BMJ Open

Using observational data to compare the effectiveness of antibiotic treatments for children hospitalised with pneumonia in Kenya

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019478
Article Type:	Research
Date Submitted by the Author:	05-Sep-2017
Complete List of Authors:	Malla, Lucas; Kenya Medical Research Institute - Wellcome Trust; University of Oxford Perera, Rafael; University of Oxford, Primary Health Care McFadden, Emily; University of Oxford, Nuffield Department of Primary Care Health Sciences English, Mike; University of Oxford, Nuffield Department of Medicine and Department of Paediatrics
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Infectious diseases, Paediatrics
Keywords:	pneumonia, missing data, propensity scores, comparative effectiveness

Using observational data to compare the effectiveness of antibiotic treatments for children hospitalised with pneumonia in Kenya

Lucas Malla[1*] (lmalla@kemri-wellcome.org)

Rafael Perera-Salazar[2] (rafael.perera@phc.ox.ac.uk)

Emily McFadden[2] (emily.mcfadden@phc.ox.ac.uk)

Mike English[1, 3] (menglish@kemri-wellcome.org)

[1]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom.

[3] Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.

***Corresponding author:**

Lucas Malla

Abstract

Objectives

Kenyan guidelines for antibiotic treatment of pneumonia recommended treatment of pneumonia characterised by indrawing with injectable penicillin alone in inpatient settings until early 2016. At this point, they were revised becoming consistent with WHO guidance after results of a Kenyan trial provided further evidence of equivalence of oral amoxicillin and injectable penicillin. This change also made possible use of oral amoxicillin for outpatient treatment in this patient group. However, given non-trivial mortality in Kenyan children with indrawing pneumonia it remained possible they would benefit from a broader spectrum antibiotic regimen. Therefore, we compared the effectiveness of injectable penicillin monotherapy with a regimen combining penicillin with gentamicin.

Setting

We used a large routine observational dataset that captures data on all admissions to 13 Kenyan county hospitals.

Participants and measures

The analyses included children aged 2 – 59 months. Selection of study population was based on inclusion criteria typical of a prospective trial, primary analysis (experiment 1, n = 4002), but we also explored more pragmatic inclusion criteria (experiment 2, n = 6420) as part of a secondary analysis. To overcome the challenges associated with the non – random allocation of treatments and missing data, we used propensity score(PS) methods and multiple imputation to minimize bias. Further, we estimated mortality risk ratios using log binomial regression and conducted sensitivity analyses using an instrumental variable and PS trimming.

Results

The estimated risk of dying, in experiment 1, in those receiving penicillin plus gentamicin was 1.46 [0.85, 2.43] compared to the penicillin monotherapy group. In experiment 2, the estimated risk was 1.04 [0.76, 1.40].

Conclusion

There is no statistical difference in the treatment of indrawing pneumonia with either penicillin or penicillin plus gentamicin. By extension it is unlikely that treatment with penicillin plus gentamicin would offer an advantage to treatment with oral amoxicillin.

Strength

- This study provides a platform to explore effectiveness of alternative treatments in routine care in a low income setting to improve health outcomes for children.

Limitations

- The analysis is limited to the variables in the observational dataset – and therefore risk bias due to unmeasured key variables.
- The influence of any resulting bias, to alter results, has however been assessed through the use of alternative methods as instrumental variables.

Introduction

World Health Organisation (WHO) recommendations guide treatment for millions of children with pneumonia every year across low and middle income countries (1). These guidelines are largely based on moderate certainty in evidence of effects (2-5). However, trials supporting recommendations for hospitalized children have included fewer participants from Africa than other settings (6) and it is suggested that African children with pneumonia have higher mortality (7). Additionally, trial populations may not always include the heterogeneous populations presenting for care, many of whom at hospital level may have co-morbidity (8). Thus despite improving access to recommended treatments and deployment of childhood vaccines at high coverage, including those against *H. influenzae* Type B and pneumococcus, clinically diagnosed pneumonia remains one of the top causes of mortality for children under five in Kenya and other countries (7). According to the Global Health Data exchange (GHDx) website (9), pneumonia caused about 212 under five deaths per 100 000 admission cases in 2015 (which was the highest compared to diarrhoea/dehydration and malaria which are the other top causes of under-five mortality in Kenya). The comparison of mortality rates between 2000 and 2015 for pneumonia, diarrhoea/dehydration and malaria is presented in the supplementary material figure A. The basic and pneumococcal vaccine coverage by 2014 for children aged 12 – 23 months in Kenya was at least 80% (10).

In a recent change to guidance it is now recommended that pneumonia characterized by lower chest wall indrawing be treated in outpatient settings with oral medication (Box 1) (11, 12). Yet it remains associated with non-trivial mortality that may be higher outside trial populations (13). Residual mortality may be associated with causes that are not prevented by currently available conjugate vaccines and organisms, which are not susceptible to the antibiotics currently recommended. Establishing whether there are benefits of alternative

treatment regimens to help reduce mortality would ideally require large, pragmatic clinical trials (14, 15). However, these remain relatively expensive and time consuming. Observational data may support comparative effectiveness analyses of alternative treatments, may be cheaper and quicker, and may enable evaluation of interventions for which randomization is difficult (16). We use observational data from Kenya to address an important contemporary question for the treatment of pneumonia, a comparison of the effectiveness of gentamicin plus penicillin versus penicillin alone for the treatment of indrawing pneumonia in routine settings. The only previous clinical trial comparing these treatments was a small study of 40 patients in Malaysia (17). In so doing we examine the potential of using data collected by providers as part of their routine practice for comparative effectiveness research in an African setting.

