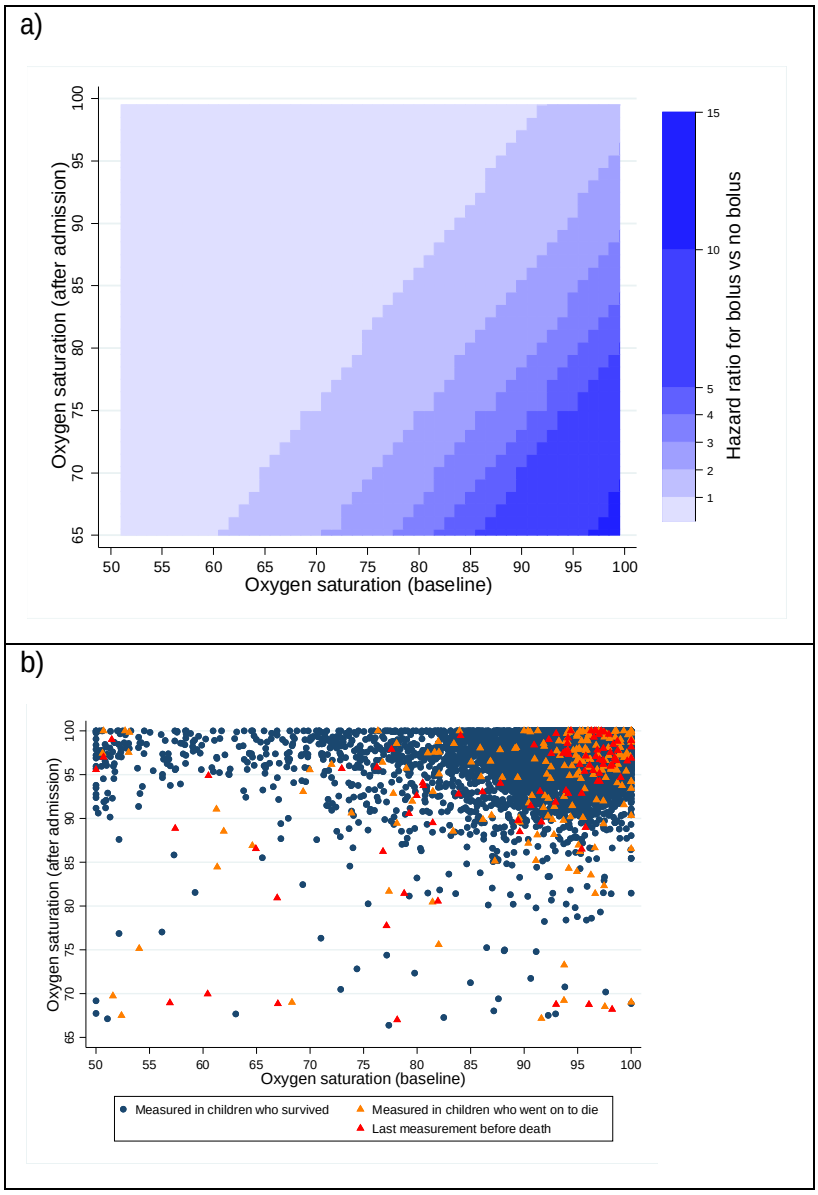
* Model for oxygen saturation only additionally adjusted for interaction between bolus and baseline measure of oxygen saturation

3.4.2.3.1 Interaction between time-updated oxygen saturation adjusted for baseline interaction

There was some evidence that the effect of bolus vs no bolus on mortality risk varied according to oxygen saturation as a time-updated covariate (i.e. an interaction between bolus and time-updated values) in the malaria group, adjusting for the previously identified interaction between baseline oxygen saturation and bolus vs no bolus (p=0.02) (allowing the effect of bolus vs bolus on mortality risk to vary by oxygen saturation at baseline as well as by time-updated oxygen saturation). There was no evidence that the effect of bolus vs no bolus on mortality risk over time-updated oxygen saturation varied differently between the malaria groups (p=0.51 for a three-way interaction). The baseline interaction remained significant in the presence of the time–updated interaction (p=0.02).

The functional form selected to best describe the relationship between time-updated oxygen saturation and mortality risk varying by bolus vs no bolus, in a model adjusted for a baseline interaction between oxygen saturation and bolus, was a quadratic function. The AIC of the selected model was estimated and compared to a model with a linear function for time-updated oxygen saturation in an interaction with bolus vs no bolus and they were found to be very close (1754 vs 1757). The interaction also still held with oxygen saturation as a linear function (p=0.01). To simplify visualising the varying effect of bolus vs no bolus on mortality risk over oxygen saturation, a model with time-updated oxygen as a linear function in the interaction with bolus was used. The bolus vs no bolus effect on mortality risk from a model with the baseline interaction and time-updated interaction was plotted as a contour plot (Figure 3.4.17). The different shades of blue on the contour plot indicate different levels of risk: the darkest represents highest risk (hazard ratio ≥10) and the lightest shade represents lowest risk (hazard ratio <1) with oxygen saturation after admission on the y-axis and oxygen saturation at baseline on the x-axis.

Figure 3.4.17: a) Contour plot of bolus vs no bolus effect on mortality risk across baseline and time-updated (after admission) oxygen saturation levels in children with malaria and b) scatter plot of distribution of oxygen saturation at baseline and oxygen saturation after admission in children with malaria.



The interaction detected in the time-updated model was that the bolus effect increased as oxygen levels decreased in children with malaria, so the children with low levels of oxygen saturation and malaria after baseline had a worse effect of bolus compared to those with high levels of oxygen saturation and malaria after baseline (Figure 17a). This interaction was not in the same direction as the baseline interaction (if they had been in the same direction then the

highest risk would have been in the top right hand corner of the contour plot). The bolus had least effect on those with poor lung function (low oxygen saturation) and malaria at baseline whether it remained poor or improved over time, whereas those with good lung function (normal oxygen saturation) and malaria at baseline which then deteriorated, had highest risk from the bolus intervention. However, the scatter plot shows that there were few data in children with high baseline oxygen saturation and low post admission oxygen saturation. There were 88/1500 (6%) children with malaria and oxygen saturation <90% after admission who had previously had ≥90% on admission (out of those with malaria and oxygen saturation ≥90% on admission), and of these 24 had oxygen saturation <80% after admission. There was only evidence for an interaction with bolus with post-admission values in the malaria group. Given the number of tests performed, this could be due to chance.

3.4.2.4 Multivariable model for estimating PTE

All the previous models looked at each clinical measure individually but it was also important to consider them in a model together and to understand if, together, they could explain more of the treatment effect. The Chen method was used to estimate PTE from a model with respiratory rate, heart rate, SBP, temperature, glucose, haemoglobin and lactate as time-updated values adjusted for their baseline values. Confidence intervals were estimated bootstrapping 10,000 replications around the Chen estimate of the PTE. An estimate of PTE calculated using the LFD method was also plotted (Figure 3.4.18).

Figure 3.4.18. Proportion of treatment effect explained by all measures combined over time from randomisation.

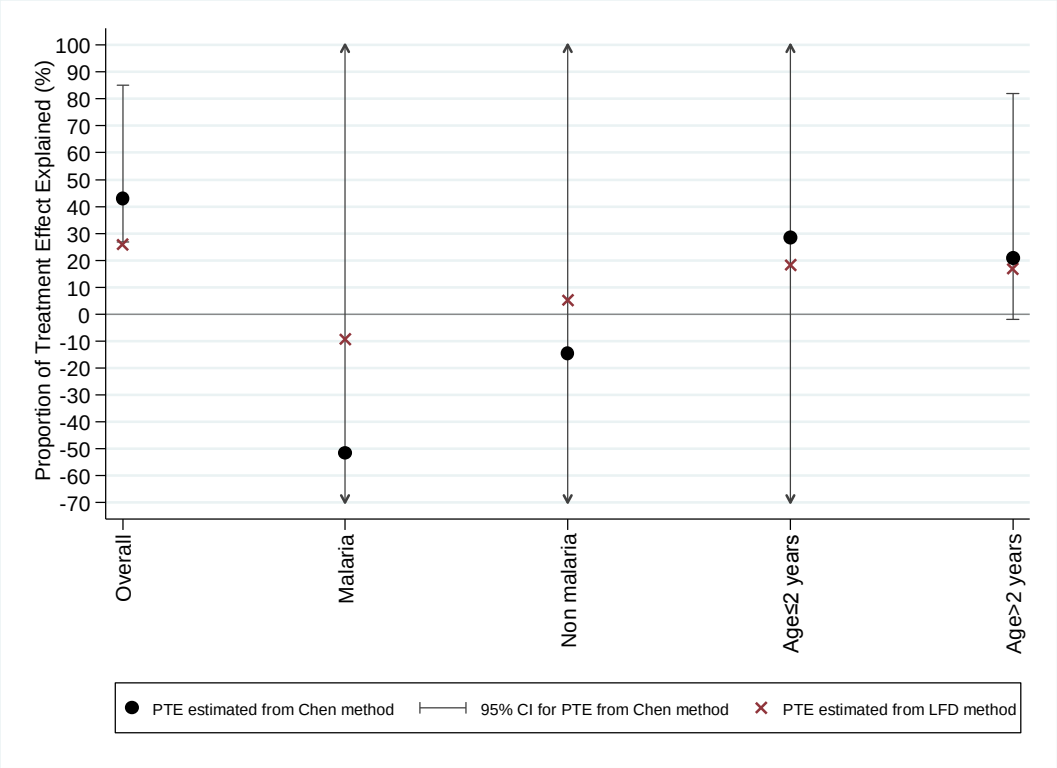


Figure 3.4.18 shows that only 43% of the treatment effect was explained in the trial population using the time-updated measures together. The malaria and non-malaria groups had a negative PTE indicating that adjusting for all the clinical measures actually increased the bolus effect in these groups. The PTEs estimated from the LFD method were all <30% and closer to 0% than the PTEs from the Chen *et al* method indicating that there was relatively little change in the hazard ratio for bolus vs no bolus after adjustment.

One reason that the estimates for the PTEs in malaria and non-malaria both had wide confidence intervals could be due to the fact that this model ignores the baseline interaction that bolus had with base excess and malaria. Thus the PTE was calculated just for the significant bolus effect for low base excess (defined as <-8mmol/L) in children with malaria as presented in Table 3.4.2 (page 175) (3.01 95% CI (1.59, 5.71)). The PTE of the significant bolus effect for low base excess in children with malaria, estimated using all the time-updated measures above, was then estimated to be -73% (95% CI -236% to -18%). This can be compared to -51% for children with malaria in Figure 3.4.18, and shows that by ignoring the interaction the model may have underestimated how much the adjustment for all the clinical

measures moved the hazard ratio **away** from the null in the malaria group (rather than towards the null as would be expected if the variables were genuinely explaining the increased risk associated with boluses). As there was no evidence for an effect of bolus in the other groups (Table 3.4.2) the PTEs were not estimated.

The model for Figure 3.4.18 also ignored the baseline interaction with oxygen saturation, and so PTE was estimated restricted to those with oxygen saturation ≥90%, for bolus vs no bolus on mortality risk from all the time-updated factors above and for the groups in which there was a significant effect of bolus. Overall, the estimated PTE by clinical measures for those with high oxygen saturation (Table 3.4.9) was less than that estimated for the trial population, but even in this subgroup all measured clinical measures combined were not able to explain much of the treatment effect. The treatment effects shown in Table 3.4.9 were calculated for each group from a model with no missing baseline values for all the clinical measures and the estimates were consistent in showing an increase in mortality risk from boluses in all groups.

Table 3.4.9: Proportion of treatment effect in children with oxygen saturation 90-100% explained by all measures combined over time from randomisation.

	Oxygen saturation 90-100%		
Group	HR for bolus vs no bolus [a]	Deaths/children in each model	PTE by clinical measures
Overall	1.83 (1.19, 2.82)	120/2104	11% (95% CI -11- 40%)
Malaria	2.50 (1.31, 4.78)	66/1270	40% (95% CI 22-75%)
Non-malaria	1.35 (0.74, 2.46)	54/831	Not estimated [b]
Age ≤ 2 years	1.84 (0.97, 3.47)	58/1037	Not estimated [b]
Age > 2 years	1.83 (1.01, 3.32)	62/1067	44% (95% CI 17-149%)

[a]These are estimated from the time-updated model with complete baseline clinical measures.

[b]The PTEs were not estimated in this group as the bolus effect was non-significant.

3.4.3 Malaria Sensitivity Analyses

Samples from the FEAST participants were tested for P*f*HRP2 protein to better describe the parasite burden in children enrolled in the trial, especially those with malaria as described in the methods section of this chapter, page 154. A substantial proportion of samples were not taken at sites during the trial and so there were only 2216/3141 (71%) available for analysis. Of these samples 1549 (70%) tested positive for any P*f*HRP2 protein. P*f*HRP2 levels of >1000ng/ml have been shown to denote a malaria attributable fraction of severe disease of 99% (95% credible interval 96%-100%) [182].

Thus, I considered an alternative definition of severe malaria (P*f*HRP2>1000ng/ml) to try to better distinguish between the malaria and non-malaria groups. The new definition is outlined below in relation to the statistical analysis plan (SAP) definition of malaria (Table 3.4.10).

Table 3.4.10: SAP defined malaria compared to a definition of severe malaria in children with P*f*HRP2 protein measured.

	Malaria by P*f*HRP2			
Malaria (SAP)	Negative (PfHRP2 per ng/ml = 0) *(% row)*	Low levels of PfHRP2 (P*f*HRP2 per ng/ml 1 - ≤1000) *(% row)*	Severe malaria (P*f*HRP2 per ng/ml>1000) *(% row)*	Total
Negative (% col)	572 (86%) *(60%)*	308 (36%) *(33%)*	70 (10%) *(7%)*	950 (43%) *(100%)*
Positive (% col)	95 (14%) *(7%)*	543 (64%) *(43%)*	628 (90%) *(50%)*	1266 (57%) *(100%)*
Total (% col)	667 (100%) *(30%)*	851 (100%) *(38%)*	698 (100%) *(32%)*	2216 (100%) *(100%)*

Mortality at different levels of P*f*HRP2 was also examined (Table 3.4.11).

Table 3.4.11: Mortality at different levels of P*f*HRP2.

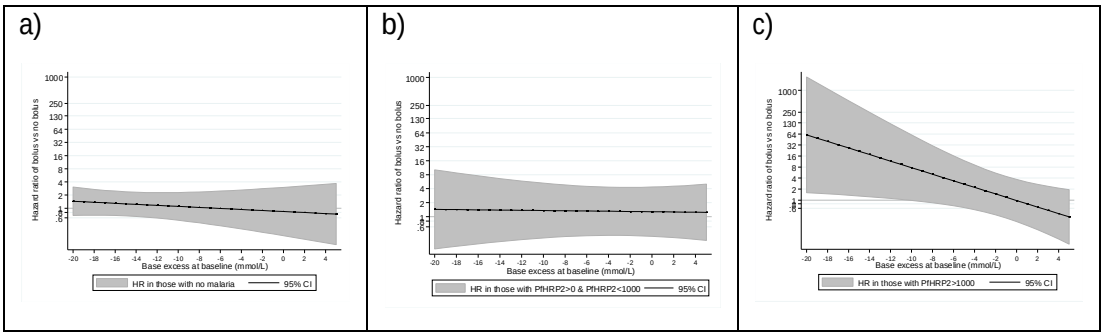
Category of P*f*HRP2	Bolus group (% deaths)	No bolus group (% deaths)	Overall (% deaths)
>1000 ng/ml	43/463 (9%)	15/235 (6%)	58/698 (8%)
1-≤1000 ng/ml	49/569 (9%)	15/282 (5%)	64/851 (8%)
None (0)	64/437 (15%)	15/230 (7%)	79/667 (12%)
Total	156/1469 (11%)	45/747 (6%)	201/2216 (9%)

A model restricted to those with P*f*HRP2-defined severe malaria showed that the impact of bolus changed over different levels of base excess (p=0.01) consistent with the previous finding in this chapter. A test of a three-way interaction between bolus, severe malaria (yes vs no, i.e. combining low and zero P*f*HRP2) and base excess confirmed that the impact of bolus changed over base excess levels in a different way in the severe malaria group to the non-severe-malaria group (p=0.04).

The categorisation of P*f*HRP2 as described in Table 3.4.11 was included in a model in all children with P*f*HRP2 measured as part of a three-way interaction with bolus and continuous base excess. There was some evidence that the impact of the bolus changed over different levels of base excess in a different way across the three categories of P*f*HRP2 (p=0.06). The plots of the hazard ratio for bolus vs no bolus at each level of base excess within each P*f*HRP2-defined group demonstrated that it was the category with the highest P*f*HRP2 levels that had the most detrimental effect of bolus at lowest levels of base excess (Figure 3.4.19).

The estimate of the harmful effect of bolus was very high for low levels of base excess when P*f*HRP2>1000 (HR=72.12 when base excess was -20) but the confidence intervals were also very wide (95% CI 2.04, 2546).

Figure 3.4.19: Hazard ratio for bolus vs no bolus over levels of base excess in a) those with no malaria defined by P*f*HRP2<0, b) those with P*f*HRP2 1-≤1000 ng/ml and c) P*f*HPR2 >1000 ng/ml.



The evidence that the impact of the bolus changed over base excess according to the level of P*f*HRP2 was also examined with the binary categorisation of base excess used earlier in the chapter, which also showed the increased effect of receiving bolus on mortality for those with severe malaria and severe acidosis (Table 3.4.12).

Table 3.4.12: Effect of bolus on mortality risk by base excess level and malaria status (defined by P*f*HRP2) at baseline.

	Base excess <-8 (acidosis)		
	Bolus (% died)	No bolus (% died)	HR for Bolus
Severe malaria (p*f*HRP2>1000ng/ml)	**27/159 (17%)**	**2/65 (3%)**	**5.98 (1.42, 25.13)**
Low levels of P*f*HRP2 (1-≤1000ng/ml)	26/147 (18%)	8/74 (11%)	1.70 (0.77, 3.75)
Non-malaria (p*f*HRP2=0)	30/117 (26%)	10/42 (24%)	1.10 (0.54, 2.26)
Total	83/423 (20%)	20/181 (11%)	
	Base excess ≥ - 8		
	Bolus (% died)	No bolus (% died)	HR for Bolus
Severe malaria (p*f*HRP2>1000ng/ml)	4/145 (3%)	2/82 (3%)	1.48 (0.38, 5.72)
Low levels of P*f*HRP2 (1-≤1000ng/ml)	10/215 (5%)	2/103 (2%)	2.43 (0.53, 11.09)
Non-malaria (p*f*HRP2=0)	7/186 (4%)	3/116 (3%)	1.12 (0.21, 6.11)
Total	21/546 (4%)	7/300 (2%)	

NB: Numbers in bold indicate a significant effect of bolus

P*f*HRP2 was also included in a model as a continuous measure, truncated at the 95th percentile, to describe severity of malaria in more detail. Fractional polynomials (fp) were used to estimate the most appropriate association between P*f*HRP2 >1000ng/ml as a continuous function and mortality, adjusting for P*f*HRP2≤1000ng/ml as a binary factor and for effect of bolus vs no bolus. The binary cut-off of 1000ng/ml was chosen to avoid issues with the mfp function when P*f*HRP2=0. This showed that a linear function of P*f*HRP2 was most appropriate. However, there was very little difference in mortality risk between high and low P*f*HRP2 above 1000ng/ml as a continuous function adjusting for base excess and whether P*f*HRP2 was ≤1000ng/ml or not. This was in contrast to a previous study of P*f*HRP2 which modelled it on the log scale and showed a U-shaped association with mortality [41]. This may be due to the authors not including binary factor to reflect when P*f*HRP2=0 in their estimation of a fp function, or that they had fewer children with P*f*HRP2≤1000ng/ml in their study compared to the FEAST trial and that children without malaria have other more severe diseases and usually have a higher mortality. A three-way interaction with P*f*HRP2 (both continuous as a linear function and binary (cut-off at ≤1000ng/ml)), bolus and continuous base excess as a linear function was tested with a likelihood ratio test and gave p=0.006.

Total parasite burden was then calculated from P*f*HRP2 levels dependent on the child's weight and haematocrit and truncated at the 95th percentile. Haematocrit in the FEAST trial was collected using the i-STAT cartridge and so there was a high level of missingness for this measure; thus calculated total parasite burden was available in only 1228/3141 (39%) children (94 deaths) (1228/2216 (55%) of those with P*f*HRP2 measured).

Total parasite burden as a binary variable (split at the 33rd percentile of parasite burden>0, giving a cut-off of 0.22 x10^{12}) and as a continuous variable (when >0.22 x10^{12}) was included in a model adjusted for the effect of bolus vs no bolus. Fractional polynomials were used to estimate the most appropriate association between parasite burden>0 as a continuous function and mortality and found a linear association between mortality risk and parasite burden, with a slight non-significant increased risk at high levels of parasite burden (HR 1.20 per 1x10^{12} increase (95%CI 0.99-1.48); p=0.07). Base excess was added to the model as a linear function (HR 0.86 (95% CI 0.84-0.89); p<0.001) and the linear association between continuous parasite burden and mortality risk remained, although with less evidence for an association (HR 1.12 per 1x10^{12} increase (95% CI 0.92-1.37); p=0.26).

A likelihood ratio test for both three-way interactions between 1) parasite burden (as binary), bolus and base excess, and 2) parasite burden (continuous), bolus and base excess gave p=0.0003 (Figure 3.4.20a). Each three-way interaction was also tested separately (p<0.001 for the binary variable three-way interaction with bolus and base excess, p=0.09 for the continuous variable three-way interaction). This confirms what was found with the SAP definition of malaria on page 154 of this chapter. Base excess was weakly inversely correlated with parasite burden (Spearman's rho=-0.14; p<0.001) and P*f*HRP2 (Spearman's rho=-0.12; p=0.006) (Figure 3.4.20b and 3.4.21) and children with P*f*HRP2>1000ng/ml had a lower mean base excess at admission (-9.3 vs -8.1; t-test p=0.003). P*f*HRP2 and parasite burden were highly correlated as expected (rho=0.97, p<0.001) (Figure 3.4.21). Analyses using either P*f*HRP2 or parasite burden both confirm the increased mortality risk from bolus at low levels of base excess when the children had malaria, but also showed that there was a dose-response impact that the more severe the malaria (i.e more parasitaemia) the higher the impact of bolus on mortality risk at low levels of base excess.

Figure 3.4.20 a) Impact of bolus vs no bolus at different levels of base excess and parasite burden, b) correlation between base excess and parasite burden.

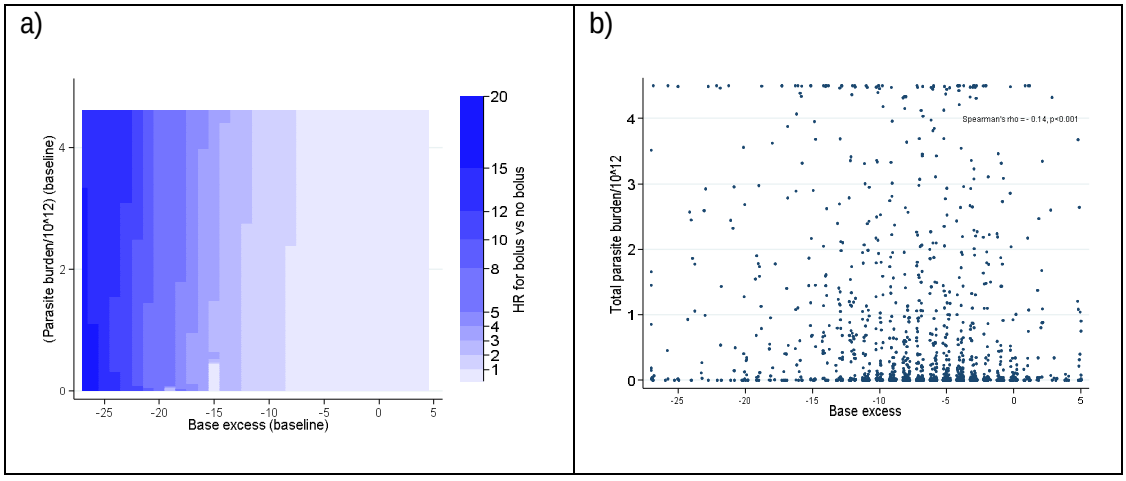
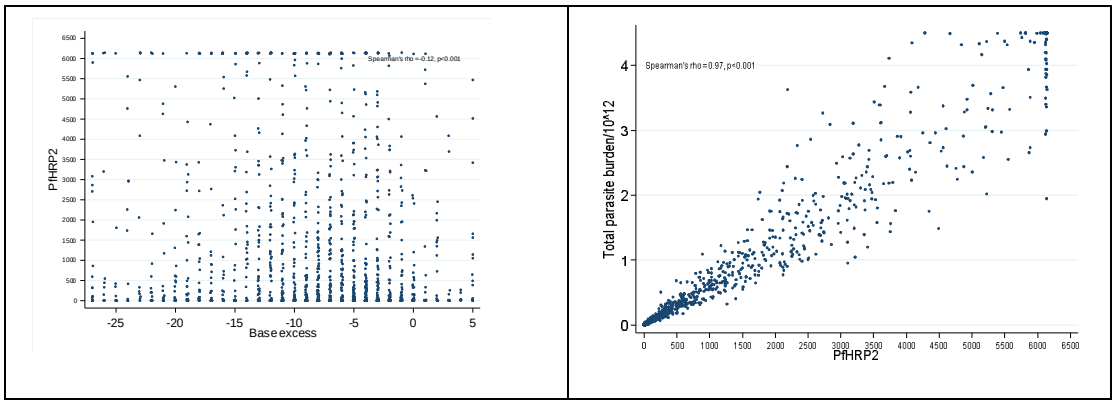


Figure 3.4.21: Correlations between P*f*HRP2 and base excess, total parasite burden and base excess, and total parasite burden and P*f*HRP2.



The Cox models for time-updated analyses were also restricted to those with severe malaria defined by P*f*HRP2>1000ng/ml. The best fitting functions to describe the association between time-updated measures and mortality risk, adjusting for baseline measures, in children in severe malaria were compared to those found for children with the SAP definition of malaria. For 6/8 measures the functions selected for the association were the same as those found in the SAP malaria group. The 2 measures for which the selected functions were different were glucose and haemoglobin. The AICs of the models with glucose and haemoglobin in severe malaria were compared to models using the same function as had been identified in the SAP malaria group and found to have a difference of <3.84, giving no evidence that the association between mortality risk and the time-updated measure was different in the severe malaria group compared to the SAP malaria group.

Restricting to the severe malaria group, the unadjusted hazard ratio was 1.46 (0.81-2.63) for the effect of bolus vs no bolus on mortality risk. As there was no evidence of an effect in this subgroup, the PTE estimates are likely to be very unreliable and less interpretable and thus were not calculated.

3.5 Discussion

The analyses in this chapter showed that no one clinical measure recorded during the first 48 hours of admission to hospital was able to explain the harmful effect of boluses in the FEAST trial. It also showed that combining the clinical measures during admission also could not explain the bolus effect that was seen in this trial.

The exploratory analyses using baseline measures showed that there was not always a linear relationship between mortality risk and values of baseline measures (especially for SBP, glucose and sodium with non-monotonic risk functions) and so modelling these with fractional polynomials was important. By analysing these measures in the most appropriate functional form, I was able to identify the values where children were at greatest risk.

The analyses confirmed that children with a high fever on admission were at lower mortality risk compared to those with hypothermia, suggesting a clinical benefit of fever which has been previously found in similar clinical settings [47]. But I also found that mortality risk at high values of temperature after admission did increase mortality risk indicating that development of fever or remaining at a high temperature after treatment had started was harmful, and this potentially identifies children for which treatment was not working leading to excess mortality risk. Thermometers are easily available in these settings and temperature can be easily monitored after admission, suggesting that continuation or development of a high fever after baseline could be a valuable tool to identify antimicrobial (for both malaria and bacteria) treatment failures (incorrect dosages/antibiotics given or antimicrobial resistance). This could be done at a specific time point (which could be 48 hours as this is when treatment failure is often defined) and identify children at increased risk of mortality after admission and could result in tailoring of treatment, especially where microbiological services are lacking.

BUN at baseline has been dichotomised for risk scores in other studies at 20, with values >20 defined as being at increased risk of mortality [103]. I found that mortality risk was high above this value and did not increase with values greater than 20, but I also detected a lot of variation in mortality risk below that value which would not be captured using a cut-off of 20.

The FEAST trial showed that boluses were harmful and the categorical subgroup analyses presented with the main trial results all supported that finding; there were no subgroups in which there was any evidence of a benefit [33]. In this chapter instead I tested interactions

with continuous variables and found that boluses were more harmful in children with good initial lung function, defined by high levels of oxygen saturation (≥90%) in all children at admission, with some weak evidence that this effect was also present within all subgroups of children analysed in this chapter (malaria, non-malaria, age≤2 years, and age>2 years). This interaction has been reported previously [157], although no reasons were put forward at that time, and there was no indication from the endpoint review committee reviews that fluid overload on the lungs was the main mechanism by which the bolus was causing harm [157]. If immediate fluid overload had been one mechanism by which the boluses were causing harm (which would correspond to a respiratory TCE close to bolus administration) then it would be expected to have an even greater impact on children with poor lungs at baseline, as indicated by low oxygen saturation (<90%). However, it was children with **high** oxygen saturation at baseline in which boluses had greatest detrimental impact, thus making clinical interpretations difficult.

The analyses in this chapter go further in describing how the bolus effect changed over different levels of oxygen saturation after admission; it was particularly those children with malaria and low values of oxygen saturation after admission that had the greatest detrimental impact of bolus, particularly in those who had high values of oxygen saturation at baseline. In terms of cause of death, children with ≥90% oxygen saturation at baseline had a higher proportion of deaths adjudicated as respiratory in the bolus arms compared to the no-bolus arm. However, they also had a higher proportion of deaths adjudicated as cardiogenic (or shock). Analyses using oxygen saturation as an outcome and looking at the impact of bolus showed that the mean oxygen saturation increased in both groups after baseline, and there was no difference between the arms; however, there was a difference between arms in respiratory rate, but this was very small in magnitude (<1 breath/min). Thus, there was not strong evidence that boluses worsened lung function (as measured by oxygen saturation) immediately in all children receiving boluses. Rather, boluses appeared to delay the rate of recovery in respiratory rate (consistent with findings in Chapter 4), and then led to a respiratory or cardiogenic TCE. The boluses may also have had a greater harmful effect in children with a normal oxygen saturation level at baseline as some of these children may have had less severe shock (as measured by the number of features of impaired perfusion (although 52% still had 2 or more)) and thus a bolus of fluid was not acting to improve perfusion as it had done in prior studies and overall in the FEAST trial [30, 157]. Although there was little indication that it was fluid overload on the lungs that directly led to deaths in the bolus arms, it

may be that some mechanism involving the lungs further along in the admission was responsible, and that it was children with good lung function at randomisation (as measured by oxygen saturation) that were most adversely impacted by bolus.

One of the findings of the paper looking to explain the mechanisms of mortality in the FEAST trial was that the cause of death with the greatest difference between arms was the cardiogenic TCE. Although base excess was not used in the definition of TCEs, because it was not sufficiently measured after baseline (67% missing at the only other timepoint (24 hours)), it was used in the definition of severe shock at baseline (base excess<-8mmol/L, or lactate≥5mmol/L, or WHO shock or moderate hypotension (SBP 50-75, 60-75, 70-85mmHg in children <12 months, 1 to 5 years and >5 years respectively)). Those with severe shock had the highest mortality risk and, consequently, largest absolute risk differences for bolus vs no bolus (7.0% compared to the overall risk difference of 3.3%) [157]. Base excess has not been modelled as a continuous variable previously, thus the evidence for an interaction found in these analyses may have been driving the previous finding. The baseline presentation of shock/acidosis, rather than respiratory or neurological syndromic presentations, was also closely aligned with the cardiogenic TCE.

The analyses in this chapter showed that the bolus effect was highest in those with lowest base excess (particularly when base excess<- 8 (acidosis)) but only when children also had malaria. For children with no malaria the harmful effect of bolus was the same across all values of base excess. Sensitivity analyses using *Pf*HRP2 showed the difference in bolus effect by level of base excess held when analyses were restricted to children with severe malaria (using a strict definition incorporating the degree of sequestered parasite load) and also held when modelled with total parasite burden (i.e a continuous measure of malaria instead of a binary indicator). The analyses also confirmed that the bolus impact increased as parasite burden increased, showing a higher impact of bolus in more severe malaria. Children needed to have had both malaria and severe acidosis for the bolus to have had the increased detrimental impact, in contrast to having had just malaria or just severe acidosis. For those with severe malaria I postulate that acidosis was secondary to deranged microcirculation in addition to the macrovascular derangements (signs of shock) and thus was more likely to be contributing to the excess mortality. However, another contributing mechanism is red blood cell deformability (occurring in both parasitised and non-parasitised red cells) [200], which has been shown to be a prognostic factor for mortality and more important than other parameters such as lactate

and acid/base status [201]. It is also correlated with base excess [201]. The bolus may have increased the ability of these cells with decreased deformability or increased rigidity (healthy red blood cells have good deformability) to circulate, reducing microcirculatory flow and oxygen delivery in the tissues.

The varying impact of the bolus by baseline excess value in children with malaria was also found to be independent of the varying impact of the bolus by oxygen saturation levels, identifying two different groups of children with increased risk of mortality from bolus administration. But where the two groups overlapped the impact of bolus on mortality risk was even higher.

The time-updated analyses showed that the risk of mortality was highest when clinical measures had not normalised after admission i.e. not returned to healthier values. For SBP, temperature, heart rate and glucose risk was increased at both high and low post-baseline values. For respiratory rate the gradient of the linear slope for mortality risk was steeper for the post admission values, showing that if respiratory rate remained high or increased since baseline then mortality risk was increased even more than equivalent differences in baseline values.