Methods

Clinical definitions of pneumonia, primary and secondary analyses.

The WHO and Kenyan pneumonia treatment guidelines are implicitly based on risk stratification of illness with children deemed at higher risk of mortality offered broader spectrum antibiotic regimens and those at lower risk narrower spectrum antibiotics (11, 18-20). We present three categories of clinically diagnosed pneumonia in Box 1. This categorization outlines previous and recently revised WHO and Kenyan pneumonia treatment guidelines (11, 19). What we refer to as indrawing pneumonia may be associated with low but clinically significant mortality rates (13, 21). Prior to March 2016 recommended treatment for this group was penicillin monotherapy and our aim is to examine whether there is any advantage of broader spectrum antibiotics in this group. Since March 2016 new guidelines recommend outpatient treatment with oral amoxicillin for this group on the basis of trials suggesting equivalence of amoxicillin and penicillin. However, as indicated above very few patients had been included in studies comparing narrow (amoxicillin or penicillin)

and broader spectrum antibiotic regimens. As indicated above, beyond the confines of clinical trials amongst all children being treated for indrawing pneumonia, clinical outcomes (including mortality) are worse than seen in the trials (7) and clinicians are often choosing not to use a single drug regime and are in fact often opting to use the combination of gentamicin and penicillin in the group meeting criteria for indrawing pneumonia in real life settings (22). As mortality is higher in real life settings than in trials and as the possibility that broad spectrum antibiotics could have an advantage over monotherapy with penicillin (or amoxicillin) has not been explored in Kenya's previous trials, we feel that examining whether broad spectrum antibiotics confer an advantage is an important question.

Box 1: Clinical Pneumonia Classifications and Treatments in use in Kenya

1. ***Severe pneumonia:*** *If a child has either oxygen saturation less than 90% or central cyanosis or is grunting or unable to drink or not alert, then s/he is classified as having severe pneumonia and is put on oxygen and treated with a combination of gentamicin and penicillin.*

   The previous WHO (41) and pre-2016 Kenyan guidelines (20) named this class as "very severe pneumonia".
2. ***Indrawing pneumonia:*** *If a child has lower chest wall indrawing (but does not have any of the qualifying signs for severe pneumonia above) and is alert then s/he is classified as having indrawing pneumonia.*

   In previous WHO (41) and pre-2016 Kenyan guidelines (20) guidelines, this class was named as "severe pneumonia" and treatment recommended was inpatient penicillin monotherapy. Our analyses are based on data from the period before March 2016 when inpatient penicillin monotherapy was recommended for this population.

   Since March 2016 in Kenya, and reflecting updated WHO guidance and results of a local trial (39), it has been recommended that this group be treated in outpatient settings with oral amoxicillin as part of an expanded group of non-severe pneumonia.

   Note: *The term indrawing pneumonia is hereafter used in this analysis to define this category of children to avoid confusion.*
3. ***Non – severe pneumonia:*** *If a child has none of the clinical signs in the 2 categories above but has cough or difficulty breathing and a respiratory rate greater than or equal to 50 breaths/minute (for age between 2 and 11 months) or respiratory rate greater than or equal to 40 breaths/minute (for age above 12 months) then Kenyan guidelines in the period pre and post March 2016 recommend s/he is classified as having non severe pneumonia and treated with oral amoxicillin as an outpatient.*

The ability to use routine data to compare treatment effects requires that patients with similar problems receive different treatments. Previous studies conducted in Kenya and elsewhere have indicated that clinicians often do not follow guideline recommendations in treating pneumonia (22). Variation from the guideline recommended approach can occur at the point of pneumonia severity assignment (clinicians do not follow a nationally approved protocol linking clinical signs and severity category outlined in Box 1) and at the point of treatment assignment (clinicians do not follow this protocol that links treatment and severity). This variability in adherence to protocols provides the opportunity for comparative effectiveness evaluation. More specifically, the adherence and non – adherence to treatment protocols by clinicians allows us to classify indrawing pneumonia admissions in two ways:

1) Those with clinical signs placing them in the group of indrawing pneumonia irrespective of the category or classification assigned to the child by the clinician.

2) Those given a clinician classification of indrawing pneumonia irrespective of the actual clinical signs observed by the clinician.

Based on these two possibilities two experiments were designed (see protocol in press (23)) with specific objectives as follows[1]:

1) **Experiment 1:** To compare effectiveness of injectable penicillin versus penicillin plus gentamicin (both injectable) in treatment of indrawing pneumonia; where the child is identified as belonging to a population of children with indrawing pneumonia on the basis of data on their recorded clinical signs. The Experiment 1 population of indrawing pneumonia is therefore consistent with pre-2016 clinical guideline recommendations.

[1] All children with danger signs were excluded from experiment 1 and in general (both in experiments 1 and 2), children with the following comorbidities were excluded: HIV, meningitis, tuberculosis and or acute severe malnutrition.