None of the clinical or laboratory measures individually explained a large proportion of the bolus treatment effect: the largest of any group was SBP in non-malaria at 34%. For a measure to be considered a candidate as a surrogate marker it is generally considered that it would need to explain around 75% of the treatment effect or higher with a lower confidence interval bound > 50% [164]. Glucose and lactate in the age > 2 years group and respiratory rate in the malaria group had positive PTEs of 30% (95%CI 18, 90%), 32% (95%CI 20, 96%) and 18% (95%CI -30%, 98%) respectively. Further, in all groups respiratory rate and lactate consistently had positive PTEs, although some confidence intervals were wide. It is important that the bolus should also have an impact on the measure considered to be explaining some of the treatment effect [202], and this was shown for respiratory rate, although with a very small absolute difference between groups (<1 breath per min). There was no evidence for a difference between arms in lactate after admission, both using GEEs for a difference in means between the two groups but also analysing the difference in the 90th percentile. But it may be that there was a short impact of bolus on lactate occurring immediately after the bolus that I could not detect, as the first post admission lactate time point was 8 hours, especially given that the small effect of treatment on respiratory rate was mainly seen at 1 and 4 hours. Thus it may be

that I failed to identify lactate as a surrogate due to a mismatch between timing of deaths and of subsequent lactate measurements. Although respiratory rate had positive PTEs across the subgroups and overall, and an association with bolus effect over time (but with minimal absolute difference between the means of <1 breath per min), there was no strong evidence this explained any of the bolus effect.

Jeeyapant *et al* suggested that changes in lactate assessed at 8 or 12 hours were valid surrogates for the treatment effect of anti-malarials on mortality in adults with severe malaria, with a PTE for absolute change in lactate of 73% at 8 hours [199], although the confidence interval for the estimate was wide (-607%, 673%). The authors suggest that this PTE was due to anti-malarials improving the microcirculatory flow which then reduces plasma lactate and mortality. There was no evidence in our malaria group that lactate correction/normalisation was a suitable surrogate for bolus treatment, but that may have been due to the differences in mechanisms of action between anti-malarials and boluses. Also, if boluses had an impact, it may have occurred prior to the first measurement at 8 hours, so it was not possible to see the change in lactate between the bolus and no bolus groups with the data that was collected. I found evidence that the parasite burden had a small but significant correlation with base excess (another measure of degree of acidosis) (rho=-0.15, p<0.001) and with lactate (rho=0.33, p<0.001) and so for children with malaria the levels of parasite burden may be explaining some of the metabolic acidosis. But across the whole trial the number of children with parasitaemia may not be large enough to show an impact of bolus on lactate as a surrogate for the effect of bolus treatment on mortality. Also, it has been shown that although lactate acid contributes to metabolic acidosis, it only accounts for one quarter of the strong anion gap [203] and that lactate acidosis of malaria has multiple aetiologies [204].

The results of the PTE analyses indicated that the mechanism by which the bolus caused harm was not by one route in one organ that can be measured easily by changes in recorded clinical measures. Even combining all measures did not explain the treatment effect, indicating there may have been unmeasured clinical changes taking place that would better reflect the mechanism.

A limitation of the analyses was that the bolus was given at baseline, and for some children again at 1 hour, but the next recorded measurements were at 4 and 8 hours. The effect of the bolus may have been better explained by the values that were not recorded in between these points, especially when 167/297 (56%) of the deaths occurred by 8 hours. Lactate and

haemoglobin were only measured at 8 and 24 hours after admission so there was even less information on these measures. Also, the measures collected and the times they were collected were focussed on clinical care and with the expectation of a benefit from boluses, and different measures at other time points could have been recorded which may have been better at explaining the harmful treatment effect.

A second limitation is missing data in some of the laboratory measures at baseline. Although multiple imputation has given more information about mortality risk at different levels of baseline measures, which showed that for most groups the complete case and MI analyses were very similar, it also had its limitations and possible interactions were not able to be explored further. The sensitivity analyses regarding total parasite burden and P*f*HRP2 were limited, due to the large amount of missing samples, and the missing information on haematocrit for the calculation of total parasite burden, even where a sample had been tested. But the analyses with P*f*HRP2 were consistent with the SAP definition of malaria, giving more evidence across malaria definitions that the impact of the bolus varied across different levels of base excess only in children with malaria, with those at lowest base excess having had greater increased risk from bolus than those with normal base excess. The sensitivity analyses also showed a dose-response relationship with greatest mortality risk from boluses in those with low base excess and highest parasite burden.

The analyses in this chapter showed that there were likely to be different mechanisms of action for the harmful effect of boluses in children with severe febrile illness, as no one clinical measure over time could explain the treatment effect that was found in the trial. The recorded clinical measures over time combined were also not able to explain the treatment effect. The strong evidence that the impact of the bolus varied over levels of base excess for children with malaria had not been shown before and was independent of the previously found interaction with oxygen saturation (where those with highest levels of oxygen saturation at baseline had greatest increased mortality risk from boluses) [157]. This suggests that base excess is an important clinical measure and should be recorded in more clinical trials in these populations in the future. Although there were difficulties in using the i-STAT machine and cartridge to measure base excess in the FEAST trial, it is a point-of-care test and so could be developed further to be used more frequently in low-income settings. Although there was no clear surrogate marker found in the analyses in this chapter, one of the more promising measures was lactate. This may show more of a relationship with the intervention if measured earlier

after baseline (for example at 1 and 4 hours instead of just 8 hours) in future trials. The findings in this chapter also highlight the need for P*f*HRP2 to be measured in more samples and a more consistent measurement of haematocrit to calculate parasite burden. This would provide vital information in future trials about the impact of fluid resuscitation in children with malaria.

4 Describing mortality risk over time from randomisation

4.1 Aims and Objectives

The aim of this chapter is to explore how the mortality risk changes with time from admission in the FEAST trial. In previous work, the Cox proportional hazards model has been used to model survival to 48 hours and estimate predictors of mortality in this population. The baseline hazard $h_0(t)$ is not directly estimated from the Cox model and is treated as a nuisance parameter. The alternative to the Cox model, parametric survival models, model the baseline hazard directly and smoothly and thus enable direct estimation of how mortality risk (hazard) changes over time. However, the disadvantage of parametric survival models is the need to assume a particular functional form. The objectives of this chapter are to use parametric survival models to: investigate more closely the underlying risk in the form of the baseline hazard; identify when the highest risk occurred; and to explore how the risk may change in those that received boluses compared to those who did not. In particular, my question is whether boluses immediately increased the mortality risk compared to the control arm or whether it took a longer time from randomisation for the hazard to return to a baseline rate in the intervention arms compared to the control arm.

4.2 Statistical Background

The main command in Stata to implement parametric survival models is *streg*, which uses one of six distributions (exponential, Weibull, Gompertz, lognormal, loglogistic, and generalized gamma). However, these distributions are not flexible enough. For example, it is not possible to find a well-fitting parametric proportional hazards (PH) model for an underlying hazard function with a turning point, because the distributions that are on the PH scale are monotonic (Weibull and Gompertz) or constant (exponential). The other models for *streg*, which would better capture a turning point or peak in hazard, only express estimates in the accelerated failure-time metric which is less well known in the medical literature.

Flexible parametric models (FPMs) (also called Royston Parmar models)[205] are able to provide smooth estimates of the hazard and survival function for any combination of covariate values and model a turning point or peak on a proportional odds or proportional hazard scale.

A simple standard parametric model is the Weibull distribution for survival, a generalisation of the most basic of parametric models – the exponential distribution. The exponential model holds the baseline hazard constant over time i.e $h_0(t) = \lambda$. The Weibull distribution generalises the constant hazard of the exponential distribution as follows:

$$h(t) = \lambda\gamma t^{\gamma-1} \tag{1}$$

$$\log h(t) = log\lambda + \log\gamma + (\gamma - 1)logt \tag{2}$$

The hazard function is then monotonic increasing when $\gamma > 1$ and monotonic decreasing when $\gamma < 1$ and is linear in log time. The log cumulative hazard is then a linear function of logt.

$$\log H(t) = log\lambda + \gamma\text{logt} \tag{3}$$

Royston and Parmar extended the Weibull parametric survival model by representing the baseline log cumulative hazard function as a restricted cubic spline function of log time instead of a linear function of log time. This gave greater flexibility to the shapes of the survival distributions they could then model.

$$\log H(t) = s(logt;\ \gamma) = \gamma_0 + \gamma_1 logt + \gamma_2 z_1\,(logt) + \gamma_3 z_2(logt) + \cdots \tag{4}$$

where $logt$, $z_1(logt)$, $z_2(logt)$ and so on, are the basis functions for cubic splines. This creates a function made up of cubic polynomial segments smoothly joined at values (referred to as knots) at which the function has a continuous second derivative.

To ensure estimates in the tails of the distribution are not unduly affected by outliers, restricted cubic splines are used which force the function to be linear before the first knot and after the last knot. Royston and Parmar (2002) suggest knot positions based on empirical centiles of the distribution of log survival time as presented in Table 4.2.1 below [205]. To avoid over fitting to the data and potentially unstable curves they found that models did not generally need more than 3 interior knots (equivalent to 4 degrees of freedom). When there are no knots the z_1, z_2....z_n basis functions in equation (4) are equal to 0 and the function reduces to the Weibull model.

Table 4.2.1: Positions of interior knots in the distribution of uncensored log-survival times.

Degrees of freedom	Number of interior knots	Position of interior knots (centiles*)
1	0	(no knots) (equivalent to Weibull distribution)
2	1	50
3	2	33, 67
4	3	25, 50, 75
5	4	20, 40, 60, 80

*Centiles of the distribution of uncensored log-survival times

FPMs can use proportional hazards (PH), proportional odds (PO) or probit scales (Probit) to give the best fit, with the coefficients of each having different interpretations (i.e hazard ratio from a proportional hazard model, odds ratio from a proportional odds model, acceleration factors on a lognormal scale from the probit model). To identify both the best fitting type of model as well as the most appropriate number of knots, the Akaike information criterion (AIC) of the three different types of unadjusted FPMs with no covariates over the number of degrees of freedom can be compared.

Another advantage of FPMs is that time-dependent effects of covariates can be fitted allowing for non-proportional hazards. This is done by forming interactions with spline terms and the covariates of interest, allowing time-dependent effects to have a different number of knots and have those knots at different locations than for the baseline time effect. Thus if there are

D time-dependent effects of covariates then the log cumulative hazard function can be written as follows:

$$logH(t|x) = s\{logt|\gamma, k_0\} + \sum_{j=1}^{D} s\{logt\,|\delta_k, k_j\}x_j + x\beta$$

where $s\{logt|\gamma, k_j\}$ is a restricted cubic spline function of log t with knots k_j, j=0,1,...D and x is the vector of covariates.

The assumption of proportional hazards can be tested using scaled Schoenfeld residuals from Cox models where the residuals have an average value approximately equal to the log hazard ratio. If there is evidence of a trend over time in the residuals then this would indicate some non-proportionality. But this tests for (monotonically) increasing or decreasing effects over time and does not test for a non-proportional effect which is not monotonic. In a FPM, the assumption can also be explored by fitting time-dependent effects and checking if the degrees of freedom for the time-dependent effect needed for a best fitting model is 1 or more, showing that the hazard is not proportional to the baseline hazard.

FPMs can also be used in competing risks analysis to examine the underlying mortality risk for specific causes of death (i.e the cause-specific hazard) where other causes of death are the competing risks [206].

4.2.1 Sensitivity analyses

Modelling the log cumulative hazard function on a log time scale may create an asymptote in the hazard going to zero or infinity when time approaches zero (personal communication M Crowther). To avoid this, the untransformed scale of time can be used as the basis for modelling and the log hazard function modelled instead of the log cumulative hazard, namely:

$$\log\{h(t|x)\} = s\{t|\gamma, k_0\} + \sum_{j=1}^{D} s\{t\,|\delta_k, k_j\}x_j + x\beta$$

where $s\{t|\gamma, k_0\}$ is a restricted cubic spline function with knots k_0, β is the vector of covariate effects, and x is the vector of covariates, D is the number of time-dependent effects, and $s\{t\ |\delta_k, k_j\}$ is the spline function for the dth time-dependent effect.

The log hazard can be represented as:

$$\log\{h(t|x)\} = \varphi(t)$$

which can then be written as:

$$h(t|x) = exp\{\varphi(t)\}$$

And the survival model follows

$$S(t) = exp\left[-\int_0^t exp\{\varphi(t)\}du\right]$$

As estimating the survival function involves integrating the hazard function, which is a complex spline function, numerical integration techniques are used alongside analytical techniques. The function is integrable prior to the first knot and after the last knot as it is restricted to be linear, and then the cubic spline section of the function in between these two knots can be integrated using Gaussian quadrature.

Although modelling the log hazard on a time scale avoids the asymptote in the estimated hazard that arises from modelling the log cumulative hazard as a function of log time, it may not have as good a fit as modelling on the transformed log time scale when using the same degrees of freedom [207]. Modelling on the log transformed time scale has been shown to be more stable to changes in the baseline degrees of freedom in the model, which is why it is the preferred default scale.

As above, the exponential model is the most basic parametric model, holding the baseline hazard constant. This can be extended to have a changing baseline hazard with the piecewise exponential model, where the baseline hazard is held constant within defined intervals of time. This is a useful model to estimate rates (i.e the baseline hazard) within these time periods but discontinuities in the function between the time periods leads to a step function which is not generally considered to be biologically plausible. It is equivalent to categorising continuous data which has been to shown to lead to loss of power, residual confounding and bias [162]. Also, if the interval lengths are too long then the function may miss important

changes in the data, but if the lengths are too short then there are too many intervals with many parameters in the model and the underlying shape of the hazard may be difficult to see due to random variation [208].

The piecewise exponential function can also be used to simulate survival data with the assumption of a constant hazard over each interval. Starting with the basic parametric distributions described as survival and failure functions:

$$S(t|x) = exp[-H(t|x)] \qquad F(t|x) = 1 - exp[-H(t|x)]$$

If T is the simulated survival time, then

$$Z = S(T) = \exp[-H(T)]$$

has the uniform distribution $U(0,1)$ over the interval (0, 1), where S and H are the survival and cumulative hazard functions respectively. It follows that

$$T = H^{-1}[-\log(Z)].$$

For the exponential distribution with hazard λ, $H(t) = \lambda t$, thus

$$T = -\log(Z)/\lambda.$$

For the piecewise exponential distribution, let $0 = t_0 < t_1 < \cdots < t_{K-1} < t_K = \infty$ be boundary time points for intervals of constant hazards $\lambda_1, \ldots, \lambda_K$. The hazard and cumulative hazard functions are

$$h(t) = \lambda_i;\ t_{i-1} < t \leq t_i\ ; i = 1, \ldots, K.$$

$$H(t) = \sum_{i=1}^{j-1} \lambda_i (t_i - t_{i-1}) + \lambda_j (t - t_{j-1})\ ;\ \text{if}\ t_{j-1} < t \leq t_j$$

Simulation from this piecewise exponential distribution proceeds by inverting $H(t)$.

Let $H_j = H(t_j) = \sum_{i=1}^{j} \lambda_i (t_i - t_{i-1});\ 1 \leq j \leq K$. Then for any real number $y > 0$, there is a unique positive integer $j(y)$, $1 \leq j(y) \leq K$ such that $H_{j(y)-1} < y \leq H_{j(y)}$. Draw z from the uniform distribution $U(0,1)$ and let $y = -log(z)$. Then

$$T = t_{j(y)-1} + (y - H_{j(y)-1})/\lambda_{j(y)}$$

4.2.2 Literature review

Patrick Royston and Mahesh Parmar first published the FPMs approach in 2002 in a paper entitled: 'Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment results'[205]. To September 2016, the paper had been cited 274 times in a variety of research areas, although most frequently in Mathematics/Statistics (48), Oncology (48), Public/Environmental/Occupational Health (44) and Computational Biology (31). The method was cited in 5 publications in Paediatrics and 12 in Infectious Diseases (with 1 publication in both categories) of which 6 modelled mortality risk, and is becoming more well known in the medical literature each year, with 43 citations in 2015 and 51 in 2016 (Table 4.4.2).

Table 4.2.2: Citations of Royston and Parmar, Stats in Medicine 2002 by year.

Publication years	Number of citations
2015-2016	94
2011-2014	111
2007-2010	48
2003-2006	18

Of the 6 papers in Paediatrics or Infectious diseases that used this method to model mortality risk, 2 showed a peak in mortality risk near time zero (i.e at diagnosis or initiation of treatment). The models outlined in one paper estimated the daily risk of death over the first year on antiretroviral treatment in HIV-infected children from Uganda and Zimbabwe with pre-ART CD4<49 cells/μL and showed a peak at 34-51 days after ART initiation [209]. The second paper used FPMs to describe the excess mortality rate over years from diagnosis for children with brain tumours and showed mortality risk peaked in the first two years from diagnosis [210]. Three of the remaining four papers looked at cumulative incidence: one used FPMs with competing risks for time to discharge or death in neonatal care [211]; a second used these models to build a prognostic model for children on ART in Southern Africa [212] and a third estimated cumulative incidence mortality curves for children with and without Autism Spectrum Disorder (ASD) [213]. In the last paper the mortality risk in HIV infected patients

aged >13 years initiating antiretroviral therapy in Uganda was actually described by Weibull models with changepoints, which was the focus of the work, and the author discussed this in relation to FPMs. Although FPMs would add flexibility to the shape of their model (as their hazard curve had discontinuities at changepoints) their view was that the spline knots do not have the same direct biological interpretation as changepoints [214].

The other papers that used FPMs in Infections or Paedatrics had non-mortality endpoints such as: incidence of HIV RNA <500 copies/ml [215]; incidence of peripheral neuropathy [216]; autism diagnosis [217]; presence of N3481 mutation in HIV-1 reverse transcriptase [218]; *Clostridium difficile* recurrence risk [219]; time to relapse in malaria for geographic regions [220]; rates of acquisition of staphylococcus aureus [221]; diagnostic delay [222]; time to viral load rebound [223]; or were not in humans [224].

Paul Lambert and Patrick Royston published an extension to FPMs in The Stata Journal in 2009 – this has been cited 136 times but with only 2 citations in paediatrics and 2 in infections. Two of these are described above as they also cited the original paper in Statistics in Medicine [209, 210]. The other two did not describe mortality risk but looked at infection-related hospitalisations [225] and at incidence of TB in children [226].

4.3 Methods

4.3.1 Data

The dataset used in this chapter is from the FEAST clinical trial and is the same dataset used in previous work but includes follow up time to 28 days, thus adding an additional 48 deaths to the 297 occurring before 48 hours used in previous analyses. Time and date of death was recorded on standardised case report forms and where time of death was missing for 9 deaths after discharge, it was given a nominal value of midday. Analysis includes all deaths in the main trial follow up period (28 days) as I am interested in both: 1) the hazard around the time of the bolus (given within the first hour of randomisation, and for some children a second bolus was given shortly after completion of the first bolus if indicated) and 2) whether, and for how long, any increased risk from receiving a bolus was maintained.

4.3.2 Statistical modelling

The overall predicted hazard in children in the FEAST trial was first examined using a Weibull model with time from randomisation to the earliest of death, lost to follow up or 28 days. The value of gamma in this model was then checked for departure from the exponential distribution. The hazard estimated from the model was plotted to show the distribution over time from randomisation. The cumulative hazard was also estimated and plotted with the widely used empirical Nelson-Aalen estimate to show the fit of the model.

FPMs were then examined to estimate the baseline hazard and improve on the fit of the Weibull model. FPMs can use proportional hazards (PH), proportional odds (PO) or probit scaled (Probit) to give the best fit, with the coefficients of each having different interpretations (i.e hazard ratio from a proportional hazard model, odds ratio from a proportional odds model, acceleration factors on a lognormal scale from the probit model). To identify both the best fitting type of model as well as the most appropriate number of knots, the AIC of the three different types of unadjusted FPMs with no covariates over the number of degrees of freedom were plotted.

The FPMs with the most appropriate scale and number of knots for the cubic spline function was identified and the estimate of cumulative hazard was plotted with the Weibull and Nelson-Aalen estimate to check the fit. The estimated mortality risk (hazard) was then plotted over time from randomisation.

A time-dependent effect of bolus was then added to the FPM, following a scaled Schoenfeld residual test for non-proportionality of the bolus effect in a Cox model. To identify the appropriate number of knots for the time-dependent effect, the AIC of models over a number of degrees of freedom for both the baseline effect and time-dependent effect were compared. The most appropriate model was selected and the mortality risk in the bolus arms and the control arm were plotted over time from randomisation, for time up to 28 days, but also focussing in on the first 4 days (96 hours).

FPMs were also used in a competing risk analysis with different causes of death. The endpoint review committee (ERC) for the FEAST trial reviewed all the deaths blinded to randomisation arm and classified them into 4 different terminal clinical events (TCE) [157] as described in Chapter 3 (page 155).

I used FPMs to estimate the mortality risk from each terminal clinical event, and the absolute difference in mortality risk between bolus and control, by modelling the cause specific hazard with competing risks analysis.

4.3.3 Sensitivity analyses

I performed sensitivity analyses to investigate the influence of early deaths both prior to randomisation and immediately after randomisation on model estimates. The analysis looking at deaths prior to randomisation included children that had died between screening and randomisation (n=11) by using time from admission and randomly allocating these children a unique death time between 1 and 11 minutes of admission (1 child per minute in the first 11 minutes). This analysis did not include treatment arm as these 11 children had not been randomised to a specific arm. The analysis looking at the influence of early deaths immediately after randomisation excluded all deaths within 1 hour of randomisation.

FPMs were also fitted to the log hazard using the untransformed time scale rather than modelling on the log time scale. The log hazard rather than the log cumulative hazard was modelled as the most appropriate command in Stata (strcs) to fit on the untransformed time scale used FPMs for the log hazard. The AICs of models with different numbers of knots were compared to identify the most appropriate number for the baseline and time-dependent effects. The difference in hazard was also estimated from these models and plotted. A bootstrapped 95% CI (with 1000 replications) was generated for estimates of the difference in hazard between the bolus and control at specific time points (12, 24 and 48 hours).

To further understand the shape of the underlying hazard function, a piecewise exponential model was fitted to the data with changepoints at 1, 2, 3, 6, 12, 24 and 48 hours from randomisation and the interval-specific rates estimated. These were then used to simulate 20 datasets with 3000 observations and survival times following this piecewise exponential distribution. Flexible parametric models for both the log cumulative hazard scale and log hazard scale were then fitted to the simulated data. The predicted hazard from the parametric models, with the same number of knots as found for the flexible parametric model fitted to the log cumulative hazard in the FEAST data (5 df, 4 internal knots, see Results section below), was estimated and plotted to explore how the models dealt with estimates of constant high hazard immediately after randomisation (when time = 0) which then dropped at each hour from randomisation.

A piecewise exponential model was also fitted to the data, including an interaction term between time (split at 1, 2, 3, 6, 12, 24 and 48 hours from randomisation) and bolus vs control and the predicted hazard estimated separately for each group to examine the risk of mortality between 1 and 4 hours.

On further consideration, the observed times of deaths may have been subject to some recording bias, in that the first observation time point for the control arm was exactly 1 hour after randomisation but the first observation time point for the bolus arm would be at 1 hour after the first bolus started which could be 15-30mins after randomisation i.e the first observation would likely be more than 1 hour after randomisation. Thus, children close to death in the control arm may have been identified slightly earlier than those in the bolus arm and when the mortality risk at 1 hour is modelled the control arm would be removed from the risk set earlier than the bolus arm deaths due to design. Interval censored survival data allows individual children to have an interval in which their death was recorded instead of an exact

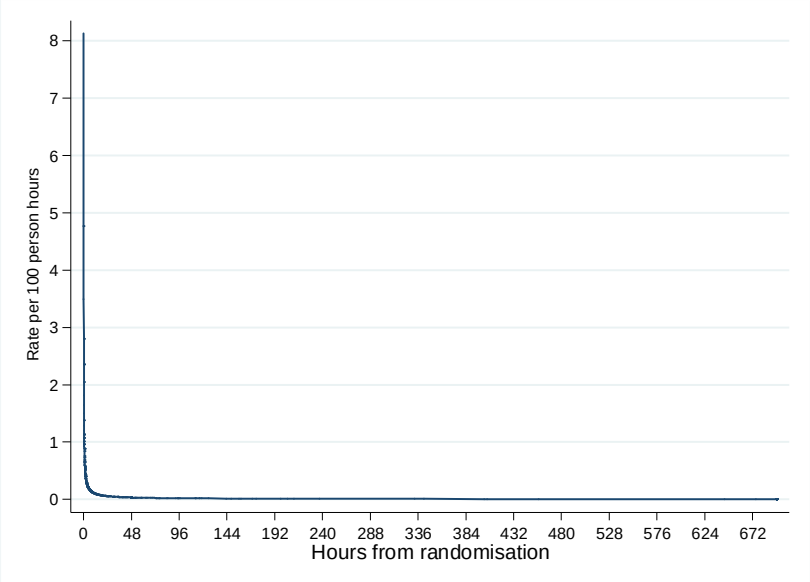
time. Parametric models were fitted to interval censored data where the death times for the first three hours were recorded as intervals instead of exact times and their AICs compared. Only basic parametric models (Weibull, exponential, Gompertz, lognormal, loglogistic and generalized gamma) were available to use with interval censored data.

4.4 Results

4.4.1 Weibull model

I first used a Weibull model to first look at mortality rate in the first 28 days after randomisation, under the assumptions of this parametric model; Figure 4.4.1 shows the overall predicted hazard. The value of γ in this model was 0.23 (95% CI 0.21, 0.26) demonstrating strong departure from the exponential distribution ($\gamma = 1$). The mortality rate was very high in the first hour, dropped steeply until around 3-4 hours after randomisation and then declined steadily until the rate was very low after 48 hours and remained low for the remainder of the follow up time.

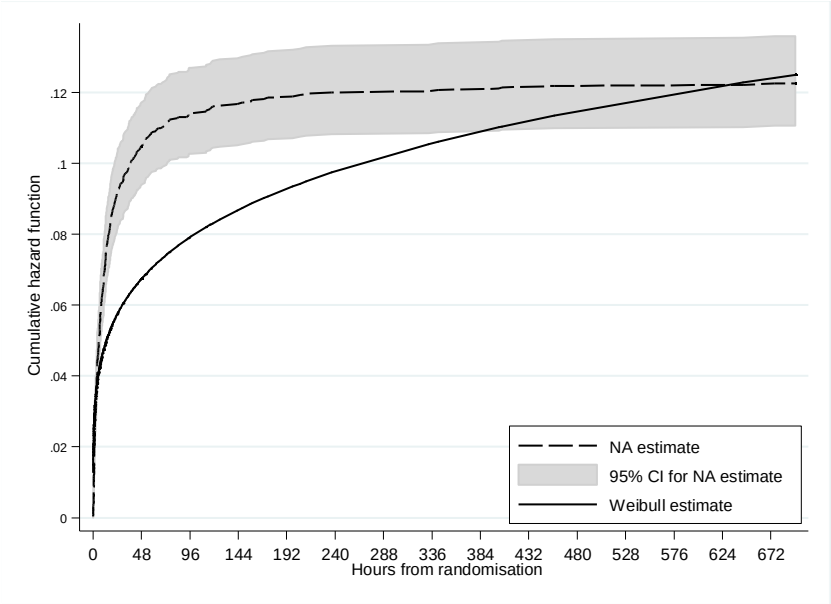
Figure 4.4.1: Overall predicted hazard over time using a Weibull model



The estimate of the cumulative hazard from the Weibull model and the Nelson-Aalen estimate are shown in el does not fit the data well.

Figure 4.4.2 below. This shows that the Weibull function was not a good fit as it over estimated the cumulative hazard between 0-2 hours and then underestimated the cumulative hazard for the rest of the time period. Thus a linear function of log time in a parametric model does not fit the data well.

Figure 4.4.2: Overall Cumulative hazard function estimated by Nelson-Aalen (NA) and Weibull model



4.4.2 Flexible parametric models

Flexible parametric models (FPMs) with time in hours from randomisation up to 672 hours (28 days) were then fitted to the data and the model AICs compared to identify the best fitting model and most appropriate scale. Figure 4.4.3a and Figure 4.4.3b show that 5 degrees of freedom (df) (4 knots) minimises the AIC of the PH and PO models, and 2 df minimises the AIC of the Probit model, and further that the PO and probit models gave AICs that were very similar to the PH model for 2-5 degrees of freedom. The 5 df model for PO and PH was however within 3.84 (chi-squared (1)) of the overall minimal value from the Probit model showing that any of these models would be a reasonable choice.

Figure 4.4.3a: AIC of three different scales for FPM

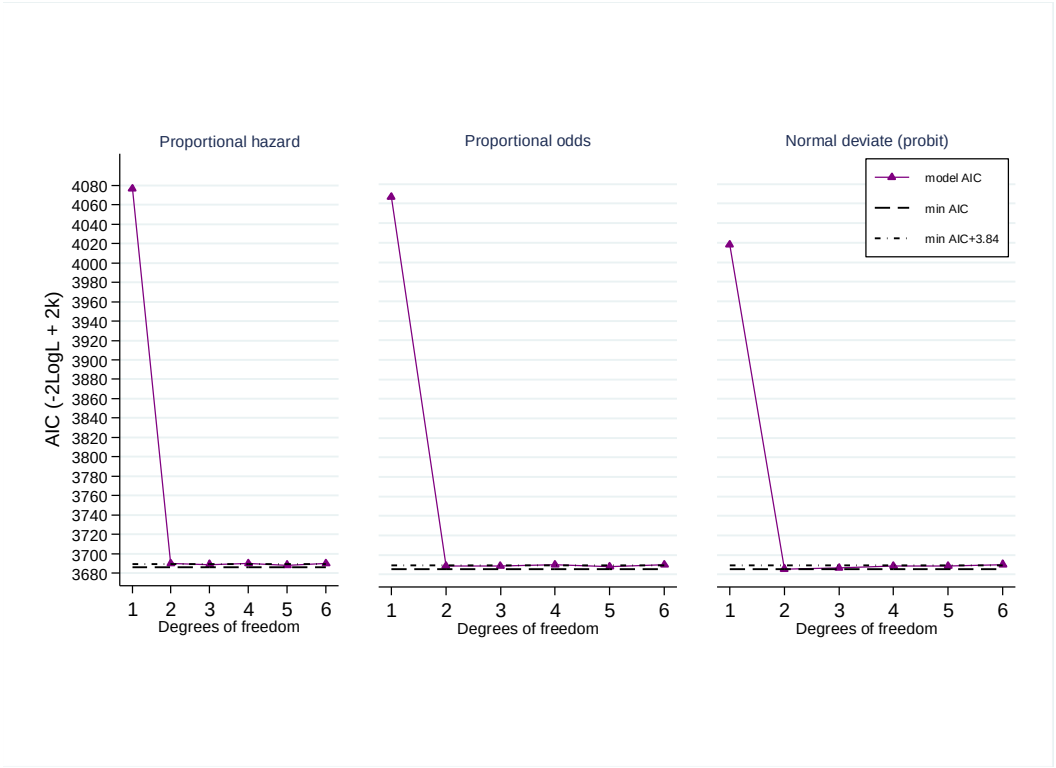
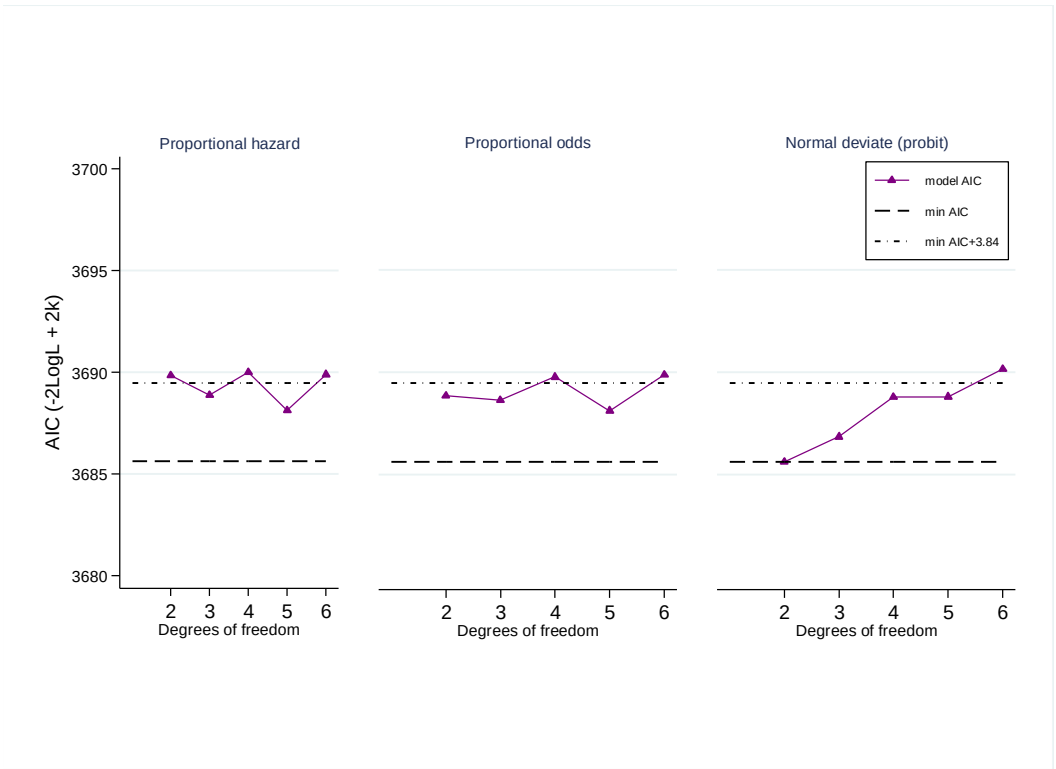


Figure 4.4.3b: AIC of three different scales for models with degrees of freedom>1 to show minimum and model AICs



As previous work looking at survival in these data has used the (proportional hazards) Cox model and as the fit of PH models and PO models were comparable, I decided to use the proportional hazards scale for ease of comparison and interpretation of hazard ratios as estimates. Figure 4.4.4 below shows that the cumulative hazard estimates from each model were very similar even though the probit model gave the lowest AIC.