2) **Experiment 2:** To compare effectiveness of injectable penicillin versus penicillin plus gentamicin in a population in which we use the clinician assigned categorisation of indrawing pneumonia, which may not be consistent with clinical guideline recommendations.

We defined Experiment 1 as our primary analysis as we propose it would identify a population similar to that recruited to a randomised trial where the inclusion criteria would be based on specified clinical signs. Experiment 2 offers a scenario that may represent a more pragmatic study design with inclusion criteria based around a clinician led classification.

Data source

We use data from the Kenyan Clinical Information Network (CIN) that was initiated to improve inpatient paediatric data availability from county (formerly district) hospitals. Thirteen county referral hospitals were purposively selected with direction from Ministry of Health (MOH) and recruited into the CIN. These hospitals were recruited into the study at different times; four in September 2013, five in October 2013 and four in February 2014. This analysis utilises data up to March 2016. On average, 25 000 paediatric admissions are captured per year. These hospitals typically have one paediatrician leading services predominantly provided by junior clinical teams. Data in these hospitals are collected prospectively post discharge by trained data clerks guided by well-defined standard operating procedures, under supervision by the hospital medical records department and the research team. Clinicians admitting patients fill standardized Paediatric Admission Record (PAR) forms (24) that have been shown to improve documentation of clinical symptoms and signs (25). Together with discharge forms, treatment sheets and laboratory reports these are all part of the patient files that are the primary data source. This data collection system has been described in detail elsewhere (26). Feedback to hospitals as part of the CIN activities has

helped improve the quality of clinical data (26). The description of hospital selection and their populations of patients is detailed elsewhere (27).

Statistical analysis

i) Defining per protocol and intention to treat populations

In typical randomised controlled trials, types of analyses to be conducted are defined beforehand – and this involves defining the type of patient populations that are included in the analyses. Intention to treat and per protocol populations derived from observational datasets have been described in Danaei (2013) (28). We defined per protocol and intention to treat populations based on the dates actual treatments were recorded as prescribed for patients included in our primary and secondary analyses (experiments 1 and 2 respectively). Within each experiment, and after applying inclusion and exclusion criteria, we define the per protocol population as those whose prescription of one of the two study regimens did not change during the admission. The intention to treat population is defined by the original treatment assignment and included children in whom treatment was subsequently changed (see Figure 1 in the Results section).

ii) Dealing with missing data and propensity score matching

As CIN comprises data from routine care settings it faces challenges of non – random treatment allocation and missing data. The missing data and propensity score methods for this analysis have been detailed in the protocol in press linked to this work (23). In brief, after exploring the patient populations, 20 datasets[2] (29) were derived using multiple imputation (with chained equations) for each experiment (all the variables in both the experiments had missing data less than 30% – see table 2 b of the supplementary material). Clinical signs and symptoms data considered were those recorded by clinicians before patients were admitted.

[2] The current literature (29) recommends the use of more than 5 imputed datasets and therefore 20 should be sufficient.

The multiple imputation excluded outcome data as guidance on the use of observational datasets for comparative effectiveness analysis recommends exclusion of outcome data in the design phase (30). Following this, those with missing outcome data were excluded from the analysis (missingness in the outcome data were 0.5% and 0.8% for experiments 1 and 2). For each imputed dataset, patients in the alternative treatment groups (penicillin monotherapy versus penicillin plus gentamicin) were then matched using propensity score (PS) methods to overcome non – random treatment allocation. Propensity scores define the probability of belonging to or being assigned a given treatment based on signs and symptoms (31). PS is a distance measure (32) which is used as a means to overcome allocation bias as treatment outcomes in children with similar propensity scores can then be compared. In these analyses we compared three approaches to reducing possible bias based on PS – optimal full matching, weighting and sub-classification (31, 32). All are aimed at creating groups of patients that are comparable in terms of the distribution of observed signs and symptoms. For each experiment, in order to select the optimum PS implementation method, absolute standardised mean differences (ASMD) were used as diagnostic checks for covariate balance and overlap (33, 34) between the alternative treatment groups. PS methods that resulted in the minimum average absolute standardised mean differences for the majority of the variables while retaining the largest number of patients in the analysis were considered the most appropriate (32).

iii) Analytic modelling and sensitivity analyses

In sample size calculations conducted prior to the experiments (presented in greater detail elsewhere (protocol)), it was estimated that a sample size of at least 4000 would be sufficient for the planned experiments to detect a minimum difference of 1.5% in mortality between the two treatment groups. The sample size for experiment one was 4002 and experiment two 6420 (including 3312 of those that were also in experiment 1). In other words, experiment 2

largely included those in the experiment 1 population but also children not meeting eligibility criteria for experiment 1. For each of the experiments, after multiple imputation, multivariable log-binomial regression models were fitted to PS weighted datasets and adjusting for all the variables also used in the PS models (also as a form of sensitivity analyses, treatment effects were estimated on PS unweighted datasets). Only pooled treatment effect estimates are reported.