Figure 4.4.4: Cumulative hazard function estimated by the best fitting model on each scale

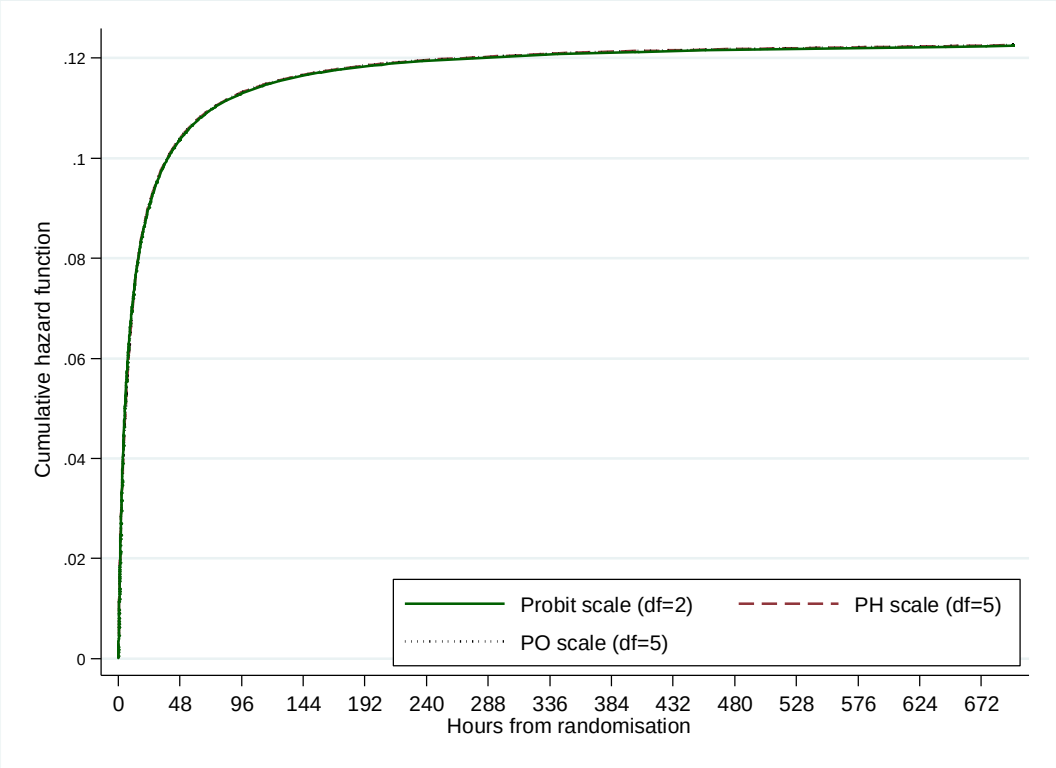


Figure 4.4.5 compares the unadjusted cumulative hazard function from the PH FPMs with 5 df (4 knots) with the non-parametric NA estimate and the Weibull model. This showed that the PH FPM 5 df estimate fitted the empirical function well, especially when compared to the Weibull model.

Figure 4.4.5: Cumulative hazard function estimated by Nelson-Aalen, Weibull and FPM with 5 degrees of freedom (4 knots)

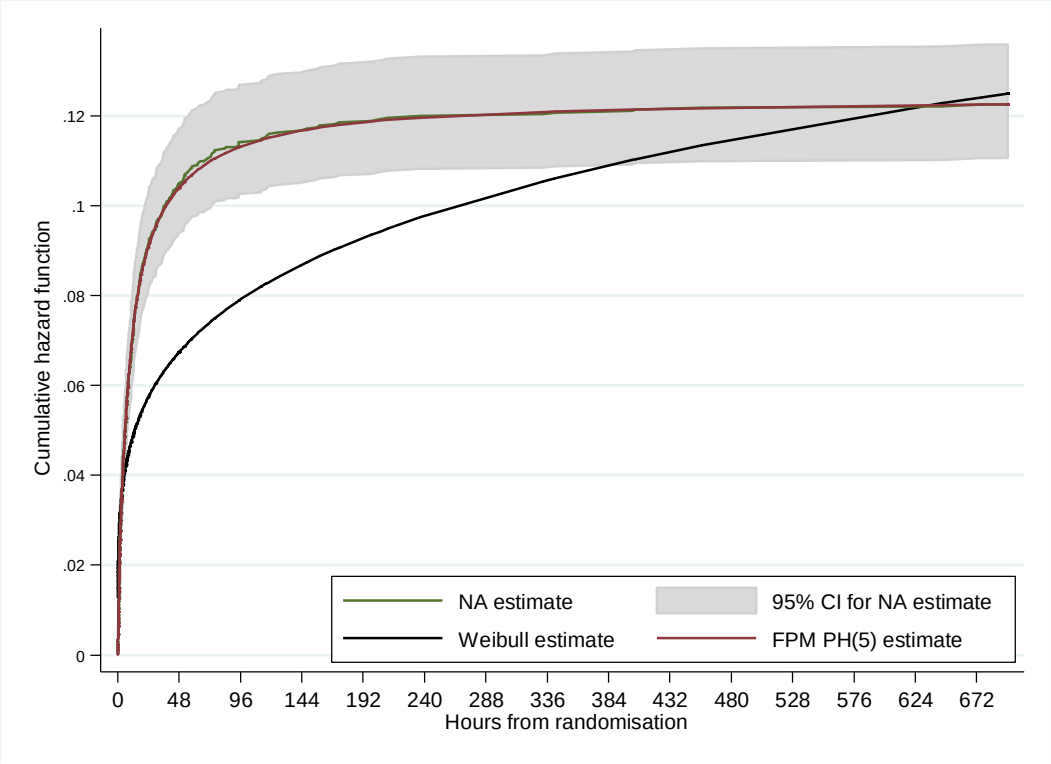
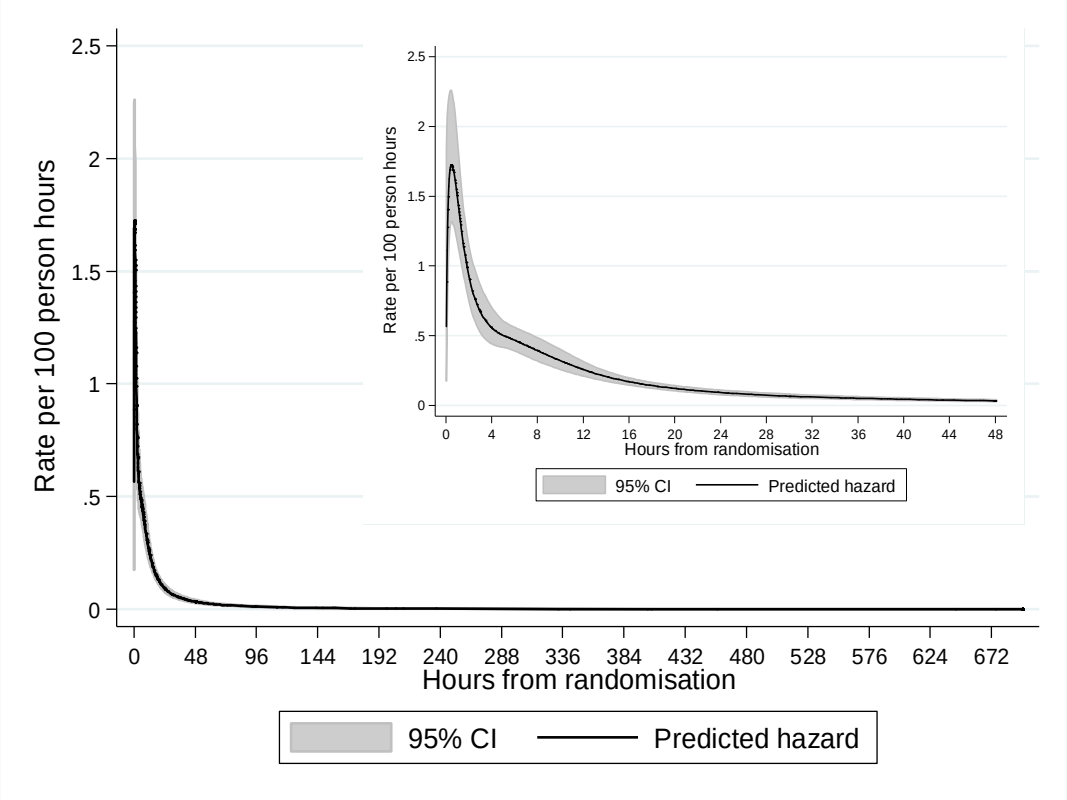


Figure 4.4.6 below shows the estimated hazard from the PH FPM 5 df both over 28 days and the inset graph focuses on the first 48 hours. The inset graph along with the knot positions in the baseline hazard show that the hazard rose to a high peak at 1.6 hours. It then declined sharply through to 5.2 hours and continued to decline but more slowly from 5.2 to 11.8 hours. Then the decline became even slower through to 31.0 hours with the large panel showing that this decline continued slowly beyond 31 hours.

Figure 4.4.6: Estimated overall hazard over 28 days (672 hours) from a model with 5 df (4 knots) in large panel, with time from randomisation restricted to 48 hours in inset panel.



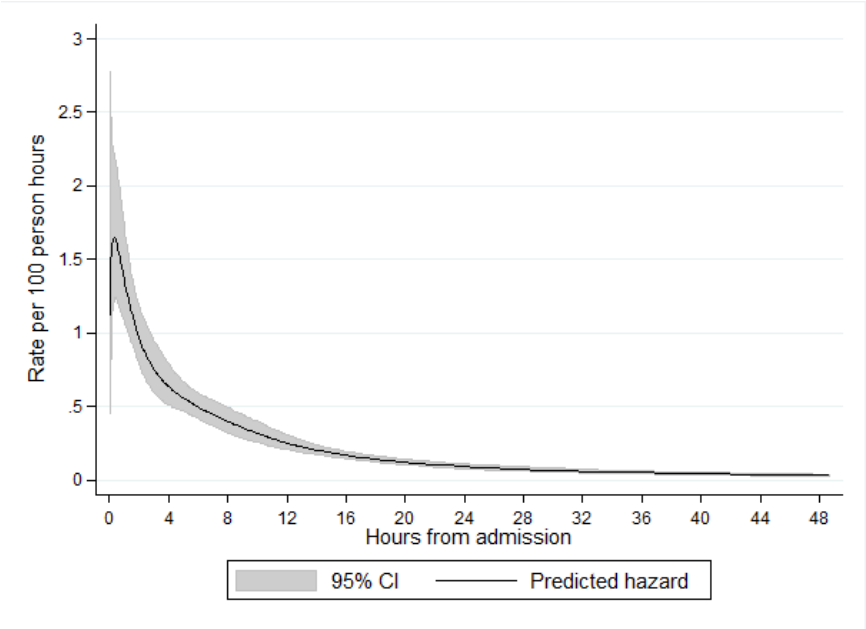
4.4.3 Sensitivity analyses for overall hazard

Two possible reasons for an overall increase in mortality risk immediately after randomisation are that it is a consequence of the design or analysis of this randomised trial, or that interventions (other than bolus) given immediately increase mortality risk. The screening and consent process in clinical trials usually prevents randomisation of children about to die and this could create an artificial peak in risk after randomisation, but in FEAST there was a verbal assent process from the parent or guardian (to enable immediate recruitment with minimal time delay) so critically ill children could get quickly randomised. Full consent was deferred to when the child had stabilised and the parent deemed to be able to make an informed choice about participating in research. It is also not clear that other immediate interventions, such as IV antibiotics, that are designed to treat the children would cause this peak so soon after randomisation, given that many children had travelled long distances before presentation (4 of 6 hospitals had regional referrals with no paramedical service) and thus survived long enough to get to hospital. Moreover, most interventions need some time to work or to cause side effects.

There were a small number of admissions (n=11) that died between screening and randomisation and were not included in the main analysis. I therefore investigated whether including these could reduce the size of the peak in early hazard I had found shortly after randomisation in Figure 4.4.6, or show that the risk simply starts high and then reduces over time. I examined this in a first sensitivity analysis by including these 11 children in the dataset, starting analysis time from time of first contact with the study team (usually through triage). These additional 11 children were randomly allocated death times (as time of death was not recorded for these screening deaths) between 1 minute and 11 minutes after admission, one per unique minute, and the overall hazard rate was plotted (Figure 4.4.7). Adding these deaths made the initial estimate of the immediate hazard of death higher (~1.1 compared to ~0.6 in Figure 4.4.6), with the maximum risk occurring at a similar magnitude and time point as the main analysis. However, in this analysis, the lower 95% confidence limit remained at about the same level as the initial estimate of the immediate hazard of death across the 1st hour from time of first contact with the study team, in contrast to Figure 4.4.6 where the lower 95% confidence limit increased markedly over the first hour at risk. Thus, this sensitivity analysis

suggests that there is relatively little evidence that mortality risk was really changing over this period.

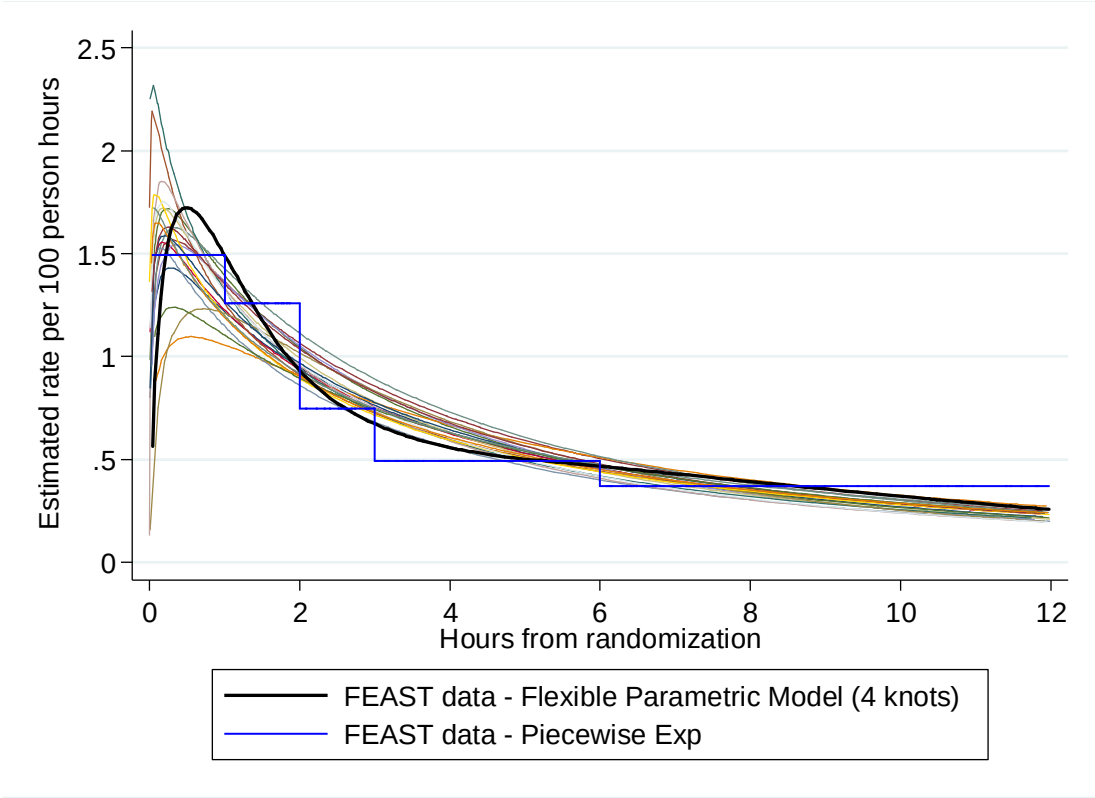
Figure 4.4.7: Overall hazard rate per 100 person hours including children dying before randomisation.



To understand whether the peak in risk may be an artefact of the statistical modelling process, rather than a clinically meaningful increase prior to a decrease, a piecewise hazard model was fitted to the FEAST data assuming high constant hazard for the first 1 hour. The interval-specific values estimated from this model were then used to simulate 20 datasets with 3000 observations and FPMs (with 5 df (4 knots)) were fitted to the simulated data. The overall estimated hazard shows that the FPM creates a peak even though the data truly has a monotonically decreasing hazard from a high rate at zero (Figure 4.4.8). Some of the datasets had a flatter peak – with a hazard of 1 per 100 person hours, and some had a very little or sharper peak with a hazard of 2 per 100 person hours. FPMs are reasonably robust to changes in knot positions and so the default knot positions used here (at centiles of the censored survival times) are not likely to be affecting this [205]. Also, using the default knot position avoids overfitting the model to the data, and if the interior knots are not too far from the median uncensored log survival time then the data will be most closely modelled in the region

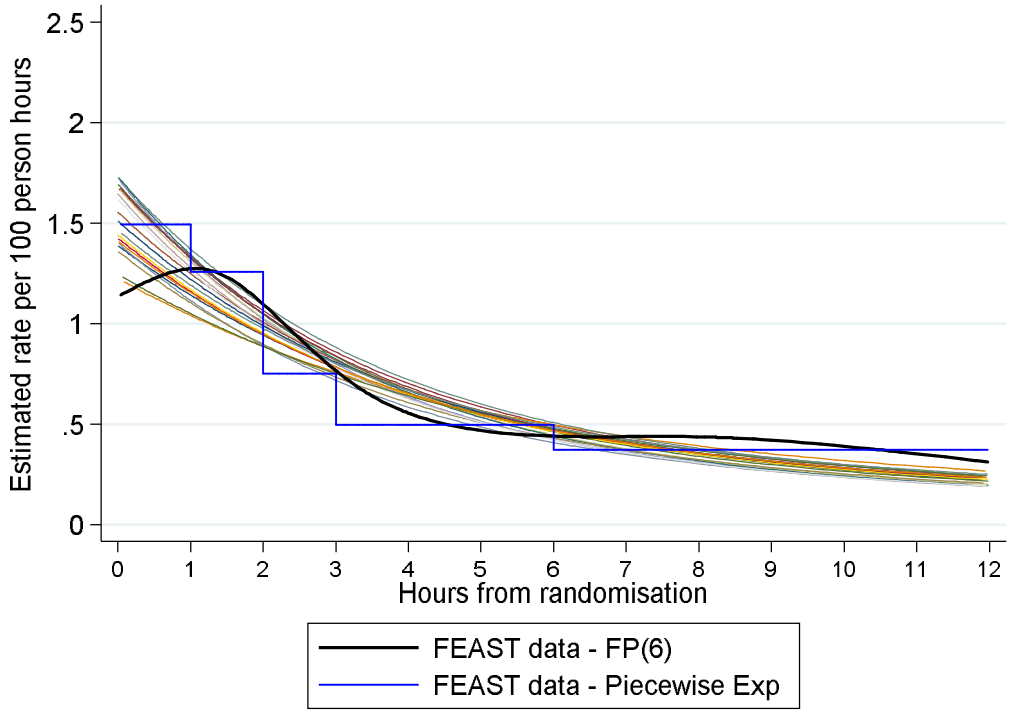
of greatest density [205]. Moving the knot positions (but keeping the same number of knots) for the flexible parametric models fitted to the simulated datasets did not remove the peak unless the first interior knot was at >48 hours compared to a median uncensored survival time of 8 hours. Moving the lowest boundary knot (and thus forcing a linear function over a longer initial time period) also did not remove the peak until the knot was at 2 hours, which was after 86 deaths (25th percentile of uncensored survival times).

Figure 4.4.8: Estimated hazard over the first 12 hours from 20 simulated datasets (modelling log cumulative hazard over log time).



As modelling on the log time scale may lead to an asymptote in the hazard function to zero (or infinity) when events times are very close to zero (minimum event time in FEAST was at 0.04 hours from randomisation), the flexible parametric models were fitted to the log hazard on the observed time scale rather than the log time scale. The AIC of models with different degrees of freedom (up to a maximum of 6 df as these models have been found to overfit with a higher number of df) were compared and the model with the lowest AIC had 6 df (5 knots). The estimated predicted hazard was then estimated from a model with 6 df using data from the 20 simulated datasets (Figure 4.4.9).

Figure 4.4.9: Estimated hazard over the first 12 hours from 20 simulated datasets (modelling log hazard over time).



The estimated hazard from the models on the untransformed time scale, instead of the log time scale, show a monotonically decreasing hazard which better reflects the piecewise exponential hazard that has created the true underlying hazard in the simulated datasets. However, the estimated hazard does appear to underestimate the underlying hazard for the first 2 hours and overestimates it from 2 to 8 hours of randomisation.

4.4.4 Effect of bolus

The test of the residuals from a Cox model with bolus as a covariate (with no other covariates) gave a p-value of 0.56 indicating no strong evidence for non-proportionality; however, this only tested for (monotonically) increasing or decreasing effects in the residuals over time and does not test for a non-proportional effect which increases and then decreases, which I hypothesize may occur here given the overall changes in mortality shown above.

To investigate the possibility of a time varying effect of bolus over 28 days, I compared the AIC of survival models, based on the original model relating log cumulative hazard to log time, including a time-dependent effect of bolus vs control with 0 to 3 interior knots (1 df for the time-dependent effect (dftvc) to 4 dftvc) and varying the baseline hazard allowing 0 to 6 interior knots (1 df to 7 df). I included boundary knots placed at the minimum and maximum of the distribution of uncensored survival times. There was some difficulty with fitting some models but initial values were obtained by fitting only the first spline basis function, and using these parameter estimates as initial values enabled the full models to fit. Several models were within 3.84 (chi-squared(1)) of the minimum AIC (Figure 4.4.10a and b).

Figure 4.4.10a: AICs from models including bolus as a time-dependent effect

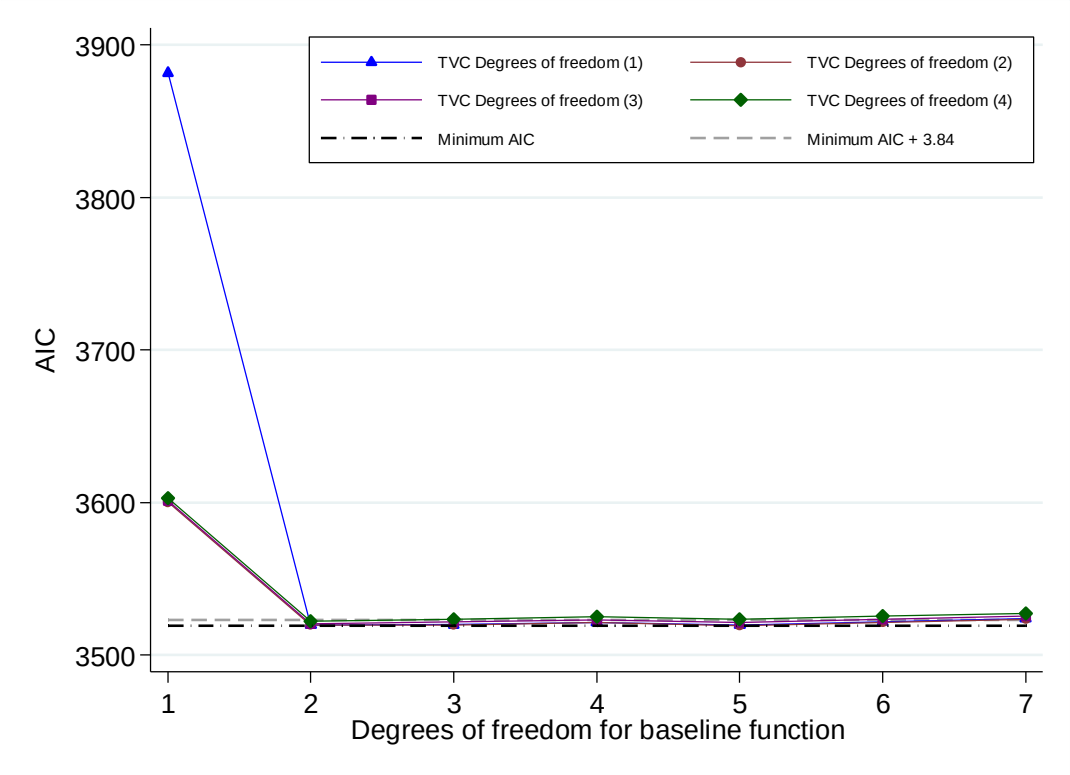
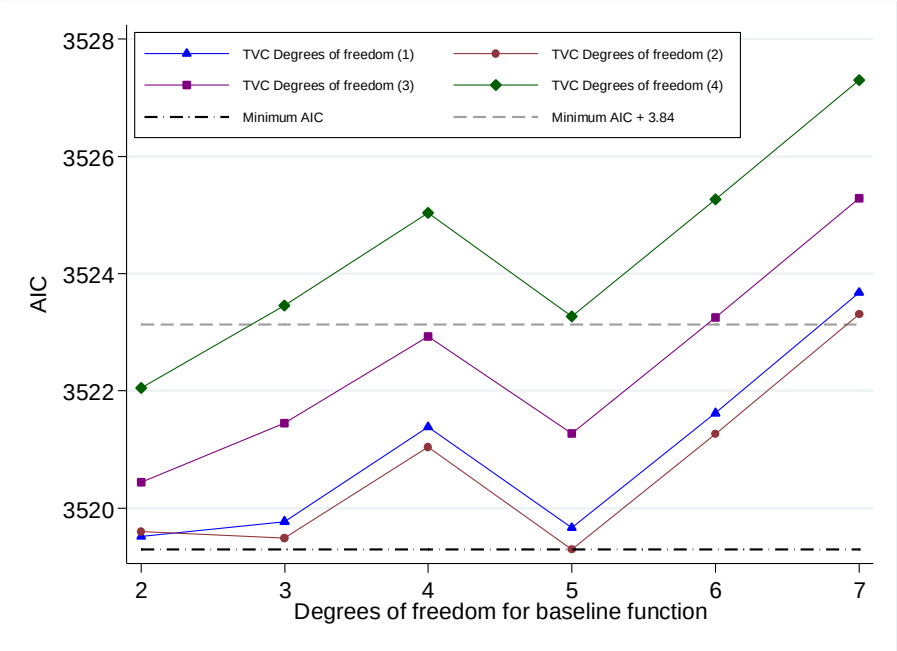
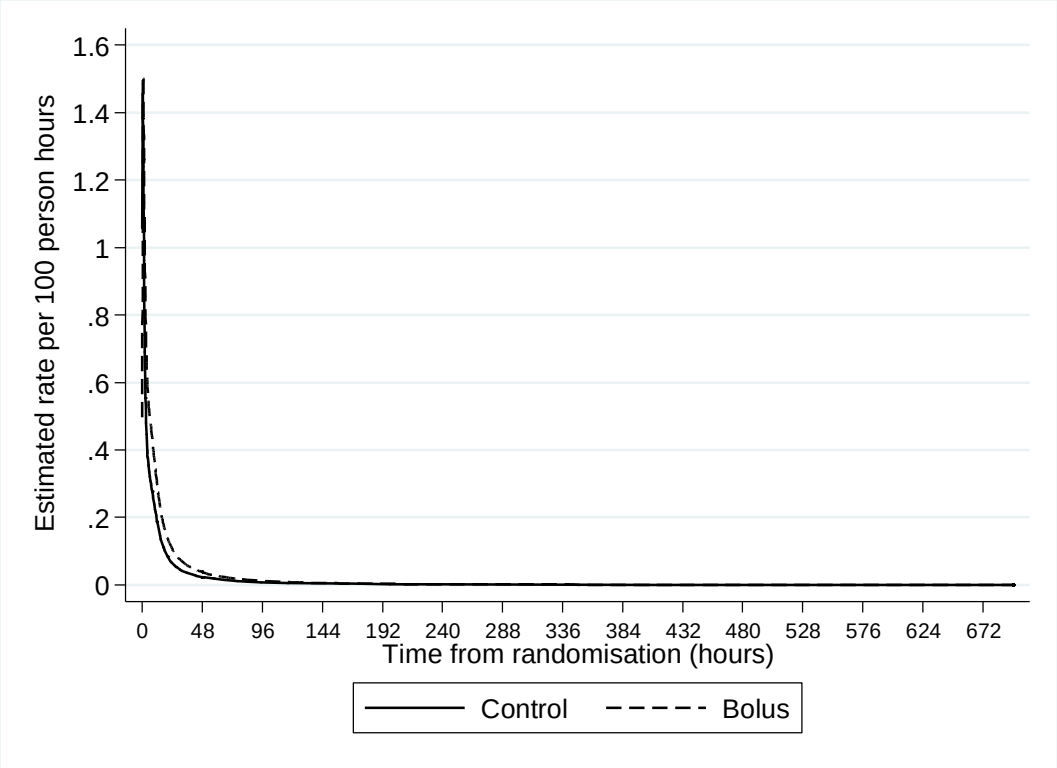


Figure 4.4.10b: AICs from models including time-dependent effects (with baseline df>1)



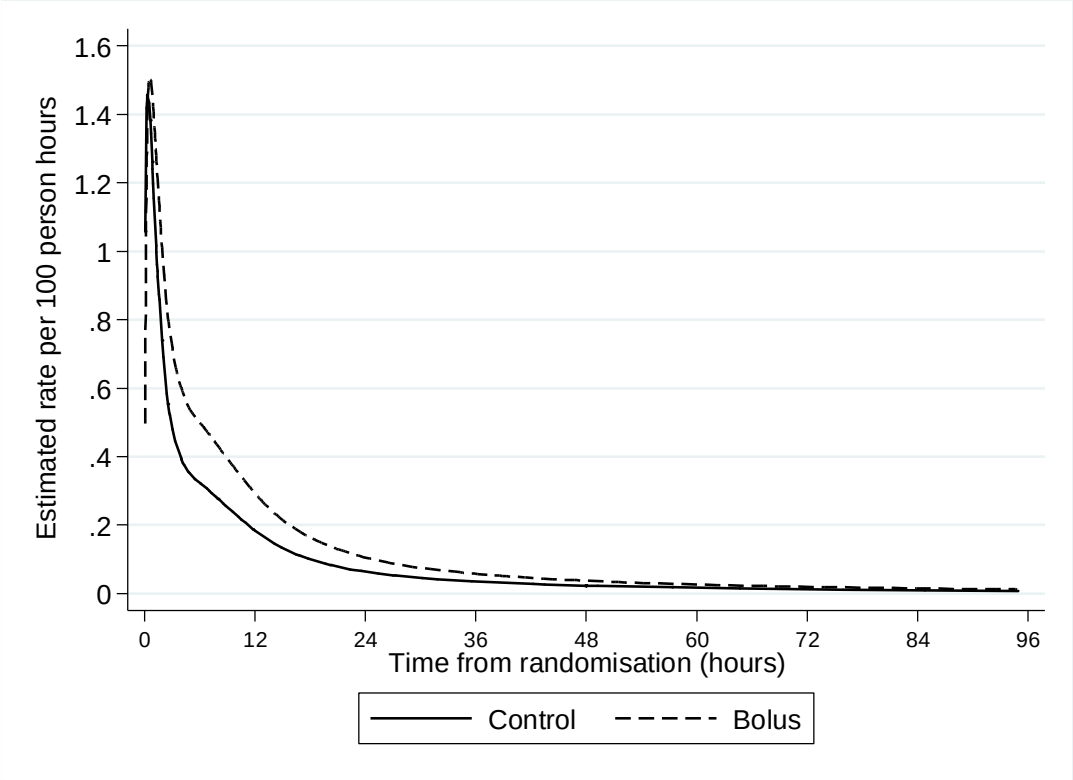
The model with 5 df for baseline function and 2 df for the time varying effect of bolus gave the lowest AIC, producing the baseline hazard below (Figure 4.4.11). The model with the next lowest AIC comes from a proportional hazards model (with 1 df for the time varying effect)) and when the time was restricted to 48 hours from randomisation this became the best model, indicating that the non-proportionality was occurring over a longer time from randomisation, i.e. that the hazards converged after 48 hours (Figure 4.4.11 and Figure 4.4.12).

Figure 4.4.11: Estimated hazard rate per 100 person hours (phrs) over 28 days by bolus vs control using 5 df for the baseline function and 2 df for the time varying effect.



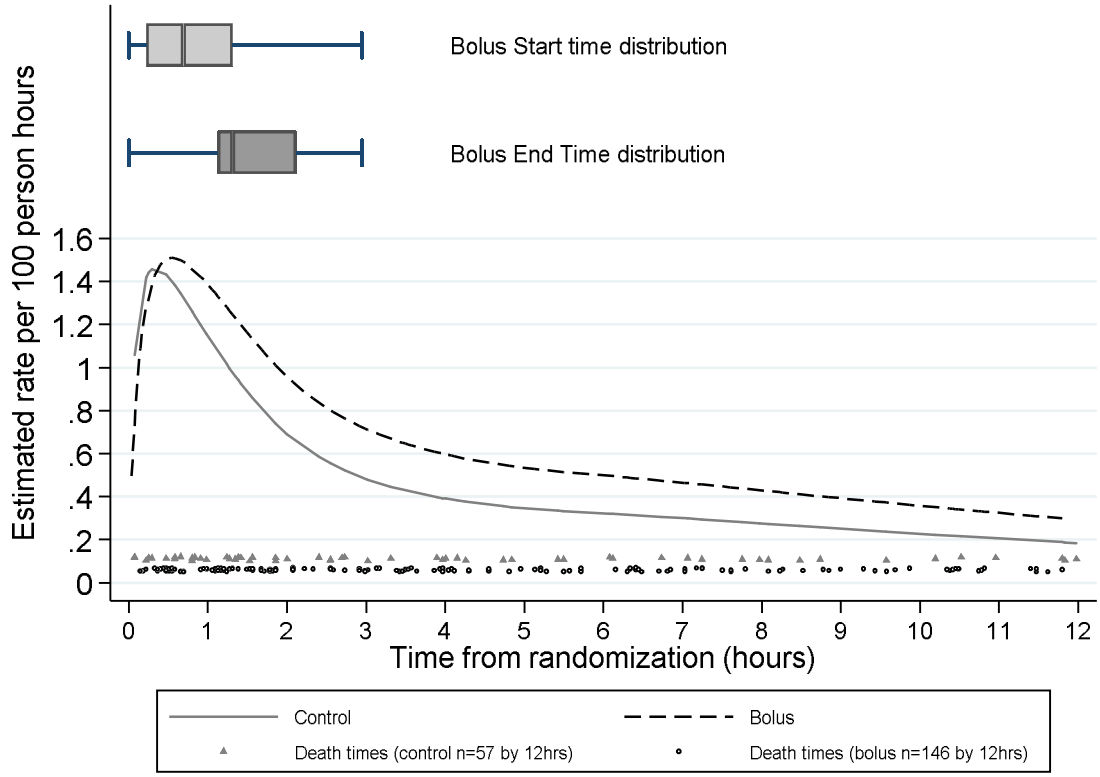
It is difficult to see the shape of the peak in Figure 4.4.11 as the x-axis goes to 672 hours (28 days) to show the full length of time in the analysis. Figure 4.4.12 restricts time from randomisation to 4 days (96 hours) as this where there is most difference between the arms. It also shows more clearly that the peak was very similar between the arms, indicating that at the time the children were receiving the bolus, their risk was similar to those not receiving a bolus. It was after that point that they appeared to have a longer period of moderately increased risk.

Figure 4.4.12: Estimated hazard over 100 phrs in the first 4 days (96 hours) of randomisation by bolus vs control.



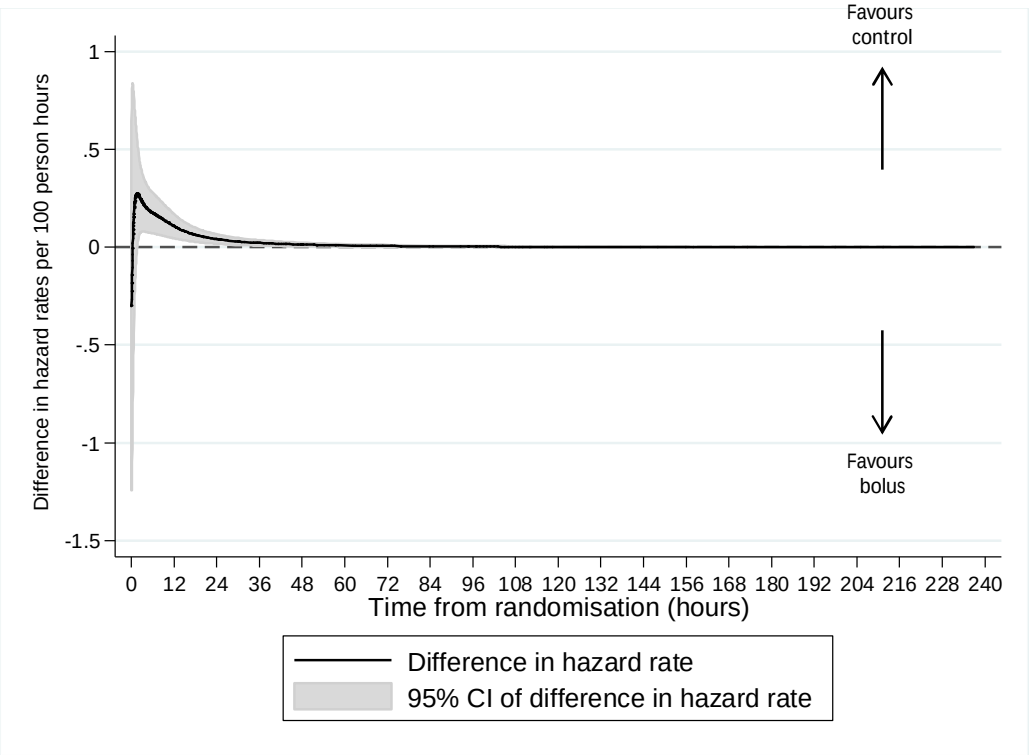
To look more closely at where risk was highest in the first few hours, the hazard rate was plotted restricted to 12 hours from randomisation in Figure 4.4.13 below. The distribution of the first bolus start and end times has been superimposed, with a median 41 minutes (IQR 13-78) from randomisation to start of bolus administration and median 78 minutes (67-126) from randomisation to the end of bolus administration. This shows that the peak in death rate was prior to when the majority of boluses were started which is consistent with the control arm also having a peak soon after randomisation and consistent with admission (and randomisation) of acutely unwell children and/or boundary issues with the modelling procedure as described above. However, the plot also indicates that the hazard remained higher in the bolus group over a long period of time, potentially preventing children from recovering at the same rate they would without a bolus, rather than them suffering a sudden increased risk of death during bolus administration. The times of deaths for each arm have also been included as scatter points on the plot to show the deaths during the first few hours from randomisation.

Figure 4.4.13: Estimated hazard per 100 phrs in the first 12 hours by bolus vs control also showing box and whisker plot of bolus start and end times (in hours from randomisation) and a scatter plot of times of death.



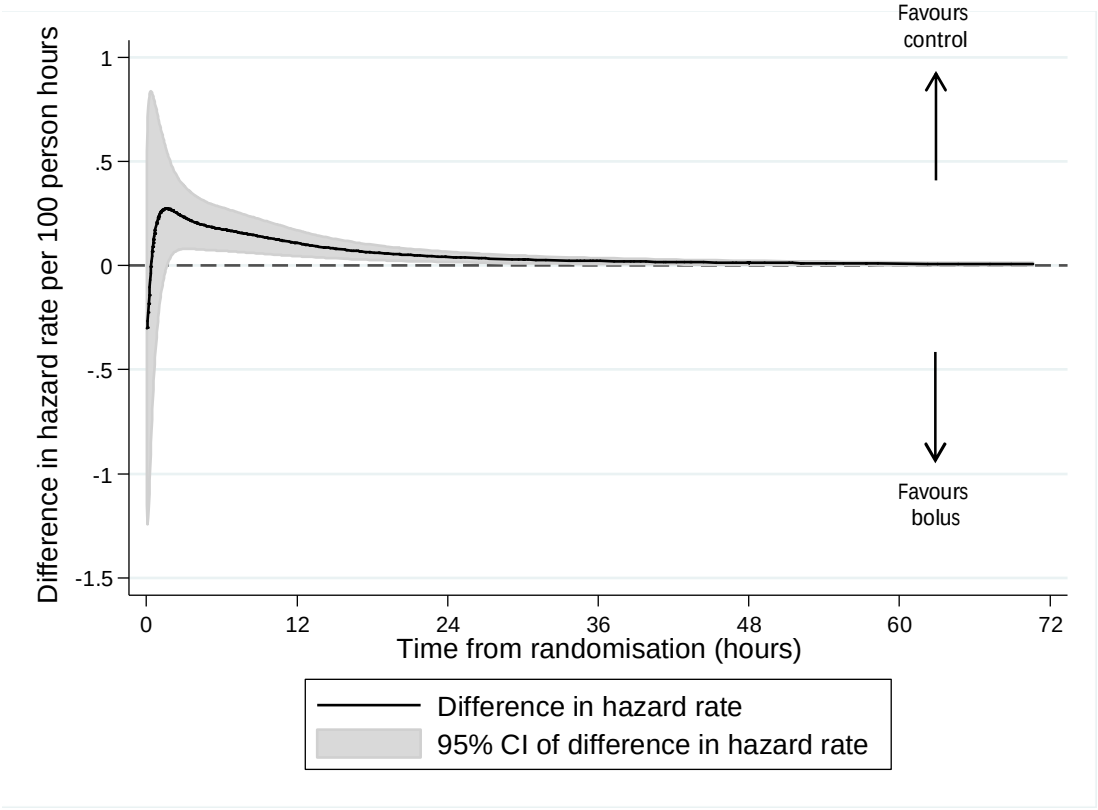
The difference in mortality risk (hazard) between the arms was also plotted (Figure 4.4.14 below) which shows the difference was zero, or very close to zero shortly after 4 days (96 hours) and before 5 days (120 hours) as mentioned above. The lower (pointwise) 95% CI bound crossed zero at 4.2 days (101 hours). This shows that the risk of mortality in the bolus arms remained higher than control arm long after the bolus itself was administered.

Figure 4.4.14: Estimated difference in hazard rates per 100 person hours between bolus and control over 240 hours from randomisation.



To investigate the difference (bolus – control) in hazard rate in more detail, it was plotted over the first 72 hours in Figure 4.4.15 below. This shows the peak of the difference more clearly which occurred 90 minutes after randomisation, where the difference was 0.274 per 100 person hours. After this, the difference decreased steadily from 4 hours to 16 hours and then slowly from that point onwards. There were also very early deaths (before 20 minutes) in both arms which may have caused the crossover from negative (favours bolus) to positive (favours control) as the model tries to estimate the changing early hazard. The difference crossed between negative and positive at 24 minutes which was 17 minutes prior to the median start time for boluses (41 minutes) and was at the 46th percentile of the distribution of time to start of the bolus. The lower (pointwise) 95% CI bound for the difference crossed at 97 minutes, showing the difference between arms became significant at a point where 86% of boluses had started and 67% had finished.

Figure 4.4.15: Difference (bolus – control) in hazard rate in the first 72 hours



The difference and pointwise 95% confidence interval was also estimated at specific time points to better demonstrate for how long the increased risk (hazard) was present for those in the bolus group Table (4.4.1).

Table 4.4.1: Difference in mortality risk (bolus-control) over time from randomisation

Time from randomisation (hours)	Difference in mortality risk per 100 person-hours (pointwise 95% CI)
1	0.24 (-0.19, 0.68)
2	0.26 (0.05, 0.48)
6	0.18 (0.07, 0.27)
12	0.11 (0.05, 0.17)
24	0.04 (0.02, 0.06)
48	0.01 (0.005, 0.02)
72	0.007 (0.001, 0.012)
96	0.003 (0.0001, 0.007)
120	0.003 (-0.0002, 0.007)

4.4.4.1 Sensitivity analyses around effect of bolus

Another question of interest is regarding the impact that very early deaths (deaths in the first hour (n=39)) had on the difference in mortality risk between the arms. As the median time to bolus start was 41 minutes, those in the intervention arm may not have even received any bolus or would not have completed a bolus if one had started. These early deaths were noted in the main trial paper and included in all analyses. Some secondary analyses with the ERC defined terminal clinical events in a subsequent paper excluded this subgroup [157]. Thus, a sensitivity analysis for this chapter excluded these early deaths.

The hazard was estimated with the same model as above but without deaths within 1 hour of randomisation. Figure 4.4.16 and Figure 4.4.17 below show that the size of the peak was reduced, as expected, given that deaths have been removed, but the shape was very similar compared to the main analyses. The difference in hazard was also very similar to the main analysis but without the initial crossover from negative to positive, supporting the crossover being due to the FPM trying to model the early deaths.

Figure 4.4.16: Estimated hazard by bolus vs control over the first 12 hours excluding deaths prior to 1 hour from randomisation.

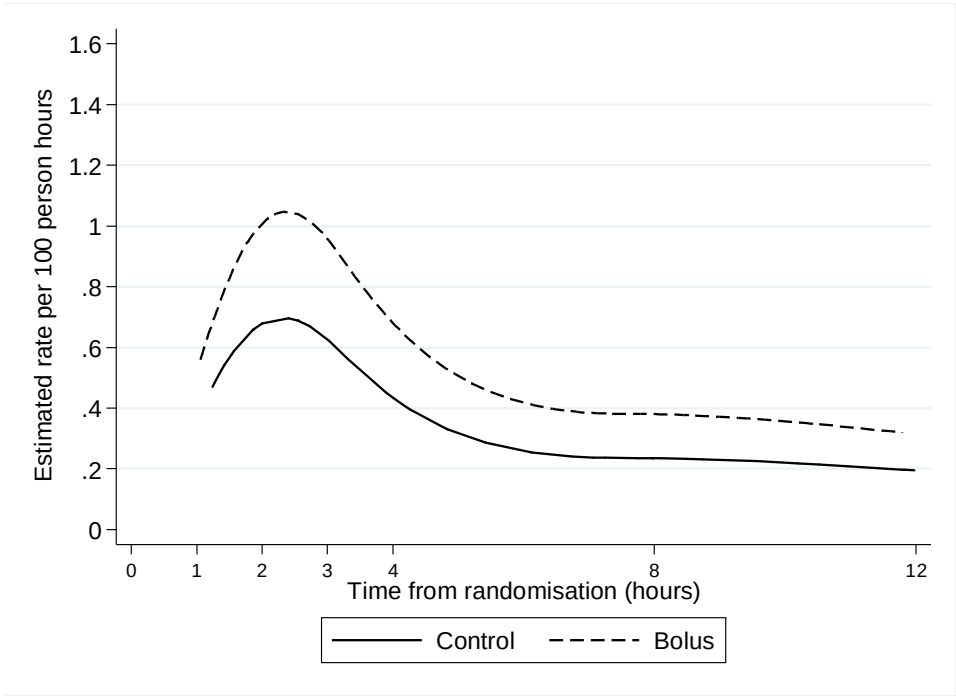
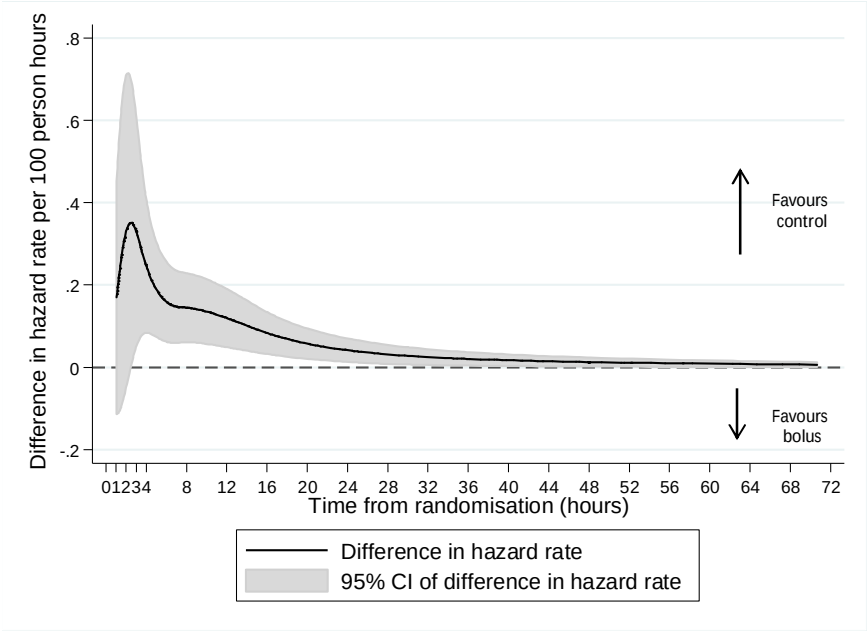
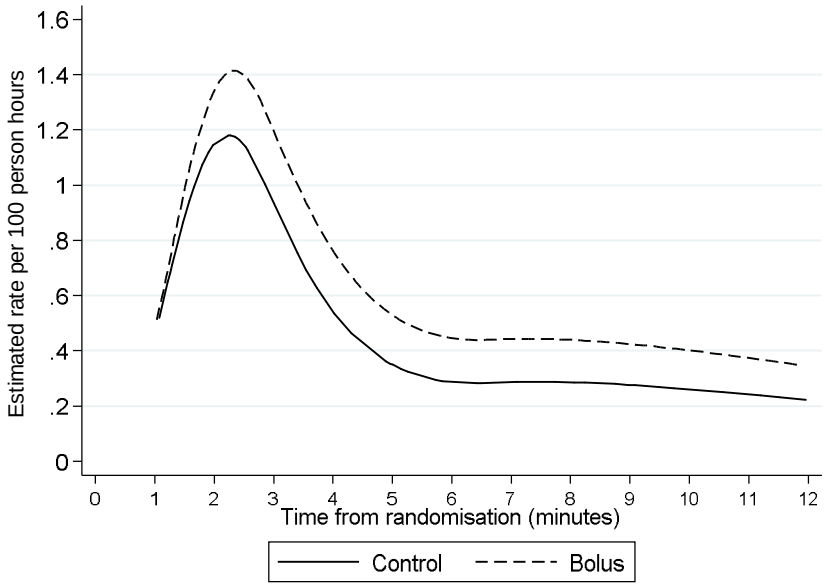


Figure 4.4.17: Difference in hazard excluding deaths within the first hour from randomisation



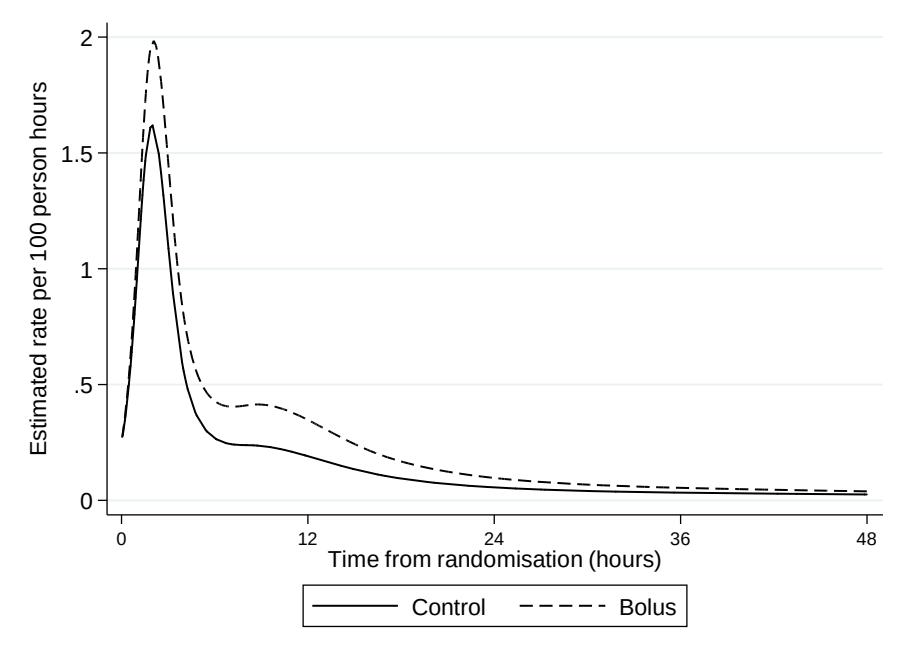
As discussed in the section on sensitivity analyses for the overall hazard (page 237), another consideration regarding the peak in hazard is whether modelling on the log time scale forced the estimates of the hazard to create an asymptote going to zero or infinity when time was less than 1 hour. A straightforward sensitivity analysis simply added 1 hour to all survival times in the FEAST data set and fitted the same flexible parametric model as the main analysis (with 5df (4 knots) and 2df (1 knot) for the time-varying covariate). This made very little impact on the overall shape of the estimated hazard (Figure 4.4.8). This perhaps might have been expected, since the knots merely shift 1 hour into the future.

Figure 4.4.18: Estimated hazard by bolus vs control over first 12 hours, having added 1 hour to all event times.



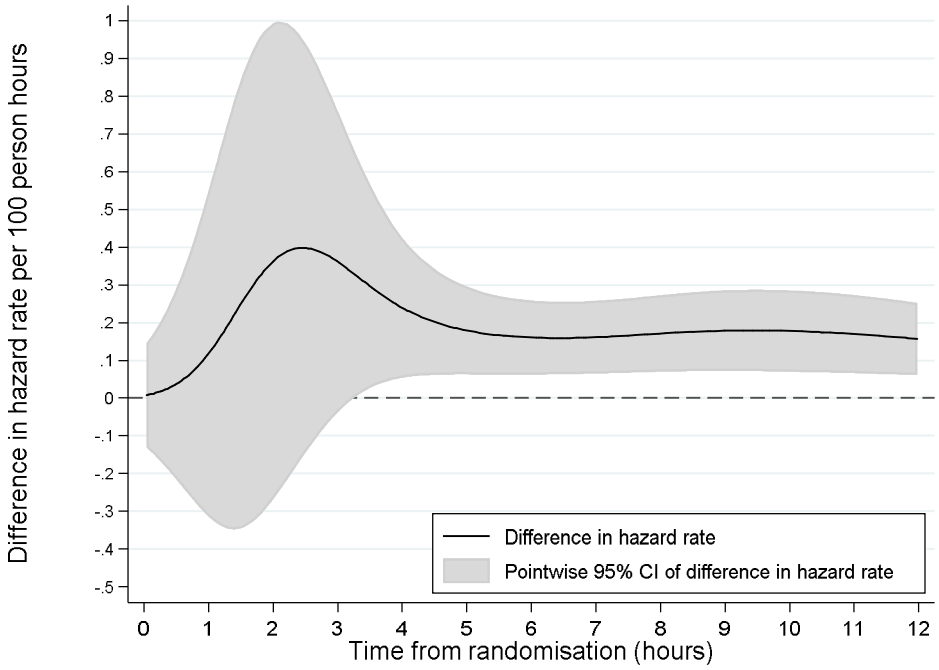
As described above, a model fitting flexible parametric models on the untransformed time scale rather than log time was also fitted to the FEAST data, and the AIC of models with different numbers of knots for the baseline and time-varying effect were informally compared. The lowest AIC was given by a model with 6 df for the baseline hazard and 3 df for the time-varying effect (Figure 4.4.19).

Figure 4.4.19: Estimated hazard per 100 person hours by bolus vs control in the best fitting model on the untransformed time scale (df(6), dftvc(3))



The estimated hazard from this model, shown in Figure 4.4.19, allowing a time-varying effect of bolus, was consistent with the model using log cumulative hazard and log time (shown in Figure 4.4.12 and Figure 4.4.13). Although the maximum of the peak in the bolus arm was higher than the control arm using this model compared to modelling with log time, the difference at the time of the maximum was not significant (Figure 4.4.20). What remained clear was the extended period of higher risk in children receiving boluses compared to maintenance fluids. Furthermore, these data did suggest a period of increased risk post randomisation was present in the FEAST data.

Figure 4.4.20: Difference in hazard per 100 person hours estimated using a flexible parametric model on the untransformed time scale.



The 95% CI in Figure 4.4.20 are pointwise and it was unclear how to best represent the uncertainty across the whole curve. To confirm these 95% CI were not unduly affected by the model fitting procedure I calculated a bootstrapped 95% CI based on percentiles with 1000 replications (Figure 4.4.2). These estimates confirmed that the mortality risk in the bolus arm remained higher than the control for longer than 48 hours after randomisation when modelling with untransformed time compared to log time (see Table 4.4.1 for pointwise model based 95% CI from a model for cumulative log hazard and log time). To determine at what time the risk become significantly different from zero, using the model on the untransformed time scale and bootstrapped confidence intervals, I used the bisection method. This showed the difference in the hazard between the two arms became non-significant at the 5% level at 53 hours.

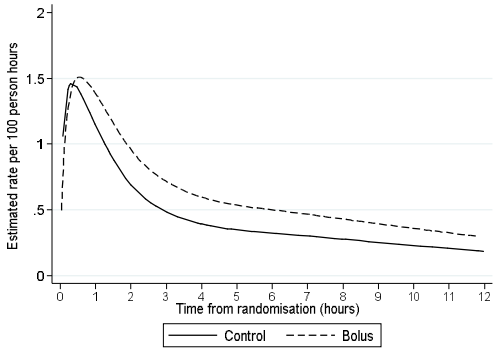
Table 4.4.2: Estimated difference in estimated hazard per 100 person hours at different time points

Time point	Difference (bootstrapped 95% CI)
12 hours	0.16 (0.06, 0.27)
24 hours	0.04 (0.02, 0.07)
48 hours	0.01 (0.002, 0.03)
52 hours	0.01 (0.003, 0.02)
53 hours	*0.01 (-0.0001, 0.02)*
54 hours	*0.01 (-0.0004, 0.02)*
72 hours	*0.005 (-0.002, 0.01)*
96 hours	0.002 (-0.005, 0.009)

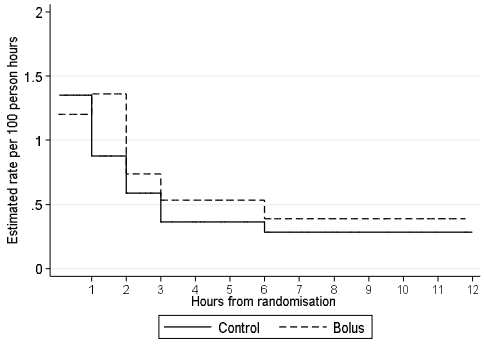
I then decided to model the hazard over time making as few assumptions as possible, using a piecewise exponential function, which assumes a constant hazard for a defined time period and then a different constant hazard for the next defined time period (Figure 4.4.19). Although this model is not generally considered to be a realistic portrayal of the underlying hazard, as the hazard is unlikely to have discontinuities and jump from one level to another, it does allow the estimated hazard to take any value for those defined time periods. Another disadvantage is that it has limited power to detect a difference between two groups within each period (with relatively few events) since this is based on a test between groups within each interval. The changepoints used here were at 1, 2, 3, 6, 12, 24, and 48 hours from randomisation and the plot below shows that there was slightly higher estimated hazard in the bolus arm after 1 hour compared to 0-1hr but that this difference was compatible with chance (p=0.27). This slightly higher estimated hazard matched the estimated hazard in the control arm from the previous time period. Then, for both arms, the hazard dropped quickly over the subsequent 5 hours from randomisation. The shape of the piecewise constant hazard in both arms in Figure 4.4.21b) was consistent with the estimated hazard from the flexible parametric model in Figure 4.4.13 (page 245) and the model in Figure 4.4.18 (page 250) and all three models were plotted in Figure 4.4.21 below. However, as expected given power issues above, I found no evidence of a difference between bolus and control in 0-1 hour (p=0.73), or 1-2 hours (p=0.25) in the piecewise model.

Figure 4.4.21: Estimated hazard per 100 person hours by bolus vs control within 12 hours of randomisation from a) model using log time scale and log cumulative hazard, b) piecewise exponential model and c) model using untransformed time scale and cumulative hazard.

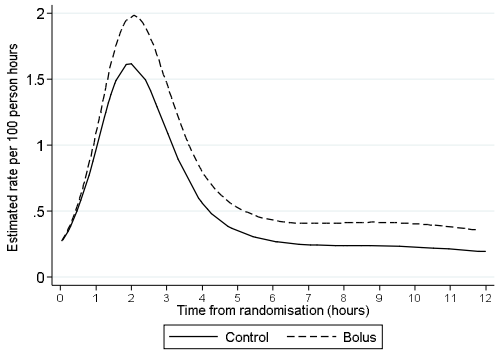
a) Model using log time scale and log cumulative hazard



b) Piecewise exponential model



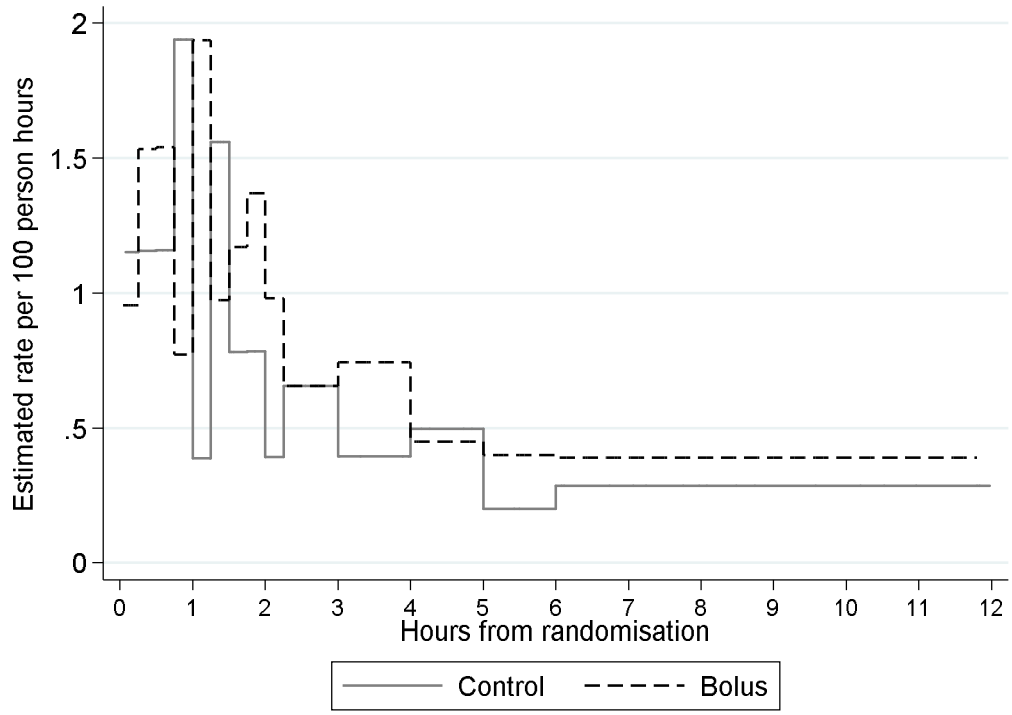
c) Model using the untransformed time scale and the cumulative hazard



The sensitivity analyses give some indication that the peak in risk immediately after randomisation was present in the FEAST data. To further understand the data without making model assumptions, a piecewise exponential model was fitted with smaller intervals in the first two hours. Cut points were at 15 minutes, 30 minutes, 45 minutes, 1 hour, 1 hour 15 minutes, 1 hour 30 minutes, 1 hour 45 minutes, 2 hours, 2 hours 15 minutes, 3 hours, 4 hours, 5 hours, 6 hours, 12 hours, 24 hours, and 48 hours. The model included an interaction between the

bolus effect and time which allowed the bolus effect to vary within each time period (Figure 4.4.22).

Figure 4.4.22: Estimated rate per 100 person hours for bolus vs control from a piecewise exponential model fitted to the FEAST data.



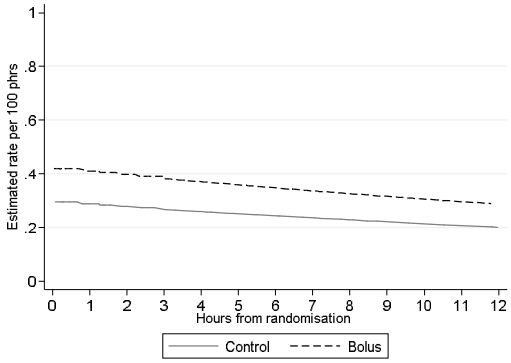
Although the piecewise exponential model had a discontinuous hazard at each cutpoint, the figure showed an interesting shape around the 1 hour point. The estimated hazard per 100 person hours in the bolus arm was low in the 15 minutes prior to the 1 hour cutpoint, and then for the 15 minutes after 1 hour rose to the level the no-bolus arm had been in the previous 15 minutes. This suggests there may have been some differences between arms in how the data were recorded. The bedside observations in the trial were done at specified intervals during the trial and the first one was at 1 hour. The hour was measured from randomisation for the control arm and from initiation of the first bolus for the bolus arm. This may be due to the fact that getting intravenous access in some children was very hard and caused delayed start of bolus therapy, and boluses were given over 1 hour so the most appropriate time to review the child was after the bolus had finished. Even though intravenous access was also required in the control arm, the review at 1 hour did not depend on this and so was more likely to be exactly at 1 hour from randomisation. This may have led to more deaths being recorded at just past

the hour from randomisation for the bolus arm as that was when they were more likely to be observed, compared to just at or prior to the hour for the control arm. In addition, there was also a digit preference for recording times as ending in 0 or 5 indicating that the recording of times of death may have been occasionally unconsciously rounded to the nearest 5 minutes. In the first two hours there were 26/76 (34%) death times ending in 0 and 23/76 (30%) ending in 5, compared to a maximum of 9% for each other digit. There may also have been delays due to ongoing resuscitation where some children may have had a delayed time to final declaration of death, although this is unlikely to have introduced any differences between the arms.

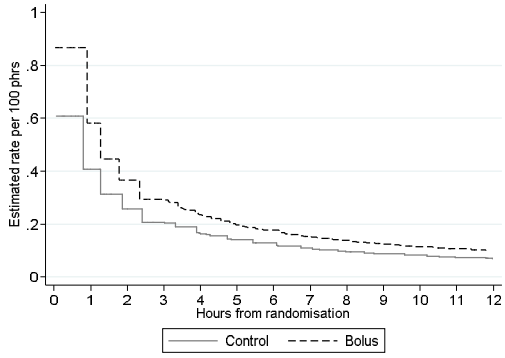
To account for possible bias in the recording of death times, especially in the first two hours, parametric models for interval censored survival time were explored. These models enabled times in the first two hours to be known within intervals but not given exactly, thus allowing an appropriate risk set to be calculated when the death occurred. The intervals I chose were from 0-45 minutes, 45 minutes to 1 hour 15 minutes, 1 hour 15 minutes to 1 hour 45 minutes, 1 hour 45 minutes to 2 hours 15 minutes and 2 hours 15 minutes to 3 hours from randomisation. The exact time of death was used when the time occurred outside of these time intervals. Interval censored survival analysis was only available for analyses in Stata using standard parametric models (exponential, Weibull, Gompertz, lognormal, loglogistic and generalised gamma) and so each of these models was applied to the interval censored FEAST data and the AICs of the models compared. The Gompertz model had the lowest AIC (3872) and gave a monotonically decreasing function over time with a higher estimated hazard in the bolus arm (Figure 4.4.23a). The other model on the proportional hazards scale (rather than the accelerated failure time scale) that could be considered with interval censored data was the Weibull model (AIC of 3922) (Figure 4.4.23b).

Figure 4.4.23: Estimated hazard for bolus vs control from interval censored data using a) Gompertz model and b) Weibull model at the upper endpoint of the time intervals.

a)



b)



Analysing the event times in the first three hours as interval censored data showed that the estimated hazard was high just after randomisation and monotonically decreased using the Weibull model, and was highest within 1 hour of randomisation with a slow steady decrease using the Gompertz model. But these parametric models were not able to model the estimated hazard over time in the same way the FPMs do allowing for increases and decreases in the same function. Thus a direct comparison with the FPM used to estimate the hazard with a complex function was not possible.