One possibility is that clinicians' treatment assignment is skewed such that patients who appear sicker (having a greater number of clinical signs of more severe illness) are assigned 'stronger' or broad spectrum treatment. In this situation as mentioned by Sturmer (2010), specific types of treatment allocation may be more likely associated with increased mortality (35). In theory, the use of propensity scores is supposed to account for such skewed assignment by comparing only outcomes of those with similar propensity scores assumed to suggest they have similar clinical profiles and thus similar risks. PS trimming attempts to tackle this problem further by excluding patients who are at the extremes of the PS distribution to create a population with clinical characteristics that are as homogeneous as possible. We use PS trimming to define a population between the 5% - 95% PS percentiles in a sensitivity analysis.

In a further sensitivity analysis, we used an instrumental variable to examine the potential influence of any unmeasured variables (36). An instrumental variable method aims to find a proxy randomised experiment in a routine or observational dataset (37). We used weekend/weekday admission as an instrumental variable as it was demonstrated in a study conducted by Berkley (2004) (38) in a Kenyan hospital that children who were admitted during the weekend experienced higher mortality compared to those admitted during the weekdays. This, in theory, implies that the type of treatment and care received depends on the day of admission – and this later determines the type of health outcome of the patient. The

process of fitting the instrumental variable models has been described in the supplementary material. The two sensitivity approaches described above were done for both primary and secondary analyses.

Results

a) Creating per protocol and intention to treat populations

Examining the dates treatments were given, five treatment arms (per experimental scenario) were defined – specifically those who received: (1) penicillin alone without changes, (2) a combination of penicillin plus gentamicin without changes, (3) penicillin but switched to a combination of penicillin plus gentamicin, (4) penicillin but switched to ceftriaxone, and (5) a combination of penicillin plus gentamicin but switched to ceftriaxone (ceftriaxone is the recommended second line treatment for severe pneumonia). Therefore, per protocol analyses would compare patients in treatment arm 1 versus 2, while intention to treat analyses would compare patients in treatment arms 1, 3, and 4 versus 2 and 5 (figure 1).

[Insert figure 1]

Figure 1: Summary of patients per treatment arm in experiments 1 – 2

In this analysis, intention to treat populations were considered primary and are reported in experiments 1 and 2 in keeping with clinical trial reporting guidelines. These analyses include a relatively larger number of patients compared to per protocol analyses. The recommended doses of penicillin and gentamicin in these hospitals are 50000 iu/Kg and 7.5 mg/Kg given 4 and once daily respectively. Additional data suggest most clinicians prescribed these doses correctly (see table 7 of the supplementary material).

b) Comparing performance of optimal full matching, weighting and PS sub-classification in experiments 1 and 2 respectively

For each experiment, the three PS implementation methods were compared to determine the one which would result in the least absolute standardised mean differences for most of the variables in the analysis (even though all the three methods resulted in variables with ASMD<=10%). For experiment 1, PS weighting performed better than PS optimal full matching and sub-classification and for experiment 2, the performance of weighting was comparable to that of optimal full matching (see figures 2 and 3). In both experiments, PS sub-classification reduced covariate imbalance the least. Thus, in the subsequent sections, outcome analyses are based on PS weighted datasets for both experiments.

[Insert Figure 2]

Figure 2: Comparing performance of the three PS implementation methods in experiment 1:

The y – axis contains all the variables used in the PS models. While x – axis shows absolute standardised mean difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value of <= 10% indicates the method has performed well in creating comparable groups.

[Insert figure 3]

Figure 3: Comparing performance of the three PS implementation methods in experiment 2

c) Outcome Analysis Results

i) Exploring mortality in raw datasets

Examining the raw datasets without PS adjustments in experiment 1, the average number of pneumonia deaths (across the 20 imputed datasets) in penicillin plus gentamicin group was 33/1363 (2.42%) and in penicillin monotherapy was 26/2639 (0.99%). And for experiment 2, the average number of deaths were 87/2296 (3.79%) and 50/4124 (1.21%) in penicillin plus gentamicin and penicillin monotherapy groups respectively. Overall, the average number of

pneumonia deaths in the penicillin plus gentamicin group was approximately two and a half to three times the number of mortality events in the penicillin monotherapy group in experiments 1 and 2 respectively.

ii) Modelling mortality risk ratios

The analysis considered penicillin monotherapy as the reference group and mortality as the outcome – and therefore a RR greater than one would be interpreted to favour penicillin over penicillin plus gentamicin. For both experiments, the treatment risk ratios (RR) estimated on the unmatched datasets were larger than the RR estimated on datasets obtained through PS weighting (see table 1 for all results). In experiment 2, the PS unadjusted analysis showed that penicillin monotherapy was significantly more effective than penicillin plus gentamicin (1.68 [1.15, 2.36]). However, the PS weighted effect estimate (1.04 [0.76, 1.40]) was much reduced and suggested that use of PS had corrected (to a degree) for allocation bias indicating that there was no statistical difference in mortality outcomes between penicillin plus gentamicin and penicillin monotherapy treatments. We also observed that the adjusted point estimate for any effect difference in experiment 2 (1.04 [0.76, 1.40]) was less than that in experiment 1 (1.46 [0.85, 2.43]). This may be due to an increase in the number of covariables available for PS weighting that could be used in Experiment 2 resulting in closer matching (see table 1 a of the supplementary material).