4.4.5 Competing risks analysis

The terminal clinical events (TCE) have been previously analysed using competing risks analysis to estimate their cumulative incidence and sub-hazard ratios for bolus vs control [157]. Here, I am interested in the hazard (instantaneous mortality risk) so instead I modelled the cause specific hazards using FPMs for each terminal clinical event (TCE) taking into account that a fatal event adjudicated as one predominant TCE logically means that it was not possible that the mode of death could be predominantly from any of the other TCEs. The same model as the main analysis with bolus as a time-dependent effect on the scale of log time was used to fit these models (with 5 df (4 knots) in the spline function for the baseline effect and 2 df (1 knot) in the time-dependent effect on a proportional hazard scale).

The mortality risk with a cardiogenic TCE (Figure 4.4.24a) peaked within 4 hours and had a slightly higher peak in the bolus arm than the control arm (up to 1.2 deaths per 100 person hours (phrs) compared to 0.9 deaths per 100 phrs respectively) with the shape of the peaks in both arms similar to those seen overall. This peak was also where there was the greatest difference between the arms (Figure 4.4.25a). The risk at four hours in the bolus arm was still significantly higher than the control arm and decreased slowly until around 24 hours but remained significantly above the control arm until 30 hours.

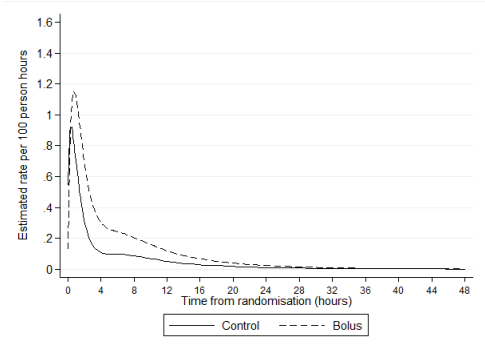
The mortality risk for neurological TCEs (Figure 4.4.24b) was increased in the control arm for the first 8 hours, but this difference was not significant as seen (Figure 4.4.25b). Of relevance, is that a series of small phase II trials showed that albumin was better than other fluids in children with cerebral malaria [32]. To investigate whether the results were consistent with those findings the mortality risk was estimated for the saline and albumin arms separately compared to the control arm (inset Figure 4.4.24b). This showed that mortality risk was increased in both the saline and control arms compared to the albumin arm but the differences between albumin and the control arm, and the albumin and saline arms were not significant (inset Figure 4.4.24b).

The mortality risk for a respiratory TCE (Figure 4.4.24c) was increased in the bolus arm similar to the overall mortality model but with a much smaller peak of mortality risk in the first four hours. There was only weak evidence for a significant difference between arms, though with the lower (pointwise) 95% confidence interval of the difference just above zero for 3 hours to 29 hours from randomisation (Figure 4.4.25c).

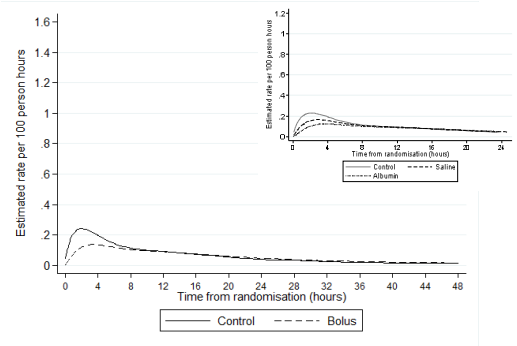
An unknown TCE could still occur during admission (essentially an unwitnessed death) and this was when there was the highest initial mortality risk. Figure 4.4.24d shows that the mortality risk was very similar between bolus and control and showed no peak in risk but simply a steep decrease following randomisation and then a slow decline after 4 hours. There was no evidence of a difference between arms for this TCE (Figure 4.4.25d).

Figure 4.4.24: Hazard rates for bolus vs control for different terminal clinical events

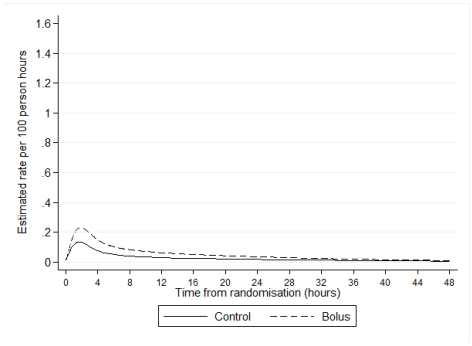
a) Cardiogenic terminal clinical event (n=146 deaths)



b) Neurological terminal clinical event (n=106 deaths) (inset showing rates by fluid)



c) Respiratory terminal clinical event (n=76 deaths)



d) Unknown terminal clinical event (n=41 deaths)

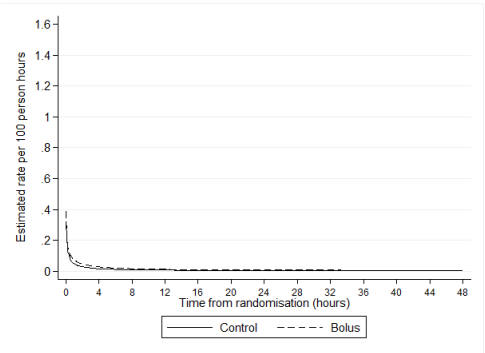
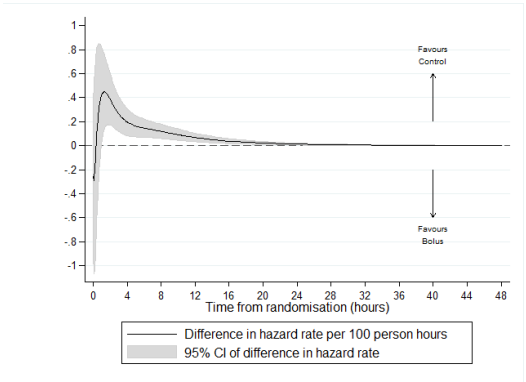
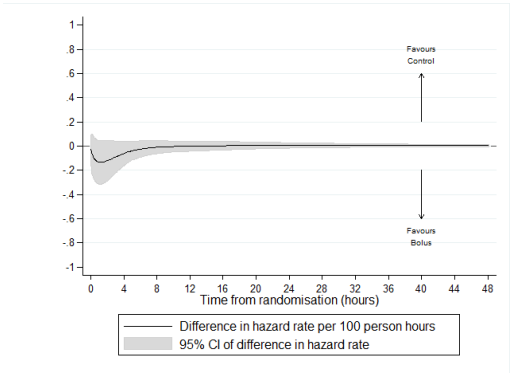


Figure 4.4.25: Difference in hazard rates for different terminal clinical events

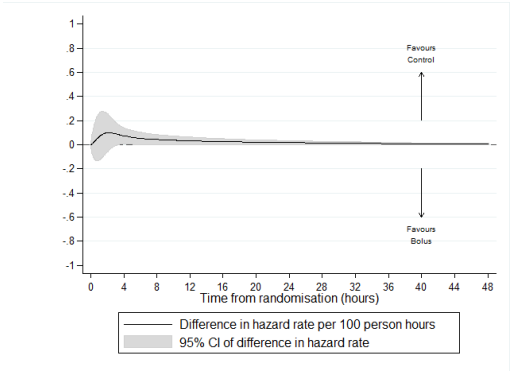
a) Cardiogenic terminal clinical event



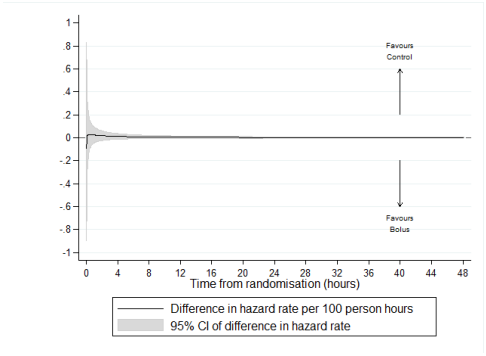
b) Neurological terminal clinical event



c) Respiratory terminal clinical event



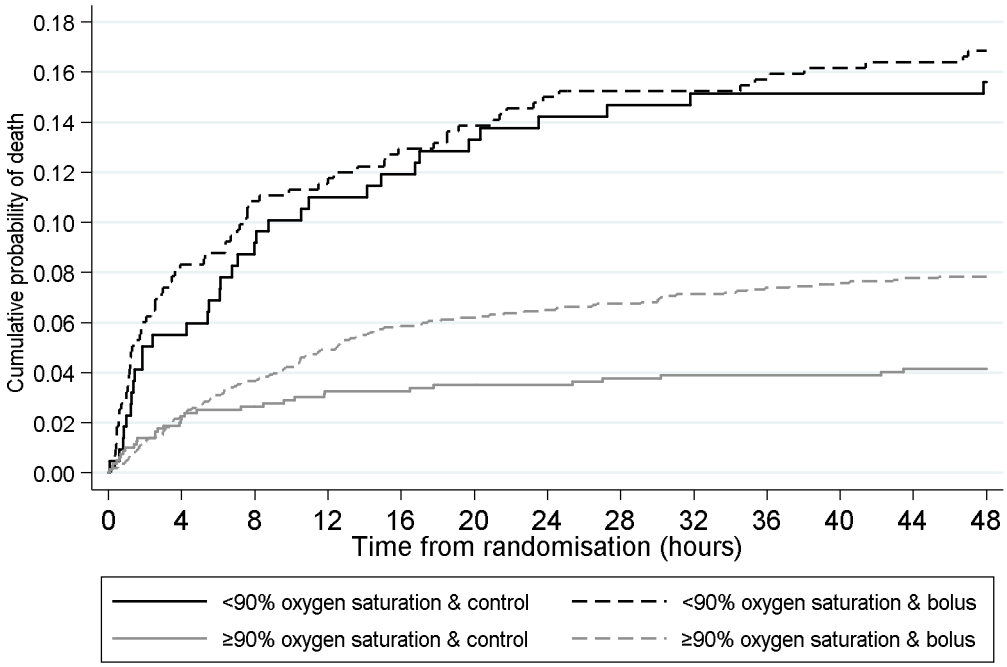
d) Unknown terminal clinical event



4.4.6 Analyses by oxygen saturation level

Chapter 3 of this thesis found that there was there was a differing impact of fluid resuscitation on mortality dependent on a child's level of oxygen saturation (SPO_2) at baseline, and those with higher oxygen saturation were at a relatively greater risk of mortality if they received boluses. To understand this interaction further, the timings of the deaths were analysed comparing between those with <90% oxygen saturation on admission and those with ≥90% oxygen saturation (judged by WHO as not a target for oxygen therapy) [227]. A Kaplan-Meier plot to compare the time to death between the groups showed that the difference between the bolus and control arms appeared after 5 hours from randomisation for those with ≥90% oxygen saturation (Figure 4.4.26 below). Those with <90% oxygen saturation had a higher mortality risk overall and the two arms (bolus vs control) were similar, although diverged a little between 3 and 8 hours.

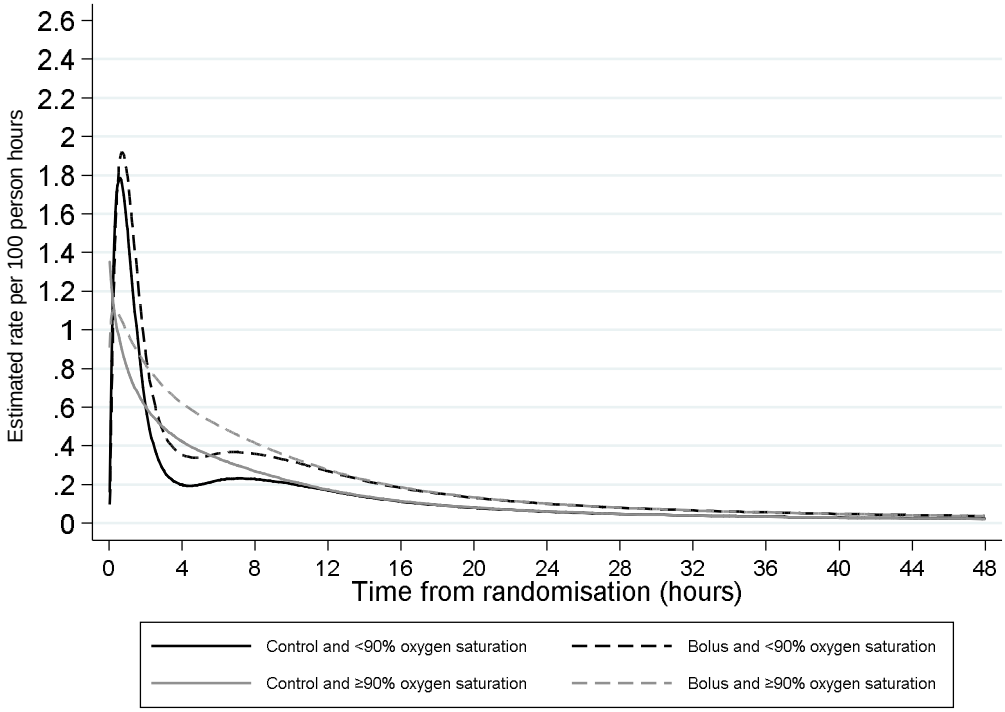
Figure 4.4.26: Kaplan-Meier failure plot split by bolus vs control and <90% vs ≥90% baseline oxygen saturation over 48 hours from randomisation.



Flexible parametric models were used to examine the underlying mortality risk by oxygen saturation group and randomisation arm. The AIC of models with differing numbers of degree of freedom (df) for the time varying effects of oxygen saturation (categorised as <90% vs

≥90%) were compared, keeping 5 df for the baseline hazard and 2 df for the bolus time-varying effect, as in previous models in this chapter, to find the most appropriate number of knots for the oxygen saturation time varying effect. The model with the lowest AIC had 4 df for the oxygen saturation subgroup effect. (Figure 4.4.27).

Figure 4.4.27: Estimated hazard per 100 phrs in the first 12 hours by bolus vs control and split by <90% and ≥90% oxygen saturation at baseline.

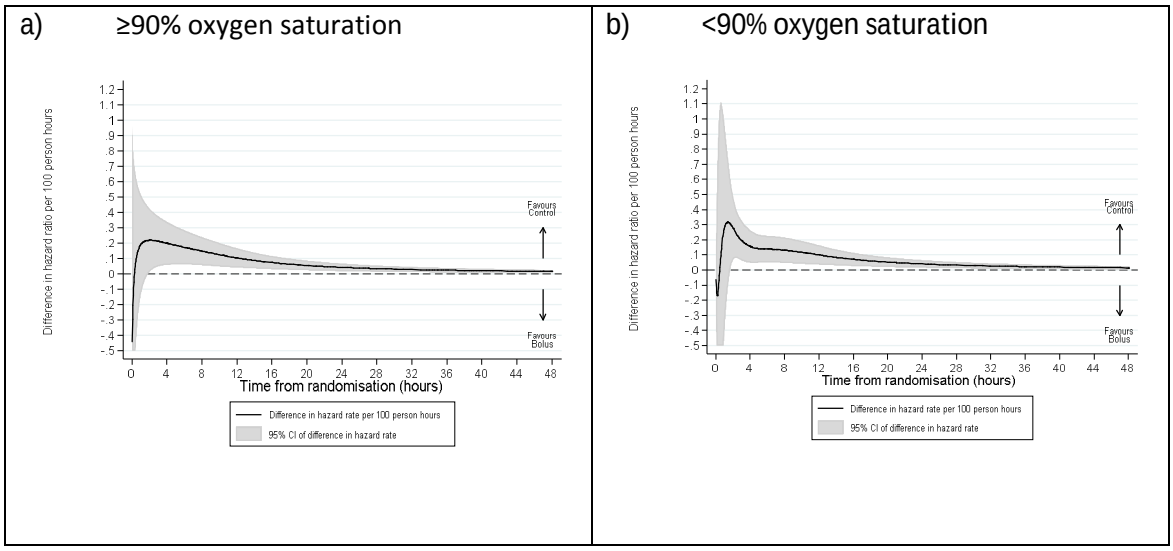


The figure above shows that the mortality risk was higher in those with <90% oxygen saturation on admission which is as expected as they were likely to be more severely ill compared to those with ≥90% oxygen saturation. However, the mortality risk for those with oxygen saturation<90% at randomisation plateaued at 0.4 per 100phrs for the bolus arm and 0.2 per 100phrs for the control arm between approximately 4 and 8 hours which was in fact lower than the mortality risk for those with ≥90% oxygen saturation during this period. This may be as some children at highest risk in the <90% oxygen saturation group had died by 4 hours, and those remaining have a lower risk. However, conditional on surviving to that timepoint, the instantaneous mortality risk in those with ≥90% oxygen saturation at baseline, and particularly in those in the bolus arm, was still higher between 4 and 8 hours.

The peak in the risk difference for those with <90% oxygen saturation at baseline led to a higher risk difference between bolus and control within the first 3 hours from randomisation compared to those with ≥90% oxygen saturation, possibly reflecting organ failure in a high risk

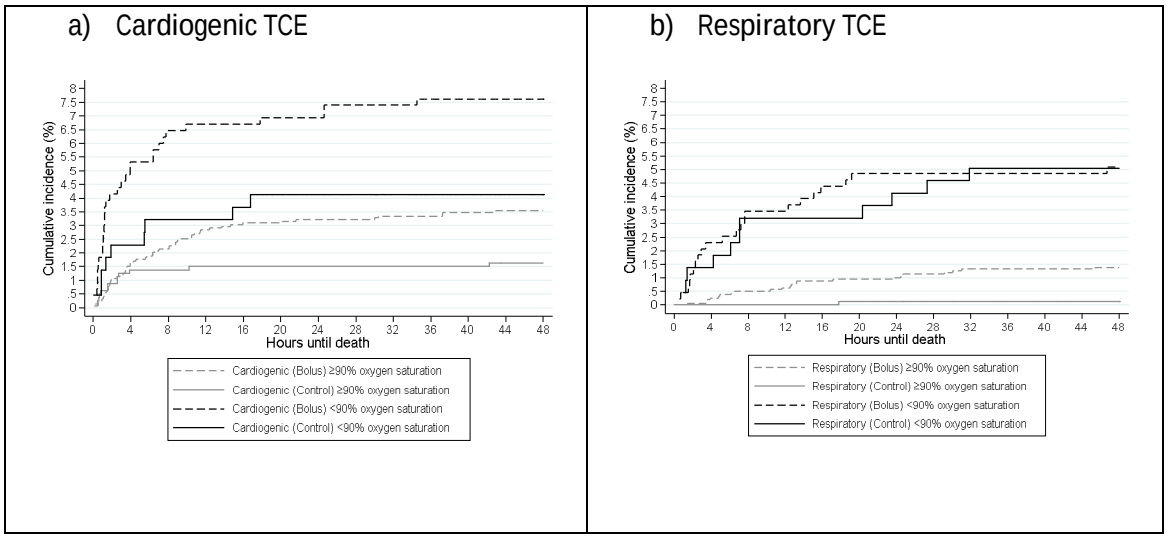
group, but then the mortality risk difference was in fact lower for the remaining 45 hours. (Figure 4.4.28a and b).

Figure 4.4.28: Difference in hazard rates for a) children with ≥90% oxygen saturation at baseline and b) children with <90% oxygen saturation at baseline.



The timing of deaths presented using a Kaplan Meier plot in Figure 4.4.26 could also be linked to terminal clinical event information to look at causes of death over time by the different oxygen saturation groups.

Figure 4.4.29: Cumulative incidence of a) cardiogenic and b) respiratory terminal clinical events by oxygen saturation levels and bolus vs control.



These show that the difference between arms in cardiogenic deaths appeared very quickly (around 1 hour) for those with <90% oxygen saturation, whereas for those with ≥90% oxygen saturation the difference appeared later from 4 hours. The eight cardiogenic deaths in children with <90% oxygen saturation in the control group also all occurred before 20 hours (out of 34 deaths in this group in total). Thus, the other causes of death reduced the difference between the bolus and control arms as seen in the overall Kaplan Meier plot in Figure 4.4.28. Figure 4.4.29b) also shows that there was a gradual increase in cumulative incidence for those in the bolus arm with ≥90% oxygen saturation, appearing after 4 hours, similar to that seen in cardiogenic deaths for children with ≥90% oxygen saturation. Children with <90% in the control arm had very similar cumulative incidence over time from both cardiogenic and respiratory TCEs; whereas children with <90% oxygen saturation in the bolus arms had a greater cumulative incidence of cardiogenic TCEs compared to respiratory TCEs.

4.5 Discussion

The FEAST clinical trial showed a 3.3% absolute increase in mortality by 48 hours (and 3.4% increase by 28 days) in the bolus arms compared to the control arm [33]. There were two plausible hypotheses regarding this increased mortality risk from the bolus intervention: either that the risk in the bolus arms increased directly following administration of the bolus or that there was a slower decrease in risk over time. Flexible models allowed me to distinguish between the two, and by estimating the mortality risk (overall hazard) directly I was able to estimate the time point the rate was highest in these children and how long the excess risk remained higher in the bolus arms relative to the control arm. I found that there was a similar peak in mortality risk for both randomised arms (bolus and control) in the first four hours from randomisation and then a slow decrease in risk. The risk in the bolus arm did not increase further but remained higher until 96 hours after randomisation, compared to the control arm. This shows that children in the bolus arms were recovering more slowly than those in the control arm and that the detrimental impact of receiving a bolus on mortality risk continued beyond admission up to 4 days after randomisation. Sensitivity analyses also supported the extended detrimental impact of receiving a bolus, although only to 53 hours from randomisation.

These findings are implicit within the Kaplan Meier (KM) plots from the main trial publication [33] which suggested that mortality risk was high soon after randomisation and peaked at a similar time in both arms; however, the specific changes in risk are not explicit just from the KM plots. If the peak in risk had been at different times for each arm then the cumulative probability of death from the KM plot would have clearly increased at different times. But from the Kaplan Meier plot alone it is difficult to distinguish between a much higher immediate peak increase in mortality risk or a slower decrease in risk.

There has been much discussion around why there was excess mortality in the bolus arm of the FEAST trial. One hypothesis is based on re-perfusion injury, where the protective response was preventing a vascular spread of pathogens [228, 229] and the fluid resuscitation could have resulted in a 'flush' mechanism in the tissues, resulting in poor outcomes further along in the admission or post-admission [157, 230]. Another hypothesis is that bolus fluid resuscitation may cause adverse effects on vascular hemodynamics and myocardial performance [33]. On admission children's organs may have been hypoperfused (so non-critical areas such as the skin, liver/gut and kidney are temporarily shut-down to conserve blood flow and the blood is diverted to the vital organs (brain, lungs, heart)), serving as a

protective physiological response. When circulation is rapidly restored the 'protective response' is diverted and cardiac output is unable to perfuse the non-critical areas (as it is at its greatest compensation) leading to cardiovascular collapse. This is supported by a study in sheep which found a marked increase in levels of cardiac troponin I at 16 hours in the sepsis group resuscitated with saline compared to a control group with sepsis not resuscitated with saline and a healthy group resuscitated with saline [231]. Cardiac troponin I is a protein released when the heart muscle has been damaged. Thus an increase in this marker following bolus resuscitation in sheep is compatible with our finding that the largest difference in mortality risk in FEAST was seen in those that have a cardiogenic TCE (i.e cardiovascular collapse). We did not have any post bolus stored samples to test this hypothesis directly in the FEAST trial however.

The peak in mortality risk was similar in both arms and very close to randomisation, occurring even before most boluses had finished. In the literature identified through the review at the beginning of this chapter, a peak in mortality risk has previously been seen in HIV-infected children starting treatment in a trial (ARROW) where mortality was highest 34-51 days after randomisation and was not accounted for by adverse treatment effects [209]. However, in the ARROW trial, the authors suggest that the peak was due to the fact there was a full two stage (i.e over a number of days) consent process in clinics and so children that were within days of death were plausibly not randomised into the trial, leading to an increase in mortality risk shortly after randomisation. A peak in excess mortality rate was also found in children with brain tumours within 2 years of diagnosis [210]. This was, however, an analysis of a population based cancer registry and thus there was not the same consent process as a clinical trial to exclude those about to die. In that paper the excess mortality rate was presented as part of relative survival analysis which is the mortality risk compared to the expected mortality risk in the general population, and it is plausible that children with cancer genuinely have a peak in mortality around this time. Although it has been well documented that child mortality rates shortly following hospital admission are high (14% of deaths were within 4 hours of admission in one study in Kenya [47]) I found that mortality risk in these settings has not been modelled over time with FPMs in the published literature, so I was not able to directly compare my findings with other studies.

The sensitivity analyses for estimating hazard overall suggested that the flexible parametric modelling on a log time scale may be artificially creating a peak, as the log hazard function creates an asymptote to zero (or infinity) when time from randomisation is very close to zero. Fitted to the simulated data with underlying initial high constant hazard, FPMs gave a peak in mortality risk even though the underlying hazard was high and then decreasing. This

hypothesis was supported by modelling on the untransformed time scale which estimated a monotonically decreasing function in the simulated datasets. This was a better reflection of the underlying shape of the hazard function used to simulate the data. I have not found discussions of this issue in the published literature, which may be due to FPMs having been widely used in diseases where there would usually be a few months or years prior to the first event, particularly cancer. But nevertheless, even using an untransformed time scale, there is evidence that there was an underlying peak in mortality risk (hazard) in the FEAST data which may be due to the way that deaths were discovered and recorded in the first two hours. The piecewise exponential model by bolus vs control suggested a small peak in mortality risk and the FPMs on the log time scale, including the time varying effect of bolus, were consistent with the piecewise model. The FPM fitted on the untransformed time scale with a time-varying effect of bolus was also consistent with an underlying peak in the data, whereas when these models were fitted to the simulated piecewise exponential models with constant high hazard initially there was no peak estimated. This suggests there may have been changing (and not constant) high hazard in the data immediately after randomisation.

The piecewise exponential models also showed how variable the hazard was when broken down into small intervals; unfortunately there was not enough power to show differences between arms for any of the time intervals in these models. Table 4.5.1 below summarises the findings when considering whether the early increase in the hazard prior to a decrease, found when using FPMs on a log cumulative hazard scale, was an artefact of the model or was present in the data. In summary, although FPMs may artificially create a peak in the hazard, there was some evidence that the peak is actually present in the FEAST data. However, what is consistent across all statistical approaches, both modelling with log time and modelling on an untransformed time scale, is that I found no immediate difference in mortality risk between randomised arms around the time of the peak, but instead the excess mortality risk in the bolus group occurred after the first 4 hours and remained up beyond 48 hours from randomisation. The models on the log time scale indicated that this excess mortality risk in the bolus group could have remained until up to 4 days from randomisation.

Table 4.5.1: Evidence for and against a true underlying peak in mortality risk (hazard) in the FEAST data

Peak present in data	**Artificially created by model**
Piecewise exponential fitted by bolus vs control suggests there was a small peak in mortality risk, possibly caused by the way deaths were recorded around the 1 hour point.	FPM fitted using log time on simulated data from an underlying initial high hazard that did not vary still gave a peak in mortality risk.
The peak remained (although was lessened) when deaths between screening and randomisation were included.	The potential for the use of splines of log time to create an asymptote (going zero or infinity) in the hazard function when event times are close to zero is a disadvantage of FPMs that has not been fully explored (personal communication – M Crowther).
Modelling on the untransformed time scale gives no peak in the simulated data with an underlying constant high initial hazard that did not vary, but did give one with FEAST data suggesting the underlying hazard was may have differed from the piecewise model used to simulate data.	Peak lessened and 95% CI widened in analysis including deaths between screening and randomisation.
Modelling bolus vs control with untransformed time and with log time both gave peaks in mortality risk in the first four hours from randomisation	

One clinical hypothesis for the peak in mortality risk is that early supportive care may have adverse consequences i.e that the child has survived to hospital and then is given various treatments which imbalances the mechanisms that are keeping them alive. There is some literature discussing this [232] but very few clinical trials or studies have taken place, as much of early supportive care is an intrinsic part of clinical practice. In the FEAST trial standard supportive treatments included oxygen therapy, rapid correction of hypoglycaemia, anti-

pyrectics and for all children parenteral antibiotics were recommended (to be given immediately to comply with surviving sepsis recommendation (ie within the first hour)). Literature on toxin release by bacteriocidal antibiotics, including the Jarisch-Herxheimer reaction (a systemic reaction believed to be caused by the release of endotoxin-like substances when harmful microorganisms are killed during antibiotic treatment) remains a controversial and debatable topic. Few trials have been prospectively designed to address this question in the critically sick patient. One in melioidosis showed differential antibiotic endotoxin release but no difference in survival, although patients were more stable than those enrolled in the FEAST trial [233]. Another study compared mortality rates before and after implementation of a new more rapid protocol for sepsis, including antibiotics and fluids within the first three hours, and found lower mortality with completion of rapid administration of antibiotics (but no difference for fluids), although many other factors may have changed in the study site in the two time periods thus potentially confounding the results [234]. Increased mortality was found with an early resuscitation sepsis protocol in adults in Zambia, but the main components of the intervention were fluid boluses and vasopressors rather than other supportive treatments [235]. In FEAST, including deaths between screening and randomisation widened the confidence intervals around the overall mortality risk, but did not impact the maximum of the peak. Therefore, although the hypothesis that early supportive care is harmful cannot be excluded, there was at least some support for the alternative, namely that children were simply arriving at hospital with high mortality risk over the first 2 hours and which steeply declined following treatment. This was also consistent with the hypothesis that the peak in mortality risk seen in the FEAST data was due to the trial processes and data collection, rather than a clinically meaningful increase immediately after randomisation.

Pointwise confidence intervals were used throughout the chapter. The alternative would be continuous confidence intervals for the function in question. Instead of using the value from a chi-squared distribution with 1 degree of freedom to estimate the confidence interval, the value would be found from a chi-squared distribution with the number of degrees of freedom based on the number of model parameters. As the flexible parametric models have at minimum 7 parameters, this would lead to a very wide confidence interval which may be uninformative and is likely to be highly over-conservative. I therefore used bootstrapping as an alternative approach to try and estimate confidence intervals.

I chose to model cause specific hazards in the competing risks analysis as these estimate an instantaneous death rate (hazard) from the specific TCE out of those that had survived to that point (as children are censored at the date of their death or the end of follow up). The alternative would be to estimate the sub-distribution hazard which corresponds to the overall

effect on death from a specific TCE. When overall event rates are low, as here, these two quantities are similar.

The results from the cause specific hazard analyses in this chapter support the findings in previous analyses, where the cumulative incidence was higher in the bolus arms for those with a cardiogenic TCE and that this TCE contributed most to the excess mortality in the bolus arms [157]. I found the greatest difference in mortality risk in this TCE and that the distribution of the difference over time is very similar to the overall difference between arms. The neurological and respiratory TCEs showed a low and straightforward decline following a peak, and the unknown TCE had small numbers with increased risk only soon after randomisation. These results fit in within the conclusions from the previous analyses that the detrimental effect of bolus was not due to fluid overload affecting the brain or lungs. Analyses splitting by oxygen saturation (<90% compared to ≥90%) indicated that it was the period of 4-8 hours where the difference in mortality risk appeared in those with ≥90% oxygen saturation, and that the excess mortality was similar for cardiogenic and respiratory TCEs. In contrast, the excess risk in those with oxygen saturation <90% was predominantly from cardiogenic TCEs.

The main strength of these analyses is the novel application of flexible parametric survival models, enabling us to further elucidate potential mechanisms. To my knowledge, no studies have applied such models to critically ill children in these settings, and our application highlights their potential. In conclusion, more detailed analysis of the FEAST trial data demonstrates that fluid boluses did not substantially increase mortality risk on administration, but rather delayed normalization of mortality risk, preventing children from recovering at the same rate they would without a bolus for longer than 48 hours and potentially up to 4 days from randomisation.

5 Discussion

5.1 Background

There are very high levels of mortality for paediatric admissions to emergency rooms in sub-Saharan African hospitals, most often from life-threating complications of common illnesses such as malaria, and pneumonia [2]. Children can deteriorate quickly, with most mortality occurring within hours of admission. Case management of childhood diseases remains an area where good intervention coverage was not achieved in all countries that were striving towards the millennium development goals by 2015 [9] and there is a continued need to improve case management and treatment for critically ill children arriving in hospitals. There are new goals that were adopted by many countries in 2015 called the Sustainable Development Goals (SDGs) which are active until 2030 [236], and which are being monitored closely. Goal number 3 'good health and well-being' includes ending preventable deaths in children under-5 years old by 2030. A review of the under-5 year old causes of mortality in 2015 indicated the leading cause of death in sub-Saharan Africa was pneumonia (17%), followed by preterm birth complications (12%), intrapartum-related events (12%), diarrhoea (10%), malaria (10%) and sepsis/meningitis (10%) [1]. The introduction of preventive strategies such as the *Haemophilus influenza* type b and pneumococcal vaccines for pneumonia, improved water and sanitation and the rotavirus vaccine for diarrhoea, and insecticide treated bednets for malaria, should have had an impact on the burden and spectrum of disease hospital admissions [1], although data has not yet been updated to reflect this. Nevertheless, there are still many children presenting to hospital who are critically ill. For example, there were estimated to be over 25,000 children in Uganda alone in 2015 who died from pneumonia, diarrhoea or malaria [1]. A major issue for many of these critical diseases is the overlap in presentation (including sepsis which is not prioritised in the WHO guideline) since these are largely defined by clinical criteria, with many children mis-classified by the syndromic classification; in addition they may occur together. An analysis of data from a large randomised clinical trial for antimalarial treatment in children (AQUAMAT) found that those in the lowest tertile of plasma P*f*HRP2 (<829ng/mL) had a low probability of malaria-attributable death, despite all being classified as having malaria to enter the trial [41]. Also, where blood cultures are available they lack sensitivity and are only positive in a small proportion of children with pneumonia; identifying the pathogen(s) causing pneumonia remains challenging despite advances in diagnostic tests [237, 238].

One specific, high-risk and frequent complication of the undifferentiated critically-ill child presenting to the emergency room in Africa is shock. There are strong recommendations regarding management of these children but until the FEAST trial, these were based on a low/weak evidence base. Whilst there is an active debate on how shock is clinically defined [239], there is limited research into the best way to treat shock, especially in low-income settings. Between 2009 and 2011, the Fluid Expansion As Supportive Therapy (FEAST) trial was conducted to determine whether treating children in shock with fluid resuscitation (boluses) would decrease the high mortality observed in these critically ill children [33]. Fluid resuscitation, worldwide, was commonly used in high income settings to treat shock. The World Health Organisation (WHO) define paediatric shock as a child having *all of the following signs*: cold extremities, capillary refill time>2s and a weak and fast pulse [227]. These signs are individually associated with poor outcome but have not been formally validated by physiological studies against evidence of macro-vascular and micro-vascular compromise and identify very few children (only 65/3141 (2%) in FEAST) and <1% in other cohorts [239]. The WHO 2013 guidelines on how best to treat this group of children include fluid boluses [227]. Other definitions of paediatric shock (such as the American Academy of Critical Care Medicine – Paediatric Advanced Life Support (ACCM-PALS), or Advanced Paediatric Life Support (APLS)) are broader and similar to the FEAST inclusion criteria, focusing on impaired circulation and for these children boluses are not recommended [227]. Internationally, there has also been research to investigate the best fluid for resuscitation (which has focused on debate around colloids versus crystalloids) [240-242]. In Kenya, small Phase II trials also addressed the safety and efficacy of different fluids in children with severe malaria in Kenya prior to the commencement of the FEAST trial [30, 31]. These Phase II trials were undertaken with the aim of investigating whether crystalloid (saline) or colloid fluid (human albumin solution, dextran or gelofusin) boluses in these settings were safe and superior as a possible supportive therapy in severe malaria.

However, since fluid resuscitation for shocked children presenting to the emergency room in sub-Saharan Africa was not standard of care, apart from the small number of children meeting the WHO definition of shock and very critically ill, there was increasing recognition of the need to conduct a large clinical trial to investigate fluid resuscitation in a wider group of children in shock (due not only to sepsis, but also to malaria). The trial needed to be powered to detect a difference in mortality between different fluids and also between fluids delivered as boluses and a control group (receiving maintenance fluids only). As fluid boluses for shock should be started very soon after admission, this treatment was hypothesised to be life-saving globally, but more especially in these settings where mortality within 24 hours of admission is high. Other treatments including anti-malarials and/or antibiotics tend to be reliant on test results

which can take time (although some point-of-care tests have made this quicker) or have to be given empirically before the diagnosis is known. The FEAST trial enrolled 3170 children presenting to hospital in shock, of whom two thirds have malaria. The trial stopped early and results showed that fluid resuscitation was not appropriate in this setting without intensive care facilities. There was an increased mortality risk in the bolus arms compared to the control arm (which received maintenance fluids only) with a risk ratio for mortality of 1.45 (95%CI 1.13-1.86; p=0.003). There was no difference between the fluid arms (risk ratio 1.01 (95%CI 0.78-1.29) for albumin bolus vs saline bolus.

5.2 Key results: summary

This thesis has applied recently developed statistical methods, not commonly used in this research area, to investigate key questions arising from the FEAST trial results and from the need for research in critically ill children presenting to hospital in sub-Saharan Africa.

Firstly, I identified children at highest risk on arrival to hospital and created and validated a prognostic risk score which is easy to use and tailored to low-income settings [161]. The FEAST Pediatric Emergency Triage (PET) score uses bedside measures, has a range of 0-10 and showed good validation on external data. Secondly, continuous measures at baseline and throughout admission were modelled with fractional polynomials to identify any interactions with bolus. I found strong evidence for interactions with bolus and oxygen saturation and bolus and base excess in children with malaria. These were further explored with other measures of malaria severity. For this analysis I included a refinement to malaria diagnosis using quantitative *Pf*HRP2 measurement which more accurately predicts true severe malaria (separating those with severe malaria from children with severe illness and incidental parasitaemia). I also investigated whether base excess (a measure of severe acidosis) may play an important part in understanding the impact of fluids on critically ill children with malaria. The mechanisms by which fluid resuscitation increased the risk of death in FEAST were explored by estimating the proportion of treatment effect explained by different physiological measures measured over time. Overall, none of the above physiological measures were found to explain any of the treatment effect, which I suggest indicates the potential difficulties in identifying harm from treatment in a group of children with diverse presentations and syndromes where one measure is not sufficient to explain the physiological impact of boluses.

Thirdly, using flexible parametric models to estimate the mortality risk over time, I also found that the 'mechanism of harm' did not translate to an immediate clinically definable reaction to the boluses (ie simultaneously or shortly following bolus therapy as has been hypothesised). Instead, I found mortality risk in children who received boluses took longer to return to normal compared to those who received maintenance fluids. In summary, the finding of the analyses in my thesis supports the hypothesis that the mechanism of harm could be the rapid reversal of a protective response (shock) to severe illness by giving large amounts of fluids fast. Future trials in this area should focus on measurement of biomarkers in children receiving fluid resuscitation (for example in intensive care settings in well-resourced settings) that may give some insight into this mechanism.

5.2.1 Key results: What prognostic indicators for death could be used in African hospitals to identify children at greatest risk? (Chapter 2)

Chapter 2 of this thesis identified the children at highest risk in the FEAST trial and created a risk score called the FEAST PET score that could be either used to stratify children in a study or for enrolment into a trial, or for triage in the admission area [161]. Triage scores need to be simple and created from easily recorded bedside measures and the PET score includes: temperature, heart rate, weak pulse, capillary refill time, conscious level, respiratory distress, lung crepitations, and severe pallor. These signs can all be measured in the critically ill child and require little equipment and no tests. Blantyre coma score, respiratory distress and capillary refill time had been shown previously to be prognostic for mortality [47, 62, 103, 104, 109, 110, 243] . Previous research has shown that improving triage systems to better identify those at highest risk of mortality can reduce mortality rates [10]. Of note, even without using a risk score, but with better triage systems in the FEAST clinical sites, the children in the control arm of the trial had reduced mortality compared to admission mortality rates prior to the trial (or recorded in other similar settings) [33]. This could also have been due to closer monitoring and more point-of-care tests than may have been offered or affordable in usual clinical practice, but implementing improved triage is very likely to have played an important role. As well as reorganising how children are assessed and admitted, it is important to collate information from basic bedside measures that are recorded regularly on all children. This is why risk scores that are simple and include easy-to-use (and to teach) signs and measurements are important to develop and validate in these settings. In this regard, the FEAST PET score uses a simple addition process with a maximum score of 10 and a minimum of 0 to help enable possible use of the score as part of triage systems. Risk scores are also useful for risk

stratification and thus enable cross-site comparisons of hospitals adjusted for the type of patient that they treat as well as ensuring particular patients are enrolled into a studies. For example, a clinical trial may only want high risk patients and so with a risk score calculated at admission then they can more easily identify those that should be enrolled. It can also help with stratification so a future trial can also investigate treatment effects in high risk and low risk patients by using a risk score calculated at admission.

The need for risk scores to be appropriate to the countries that they are being used in was also highlighted in Chapter 2. Thus, risk scores that involve a lot of monitoring or laboratory tests are not as useful in low-income settings where those facilities are not available and scores from high-income settings are not usually generalisable to low-income settings. It is also difficult to properly validate such scores in low and middle-income country (LMIC) settings as many score components are usually not recorded, and so values for missing components are set to clinically normal levels (as directed by the researchers that developed the score). This has also been found with adult prognostic scoring systems developed in high income countries where they were found to be at best performing moderately in LMIC ICUs with limitations in calibration [244]. Many of the laboratory tests included in scores such as PRISM and PRISMIII (for example: total bilirubin, calcium, potassium, arterial oxygen tension, creatinine, and prothrombin time) that are commonly done in high-income settings are not measured/available routinely at admission in most African hospitals. Care must also be taken in applying adult risk scores to paediatric populations, even within high income countries. The quick Sequential Organ Dysfunction Assessment (qSOFA) score was developed as part of the Sepsis-3 guidelines [245] and only uses three components (hypotension, tachypnoea and conscious level) but was found to have poor performance to identify children at risk for worst outcomes in a large dataset of PICU admissions in Australia [246]. It is also important for risk scores to be validated to show good generalisability: a strength of the PET score was that it was validated on data from similar settings to the trial (Kilifi admissions data) and was shown to have good discriminatory ability. The FEAST PET score had slightly better discriminatory ability in the more general dataset from ward admissions, indicating it could identify high risk patients amongst a more general population compared to when restricted to those admitted to the KEMRI high dependency ward who were a selected group at higher mortality risk. Importantly some of these children had deteriorated in hospital prompting admission to the KEMRI ward; admission characteristics might be expected to perform more poorly in predicting overall mortality risk in these children. Even though many laboratory tests are not available in these settings, improvements in technology mean that there are more point-of-care tests becoming available and more robust and smaller machines for testing. This is why I also considered some laboratory measures, in order to test whether they could provide additional

discrimination. Three laboratory measures (BUN, lactate and pH) were added to the PET score based on their NRI statistics to create the Paediatric Emergency Triage and Laboratory (PETaL) score. This score includes lactate, which is more likely than other tests to become part of routine care; hyperlactataemia has been shown to be predictive of mortality in FEAST both in analyses in this thesis and in a separate analysis [247].

5.2.2 Key results: Exploring mechanisms of action of the boluses (Chapter 3)

The FEAST trial findings were unexpected and the questions around volume of fluids and speed of administration to the critically ill child have been debated widely since the trial results were published [248-252]. In the FEAST trial, most children who received bolus-therapy largely received 20 mls/kg over one hour. This was only repeated over the following hour after admission in those with unresolved signs of shock (37% children in the bolus arms combined). Comparing this to the WHO guidelines at the time [18], in which WHO recommended 20mls/kg as fast as possible with up to two repeats as fast as possible, and international guidelines recommending 60mls/kg over 15 minutes, the FEAST trial volumes and rates of infusion were exceptionally conservative. The amounts given in the FEAST trial were even reviewed during the trial and a protocol amendment was made to increase the amount of fluids for the first bolus to 40mls/kg as it was thought the hypothesised beneficial effect may not been seen due to the conservative amounts. The amendment was in place for the last 606/3141 (19%) children enrolled. Whilst international recommendations for fluid-bolus therapy prior to FEAST had not been tested in clinical trials, despite weak evidence supporting the recommendations in guidelines, as a result of FEAST there are now a number of trials that have been proposed in high-income countries to investigate a similar question [253-255]. Choice of endpoints is difficult for these trials; because primary mortality rates are so low in high-income Paediatric Intensive Care Units (PICUs) a very large trial would be needed.

Two of the trials evaluating fluid boluses in children in hospital in high income settings are Fluids in Shock (FiSH) in the UK (ISRCTN 15244462) and a Trial to Determine Whether Septic Shock Reversal is Quicker in Pediatric Patients Randomized to an Early Goal Directed Fluid-sparing Strategy vs. Usual Care (SQUEEZE) in Canada (NCT01973907) [254]. There is also one trial registered in adults - Restricted Fluid Resuscitation in Sepsis-associated Hypotension (REFRESH) in Australia (ACTRN12616000006448) [255]. Both paediatric trials were enrolling children who have already received boluses of fluids and the randomisation is either to the amount given in second boluses (10mls/kg vs 20mls/kg) or to a fluid sparing strategy instead of

standard of care for subsequent boluses. FiSH was a pilot study with two different elements, one a qualitative study in 21 children which was looking at whether parents would be happy with deferred consent [256], and the second looked at recruitment and adherence to fluid resuscitation strategies in 108 children. The FiSH trial protocol does not directly mention any blood tests although they describe 'physiology' as being measured every 15 minutes for the first 4 hours. The statistical analysis plan (SAP) for the study indicates a PIM2 score will be calculated which involves measuring PaO_2 and base excess. The SQUEEZE trial is also a pilot trial comparing standard of care to a fluid sparing resuscitation strategy in 50 children, with recruitment and adherence to fluid resuscitation strategies as primary outcomes. The proposed primary outcome for a subsequent full trial is time to shock reversal as defined by meeting all of the following criteria: free from vasoactive medication support; normalisation of heart rate (above the 5th and below the 95th percentile for age); normalisation of blood pressure (SBP and MBP above the 5th percentile for age); and capillary refill <3s. No specific biomarker or laboratory tests are indicated in the protocol but there will be continuous cardiac monitoring.

REFRESH is a phase II clinical feasibility trial in 100 adults with a primary outcome measure of total volume administered in the first 6 hours. The REFRESH trial is measuring a variety of biomarkers (including troponin, atrial natriuretic peptide, inflammatory cytokines, NGAL (a marker of renal injury), endothelial cell activation biomarkers and soluble markers of glycocalyx degradation) and plans to look at peak values as well as patterns over time at 0, 3, 6 and 24 hours. By comparing the protocols of these three trials it appears that REFRESH may be able to investigate more biomarkers to understand the mechanism of bolus resuscitation in more detail compared to the paediatric trials, but all trials are initially looking at management factors and protocol adherence as outcomes. They are also small trials and thus will not have power to correlate the biomarkers with more clinically serious events.

There has been debate as to the most appropriate non-fatal outcome for fluid trials in high-income settings and the subjectivity of such non-fatal outcomes, with only one clinical outcome for large trial having been proposed so far (SQUEEZE trial). A commentary published in Intensive Care Medicine in 2016 indicated that there is a need to find non-invasive, reproducible and reliable haemodynamic measures that can be suitable outcomes in studies looking at fluid resuscitation in children [257]. With respect to this challenge Chapter 3 aimed to identify if there were any surrogate markers which may have been useful to measure and then validate for use in smaller trials where mortality rate is very low.

Firstly, Chapter 3 modelled physiological measures over time looking at their associations with mortality, importantly modelling these associations as continuous variables using fractional

polynomials. This is an approach not often taken as much research categorises bedside measures. However, it is important to retain all the information the measurements provide without categorisation or needing to assume a linear relationship [162], both to adequately adjust for confounding and to represent biologically plausible associations. This also ensured the appropriate functional form of the physiological measure was included in models looking at the proportion of treatment effect explained. A key finding from these analyses was that children with a high fever on admission were at lower risk of mortality compared to those with hypothermia. Temperature was also identified as part of Berkley *et al*'s prognostic score, with fever shown to be beneficial and hypothermia to predict higher mortality [47]. I also showed that children with a persistent high temperature or developing a fever after the initial admission were at higher risk of mortality. This potentially identifies children for which treatment may not be working (for example antimicrobial treatment failures – where they may have had the incorrect dose or incorrect antibiotic or there was antimicrobial resistance). This could be a potential future cheap and simple point of care test for identifying high risk children during their hospital stay but after initial admission to the ward, and highlights the importance of basic clinical monitoring.

One method to identify surrogate markers is by calculating the proportion of treatment effect (PTE) explained by a physiological measure. This is a reasonably intuitive measure to understand and it has been suggested that if it is above 75% (with a lower confidence interval bound of >50%) then the measure may be considered either a surrogate marker by itself or worth further investigation [164]. There is some debate as to the best way to calculate the proportion of treatment effect explained and other similar measures have been proposed, but the method used in this thesis is widely known and understood and overcomes some methodological issues identified in initial methods (such as the hypothesis that two different statistical models are both the true models) [167]. Jeeyapant *et al* identified absolute lactate value measured at 8 hours as a possible surrogate marker for the effect of antimalarials on mortality in hyperlactataemic adults [199]. They showed that artemether had a greater impact on absolute change from baseline in lactate (-1.7 (95% CI -2.9, -0.5) vs quinine (+0.7 (95% CI -0.4, +1.8) at 8 hours, and the PTE of artemether by lactate was 73%. However, in that study the 95% CI was very wide (-607%, 673%) and thus covered all possible interpretations. In the analyses presented in my thesis the PTE by lactate for each subgroup was positive for all subgroups considered (malaria, non-malaria, age≤2 years, age>2 years) but the estimates were low, all between 10% and 20%, with confidence intervals within the interval [-100, 100]. There was also no evidence in the FEAST data for a difference between bolus and no bolus groups for absolute lactate value at 8 hours adjusted for the baseline value, which is also necessary for a measure to be considered a possible surrogate marker.

Jeeyapant *et al* recommended change in plasma lactate at 8 or 12 hours could be considered as an appropriate endpoint for trials in severe malaria [199], but analysis of the FEAST data did not provide evidence to support this in critically ill children overall or in the subgroup with malaria. This may be due to a large number of deaths in the FEAST trial occurring prior to 8 hours (167/345 (48%)) thus meaning that any true surrogacy was missed, or that the mechanisms of action of anti-malarials and boluses were different, meaning that lactate may be a surrogate for the specific antimalarial treatment effect. However, in the AQ study used to assess the proportion of treatment effect explained by lactate, 15/38 (40%) deaths occurred by 8 hours which is similar to the FEAST data, and the 8 hour lactate measurement was available for 158/173 (91%). In terms of mechanism of action, the Jeeyapant *et al* study compared artesunate, which has an action against all stages of parasites (and thus the deep sequestered parasite load hypothesised to cause microvascular obstruction (and anaerobic respiration)), with quinine, which only acts on late stages. It is therefore plausible that lactate acts as a good surrogate for the metabolic effects of malaria. Also, even though lactate clearance is putatively recommended by the surviving sepsis guidelines as being a measure of improved perfusion, and thus a marker that would change with the receipt of fluids, I did not find that this was the case in the FEAST dataset.

The only other physiological measure with positive PTEs across the subgroups and overall in FEAST was respiratory rate, which did have an association with receiving boluses as there was a small but significant difference (<1 breath per min) from 1 to 48 hours between bolus arms and control (global p-value = 0.001). However, the PTEs were all <35% indicating lack of strong evidence that respiratory rate could be a useful surrogate endpoint. The PTE analyses overall showed there was no single physiological measure, nor any combination of measures, recorded over time that could explain the mechanism by which boluses caused harm.

In the FEAST trial, bolus resuscitation (irrespective of fluid type) in the African setting was shown to be of no benefit in any of the pre-planned and subsequent subgroup analyses; conversely most analyses showed evidence of substantial and consistent harm [33, 157]. In looking to understand the mechanisms of the detrimental impact of boluses, physiological measures at baseline and over time were examined as possible effect modifiers in the four groups of interest – malaria, non-malaria, age≤2 years, and age>2 years. I examined the measures separately in the subgroups as their associations with mortality may have differed in ways that would not be identified through simple heterogeneity tests. This showed strong evidence that the impact of bolus varied over levels of base excess for children with malaria, indicating an especially harmful impact of boluses in those with severe acidosis and malaria. The harmful effect of the bolus was constant across levels of base excess if the child did not

have malaria. Sensitivity analyses also showed that the difference in bolus effect across levels of base excess held when analyses were restricted to those with 'true' severe malaria, as defined by P*f*HRP2, and if modelled with a continuous measure of malaria (total parasite burden). This was in addition to, and independent of, a varying impact of boluses over levels of oxygen saturation (where children with the highest oxygen saturation (ie largely normal levels of oxygen saturation) that were at least risk of mortality overall were counterintuitively at most risk from boluses).

Although bolus resuscitation was shown to be deleterious, with evidence of harm in both the non-malaria and malaria subgroups (as defined by the main trial statistical analysis plan) [33], the heterogeneity of the impact of boluses according to base excess in those with malaria points to one possible mechanism of harm. This was consistent with and extends an analysis where base excess was used to define presentation syndromes and then the risk of mortality was compared between bolus and no bolus groups in those with different clinically identifiable presentation syndromes [157]. Importantly, those with just a severe shock/acidosis presentation (defined by lactate≥5mmol/L or base excess≤-8 or WHO definition of shock or moderate hypotension as included in most international definitions of shock) had a much greater risk difference for mortality in boluses vs no bolus (10% vs 3%) compared to other presentation groups (neurological, or respiratory) and even to any child with more than one presentation syndrome [157]. A commentary on the paper describing presentation syndromes hypothesised that the boluses may have induced hyperchloremic metabolic acidosis [258], which would have been exacerbated in those presenting with severe acidosis. Nevertheless, the dose of 'acidemic fluids' is substantially less than would have been given under other guidelines, and I found no evidence that children with high chloride on admission were at more harm from boluses compared to those with low chloride.

The presence of malaria parasites in children with severe acidosis at baseline is also important. I found a correlation between base excess and P*f*HRP2, and children with P*f*HRP2>1000ng/ml had a lower base excess at admission. As high P*f*HRP2 is a marker of deep sequestration this could suggest that the P*f*HRP2 is a possible cause of severe acidosis in these particular children. Severe acidosis caused by deep sequestration of parasitaemia may be better treated with sevuparin [259] compared to severe acidosis as part of shock. Another contributing mechanism could be red blood cell deformability (both in parasitised and non-parasitised red cells) in those with severe malaria which has been shown to be a prognostic factor for mortality and more important than other parameters such as lactate and acid/base status [200, 201]. Red blood cell deformability is also correlated with base excess [201]. The bolus may have increased the ability of these cells with decreased deformability to circulate,

reducing microcirculatory flow and oxygen delivery in the tissues. Children without malaria would not have been subject to these effects.

The other important bedside measure explored in Chapter 3 was oxygen saturation, as my analyses confirmed a previously identified interaction between levels of oxygen saturation at baseline and the effect of the bolus. This had been described with a binary categorisation of oxygen saturation with a cut-off at <92% but also as a continuous variable [157]. This was extended in this thesis by considering levels of oxygen saturation after baseline which also changed the effect of the bolus, namely children with low oxygen saturation levels after baseline and with malaria had the higher mortality risk after receiving boluses compared to those in the no bolus arm. However, putting the two factors together, it was children with malaria with high (normal) levels of oxygen saturation (≥90%) at baseline that then dropped at any timepoint after baseline, the bolus led to the highest mortality risk compared to the no bolus arm. For children without malaria, it was simply high oxygen saturation at baseline that increased the harmful effect of boluses and there was no evidence that the levels of oxygen saturation after baseline had an impact. The endpoint review committee adjudicated all causes of death as well as mode of death/terminal clinical event (cardiogenic, respiratory, neurological or unknown) blinded to randomised allocation, and they found no pulmonary oedema events possibly or probably related to fluid in children that died in the bolus arms and only one possibly or probably related in the no bolus arms, with only 19 pulmonary oedema events possibly present and contributing to death in the bolus arms overall (5 unlikely related to fluids, 12 uncertainly related and 2 with insufficient information) [157]. However, the adjudication was not done by malaria subgroup, nor were analyses in the paper describing terminal clinical events.

Children in the bolus arms with high oxygen saturation (≥90%) had a higher proportion of deaths adjudicated as respiratory TCEs (1.6% (22 deaths) in bolus arms vs 0.1% (1 death) in no bolus arm) but also a higher proportion adjudicated as cardiogenic TCEs (3.7% (56 deaths) in bolus arms vs 1.9% (14 deaths) in no bolus arm); that is, children with high oxygen saturation had similar excess mortality from both respiratory and cardiogenic causes. In contrast, excess mortality risk associated with boluses in children with low oxygen saturation was predominately from cardiogenic causes. There was no evidence of a difference in deaths adjudicated as neurological TCEs between arms for those with high oxygen saturation. When analysing oxygen saturation as an outcome there was no evidence of a difference between the arms and the mean oxygen saturation increased in both arms after admission; however, as discussed above, there was a small difference in the respiratory rate with a slower reduction to normal levels in the bolus group.

5.2.3 Key results: Describing mortality risk over time (Chapter 4)

Chapter 4 showed the bolus and no bolus groups had a similar high mortality risk within the first 1.5 hours of admission to hospital, which subsequently decreased in both groups. However, excess mortality risk in the bolus arms persisted for up to 96 hours, indicating that the detrimental effect of fluids continued beyond the primary outcome endpoint of 48 hours and thus children in the bolus arms were recovering more slowly than those in the no bolus arm. This result supports two hypotheses. One hypothesis is that it could be that there is a protective physiological response (where non-vital organs (for example skin, liver/gut, kidney) are hypo-perfused to conserve blood flow) but when circulation is restored rapidly the protective response is diverted and the cardiac output is unable to perfuse the non-critical areas (as it is at its greatest compensation) and this leads to cardiovascular collapse [33]. This should be identified as a cardiogenic terminal clinical event. Assessment of shock usually includes assessment of features that pertain to limb perfusion, for example capillary refill time or temperature gradient, and shock reversal (in relation to peripheral perfusion) is what guidelines suggest is associated with better outcomes; however, we did not find that time-updated values of these physiological measures were associated with survival. By reversing the impaired perfusion quickly it may be that the protective physiological response can no longer sustain the perfusion of all the organs and cardiovascular collapse occurs; this was shown to be the terminal clinical event with the largest cumulative incidence difference between children receiving boluses and no boluses [157], and the largest difference over time from the competing risk analyses using flexible parametric models presented in this thesis. This is also supported by a study in sheep which found a marked increase in levels of cardiac troponin I at 16 hours in the sepsis group resuscitated with saline compared to a control group with sepsis not resuscitated with saline and a healthy group resuscitated with saline [231]. Cardiac troponin I is a protein released when the heart muscle has been damaged. A research group in Australia is also examining this hypothesis in children by monitoring cardiac index closely through ultrasound every 5 minutes during fluid resuscitation (and 60 minutes after fluids have finished) to better understand physiological changes [260]. There is also a second hypothesis, although this has not been possible to explore as much in my analyses, which is that impaired perfusion could be a protective response preventing a vascular systemic spread of pathogens and there is re-perfusion injury when there is a ‘flush’ mechanism in the tissues, probably damaging the glycocalyx which is critical for microcirculation health, resulting in poorer outcomes further at a later time after bolus administration [228, 229].

Different modelling techniques for estimating the mortality risk over time were also examined, using flexible parametric models on the untransformed time scale (to avoid an asymptote in the risk as time approaches zero) and the log time scale (which is more common and more stable to changes in baseline degrees of freedom) [207]. Both methods suggested that there may have been a sharp increase in risk immediately after randomisation followed by a drop by 2 hours indicating changing high risk in this period. One clinical hypothesis for a peak in mortality risk at this time is that early supportive care may have adverse consequences, but this is intrinsic to clinical practice so very little research has been done in this area [232]. Early goal directed therapy (EGDT) has been compared in previous studies to standard of care (where treating clinicians determined rates and amounts of fluid administration) especially within sepsis protocols (where a bundle of interventions has been mandated to start very quickly after admission) and has been reported to have discordant results and to be both effective and not effective in different settings. However, of importance, 3 large clinical trials undertaken between 2008 and 2015 and comparing EGDT to standard of care in adults showed no difference in mortality by 90 days in the US (ProCESS [261]), Australia (ARISE [262]) or the UK (ProMISe [263]). An individual patient-level meta-analysis of these trials also showed that EGDT was associated with a greater mean use of intensive care and cardiovascular support (in days) compared to usual care [264]. In low-income settings increased mortality was found with early goal directed therapy in a clinical trial comparing an early sepsis protocol (including fluid boluses and vasopressors) to standard of care in 212 adults in Zambia [235]. In contrast, no difference in mortality was found in 328 children with sepsis in Bangladesh comparing pre and post protocol implementation of an EGDT bundle, although interestingly there was an increase in fluid overload, heart failure and increased length of stay post implementation [265]. There was also no evidence of a difference in mortality in a trial in 96 children in India comparing 20mls/kg boluses given over 5-10 minutes compared to the same amount over 15-20 minutes, but there was increased need for mechanical ventilation in the first 6 and 24 hours in the 5-10 minutes group [266]. More rapid completion of a 3-hour bundle of care and rapid administration of antibiotics was associated with lower mortality in 40,696 adults in the USA, but rapid completion of a bolus of fluids was not associated with mortality [234]. However, two of these studies were not randomised and either compared two different time periods in the same hospital or two different bundles of care and so are open to unmeasured confounding by other factors changing between the time periods (Bangladesh) and between those who did and did not receive the 3 hour bundle in the study (USA). The study in India was small and stopped early (sample size calculations were based on 210 children) due to the researchers' Institutional Ethics committee deciding *a-priori* that a significant increase in

ventilatory requirements in either group would be a potential trigger for termination of the trial.

Although the hypothesis that early supportive care could be harmful cannot be excluded, there was also support in the sensitivity analyses in Chapter 4 for the alternative - that children were simply arriving in hospital with high mortality risk over the first 2 hours which steeply declined following treatment. Of note, including deaths between screening and randomisation reduced the peak in the mortality risk estimate in the first 2 hours, and the 95% confidence band was consistent with constant risk over this period. There was also a suggestion that a peak in mortality risk may have been caused by the way deaths were recorded around the 1 hour time point which was the first mandated observation time. The strength of the analyses in Chapter 4 is the consistent finding that boluses delayed normalisation of risk, preventing children from recovering at the same rate as those not receiving boluses. Of importance, this finding was present, regardless of analysis method used.

5.3 Limitations and challenges

A limitation of the analyses in this thesis is missing data in the dataset, which varies from <1% to 33% depending on the variable recorded. Baseline clinical measures were well recorded and there were very low levels of missing data in bedside measures recorded over time. There was slightly more missing data for point-of-care tests such as haemoglobin, lactate and glucose. Despite being provided at sites by the trial, there was up to 34% missing data for tests included in the i-STAT cartridge (sodium, potassium, chloride, BUN, pH, base excess) due to difficulty in maintaining the machine, cartridge stock-outs and some issues with very haemolysed blood samples (possibly caused by high parasitaemia). Multiple imputations were used, where appropriate, to take into account the missing data but had limitations when investigating interactions with baseline measures on the i-STAT cartridge when the baseline measure did not have a linear association with mortality (models would not converge). However, the multiple imputation analyses, when used, all supported the complete cases analyses with no indication of large biases. Another limitation of the analyses in this thesis is that there are data of interest that were never recorded or not recorded frequently enough, such as lactate, which was only measured at 8 hours after baseline. There may also have been biomarkers of interest, such as cardiac troponin I as discussed in Chapter 4, which if measured, may have given some further insight into the mechanism of action. Unfortunately, there were no samples stored after randomisation to investigate this. There were also limited baseline samples available for

quantifying P*f*HRP2 levels which has been shown to be a good marker of severe malaria [41] and has refined the definition of severe malaria enabling better distinction between children with severe malaria and children with a different severe illness but with incidental parasitaemia. Complete case analyses showed P*f*HRP2 levels were important when combined with extreme values of base excess in identifying the subgroup at greatest harm from the boluses. Due to missing data in both the P*f*HRP2 measure and in the haematocrit recorded total parasite burden could be estimated in only 1228/2216 (55%) of those with P*f*HRP2 measured, 1228/3170 (39%) of the trial as a whole.

Data from clinical trials are good for creating risk scores as there are few missing data, thus reducing potential bias even when a model is built from complete cases; trial data also provide a large number of events to be included in the model building process. However, trial data can also be criticised for being non-generalisable as the amount of treatment or monitoring post randomisation may not be representative of clinical practice, and the trial population may be a selected subset of the target population for the risk score. FEAST was a large pragmatic trial in febrile children with severe infection using broad eligibility criteria, and was designed to include as many children as possible without focusing on particular diseases or presentations. The broad eligibility criteria were informed by a critical review of international shock criteria and also applied to clinical surveillance dataset from Kilifi [239]. All children received standard of care and additional medications where necessary following national guidelines, with the only difference between the arms being the administration of boluses. Although there was a higher level of monitoring of children in the trial compared to those not enrolled, this was done by routine hospital nursing staff rather than from equipment not usually available in these settings; I, therefore, am confident that the data used to create the PET score are still generalisable. In addition, half (3/6) the study sites had not undertaken large research studies before and did not have a higher standard of laboratory facilities or hospital facilities compared to other hospitals in Uganda. The FEAST PET score could also be criticised for having measures that could be subject to variability depending on the healthcare professional recording it [267]. In FEAST the staff were trained at the beginning of the trial and monitored during it, but this could be something that limits implementation in other hospitals. Although the PET score was validated on the Kilifi data, its ease-of-use and implementation as part of triage can only be tested in future research. The FEAST PET score was expanded to include some measures that were available from point-of-care tests which will hopefully become even more widely available at the bedside, but unfortunately the FEAST PETaL score could not be externally validated as the measures were not available in the Kilifi datasets.

The proportion of treatment effect (PTE) measure used to find potential surrogate markers to examine the mechanism of effect also has limitations. It does not have a causal interpretation, as the hazard ratio of bolus vs no bolus adjusted for the potential surrogate marker cannot be interpreted as the direct effect of the treatment (i.e an effect not mediated through the surrogate marker) unless there are no confounders for the association between the biomarker and the outcome. Also, even if all surrogate marker outcome confounders were measured, the adjustment for these confounders may not be straightforward unless we can assume the confounders are not affected by the intervention. Furthermore, PTE often has very wide confidence intervals, as it requires a large number of events for better precision; these can also be outside of the [-100%, 100%] interval making interpretation more difficult.

5.4 Areas for further exploration

In trials going forward, I feel that my analyses highlight the importance of including quantitative P*f*HRP2 testing in malaria endemic areas for refining diagnosis of severe malaria (including the recording of haematocrit) to calculate parasite burden. In FEAST this would have helped to investigate what was causing greater harm from fluid resuscitation in those with high parasite burden and extreme values of base excess. A quantitative P*f*HRP2 test is planned to be incorporated into point-of-care diagnostics with a malaria rapid diagnostic test (RDT) so may be able to be more widely completed on all children enrolled in future trials in these settings, where the mechanisms of malaria are being targeted by specific therapies [268]. In general, point-of-care tests are increasingly being used in intensive care facilities globally as they provide much quicker test results [269] and are generally easier to implement in countries where laboratory facilities are sparse. However, many point-of-care tests' still require some machinery which needs maintaining and issues of quality control, training of healthcare workers and flow of patients in facilities all require attention, so there is a need for more robust point-of-care systems that can withstand the conditions in sub-Saharan Africa.