d) Sensitivity analysis through trimming using 5 – 95% PS population restriction

After excluding 10% of the populations as a result of PS trimming in sensitivity analyses for experiments 1 and 2, the resulting sample sizes were 3583 and 5778. The skewed assignment of children to treatment with gentamicin and penicillin is demonstrated by their higher PS

scores in figure 4 for experiment 1 (and figure 3 in supplementary material for experiment 2). As higher PS scores are associated with the presence of a greater number of clinical signs of illness this also suggests an association between more severe illness and treatment with gentamicin and penicillin. For experiment 1, the estimated average mortality events (on PS unadjusted datasets) were 26/1201 (2.16%) and 24/2382 (1.01%) for penicillin plus gentamicin and penicillin monotherapy groups. While the estimated events in experiment 2 were 62/2026 (3.06%) and 46/3752 (1.22%). Thus in sensitivity analyses for both experiments, trimming excluded more mortality events in the penicillin plus gentamicin group compared with the penicillin monotherapy group. The treatment effects estimated using PS weighted models for the restricted populations as a result of PS trimming showed no statistical difference between the two treatments (table 1).

[Insert figure 4]

Figure 4: Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those who were given penicillin plus gentamicin. While the continuous black line shows the PS distribution for those who received penicillin alone.

e) Sensitivity Analysis through the use of weekend/weekday as an instrumental variable

In order to assess whether a timing of admission variable would form a natural and random experiment, the distributions of covariates were examined across the levels of the instrumental variable (weekend/weekday) in experiments 1 and 2. The distribution of each of the patient characteristics between weekend and weekday admissions was approximately similar (table 4 in supplementary material) suggesting that weekend/weekday admission

satisfactorily satisfies one of the criteria as a valid IV (also see supplementary material for the set of criteria for a valid IV).

The estimated treatment effects, both in experiments 1 and 2, suggest there is no statistical difference in treating indrawing pneumonia with either penicillin alone or penicillin plus gentamicin. The effect estimates obtained using our IV in both experiments are less than one as compared to those obtained with PS weighting which are greater than one. Biologically, the effectiveness of gentamicin plus penicillin (when administered in correct doses) is expected to be the same or greater than that of penicillin monotherapy. Based on the magnitude and direction of effects, the use of the IV seems to demonstrate that the effects obtained through PS weighting may have had some residual bias. However, it is important to highlight that for all analyses the 95%CI obtained are consistent with the Null Hypothesis of no different effect for the treatments.

Table 1: Treatment effect estimates (RR (95% C.I))

	Experiment 1	**Experiment 2**
Regression without PS adjustment	1.75 [0.94, 2.77]	1.68 [1.15, 2.36]
PS Weighting	1.46 [0.85, 2.43]	1.04 [0.76, 1.40]
PS trimming (5% – 95% restriction)	1.39 [0.74, 2.15]	1.05 [0.74, 1.41]
Instrumental variable	0.91 [0.41, 2.20]	0.44 [0.34, 1.32]

Discussion

We compared penicillin alone with penicillin plus gentamicin in treatment of indrawing pneumonia in populations with overall mortality of 1.5% and 2% in experiments 1 and 2 respectively. There were more fatal events in the penicillin plus gentamicin group than the penicillin group (approximately 2.5 times) and unadjusted analyses pointed, therefore, to a protective effect of penicillin treatment. However, adjusted analyses, both in experiments 1 and 2, that aim to account for allocation bias that can result from non-random treatment allocation suggest that there is no appreciable difference in outcomes between penicillin and

gentamicin plus penicillin treatment of indrawing pneumonia. Such adjusted analyses were conducted with multiple propensity score methods (PS weighting, optimal full matching and sub-classification) and both intention to treat and per protocol populations, all of which showed similar findings (see the provided supplementary material). Propensity score trimming and instrumental variable analyses further support the suggestion that poor outcome in this population is not associated with the antibiotic regimen received.

Our analyses were conducted using data from over 4,000 children, one hundred times more participants than were included in the only prior randomised controlled trial of penicillin monotherapy and penicillin plus gentamicin in treatment of pneumonia in an Asian population (17). There are continuing concerns of clinically important mortality in children with indrawing pneumonia in Africa (21). This has led to hesitation to adopt new WHO and Kenyan guidelines that now recommend the treatment of indrawing pneumonia as an outpatient using amoxicillin (11, 19). Our results suggest that there are likely to be two distinct issues. Firstly, they suggest that offering broader spectrum injectable antibiotic treatment to children with indrawing pneumonia may not improve outcomes compared to treatment with penicillin monotherapy. As other studies have suggested equivalence between oral (high dose) amoxicillin therapy and injectable penicillin therapy (2-5, 39) it seems likely therefore that oral amoxicillin and penicillin plus gentamicin combination therapy would result in similar outcomes when used to treat indrawing pneumonia. Clinicians should therefore carefully adhere to guidelines for treatment of indrawing pneumonia and avoid using gentamicin helping to prevent any possible toxicity.

Secondly, however, our results suggest that children fulfilling a definition of indrawing pneumonia based on clinical signs, and having excluded serious co-morbidities, may still have an appreciable risk of mortality irrespective of their antibiotic treatment (1.5% in all children in experiment 1). When clinicians categorise children with indrawing pneumonia

and imperfectly adhere to clinical sign based guidance mortality tends to be higher (2% in all children in experiment 2). These findings point to as yet uncharacterised risk factors that could be important in determining which children need admission to hospital as current guidance indicates that all those with indrawing pneumonia can be treated as an outpatient. While offering an alternative antibiotic to amoxicillin to this group may not improve outcomes it is possible that closer and continuing observation in hospital may help identify co-morbid or alternative conditions that are contributing to this mortality and that may be treated.