It would be beneficial for the FEAST PET score to be further externally validated in other African populations outside of the FEAST trial centres, but even without this, it could be used as part of screening for clinical trials and to stratify children into groups or perform risk-adjusted comparisons of emergency care. The advantage of the FEAST PET score is that when used as part of a triage system it would work across specific syndromes and diseases to identify those that need prioritisation of any supportive therapies available. Further validation leads to wider implementation and future work could include validating this score and the

FEAST PETaL score on data collected through ongoing large clinical trials in these settings, such as Transfusion and Treatment of severe Anaemia in African Children Trial (TRACT) [270] and Children's Oxygen Administration Strategies Trial (COAST) [271]. The FEAST PET score is also something that could be tested as part of research into implementing new triage systems, and/or developed into a mobile application in a similar way to other triage measurements; for example, researchers at British Columbia's Child and Family Research Institute have developed a way to measure respiratory rate on a mobile app [272]. The FEAST PETaL score included lactate which could be very important to monitor during admission, along with temperature, as a way to look for treatment failure as continued or recurrence of high temperatures after admission were shown to have high mortality risk. Lactate has been shown to be an important prognostic factor when measured at admission in this thesis, in addition to failure to clear lactate by 8 hours; both identified children at persistent high risk in FEAST [247]. Single lactate measurements have also been used in other studies in critically ill children [243, 273-275] and in adults [276], and is an area on which future research could focus. Monitoring temperature at certain timepoints after admission, easily done in this setting with thermometers, could result in tailoring of treatment, especially where microbiological services are lacking. This hypothesis could be explored more fully in future work considering what other bedside measures impact the association between high temperature and high mortality risk only after admission; an easy-to-use tool could be developed for identifying possible treatment failure in these settings.

In several recent PICU studies, early fluid overload (defined as fluid accumulation of at least 10% of admission body weight) has been identified as associated with poor outcomes such as increased ventilation days, length of stay and mortality [257]. Children with early fluid overload were also shown to be more likely to develop respiratory failure [277]. Those studies were in high income settings where the amount of fluid given is substantially greater than that received by children in the FEAST trial (for example, standard of care in the FiSH trial is 20mls/kg in 15 mins plus up to maximum of 120mls/kg in 20mls/kg/15mins boluses). In general, children who receive more fluids in PICUs are those who are sicker patients so it may not be unexpected that those who receive more fluid are at higher risk of mortality; however, they also have to survive for long enough to receive more fluids so this association is subject to time-dependent confounding. Future work may be able investigate the dose-response relationship between receiving fluids and mortality in the FEAST data using probability weighting methods (so called g-methods) that adjust for time-dependent confounding. This work could include bedside observations as predictors of receipt of fluid but could also have a focus on oxygen saturation in order to explore this mechanism further.

The analyses using PTE did not identify any one measure that explained a large proportion of the treatment effect of boluses on mortality. Given the limitations of PTE, an alternative analysis could be to use mediation methods to try and better identify surrogate markers that are on the causal pathway between the bolus administration and death. Causal diagrams could be used to display indirect and direct effects, and the indirect estimates estimated. This would be of particular interest for oxygen saturation and the models could consider how much of the effect of boluses is mediated through oxygen saturation and could calculate the proportion net effect explained.

Options for treating critically ill children presenting to hospital in low-income settings are still limited to the mainstays of antimalarials and antibiotics and there is a continuing need to identify suitable supportive therapies that can be applied across underlying diseases in these settings. A search of ISRCTN registered trials in children in the 'infections and infestations' or 'haematological' categories completing after 2011 brought up 104 trials of which 8 were in sub-Saharan settings. A similar search on Clinicaltrials.gov brought up 793 trials that completed after 2011 in children using 'infections' and 'hospitals' as search terms. 73/793 were in sub-Saharan Africa, of which 14 trials were in critically ill children (aged >2 months) in hospital (Table 5.4.1)..

Table 5.4.1: Trials researching treatment in critically ill children in hospitals in Africa between 2011 and 2018.

Trial number (and name of trial if known)	Location	Disease area	Comparison	Sample size	Status
ISRCTN 49726849 (TABS-PKPD)	Uganda	Bacterial infections	Dosing of azithromycin	105	Not yet enrolling
ISRCTN 18051843 (FLACSAM)	Uganda, Kenya	Severe acute malnutrition	IV ceftriaxone to IV benzyl penicillin plus gentamicin (usual care)	2000	Enrolling
ISRCTN 15622505 (COAST) [271]	Uganda, Kenya	Pneumonia	High flow vs low flow vs permissible hypoxia (no oxygen).	4200	Enrolling
ISRCTN 84086586 (TRACT) [270]	Uganda, Malawi	Severe anaemia	20mls/kg vs 30mls/kg vs no blood transfusion	3960	Completed enrolment
ISRCTN 11594437 (PAC study)	Uganda, Congo	Malaria	Primaquine vs placebo	1600	Enrolling
ISRCTN 91805477 (Coco Trial)	Rwanda	Malaria	Coartem vs coarnate	900	Completed
ISRCTN 17472707 [278]	Burkina Faso, Kenya, Tanzania	Malaria	artesunate-mefloquine vs artemether-lumefantrine	945	Completed
ISRCTN 96891086 [279]	Ghana	Sickle cell disease and malaria	artesunate-amodiaquine vs artemether-lumefantrine	119	Completed
NCT02760420	Malawi	Fast breathing and Pneumonia	3 days amoxicillin vs placebo	1126	Not yet enrolling
NCT02678195	Malawi	Chest indrawing and pneumonia	3 days vs 5 days amoxicillin	2000	Enrolling
NCT02414399 (Tota Bora) [280]	Kenya	Hospitalised children	Azithromycin vs placebo post discharge	1400	Enrolling
NCT00934492 [281]	Kenya	Severely malnourished children	Cotrimoxazole vs placebo	1778	Completed
NCT01247909 [282]	Malawi	Sepsis	Ceftriaxone vs penicillin	348	Completed
NCT01868113	Uganda	Acute respiratory infection	Inhaled corticosteroids vs placebo	1010	Completed

NCT01255215 [283]	Uganda	Severe malaria	Inhaled nitric oxide vs placebo	120	Completed
NCT02025452	Botswana	Acute gastroenteritis	Probiotics vs placebo	76	Completed
NCT01014988 [284]	South Africa	Flu	IV Zanamir (single arm)	202	Completed
NCT01540838 [285]	Angola	Bacterial meningitis	beta-lactam infusion with vs without concomitant oral paracetamol	723	Completed
NCT01461590 (Tx-30) [286]	Uganda	Severe anaemia	20mls vs 30mls blood transfusion	160	Completed
NCT01258049 [287]	Rwanda	Malaria	ArTiMist (sublingual artemether) vs quinine	182	Completed

My PhD supervisors and I are involved in TABS-PKPD, COAST and TRACT (which are not yet completed). We are also planning to investigate how best to use adjunctive supportive therapy and other interventions to optimise the treatment pathway for children with severe malaria through a new consortium for research and trials called SMAART. There is ongoing discussion and need for more evidence on how best to treat paediatric sepsis in low-income settings [288] and how best to treat children with blackwater fever (also known as haemoglobinurea) which is prevalent in Eastern Uganda [289]. The trials above that have completed and published have not identified any specific adjunctive supportive therapy that is able to treat critically ill children quickly in LMIC settings, and thus more research seems urgently needed.

6 References

1. Liu, L., et al., *Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals.* Lancet, 2016. **388**(10063): p. 3027-3035.
2. Nolan, T., et al., *Quality of hospital care for seriously ill children in less-developed countries.* Lancet, 2001. **357**(9250): p. 106-110.
3. English, M., et al., *Assessment of inpatient paediatric care in first referral level hospitals in 13 districts in Kenya.* Lancet, 2004. **363**(9425): p. 1948-1953.
4. Mulholland, K. and R. Adegbola, *Bacterial infections--a major cause of death among children in Africa.* The New England journal of medicine, 2005. **352**(1): p. 75-77.
5. Molyneux, E., *Paediatric emergency care in developing countries.* Lancet, 2001. **357**(9250): p. 86-87.
6. Ralston, M., et al., *Global paediatric advanced life support: improving child survival in limited-resource settings.* The Lancet, 2013. **381**(9862): p. 256-265.
7. United Nations Non-Governmental Liaison Service. *MDG Targets and Indicators*. [cited 2018 22nd March 2018]; Available from: https://www.un-ngls.org/index.php/background-mgd10/1386-mdg-targets-and-indicators.
8. United Nations General Assembly., *Road map towards the implementation of the United Nations Millennium Declaration. Report of the Secretary General.* . 2001, United Nations: Geneva.
9. Requejo, J.H., et al., *Countdown to 2015 and beyond: fulfilling the health agenda for women and children.* Lancet, 2015. **385**(9966): p. 466-76.
10. Molyneux, E., S. Ahmad, and A. Robertson, *Improved triage and emergency care for children reduces inpatient mortality in a resource-constrained setting.* Bulletin of the World Health Organization, 2006. **84**(4): p. 314-319.
11. Baker, T., *Pediatric emergency and critical care in low-income countries.* Pediatric Anesthesia, 2009. **19**(1): p. 23-27.
12. Gove, S., et al., *Development and technical basis of simplified guidelines for emergency triage assessment and treatment in developing countries. WHO Integrated Management of Childhood Illness (IMCI) Referral Care Project.* Archives of disease in childhood, 1999. **81**(6): p. 473-477.
13. Organisation, W.H., *Serious childhood problems in countries with limited resources*. 2004, World Health Orgaisation: Geneva.
14. Duke, T., *Child survival and IMCI: in need of sustained global support.* The Lancet, 2009. **374**(9687): p. 361-362.
15. Organisation, W.H., *Integrated Management of Childhood Illnesses*. 2000.
16. English, M., et al., *Delivery of paediatric care at the first-referral level in Kenya.* Lancet, 2004. **364**(9445): p. 1622-1629.
17. Peterson, S., et al., *Coping with paediatric referral--Ugandan parents' experience.* Lancet, 2004. **363**(9425): p. 1955-1956.
18. Organisation, W.H., *Pocket Book of Hospital Care for Children — Guidelines for Management of Common Childhood Illnesses in Rural and District Hospitals with Limited Resources.* 2005.
19. Li, M., et al., *Global use of the WHO pocket book of hospital care for children.* Paediatrics and international child health, 2013. **33**(1): p. 4-17.
20. Tamburlini, G., et al., *Evaluation of guidelines for emergency triage assessment and treatment in developing countries.* Archives of disease in childhood, 1999. **81**(6): p. 478-482.
21. Molyneux, E.M., *Triaging children - keep it simple, swift and safe.* South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde, 2013. **103**(3): p. 158-159.

22. Gray, A. and C. Maclennan, *What are the pre-requisites/pre-conditions for emergency triage and treatment (ETAT) to be beneficial?* ICHRC, 2008.
23. Nyirasafari, R., et al., *Predictors of mortality in a paediatric intensive care unit in Kigali, Rwanda.* Paediatr Int Child Health, 2016: p. 1-7.
24. Irimu, G., et al., *Developing and introducing evidence based clinical practice guidelines for serious illness in Kenya.* Archives of disease in childhood, 2008. **93**(9): p. 799-804.
25. Buys, H., et al., *An adapted triage tool (ETAT) at Red Cross War Memorial Children's Hospital Medical Emergency Unit, Cape Town: an evaluation.* South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde, 2013. **103**(3): p. 161-165.
26. World Health Organisation, *Guideline: Updates on Paediatric Emergency Triage, Assessment and Treatment: Care of Critically-Ill Children.* 2016, World Health Organisation: Geneva.
27. Duke, T. and B. Cheema, *Paediatric emergency and acute care in resource poor settings.* J Paediatr Child Health, 2016. **52**(2): p. 221-6.
28. Dondorp, A., et al., *Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): an open-label, randomised trial.* Lancet, 2010. **376**(9753): p. 1647-1657.
29. Dellinger, P., et al., *Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008.* Critical care medicine, 2008. **36**(1): p. 296-327.
30. Maitland, K., et al., *Randomized trial of volume expansion with albumin or saline in children with severe malaria: preliminary evidence of albumin benefit.* Clin Infect Dis, 2005. **40**(4): p. 538-45.
31. Maitland, K., et al., *Response to volume resuscitation in children with severe malaria.* Pediatr Crit Care Med, 2003. **4**(4): p. 426-31.
32. Akech, S., H. Ledermann, and K. Maitland, *Choice of fluids for resuscitation in children with severe infection and shock: systematic review.* BMJ (Clinical research ed.), 2010. **341**.
33. Maitland, K., et al., *Mortality after fluid bolus in African children with severe infection.* New England Journal of Medicine, 2011. **364**(26): p. 2483-2495.
34. Molyneux, S., et al., *'The words will pass with the blowing wind': staff and parent views of the deferred consent process, with prior assent, used in an emergency fluids trial in two african hospitals.* PloS one, 2013. **8**(2).
35. Nadjm, B., et al., *WHO guidelines for antimicrobial treatment in children admitted to hospital in an area of intense Plasmodium falciparum transmission: prospective study.* BMJ (Clinical research ed.), 2010. **340**.
36. Cserti-Gazdewich, C., et al., *Inter-relationships of cardinal features and outcomes of symptomatic pediatric Plasmodium falciparum MALARIA in 1,933 children in Kampala, Uganda.* The American journal of tropical medicine and hygiene, 2013. **88**(4): p. 747-756.
37. Nakawesi, J., et al., *Prevalence and factors associated with rotavirus infection among children admitted with acute diarrhea in Uganda.* BMC pediatrics, 2010. **10**: p. 69.
38. Ogwang, M., et al., *Prevalence of hospital-associated infections can be decreased effectively in developing countries.* The Journal of hospital infection, 2013. **84**(2): p. 138-142.
39. Todd, J., et al., *When enough is enough: how the decision was made to stop the FEAST trial: data and safety monitoring in an African trial of Fluid Expansion As Supportive Therapy (FEAST) for critically ill children.* Trials, 2013. **14**: p. 85.
40. English, M., et al., *Clinical overlap between malaria and severe pneumonia in Africa children in hospital.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 1996. **90**(6): p. 658-662.
41. Hendriksen, I.C., et al., *Diagnosing severe falciparum malaria in parasitaemic African children: a prospective evaluation of plasma PfHRP2 measurement.* PLoS Med, 2012. **9**(8): p. e1001297.

42. Scott, A., et al., *Profile: The Kilifi Health and Demographic Surveillance System (KHDSS).* International journal of epidemiology, 2012. **41**(3): p. 650-657.
43. Berkley, J., et al., *HIV infection, malnutrition, and invasive bacterial infection among children with severe malaria.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2009. **49**(3): p. 336-343.
44. Aiken, A., et al., *Risk and causes of paediatric hospital-acquired bacteraemia in Kilifi District Hospital, Kenya: a prospective cohort study.* Lancet, 2011. **378**(9808): p. 2021-2027.
45. Okiro, E., et al., *The decline in paediatric malaria admissions on the coast of Kenya.* Malaria journal, 2007. **6**(1): p. 151.
46. Scott, A., et al., *Relation between falciparum malaria and bacteraemia in Kenyan children: a population-based, case-control study and a longitudinal study.* Lancet, 2011. **378**(9799): p. 1316-1323.
47. Berkley, J.A., et al., *Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study.* BMJ (Clinical research ed.), 2003. **326**(7385).
48. English, M., et al., *Interobserver variation in respiratory signs of severe malaria.* Archives of disease in childhood, 1995. **72**(4): p. 334-336.
49. Moons, K., et al., *Prognosis and prognostic research: what, why, and how?* BMJ, 2009. **338**.
50. Wyatt, J.C. and D.G. Altman, *PROGNOSTIC MODELS - CLINICALLY USEFUL OR QUICKLY FORGOTTEN - COMMENTARY.* British Medical Journal, 1995. **311**(7019): p. 1539-1541.
51. F, C. and I.N. Network, *The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. The International Neonatal Network.* Lancet, 1993. **342**(8865): p. 193-198.
52. Knaus, W.A., et al., *The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults.* Chest, 1991. **100**(6): p. 1619-1636.
53. Gunning, K. and K. Rowan, *ABC of intensive care: outcome data and scoring systems.* BMJ (Clinical research ed.), 1999. **319**(7204): p. 241-244.
54. Narci, A., et al., *The prognostic importance of trauma scoring systems in pediatric patients.* Pediatric surgery international, 2009. **25**(1): p. 25-30.
55. Justice, A.C., K.E. Covinsky, and J.A. Berlin, *Assessing the generalizability of prognostic information.* Annals of internal medicine, 1999. **130**(6): p. 515-524.
56. Seymour, C.W., et al., *Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).* JAMA, 2016. **315**(8): p. 762-74.
57. Altman, D.G. and P. Royston, *What do we mean by validating a prognostic model?* Statistics in medicine, 2000. **19**(4): p. 453-473.
58. Altman, D.G., *Systematic reviews of evaluations of prognostic variables.* BMJ (Clinical research ed.), 2001. **323**(7306): p. 224-228.
59. Altman, D., et al., *Prognosis and prognostic research: validating a prognostic model.* BMJ, 2009. **338**(may28 1): p. b605-b605.
60. Zhou, D. and M. Frize, *Predicting probability of mortality in the neonatal intensive care unit.* Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2006. **1**: p. 2308-2311.
61. Liestbl, K., P. Andersen, and U. Andersen, *Survival analysis and neural nets.* Statist. Med., 1994. **13**(12): p. 1189-1200.
62. Evans, J., et al., *Capillary refill time as an independent prognostic indicator in severe and complicated malaria.* The Journal of pediatrics, 2006. **149**(5): p. 676-681.
63. Hart, A. and J. Wyatt, *Evaluating black-boxes as medical decision aids: issues arising from a study of neural networks.* Informatics for Health and Social Care, 1990. **15**(3): p. 229-236.

64. Freeman, R., et al., *Comparison of artificial neural networks with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty.* American Heart Journal, 2000. **140**(3): p. 511-520.
65. Tu, J.V., *Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes.* Journal of clinical epidemiology, 1996. **49**(11): p. 1225-1231.
66. Ennett, C.M., M. Frize, and C.R. Walker, *Influence of missing values on artificial neural network performance.* Studies in health technology and informatics, 2001. **84**(Pt 1): p. 449-453.
67. Aitchison, T.C., et al., *Prognostic trees to aid prognosis in patients with cutaneous malignant melanoma. Scottish Melanoma Group.* BMJ (Clinical research ed.), 1995. **311**(7019).
68. Segal, M.R. and D.A. Bloch, *A comparison of estimated proportional hazards models and regression trees.* Statistics in medicine, 1989. **8**(5): p. 539-550.
69. Werneck, G.L., et al., *Classification trees and logistic regression applied to prognostic studies: a comparison using meningococcal disease as an example.* Journal of tropical pediatrics, 1999. **45**(4): p. 248-251.
70. Ambler, G., A. Brady, and P. Royston, *Simplifying a prognostic model: a simulation study based on clinical data.* Statistics in medicine, 2002. **21**(24): p. 3803-3822.
71. Royston, P. and W. Sauerbrei, *Multivariable Model Building: A Pragmatic Approach to Regression Anaylsis based on Fractional Polynomials for Modelling Continuous Variables*. 2008, Chichester: John Wiley and Sons.
72. Harrell, F.E., K.L. Lee, and D.B. Mark, *Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.* Statistics in medicine, 1996. **15**(4): p. 361-387.
73. Cook, N. and P. Ridker, *Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures.* Annals of internal medicine, 2009. **150**(11): p. 795-802.
74. Hosmer, D.W. and S. Lemeshow, *Applied Logistic Regression*. 1989, Chichester: John Wiley and Sons.
75. Cook, N.R., *Use and misuse of the receiver operating characteristic curve in risk prediction.* Circulation, 2007. **115**(7): p. 928-935.
76. Moons, K., et al., *Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker.* Heart (British Cardiac Society), 2012. **98**(9): p. 683-690.
77. Reilly, B. and A. Evans, *Translating clinical research into clinical practice: impact of using prediction rules to make decisions.* Annals of internal medicine, 2006. **144**(3): p. 201-209.
78. Riordan, A., et al., *Prospective validation of the Glasgow Meningococcal Septicaemia Prognostic Score. Comparison with other scoring methods.* European journal of pediatrics, 2002. **161**(10): p. 531-537.
79. Barquet, N., et al., *Prognostic factors in meningococcal disease. Development of a bedside predictive model and scoring system. Barcelona Meningococcal Disease Surveillance Group.* JAMA : the journal of the American Medical Association, 1997. **278**(6): p. 491-496.
80. Cervantes, F., et al., *New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment.* Blood, 2009. **113**(13): p. 2895-2901.
81. Teoh, G.S., et al., *APACHE II: preliminary report on 100 intensive care unit cases in University Hospital, Kuala Lumpur.* The Medical journal of Malaysia, 1991. **46**(1): p. 72-81.
82. Rhee, K.J., C.J. Fisher, and N.H. Willitis, *The Rapid Acute Physiology Score.* The American journal of emergency medicine, 1987. **5**(4): p. 278-282.

83. Marcin, J.P., et al., *Certainty and mortality prediction in critically ill children.* Journal of medical ethics, 2004. **30**(3): p. 304-307.
84. Dramaix, M., et al., *Prognostic indices for mortality of hospitalized children in central Africa.* American journal of epidemiology, 1996. **143**(12): p. 1235-1243.
85. Bairagi, R., M.A. Koenig, and K.A. Mazumder, *Mortality-discriminating power of some nutritional, sociodemographic, and diarrheal disease indices.* American journal of epidemiology, 1993. **138**(5): p. 310-317.
86. Charnigo, R., et al., *Thinking outside the curve, part I: modeling birthweight distribution.* BMC pregnancy and childbirth, 2010. **10**: p. 37.
87. Charnigo, R., et al., *Thinking outside the curve, part II: modeling fetal-infant mortality.* BMC pregnancy and childbirth, 2010. **10**: p. 44.
88. Cole, T.J., E. Hey, and S. Richmond, *The PREM score: a graphical tool for predicting survival in very preterm births.* Archives of disease in childhood. Fetal and neonatal edition, 2010. **95**(1).
89. Marcin, J.P., et al., *Decision support issues using a physiology based score.* Intensive care medicine, 1998. **24**(12): p. 1299-1304.
90. Ruttimann, U.E. and M.M. Pollack, *Objective assessment of changing mortality risks in pediatric intensive care unit patients.* Critical care medicine, 1991. **19**(4): p. 474-483.
91. Heard, C.M.B., J.E. Fletcher, and M.C. Papo, *A report of the use of the Dynamic Objective Risk Assessment (DORA) score in the changing pediatric intensive care environment.* Critical Care Medicine, 1998. **26**(9): p. 1593-1595.
92. Ruttimann, U.E., M.M. Pollack, and D.H. Fiser, *Prediction of three outcome states from pediatric intensive care.* Critical care medicine, 1996. **24**(1): p. 78-85.
93. Pollock, E., et al., *Use of the Pediatric Risk of Mortality score to predict nosocomial infection in a pediatric intensive care unit.* Critical care medicine, 1991. **19**(2): p. 160-165.
94. Castello, F.V., et al., *The Pediatric Risk of Mortality (PRISM) Score and Injury Severity Score (ISS) for predicting resource utilization and outcome of intensive care in pediatric trauma.* Critical care medicine, 1999. **27**(5): p. 985-988.
95. Nacul, L., et al., *Aetiology and clinical presentation of pneumonia in hospitalized and outpatient children in Northeast Brazil and risk factors for severity.* Journal of health, population, and nutrition, 2005. **23**(1): p. 6-15.
96. Wu, E.-T., et al., *Survey of outcome of CPR in pediatric in-hospital cardiac arrest in a medical center in Taiwan.* Resuscitation, 2009. **80**(4): p. 443-448.
97. Gemke, R. and J. van Vught, *Scoring systems in pediatric intensive care: PRISM III versus PIM.* Intensive care medicine, 2002. **28**(2): p. 204-207.
98. Garcia, P. and J. Piva, *Pediatric Index of Mortality 2 (PIM2). A prognostic tool for developing countries: Easy, efficient, and free!* Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 2007. **8**(1): p. 77-78.
99. Cross Continents Collaboration for Kids 3Cs4kids, A., C. Writing, and D. Gibb, *Markers for predicting mortality in untreated HIV-infected children in resource-limited settings: a meta-analysis.* AIDS (London, England), 2008. **22**(1): p. 97-105.
100. Gogia, S. and H.S. Sachdev, *Neonatal vitamin A supplementation for prevention of mortality and morbidity in infancy: systematic review of randomised controlled trials.* BMJ (Clinical research ed.), 2009. **338**.
101. Doctor, H., *Does living in a female-headed household lower child mortality? The case of rural Nigeria.* Rural and remote health, 2011. **11**(2).
102. Gérardin, P., et al., *Outcome of life-threatening malaria in African children requiring endotracheal intubation.* Malaria journal, 2007. **6**: p. 51.
103. von Seidlein, L., et al., *Predicting the Clinical Outcome of Severe Falciparum Malaria in African Children: Findings From a Large Randomized Trial.* Clinical Infectious Diseases, 2012. **54**(8): p. 1080-1090.

104. Newton, C., et al., *The Prognostic Value of Measures of Acid/Base Balance in Pediatric Falciparum Malaria, Compared with Other Clinical and Laboratory Parameters.* Clinical Infectious Diseases, 2005. **41**(7): p. 948-957.
105. Maitland, K., et al., *Severe P. falciparum malaria in Kenyan children: evidence for hypovolaemia.* QJM, 2003. **96**(6): p. 427-434.
106. Allen, S.J., et al., *Severe malaria in children in Papua New Guinea.* QJM : monthly journal of the Association of Physicians, 1996. **89**(10): p. 779-788.
107. De León, A.L.P.-P., et al., *Simplified PRISM III score and outcome in the pediatric intensive care unit.* Pediatrics international, 2005. **47**(1): p. 80-83.
108. Marsh, K., et al., *Indicators of life-threatening malaria in African children.* The New England journal of medicine, 1995. **332**(21): p. 1399-1404.
109. Planche, T., et al., *A prospective comparison of malaria with other severe diseases in African children: prognosis and optimization of management.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2003. **37**(7): p. 890-897.
110. Roine, I., et al., *Influence of admission findings on death and neurological outcome from childhood bacterial meningitis.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2008. **46**(8): p. 1248-1252.
111. Werneck, G.L., et al., *Prognostic factors for death from visceral leishmaniasis in Teresina, Brazil.* Infection, 2003. **31**(3): p. 174-177.
112. Molyneux, M.E., et al., *CLINICAL-FEATURES AND PROGNOSTIC INDICATORS IN PEDIATRIC CEREBRAL MALARIA - A STUDY OF 131 COMATOSE MALAWIAN CHILDREN.* Quarterly Journal of Medicine, 1989. **71**(265): p. 441-459.
113. Maitland, K., et al., *Absence of malaria-specific mortality in children in an area of hyperendemic malaria.* Trans R Soc Trop Med Hyg, 1997. **91**(5): p. 562-6.
114. Spiegelhalter, D. and R. Knilljones, *STATISTICAL AND KNOWLEDGE-BASED APPROACHES TO CLINICAL DECISION-SUPPORT SYSTEMS, WITH AN APPLICATION IN GASTROENTEROLOGY.* JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY, 1984. **147**: p. 35-77.
115. Maitland, K., et al., *Severe P. falciparum malaria in Kenyan children: evidence for hypovolaemia.* QJM, 2003. **96**(6): p. 427-34.
116. Nadjm, B., et al., *Inter-observer variation in paediatric clinical signs between different grades of staff examining children admitted to hospital in Tanzania.* Tropical Medicine & International Health, 2008. **13**(9): p. 1213-1219.
117. Pollack, M.M., U.E. Ruttimann, and P.R. Getson, *Pediatric risk of mortality (PRISM) score.* Critical care medicine, 1988. **16**(11): p. 1110-1116.
118. Goddard, J.M., *Pediatric risk of mortality scoring overestimates severity of illness in infants.* Critical care medicine, 1992. **20**(12): p. 1662-1665.
119. Wells, M., et al., *Poor discriminatory performance of the Pediatric Risk of Mortality (PRISM) score in a South African intensive care unit.* Critical care medicine, 1996. **24**(9): p. 1507-1513.
120. Gemke, R.J., G.J. Bonsel, and A.J. van Vught, *Effectiveness and efficiency of a Dutch pediatric intensive care unit: validity and application of the Pediatric Risk of Mortality score.* Critical care medicine, 1994. **22**(9): p. 1477-1484.
121. Martinot, A., B. Grandbastien, and F. Leclerc, *Performance of the Pediatric Risk of Mortality Score.* Critical care medicine, 1997. **25**(4): p. 716-717.
122. Pollack, M.M., K.M. Patel, and U.E. Ruttimann, *PRISM III: an updated Pediatric Risk of Mortality score.* Critical care medicine, 1996. **24**(5): p. 743-752.
123. Pollack, M.M., K.M. Patel, and U.E. Ruttimann, *The Pediatric Risk of Mortality III--Acute Physiology Score (PRISM III-APS): a method of assessing physiologic instability for pediatric intensive care unit patients.* The Journal of pediatrics, 1997. **131**(4): p. 575-581.

124. Bertolini, G., et al., *Pediatric risk of mortality: an assessment of its performance in a sample of 26 Italian intensive care units.* Critical care medicine, 1998. **26**(8): p. 1427-1432.
125. Slater, A. and F. Shann, *The suitability of the Pediatric Index of Mortality (PIM), PIM2, the Pediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand.* Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 2004. **5**(5): p. 447-454.
126. Choi, K.M., et al., *Assessment of the Pediatric Index of Mortality (PIM) and the Pediatric Risk of Mortality (PRISM) III score for prediction of mortality in a paediatric intensive care unit in Hong Kong.* Hong Kong medical journal = Xianggang yi xue za zhi / Hong Kong Academy of Medicine, 2005. **11**(2): p. 97-103.
127. Singhal, D., et al., *Prediction of mortality by application of PRISM score in intensive care unit.* Indian pediatrics, 2001. **38**(7): p. 714-719.
128. Shann, F., et al., *Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care.* Intensive care medicine, 1997. **23**(2): p. 201-207.
129. Ozer, E.A., et al., *The Comparison of PRISM and PIM scoring systems for mortality risk in infantile intensive care.* Journal of tropical pediatrics, 2004. **50**(6): p. 334-338.
130. Taori, R., K. Lahiri, and M. Tullu, *Performance of PRISM (Pediatric Risk of Mortality) score and PIM (Pediatric Index of Mortality) score in a tertiary care pediatric ICU.* Indian journal of pediatrics, 2010. **77**(3): p. 267-271.
131. Slater, A., F. Shann, and G. Pearson, *PIM2: a revised version of the Paediatric Index of Mortality.* Intensive care medicine, 2003. **29**(2): p. 278-285.
132. Shann, F., *Are we doing a good job: PRISM, PIM and all that.* Intensive Care Medicine, 2002. **28**(2): p. 105-107.
133. Balakrishnan, G., et al., *Prospective evaluation of the Paediatric Risk of Mortality (PRISM) score.* Archives of disease in childhood, 1992. **67**(2): p. 196-200.
134. Tan, G.H., et al., *Risk factors for predicting mortality in a paediatric intensive care unit.* Annals Academy of Medicine Singapore, 1998. **27**(6): p. 813-818.
135. Britto, J., et al., *Pediatric Risk of Mortality Score underestimates the requirement for intensive care during interhospital transport.* Critical care medicine, 1994. **22**(12): p. 2029-2030.
136. Gérardin, P., et al., *Evaluation of Pediatric Risk of Mortality (PRISM) scoring in African children with falciparum malaria.* Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies, 2006. **7**(1): p. 45-47.
137. McGinn, T.G., et al., *Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group.* JAMA : the journal of the American Medical Association, 2000. **284**(1): p. 79-84.
138. Conroy, A., et al., *Prospective validation of pediatric disease severity scores to predict mortality in Ugandan children presenting with malaria and non-malaria febrile illness.* Critical Care, 2015. **19**(1): p. 47.
139. Royston, P., et al., *Prognosis and prognostic research: Developing a prognostic model.* BMJ (Clinical research ed.), 2009. **338**.
140. Harrell, F.E., et al., *REGRESSION MODELING STRATEGIES FOR IMPROVED PROGNOSTIC PREDICTION.* Statistics in Medicine, 1984. **3**(2): p. 143-152.
141. Peduzzi, P., et al., *Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates.* Journal of clinical epidemiology, 1995. **48**(12): p. 1503-1510.
142. Sauerbrei, W., P. Royston, and H. Binder, *Selection of important variables and determination of functional form for continuous predictors in multivariable model building.* Statistics in medicine, 2007. **26**(30): p. 5512-5528.
143. Mantel, N., *Why Stepdown procedures in Variable Selection.* Technometrics, 1970. **12**(3): p. 621-625.

144. Sauerbrei, W., *The Use of Resampling Methods to Simplify Regression Models in Medical Statistics.* Journal of the Royal Statistical Society: Series C (Applied Statistics), 1999. **48**(3): p. 313-329.
145. Vergouwe, Y., et al., *Development and validation of a prediction model with missing predictor data: a practical approach.* Journal of clinical epidemiology, 2010. **63**(2): p. 205-214.
146. Qureshi, A.U., A.S. Ali, and T.M. Ahmad, *Comparison of three prognostic scores (PRISM, PELOD and PIM 2) at pediatric intensive care unit under Pakistani circumstances.* Journal of Ayub Medical College, Abbottabad : JAMC, 2007. **19**(2): p. 49-53.
147. Gordon, T. and W. Kannel, *Multiple risk functions for predicting coronary heart disease: The concept, accuracy, and application.* American Heart Journal, 1982. **103**(6): p. 1031-1039.
148. May, M., et al., *Prognosis of HIV-1-infected patients up to 5 years after initiation of HAART: collaborative analysis of prospective studies.* AIDS (London, England), 2007. **21**(9): p. 1185-1197.
149. Pencina, M., et al., *Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond.* Statist. Med., 2008. **27**(2): p. 157-172.
150. Greenland, S., *The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929).* Statistics in medicine, 2008. **27**(2): p. 199-206.
151. Mogeni, P., et al., *Diagnostic performance of visible severe wasting for identifying severe acute malnutrition in children admitted to hospital in Kenya.* Bulletin of the World Health Organization, 2011. **89**(12): p. 900-906.
152. Royston, P. and D.G. Altman, *Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling.* APPLIED STATISTICS-JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C, 1994. **43**(3): p. 429-467.
153. de Onis, M., et al., *Development of a WHO growth reference for school-aged children and adolescents.* Bulletin of the World Health Organization, 2007. **85**(9): p. 660-667.
154. Group, W.H.O.M.G.R.S., *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development*. 2006. 312.
155. Morris, T.P., et al., *Multiple imputation for an incomplete covariate that is a ratio.* Statistics in Medicine, 2014. **33**(1): p. 88-104.
156. DB, R., *Multiple Imputation for non-response in surveys*. 1987, Chichester: Wiley.
157. Maitland, K., et al., *Exploring mechanisms of excess mortality with early fluid resuscitation: insights from the FEAST trial.* BMC medicine, 2013. **11**: p. 68.
158. Mwaniki, M., et al., *Emergency triage assessment for hypoxaemia in neonates and young children in a Kenyan hospital: an observational study.* Bulletin of the World Health Organization, 2009. **87**(4): p. 263-270.
159. Banajeh, S.M., *Outcome for children under 5 years hospitalized with severe acute lower respiratory tract infections in Yemen: a 5 year experience.* Journal of tropical pediatrics, 1998. **44**(6): p. 343-346.
160. Moons, K., et al., *Prognosis and prognostic research: application and impact of prognostic models in clinical practice.* BMJ (Clinical research ed.), 2009. **338**.
161. George, E.C., et al., *Predicting mortality in sick African children: the FEAST Paediatric Emergency Triage (PET) Score.* BMC Med, 2015. **13**: p. 174.
162. Royston, P., D. Altman, and W. Sauerbrei, *Dichotomizing continuous predictors in multiple regression: a bad idea.* Statistics in medicine, 2006. **25**(1): p. 127-141.
163. Prentice, R.L., *Surrogate endpoints in clinical trials: definition and operational criteria.* Stat Med, 1989. **8**(4): p. 431-40.
164. Freedman, L.S., B.I. Graubard, and A. Schatzkin, *Statistical validation of intermediate endpoints for chronic diseases.* Stat Med, 1992. **11**(2): p. 167-78.

165. Lin, D.Y., T.R. Fleming, and V. De Gruttola, *Estimating the proportion of treatment effect explained by a surrogate marker.* Stat Med, 1997. **16**(13): p. 1515-27.
166. Chen, C., H. Wang, and S.M. Snapinn, *Proportion of treatment effect (PTE) explained by a surrogate marker.* Stat Med, 2003. **22**(22): p. 3449-59.
167. Kobayashi, F. and M. Kuroki, *A new proportion measure of the treatment effect captured by candidate surrogate endpoints.* Stat Med, 2014. **33**(19): p. 3338-53.
168. Cowles, M.K., *Bayesian estimation of the proportion of treatment effect captured by a surrogate marker.* Stat Med, 2002. **21**(6): p. 811-34.
169. Wang, Y. and J.M. Taylor, *A measure of the proportion of treatment effect explained by a surrogate marker.* Biometrics, 2002. **58**(4): p. 803-12.
170. Parast, L., M.M. McDermott, and L. Tian, *Robust estimation of the proportion of treatment effect explained by surrogate marker information.* Stat Med, 2016. **35**(10): p. 1637-53.
171. Qu, Y. and M. Case, *Quantifying the effect of the surrogate marker by information gain.* Biometrics, 2007. **63**(3): p. 958-62; author reply 962-3.
172. Alonso, A., et al., *Quantifying the effect of the surrogate marker by information gain - Reply.* Biometrics, 2007. **63**(3): p. 962-963.
173. Huang, J. and B. Huang, *Evaluating the Proportion of Treatment Effect Explained by a Continuous Surrogate Marker in Logistic or Probit Regression Models.* Stat Biopharm Res, 2010. **2**(2): p. 229-238.
174. Li, W., Y. Qu, and P.M. Kulkarni, *Sensitivity in statistical evaluation of biomarkers.* Stat Med, 2013. **32**(21): p. 3636-45.
175. Royston, P. and W. Sauerbrei, *Multivariable modeling with cubic regression splines: A principled approach.* Stata Journal, 2007. **7**(1): p. 45-70.
176. Wood, A.M., I.R. White, and P. Royston, *How should variable selection be performed with multiply imputed data?* Stat Med, 2008. **27**(17): p. 3227-46.
177. Morris, T.P., et al., *Combining fractional polynomial model building with multiple imputation.* Stat Med, 2015. **34**(25): p. 3298-317.
178. Dondorp, A.M., et al., *Estimation of the total parasite biomass in acute falciparum malaria from plasma PfHRP2.* PLoS Med, 2005. **2**(8): p. e204.
179. Yeo, T.W., et al., *Angiopoietin-2 is associated with decreased endothelial nitric oxide and poor clinical outcome in severe falciparum malaria.* Proc Natl Acad Sci U S A, 2008. **105**(44): p. 17097-102.
180. Rubach, M.P., et al., *Plasma Plasmodium falciparum histidine-rich protein-2 concentrations are associated with malaria severity and mortality in Tanzanian children.* PLoS One, 2012. **7**(5): p. e35985.
181. White, N.J., et al., *Lethal malaria: Marchiafava and Bignami were right.* J Infect Dis, 2013. **208**(2): p. 192-8.
182. Hendriksen, I.C., et al., *Defining falciparum-malaria-attributable severe febrile illness in moderate-to-high transmission settings on the basis of plasma PfHRP2 concentration.* J Infect Dis, 2013. **207**(2): p. 351-61.
183. Newton, C.R., et al., *Intracranial hypertension in Africans with cerebral malaria.* Arch Dis Child, 1997. **76**(3): p. 219-26.
184. Newton, C.R., et al., *Perturbations of cerebral hemodynamics in Kenyans with cerebral malaria.* Pediatr Neurol, 1996. **15**(1): p. 41-9.
185. Waller, D., et al., *Intracranial pressure in childhood cerebral malaria.* Trans R Soc Trop Med Hyg, 1991. **85**(3): p. 362-4.
186. El-Radhi, A.S., *Fever management: Evidence vs current practice.* World J Clin Pediatr, 2012. **1**(4): p. 29-33.
187. Investigators, N.-S.S., et al., *Intensive versus conventional glucose control in critically ill patients.* N Engl J Med, 2009. **360**(13): p. 1283-97.
188. Investigators, N.-S.S., et al., *Hypoglycemia and risk of death in critically ill patients.* N Engl J Med, 2012. **367**(12): p. 1108-18.

189. Macrae, D., et al., *A randomized trial of hyperglycemic control in pediatric intensive care.* N Engl J Med, 2014. **370**(2): p. 107-18.
190. Nadjm, B., et al., *Blood glucose as a predictor of mortality in children admitted to the hospital with febrile illness in Tanzania.* The American journal of tropical medicine and hygiene, 2013. **89**(2): p. 232-237.
191. Dondorp, A.M., et al., *Direct in vivo assessment of microcirculatory dysfunction in severe falciparum malaria.* J Infect Dis, 2008. **197**(1): p. 79-84.
192. Maitland, K., et al., *Hypokalemia in children with severe falciparum malaria.* Pediatr Crit Care Med, 2004. **5**(1): p. 81-5.
193. Organisation, W.H., *Oxygen therapy for children: a manual for health workers.* 2016.
194. World Health Organisation, *Hospital Care for Children: guidelines for the management of common illnesses with limited resources*. 2005.
195. Das, B.S., et al., *Hypoglycaemia in severe falciparum malaria.* Trans R Soc Trop Med Hyg, 1988. **82**(2): p. 197-201.
196. Ogetii, G.N., et al., *Hypoglycaemia in severe malaria, clinical associations and relationship to quinine dosage.* BMC Infect Dis, 2010. **10**: p. 334.
197. Kiguli, S., et al., *Anaemia and blood transfusion in African children presenting to hospital with severe febrile illness.* BMC Med, 2015. **13**: p. 21.
198. Tran, T.H., et al., *A controlled trial of artemether or quinine in Vietnamese adults with severe falciparum malaria.* N Engl J Med, 1996. **335**(2): p. 76-83.
199. Jeeyapant, A., et al., *Defining Surrogate Endpoints for Clinical Trials in Severe Falciparum Malaria.* PLoS One, 2017. **12**(1): p. e0169307.
200. Dondorp, A.M., et al., *Abnormal blood flow and red blood cell deformability in severe malaria.* Parasitol Today, 2000. **16**(6): p. 228-32.
201. Dondorp, A.M., et al., *Prognostic significance of reduced red blood cell deformability in severe falciparum malaria.* Am J Trop Med Hyg, 1997. **57**(5): p. 507-11.
202. Buyse, M., et al., *Statistical evaluation of surrogate endpoints with examples from cancer clinical trials.* Biom J, 2016. **58**(1): p. 104-32.
203. Sriboonvorakul, N., et al., *Liquid chromatographic-mass spectrometric method for simultaneous determination of small organic acids potentially contributing to acidosis in severe malaria.* J Chromatogr B Analyt Technol Biomed Life Sci, 2013. **941**: p. 116-22.
204. Sasi, P., et al., *Metabolic acidosis and other determinants of hemoglobin-oxygen dissociation in severe childhood Plasmodium falciparum malaria.* Am J Trop Med Hyg, 2007. **77**(2): p. 256-60.
205. Royston, P. and M.K. Parmar, *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.* Stat Med, 2002. **21**(15): p. 2175-97.
206. Hinchliffe, S.R. and P.C. Lambert, *Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions.* BMC Med Res Methodol, 2013. **13**: p. 13.
207. Bower, H., M.J. Crowther, and P.C. Lambert, *strcs: A command for fitting flexible parametric survival models on the log-hazard scale.* Stata Journal, 2016. **16**(4): p. 989-1012.
208. Royston, P. and P.C. Lambert, *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. 2011: Stata Press.
209. Walker, S., et al., *Mortality in the year following antiretroviral therapy initiation in HIV-infected adults and children in Uganda and Zimbabwe.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2012. **55**(12): p. 1707-1718.
210. Samaan, M.C. and N. Akhtar-Danesh, *The impact of age and race on longevity in pediatric astrocytic tumors: A population-based study.* Pediatr Blood Cancer, 2015. **62**(9): p. 1567-71.
211. Hinchliffe, S.R., et al., *Modelling time to death or discharge in neonatal care: an application of competing risks.* Paediatr Perinat Epidemiol, 2013. **27**(4): p. 426-33.

212. Davies, M.A., et al., *Prognosis of children with HIV-1 infection starting antiretroviral therapy in Southern Africa: a collaborative analysis of treatment programs.* Pediatr Infect Dis J, 2014. **33**(6): p. 608-16.
213. Schendel, D.E., et al., *Association of Psychiatric and Neurologic Comorbidity With Mortality Among Persons With Autism Spectrum Disorder in a Danish Population.* JAMA Pediatr, 2016. **170**(3): p. 243-50.
214. Yiannoutsos, C.T., *Modeling AIDS survival after initiation of antiretroviral treatment by Weibull models with changepoints.* J Int AIDS Soc, 2009. **12**: p. 9.
215. Pillay, D., et al., *The impact of transmitted drug resistance on the natural history of HIV infection and response to first-line therapy.* AIDS, 2006. **20**(1): p. 21-8.
216. Arenas-Pinto, A., et al., *The risk of developing peripheral neuropathy induced by nucleoside reverse transcriptase inhibitors decreases over time: evidence from the Delta trial.* Antivir Ther, 2008. **13**(2): p. 289-95.
217. Parner, E.T., D.E. Schendel, and P. Thorsen, *Autism prevalence trends over time in Denmark: changes in prevalence and age at diagnosis.* Arch Pediatr Adolesc Med, 2008. **162**(12): p. 1150-6.
218. Price, H., et al., *Positive and negative drug selection pressures on the N348I connection domain mutation: new insights from in vivo data.* Antivir Ther, 2010. **15**(2): p. 203-11.
219. Eyre, D.W., et al., *Predictors of first recurrence of Clostridium difficile infection: implications for initial management.* Clin Infect Dis, 2012. **55 Suppl 2**: p. S77-87.
220. Lover, A.A. and R.J. Coker, *Quantifying effect of geographic location on epidemiology of Plasmodium vivax malaria.* Emerg Infect Dis, 2013. **19**(7): p. 1058-65.
221. Miller, R.R., et al., *Dynamics of acquisition and loss of carriage of Staphylococcus aureus strains in the community: the effect of clonal complex.* J Infect, 2014. **68**(5): p. 426-39.
222. Nunes, C. and B.M. Taylor, *Modelling the time to detection of urban tuberculosis in two big cities in Portugal: a spatial survival analysis.* Int J Tuberc Lung Dis, 2016. **20**(9): p. 1219-25.
223. Stohr, W., et al., *Factors associated with virological rebound in HIV-infected patients receiving protease inhibitor monotherapy.* AIDS, 2016. **30**(17): p. 2617-2624.
224. Lover, A.A. and R.J. Coker, *Do mixed infections matter? Assessing virulence of mixed-clone infections in experimental human and murine malaria.* Infect Genet Evol, 2015. **36**: p. 82-91.
225. Brand, J.S., et al., *Infection-related hospitalizations in breast cancer patients: Risk and impact on prognosis.* J Infect, 2016. **72**(6): p. 650-8.
226. Crook, A.M., et al., *Tuberculosis incidence is high in HIV-infected African children but is reduced by co-trimoxazole and time on antiretroviral therapy.* BMC Med, 2016. **14**: p. 50.
227. World Health Organisation, *Pocket book of hospital care for children: Second edition Guidelines for the management of common childhood illnesses*. 2013: Geneva.
228. Melican, K., et al., *Bacterial infection-mediated mucosal signalling induces local renal ischaemia as a defence against sepsis.* Cellular Microbiology, 2008. **10**(10): p. 1987-1998.
229. Singer, M., et al., *Multiorgan failure is an adaptive, endocrine-mediated, metabolic response to overwhelming systemic inflammation.* Lancet, 2004. **364**(9433): p. 545-548.
230. Ince, C., *The microcirculation is the motor of sepsis.* Crit Care, 2005. **9 Suppl 4**: p. S13-9.
231. Obonyo, N., et al., *Cardiac Metabolism and Mycocardial Injury in Volume Resuscitated Septic Shock*, in *ANZICS Clinical Trials Group meeting*. 2017: Noosa, Queensland, Australia.
232. Singer, M. and P. Glynne, *Treating critical illness: the importance of first doing no harm.* PLoS Med, 2005. **2**(6): p. e167.
233. Simpson, A.J., et al., *Differential antibiotic-induced endotoxin release in severe melioidosis.* J Infect Dis, 2000. **181**(3): p. 1014-9.

234. Seymour, C.W., et al., *Time to Treatment and Mortality during Mandated Emergency Care for Sepsis.* N Engl J Med, 2017. **376**(23): p. 2235-2244.
235. Andrews, B., et al., *Effect of an Early Resuscitation Protocol on In-hospital Mortality Among Adults With Sepsis and Hypotension: A Randomized Clinical Trial.* JAMA, 2017. **318**(13): p. 1233-1240.
236. Assembly, U.N.G. *Transforming our world: the 2030 Agenda for Sustainable Development*. 2015 31st Jan 2018]; Available from: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.
237. Feikin, D.R., et al., *The Enduring Challenge of Determining Pneumonia Etiology in Children: Considerations for Future Research Priorities.* Clin Infect Dis, 2017. **64**(suppl_3): p. S188-S196.
238. Hammitt, L.L., et al., *Addressing the Analytic Challenges of Cross-Sectional Pediatric Pneumonia Etiology Data.* Clin Infect Dis, 2017. **64**(suppl_3): p. S197-S204.
239. Houston, K.A., E.C. George, and K. Maitland, *Implications for paediatric shock management in resource-limited settings: a perspective from the FEAST trial.* Crit Care, 2018 [in press].
240. Finfer, S., et al., *A comparison of albumin and saline for fluid resuscitation in the intensive care unit.* N Engl J Med, 2004. **350**(22): p. 2247-56.
241. Myburgh, J.A., et al., *Hydroxyethyl starch or saline for fluid resuscitation in intensive care.* N Engl J Med, 2012. **367**(20): p. 1901-11.
242. Semler, M.W., et al., *Balanced Crystalloids versus Saline in Critically Ill Adults.* N Engl J Med, 2018. **378**(9): p. 829-839.
243. Marsh, K., et al., *Indicators of life-threatening malaria in African children.* N Engl J Med, 1995. **332**(21): p. 1399-404.
244. Haniffa, R., et al., *Performance of critical care prognostic scoring systems in low and middle-income countries: a systematic review.* Crit Care, 2018. **22**(1): p. 18.
245. Singer, M., et al., *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).* JAMA, 2016. **315**(8): p. 801-10.
246. Schlapbach, L.J., et al., *Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit.* Intensive Care Med, 2018. **44**(2): p. 179-188.
247. Aramburo, A., et al., *Lactate clearance as a prognostic marker of mortality in severely ill febrile children in East Africa.* BMC Med, 2018. **16**(1): p. 37.
248. Brewster, D., *African fluid bolus study: implications for practice.* J Paediatr Child Health, 2012. **48**(7): p. 548-50.
249. Hilton, A.K. and R. Bellomo, *Totem and taboo: fluids in sepsis.* Crit Care, 2011. **15**(3): p. 164.
250. Duke, T., *What the African fluid-bolus trial means.* Lancet, 2011. **378**(9804): p. 1685-7.
251. Southall, D.P. and M.P. Samuels, *Treating the wrong children with fluids will cause harm: response to 'mortality after fluid bolus in African children with severe infection'.* Arch Dis Child, 2011. **96**(10): p. 905-6.
252. Russell, M.J. and H.K. Kanthimathinathan, *Is There an Optimum Duration of Fluid Bolus in Pediatric Septic Shock? A Critical Appraisal of "Fluid Bolus Over 15-20 Versus 5-10 Minutes Each in the First Hour of Resuscitation in Children With Septic Shock: A Randomized Controlled Trial" by Sankar et al (Pediatr Crit Care Med 2017; 18:e435-e445).* Pediatr Crit Care Med, 2018.
253. Inwald, D.P., et al., *Emergency management of children with severe sepsis in the United Kingdom: the results of the Paediatric Intensive Care Society sepsis audit.* Archives of Disease in Childhood, 2009. **94**(5): p. 348-353.
254. Parker, M.J., et al., *A trial to determine whether septic shock-reversal is quicker in pediatric patients randomized to an early goal-directed fluid-sparing strategy versus usual care (SQUEEZE): study protocol for a pilot randomized controlled trial.* Trials, 2016. **17**(1): p. 556.

255. Macdonald, S.P.J., et al., *REstricted Fluid REsuscitation in Sepsis-associated Hypotension (REFRESH): study protocol for a pilot randomised controlled trial.* Trials, 2017. **18**(1): p. 399.
256. O'Hara, C.B., et al., *A qualitative feasibility study to inform a randomised controlled trial of fluid bolus therapy in septic shock.* Arch Dis Child, 2018. **103**(1): p. 28-32.
257. Inwald, D.P., W. Butt, and R.C. Tasker, *Fluid resuscitation of shock in children: what, whence and whither?* Intensive Care Med, 2015. **41**(8): p. 1457-9.
258. Myburgh, J. and S. Finfer, *Causes of death after fluid bolus resuscitation: new insights from FEAST.* BMC medicine, 2013. **11**: p. 67.
259. Leitgeb, A.M., et al., *Inhibition of merozoite invasion and transient de-sequestration by sevuparin in humans with Plasmodium falciparum malaria.* PLoS One, 2017. **12**(12): p. e0188754.
260. Long, E., et al., *An observational study using ultrasound to assess physiological changes following fluid bolus administration in paediatric sepsis in the emergency department.* BMC Pediatr, 2016. **16**: p. 93.
261. Pro, C.I., et al., *A randomized trial of protocol-based care for early septic shock.* N Engl J Med, 2014. **370**(18): p. 1683-93.
262. Investigators, A., et al., *Goal-directed resuscitation for patients with early septic shock.* N Engl J Med, 2014. **371**(16): p. 1496-506.
263. Mouncey, P.R., et al., *Trial of early, goal-directed resuscitation for septic shock.* N Engl J Med, 2015. **372**(14): p. 1301-11.
264. Investigators, P., et al., *Early, Goal-Directed Therapy for Septic Shock - A Patient-Level Meta-Analysis.* N Engl J Med, 2017. **376**(23): p. 2223-2234.
265. Kortz, T.B., et al., *Clinical outcomes and mortality before and after implementation of a pediatric sepsis protocol in a limited resource setting: A retrospective cohort study in Bangladesh.* PLoS One, 2017. **12**(7): p. e0181160.
266. Sankar, J., et al., *Fluid Bolus Over 15-20 Versus 5-10 Minutes Each in the First Hour of Resuscitation in Children With Septic Shock: A Randomized Controlled Trial.* Pediatr Crit Care Med, 2017. **18**(10): p. e435-e445.
267. Otieno, H., et al., *Are bedside features of shock reproducible between different observers?* Archives of Disease in Childhood, 2004. **89**(10): p. 977-979.
268. Boyce, R., et al., *Use of a Dual-Antigen Rapid Diagnostic Test to Screen Children for Severe Plasmodium falciparum Malaria in a High-Transmission, Resource-Limited Setting.* Clin Infect Dis, 2017. **65**(9): p. 1509-1515.
269. Kapoor, D., M. Srivastava, and P. Singh, *Point of care blood gases with electrolytes and lactates in adult emergencies.* Int J Crit Illn Inj Sci, 2014. **4**(3): p. 216-22.
270. Mpoya, A., et al., *Transfusion and Treatment of severe anaemia in African children (TRACT): a study protocol for a randomised controlled trial.* Trials, 2015. **16**: p. 593.
271. Maitland, K., et al., *Children's Oxygen Administration Strategies Trial (COAST): A randomised controlled trial of high flow versus oxygen versus control in African children with severe pneumonia.* Wellcome Open Res, 2017.
272. Karlen, W., et al., *Improving the accuracy and efficiency of respiratory rate measurements in children using mobile devices.* PLoS One, 2014. **9**(6): p. e99266.
273. Scott, H.F., et al., *Association Between Early Lactate Levels and 30-Day Mortality in Clinically Suspected Sepsis in Children.* JAMA Pediatr, 2017. **171**(3): p. 249-255.
274. Scott, H.F., A. Kempe, and L. Bajaj, *Venous vs Arterial Lactate and 30-Day Mortality in Pediatric Sepsis-Reply.* JAMA Pediatr, 2017. **171**(8): p. 813-814.
275. Schlapbach, L.J., G. MacLaren, and L. Straney, *Venous vs Arterial Lactate and 30-Day Mortality in Pediatric Sepsis.* JAMA Pediatr, 2017. **171**(8): p. 813.
276. Reed, L., et al., *Serum lactate as a screening tool and predictor of outcome in pediatric patients presenting to the emergency department with suspected infection.* Pediatr Emerg Care, 2013. **29**(7): p. 787-91.
277. Bhaskar, P., et al., *Early fluid accumulation in children with shock and ICU mortality: a matched case-control study.* Intensive Care Med, 2015. **41**(8): p. 1445-53.

278. Sirima, S.B., et al., *Comparison of artesunate-mefloquine and artemether-lumefantrine fixed-dose combinations for treatment of uncomplicated Plasmodium falciparum malaria in children younger than 5 years in sub-Saharan Africa: a randomised, multicentre, phase 4 trial.* Lancet Infect Dis, 2016. **16**(10): p. 1123-1133.
279. Adjei, G.O., et al., *A randomized trial of artesunate-amodiaquine versus artemether-lumefantrine in Ghanaian paediatric sickle cell and non-sickle cell disease patients with acute uncomplicated malaria.* Malar J, 2014. **13**: p. 369.
280. Pavlinac, P.B., et al., *Azithromycin to prevent post-discharge morbidity and mortality in Kenyan children: a protocol for a randomised, double-blind, placebo-controlled trial (the Toto Bora trial).* BMJ Open, 2017. **7**(12): p. e019170.
281. Berkley, J.A., et al., *Daily co-trimoxazole prophylaxis to prevent mortality in children with complicated severe acute malnutrition: a multicentre, double-blind, randomised placebo-controlled trial.* Lancet Glob Health, 2016. **4**(7): p. e464-73.
282. Molyneux, E.M., et al., *The Treatment of Possible Severe Infection in Infants: An Open Randomized Safety Trial of Parenteral Benzylpenicillin and Gentamicin Versus Ceftriaxone in Infants <60 days of Age in Malawi.* Pediatr Infect Dis J, 2017. **36**(12): p. e328-e333.
283. Bangirana, P., et al., *Inhaled nitric oxide and cognition in pediatric severe malaria: A randomized double-blind placebo controlled trial.* PLoS One, 2018. **13**(1): p. e0191550.
284. Marty, F.M., et al., *Safety and pharmacokinetics of intravenous zanamivir treatment in hospitalized adults with influenza: an open-label, multicenter, single-arm, phase II study.* J Infect Dis, 2014. **209**(4): p. 542-50.
285. Pelkonen, T., et al., *Slow initial beta-lactam infusion and oral paracetamol to treat childhood bacterial meningitis: a randomised, controlled trial.* Lancet Infect Dis, 2011. **11**(8): p. 613-21.
286. Olupot-Olupot, P., et al., *Phase II trial of standard versus increased transfusion volume in Ugandan children with acute severe anemia.* BMC Med, 2014. **12**: p. 67.
287. Bendel, D., et al., *Efficacy of a novel sublingual spray formulation of artemether in African children with Plasmodium falciparum malaria.* Antimicrob Agents Chemother, 2015. **59**(11): p. 6930-8.
288. Musa, N., S. Murthy, and N. Kissoon, *Pediatric sepsis and septic shock management in resource-limited settings.* Intensive Care Med, 2016. **42**(12): p. 2037-2039.
289. Olupot-Olupot, P., et al., *High Frequency of Blackwater Fever Among Children Presenting to Hospital With Severe Febrile Illnesses in Eastern Uganda.* Clin Infect Dis, 2017. **64**(7): p. 939-946.

7 Appendix

Table A1: The distribution of SBP at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

SBP at baseline	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	≥160	Total	Missing
Overall	46	324	853	949	576	207	79	13	12	6	35	3100	41
Malaria	19	192	527	534	325	100	39	4	7	4	24	1775	20
Non-malaria	27	131	321	413	246	106	39	8	5	2	11	1309	21
Age≤2 years	26	146	426	485	276	104	45	9	5	2	18	1542	29
Age>2 years	20	178	427	464	300	103	34	4	7	4	17	1558	12

Table A2: The distribution of respiratory rate at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Respiratory rate at baseline	20-29	30-39	40-49	50-59	60-69	70-79	80-89	≥90	Total	Missing
Overall	53	265	576	805	784	375	179	87	3124	17
Malaria	19	147	341	498	473	199	73	33	1783	12
Non-malaria	33	116	233	306	306	173	105	54	1326	4
Age≤2 years	13	48	173	360	481	283	135	73	1566	5
Age>2 years	40	217	403	445	303	92	44	14	1558	12

Table A3: The distribution of heart rate at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Heart rate at baseline	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	170-179	180-189	190-199	200-210	≥210	Total	Missing
Overall	35	22	34	52	103	165	268	264	666	530	546	268	128	54	3135	6
Malaria	19	10	12	21	52	88	154	159	377	330	318	153	72	25	1790	5
Non-malaria	16	12	21	30	49	76	111	104	287	198	227	113	56	29	1329	1
Age≤2 years	15	6	6	9	21	21	55	71	283	287	428	212	107	47	1568	3
Age>2 years	20	16	28	43	82	144	213	193	383	243	118	56	21	7	1567	3

Table A4: The distribution of oxygen saturation at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Oxygen saturation at baseline	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-100	Total	Missing
Overall	41	20	23	35	53	76	156	247	715	1669	3035	106
Malaria	19	10	10	16	22	32	65	112	386	1067	1739	56
Non-malaria	22	10	13	19	31	42	91	133	325	596	1282	48
Age≤2 years	25	8	14	20	35	45	74	134	395	782	1532	39
Age>2 years	16	12	9	15	18	31	82	113	320	887	1503	67

Table A5: The distribution of temperature at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Temperature at baseline (°C)	35.0-35.4	35.5-35.9	36.0-36.4	36.5-36.9	37.0-37.4	37.5-37.9	38.0-38.4	38.5-38.9	39.0-39.4	39.5-39.9	≥40.0	Total	Missing
Overall	44	109	142	235	301	484	512	461	375	288	148	3135	42
Malaria	22	51	70	118	165	263	311	272	235	181	89	1792	18
Non-malaria	22	10	13	19	31	42	91	133	325	529	67	1328	48
Age≤2 years	15	38	44	94	148	258	277	258	213	155	59	1569	12
Age>2 years	29	71	98	141	153	226	235	203	162	133	89	1566	30

Table A6: The distribution of glucose at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Glucose at baseline	1.0-1.9	2.0-2.9	3.0-3.9	4.0-4.9	5.0-5.9	6.0-6.9	7.0-7.9	8.0-8.9	9.0-9.9	10.0-10.9	11.0-11.9	12.0-12.9	13.0-13.9	14.0-14.9	15.0-15.9	16.0-16.9	17.0-17.9	18.0-18.9	19.0-19.9	≥20.0	Total	Missing
Overall	85	102	149	260	420	518	444	339	223	154	86	48	32	26	14	10	16	7	4	30	2967	174
Malaria	57	74	90	149	226	302	251	197	129	79	35	23	17	14	10	4	5	3	3	19	1687	108
Non-malaria	28	28	59	111	193	216	190	142	94	74	51	25	15	12	4	6	11	4	1	11	1275	55
Age≤2 years	46	46	75	137	220	261	224	174	119	72	34	19	15	14	6	5	5	4	1	12	1489	82
Age>2 years	39	56	74	123	200	257	220	165	104	82	52	29	17	12	8	5	11	3	3	18	1478	92

Table A7: The distribution of haemoglobin at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Haemoglobin at baseline	1.0-1.9	2.0-2.9	3.0-3.9	4.0-4.9	5.0-5.9	6.0-6.9	7.0-7.9	8.0-8.9	9.0-9.9	10.0-10.9	11.0-11.9	12.0-12.9	≥13.0	Total	Missing
Overall	69	249	342	329	255	242	289	267	281	298	245	122	66	3054	87
Malaria	23	117	211	235	193	182	176	153	131	139	109	44	23	1736	59
Non-malaria	46	129	131	93	62	60	110	113	148	159	135	77	43	1306	24
Age≤2 years	32	96	155	164	136	119	176	148	153	177	122	42	13	1533	38
Age>2 years	37	153	187	165	119	123	113	119	128	121	123	80	53	1521	49

Table A8: The distribution of lactate at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Lactate at baseline	1.0-1.9	2.0-2.9	3.0-3.9	4.0-4.9	5.0-5.9	6.0-6.9	7.0-7.9	8.0-8.9	9.0-9.9	10.0-10.9	11.0-11.9	12.0-12.9	13.0-13.9	≥14.0	Total	Missing
Overall	513	630	418	260	173	129	112	106	71	72	60	102	120	110	2980	265
Malaria	202	308	245	174	124	91	76	81	46	45	40	70	77	68	1707	148
Non-malaria	311	322	173	83	49	38	36	25	25	26	20	32	41	42	1267	107
Age≤2 years	280	328	205	109	81	51	61	55	34	38	34	48	65	54	1483	128
Age>2 years	233	302	213	151	92	78	51	51	37	34	26	54	55	56	1497	137

Table A9: The distribution of sodium at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Sodium at baseline	110-114	115-119	120-124	125-129	130-134	135-139	≥140	Total	Missing
Overall	36	31	78	266	769	743	185	2108	1033
Malaria	23	21	46	156	477	365	77	1165	630
Non-malaria	13	10	32	110	288	378	107	938	392
Age≤2 years	13	15	28	125	403	386	76	1046	525
Age>2 years	23	16	50	141	366	357	109	1062	508

Table A10: The distribution of chloride at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Chloride at baseline	90-94	95-99	100-104	105-109	110-104	105 -109	110-114	115-119	≥120	Total	Missing
Overall	58	245	775	698	204	39	15	10	23	2067	1074
Malaria	31	166	470	332	89	19	8	8	15	1138	657
Non-malaria	27	78	303	366	114	20	7	2	7	924	406
Age≤2 years	21	90	364	388	117	20	8	2	16	1026	545
Age>2 years	37	155	411	310	87	19	7	8	7	1041	529

Table A11: The distribution of base excess at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Base excess at baseline	<-20	-20 to -16	-15 to -11	-10 to -4	-5 to -1	0 to 4	5 to 10	Total	Missing
Overall	53	101	184	361	651	567	124	2068	1100
Malaria	32	64	112	210	360	297	51	1139	669
Non-malaria	21	36	72	150	291	269	72	925	419
Age≤2 years	31	51	78	194	351	252	52	1020	562
Age>2 years	22	50	106	167	300	315	72	1048	538

Table A12: The distribution of PH at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

PH at baseline	<6.90	6.91-6.99	7.00-7.09	7.10-7.19	7.20-7.29	7.30-7.39	7.40-7.49	≥7.50	Total	Missing
Overall	34	44	88	250	704	736	168	22	2082	1095
Malaria	21	24	53	138	368	417	91	11	1147	672
Non-malaria	13	20	35	112	334	318	76	11	931	411
Age≤2 years	22	25	41	126	377	348	64	8	1027	560
Age>2 years	12	19	47	124	327	388	104	14	1055	535

Table A13: The distribution of BUN at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

BUN at baseline	<5	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	≥85	Total	Missing
Overall	153	1144	352	142	73	28	19	14	7	23	1955	1186
Malaria	49	664	241	86	37	19	11	7	5	9	1128	667
Non-malaria	104	477	111	56	35	9	8	7	2	14	823	507
Age≤2 years	99	594	158	56	20	10	6	3	1	3	950	621
Age>2 years	54	550	194	86	53	18	13	11	6	20	1005	565

Table A14: The distribution of potassium at baseline for all children and within malaria, non-malaria, age≤2 years and age>2 year subgroups.

Potassium at baseline	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	≥70	Total	Missing
Overall	31	55	203	484	571	387	144	69	33	19	33	2029	1112
Malaria	11	27	126	295	306	198	68	37	15	11	17	1111	684
Non-malaria	19	28	77	187	265	189	75	32	18	8	15	913	417
Age≤2 years	13	20	86	212	294	211	80	41	18	9	20	1004	567
Age>2 years	18	35	117	272	277	176	64	28	15	10	13	1025	545