The trials that informed the basis for the revised WHO guidelines (2-5) showed extremely low mortality (0 – 0.2%) suggesting that the populations included in such trials may not be directly representative of all those to whom guidelines are applied in routine settings. In the trial by Agweyu (2015) conducted in Kenya (which compared penicillin versus oral amoxicillin for indrawing pneumonia) overall mortality was 0.8% (39). In a parallel observational cohort providing data from the same hospitals over the same time period for children treated with penicillin alone but not included in the Kenyan trial mortality was not significantly different but marginally higher at 1.2% (Agweyu (2017), submitted) perhaps suggesting that even the limited exclusion criteria in this pragmatic trial might result in exclusion of some sicker children. Taken together with data from the analyses presented here it does appear there is a need to explore whether guidelines might be modified to accommodate additional clinical risk factors for possible life-threatening illness that should prompt admission. In a population with high coverage with conjugate vaccines this may more usefully be for more rigorous evaluation to identify alternative diagnoses or for improved supportive care than for different antibiotics.

Strengths and limitations

Conducting comparative effectiveness analyses using observational datasets can offer the advantage of larger sample sizes at lower cost than randomised controlled clinical trials. They also include patients that may not qualify for enrolment in a typical explanatory randomised controlled trial – and therefore perhaps provide more true to life estimates of treatment effects similar to those observed in highly pragmatic trials (15). However, as most observational datasets are not meant for research, they have challenges of non-random treatment allocation and missing data. We employed a rigorous 'experimental design' strategy as is recommended when using observational data (30). We used PS and multiple imputation methods in an effort to minimise bias due to non-random treatment allocation and missing data and analyses suggested no appreciable difference in outcomes of indrawing pneumonia treated with penicillin alone compared with penicillin plus gentamicin. This was in contrast to unadjusted regression analyses that pointed towards better outcomes with penicillin alone suggesting presence of allocation bias. As most observational datasets are limited to observed variables, it is important to conduct sensitivity analysis to explore if the estimated effects are potentially sensitive to unmeasured variables. We used an instrumental variable and PS trimming, both supported the idea of no appreciable difference regimens when treating indrawing pneumonia. While there are differences (in terms of magnitude) in the mortality observed in the different groups that suggest some residual bias in treatment allocation, these mortality differences are no greater than might occur by chance after PS adjustment (with the type 1 and 2 errors specified in the protocol). In that sense the PS approach may still have limitations but it does allow us to conclude no statistical difference in mortality outcomes between the two treatment arms.

The WHO recommended guidelines for treating pneumonia have considerable influence on policy and practice in low and middle income countries. While the evidence base and rigour of guideline development have improved considerably there remain few data on their

effectiveness when implemented in non-trial settings. Even though well-designed, large pragmatic trials would be preferred, we demonstrate that carefully collected routine data may be useful for assessing the effectiveness of alternative treatments (15). Such analyses may become increasingly possible as electronic medical records are deployed in low and middle income countries (40) but it is important that such studies are carefully designed to limit as far as possible the biases that arise from non-random treatment allocation (30). Our results suggest that children with indrawing pneumonia may gain little benefit from treatment with broader spectrum antibiotic regimens. However, they also suggest that further work is needed to identify those who are at higher risk of death who might be prioritised for an inpatient diagnostic work up and improved supportive care rather than treated as outpatients.

Additional files

Additional file 1: Analysis protocol (published in BMJ Open)

Additional file 2: Supplementary material

Acknowledgement

We would like to thank Ambrose Agweyu for his comments that helped improve this manuscript. We also appreciate the valuable contribution offered by the CIN team: Lydia Thuranira & Grace Ochieng' (Kiambu County Hospital), Barnabas Kigen (Busia County Hospital), Melab Musabi & Rachel Inginia (Kitale County Hospital), Anne Kamunya & Sam Otido (Embu County Hospital), Margaret Kuria (Kisumu East County Hospital), Agnes Mithamo & Francis Kanyingi (Nyeri County Hospital), Celia Muturi, Caren Emadau & Cecilia Mutiso (Mama Lucy Kibaki County Hospital), David Kimutai & Loice Mutai (Mbagathi County Hospital), Nick Aduro (Kakamega County Hospital), Samuel Ng'arng'ar (Vihiga County Hospital). Fred Were & David Githanga (Kenya Paediatric Association). Rachel Nyamai (Ministry of Health).

Funding

We are grateful for the funds from the Wellcome Trust (#097170) that support ME through a fellowship and additional funds from a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (#092654) that supported this work. LM is supported by a Nuffield Department of Medicine Prize DPhil Studentship and Clarendon Scholarship (Oxford University). The funders had no role in drafting or submitting this manuscript.

Availability of Data

The hospital specific datasets are in custody of the hospitals participating in CIN (these datasets have been de-identified).

Authors Contributions

The contributions of the authors were as follows: LM did an initial draft of this manuscript with the support of RP, EM and ME. Thereafter, all authors edited subsequent versions and approved the final copy.

Competing Interests

The authors declare they have no competing interests.

Ethics approval

This analysis is based on a larger project (CIN) which was cleared by the Kenya Medical Research Institute ethics and review board (Protocol number: 2465).

References

1. Olayemi E, Asare EV, Benneh-Akwasi Kuma AA. Guidelines in lower-middle income countries. Br J Haematol. 2017.

2. Atkinson M, et al. A multicentre randomised controlled equivalence trial comparing oral amoxicillin and intravenous benzyl penicillin for community acquired pneumonia in children PIVOT Trial. Thorax. 2007;61:1102-6.

3. Hazir T, et al. Ambulatory short-course high-dose oral amoxicillin for treatment of severe pneumonia in children: a randomised equivalency trial. Lancet. 2008;371:49-56.

4. Addo-Yobo E, Chisaka N, Hassan M, Hibberd P, Lozano JM, Jeena P, et al. Oral amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59 months: a randomised multicentre equivalency study. Lancet. 2004;364:1141-8.

5. Addo-Yobo E, et al. Outpatient treatment of children with severe pneumonia with oral amoxicillin in four countries: the MASS study. Tropical Medicine and International Health. 2011;16:995-1006.

6. Mulholland K, Carlin JB, Duke T, Weber M. Challenges of trials of antibiotics for pneumonia in low-income countries. The Lancet Respiratory Medicine. 2014;2(12):952-4.

7. IVAC. Pneumonia and Diarrhoea Progress Report 2014. John Hopkins Bloomberg School of Public Health; 2014.

8. Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, et al. Bacteremia among Children Admitted to a Rural Hospital in Kenya. The new england journal of medicine. 2005;352:39-47.

9. Global Health Data Exchange [Internet].; 2015 []. Available from: http://ghdx.healthdata.org/gbd-results-tool?params=querytool-permalink/ee043b99b22a223f41b3e9d38c5c596a.

10. Kenya Demographic and Health Survey [Internet].; 2014 []. Available from: http://dhsprogram.com/pubs/pdf/FR308/FR308.pdf.

11. Basic paediatric protocols [Internet].: Kenyan Ministry of Health; 2016 []. Available from: https://www.tropicalmedicine.ox.ac.uk/_asset/file/basic-paediatric-protocols-2016.pdf.

12. WHO. Hospital Care for Children. 2013.

13. UNICEF DATA: Monitoring the situation of women and children [Internet].; 2017 []. Available from: https://data.unicef.org/topic/child-health/pneumonia/.

14. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. BMJ Open Research Methods. 2014.

15. Thadhani R. Formal trials versus observational studies.Oxford PharmaGenesis. 2006.

16. West SG, Duan N, Pequegnat W, Gaist P, Jarlais DCD, Holtgrave D, et al. Alternatives to Randomised Controlled Trials. American Public Health Association. 2008;98(8):1359-66.

17. Hasali MAA, Ibrahim MIM, Sulaiman SAS, Ahmad Z, Hasali JBA. A clinical and economic study of community-acquired pneumonia between single versus combination therapy. Pharm World Sci. 2005;27(3):249-53.

18. Ayieko P, English M. Case Management of Childhood Pneumonia in Developing Countries. Pediatr Infect Dis J. 2007;26:432-40.

19. Hospital Care for Children [Internet].; 2013 []. Available from: http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf.

20. Basic Paediatric Protocols: November 2013 Edition [Internet].: Ministry of Health; 2013 []. Available from: http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition.

21. Koh JW, Wong JJ, Sultana R, Wong PP, Mok YH, Lee JH. Risk factors for mortality in children with pneumonia admitted to the pediatric intensive care unit. Pediatric pulmonology. 2017.

22. Agweyu A, Kibore M, Digolo L, Kosgei C, Maina V, Mugane S, et al. Prevalence and correlates of treatment failure among Kenyan children hospitalised with severe community-acquired pneumonia: a prospective study of the clinical effectiveness of WHO pneumonia case management guidelines. Tropical Medicine & International Health. 2014.

23. Malla L, Perera R, McFadden E, English M. Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in Kenya: A protocol for an observational study. BMJ Open (in press) [Internet]. 2017.

24. Mwakyusa S, Wamae A, Wasunna A, Were F, Esamai F, Ogutu B, et al. Implementation of a structured paediatric admission record for district hospitals in Kenya – results of a pilot study. BMC International Health and human rights. 2006.

25. Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A Multifaceted Intervention to Implement Guidelines and Improve Admission Paediatric Care in Kenyan District Hospitals: A Cluster Randomised Trial. PLoS Medicine. 2011.

26. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation of clinical care within a clinical information network: an essential initial step in efforts to understand and improve care in Kenyan hospitals. BMJ Global Health. 2016.

27. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in Kenyan hospitals within a new Clinical Information Network. Archives of Disease in Childhood. 2015.

28. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. Stat.Methods Med.Res. 2013.

29. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.

30. Rubin D.B. For Objective Causal Inference, Design Trumps Analysis. The Annals of Applied Statistics. 2008;2(3):808-40.

31. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011.

32. Stuart EA. Matching methods for causal inference:A review and a look forward. Statist. Sci. 2010;25(1):1-21.

33. Austin PC. Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. Pharmacoepidemiol Drug Saf. 2008 2008;17(12):1218-25.

34. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009 10 Nov 2009;28(25):3083-107.

35. Stumer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the presence of Unmeasured Confounding: Dealing with observations in the Tails of the Propensity Score Distribution - A Simulation Study. American Journal of Epidemiology. 2010.

36. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014 15 Jun 2014;33(13):2297-340.

37. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference. Statistics in Medicine. 2014;33(13):2297-340.

38. Berkley JA, Brent A, Mwangi I, English M, Maitland K, Marsh K, et al. Mortality among Kenyan children admitted to a rural district hospital on weekends as compared with weekdays . Pediatrics. 2004;114:1737-8.

39. Agweyu A, Gathara D, Oliwa J, Muinga N, Edwards T, Allen E, et al. Oral amoxicillin versus benzyl penicillin for severe pneumonia among kenyan children: a pragmatic randomised controlled noninferiority trial. Clin Infect Dis. 2015.

40. English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al. Building Learning Health Systems to Accelerate Research and Improve Outcomes of Clinical Care in Low-and Middle-Income Countries. PLOS Medicine. 2016.

41. Hospital Care for Children [Internet].; 2005 []. Available from: http://apps.who.int/iris/bitstream/10665/43206/1/9241546700.pdf.

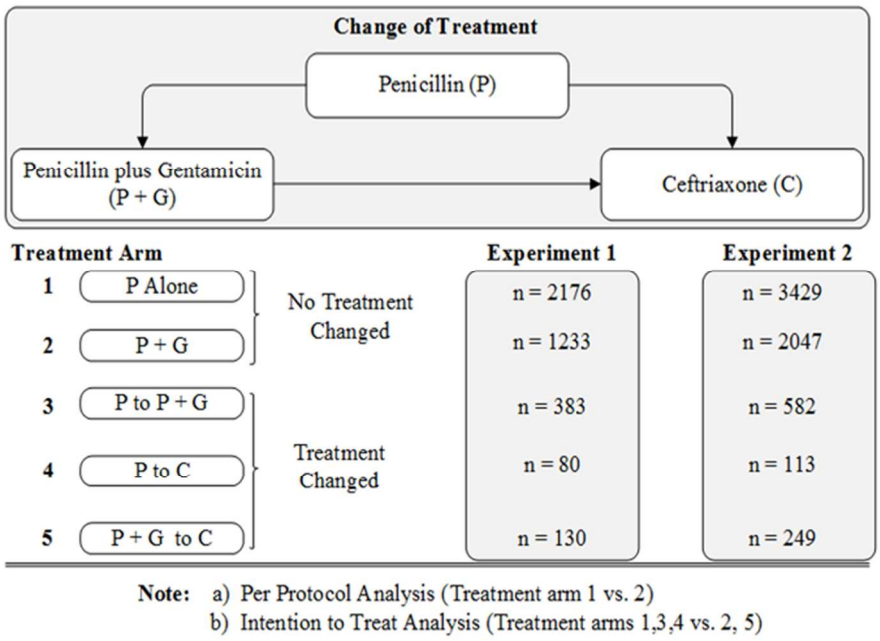


Figure 1: Summary of patients per treatment arm in experiments 1 – 2

63x45mm (600 x 600 DPI)

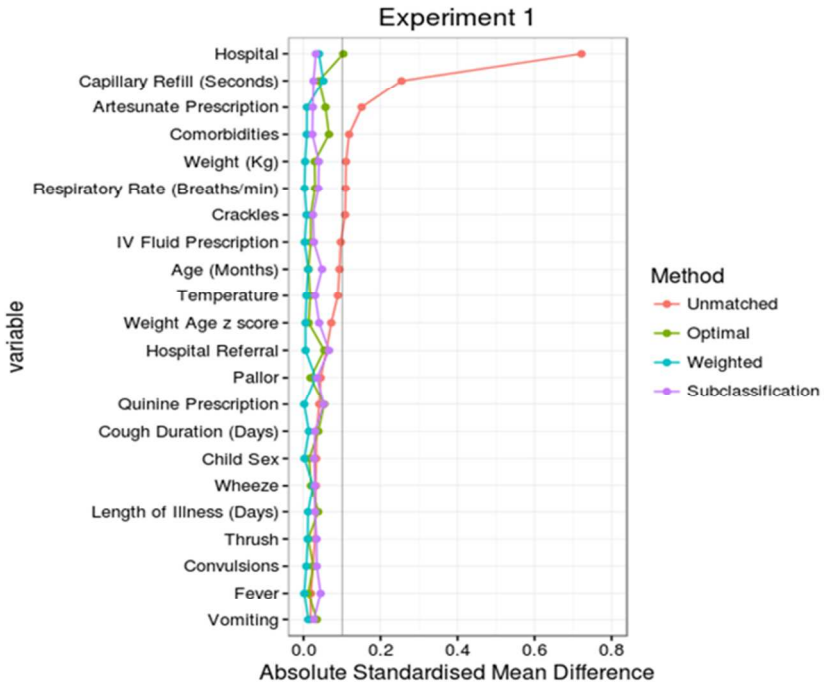


Figure 2: Comparing performance of the three PS implementation methods in experiment 1: The y – axis contains all the variables used in the PS models. While x – axis shows absolute standardised mean difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value of <= 10% indicates the method has performed well in creating comparable groups.

63x45mm (600 x 600 DPI)

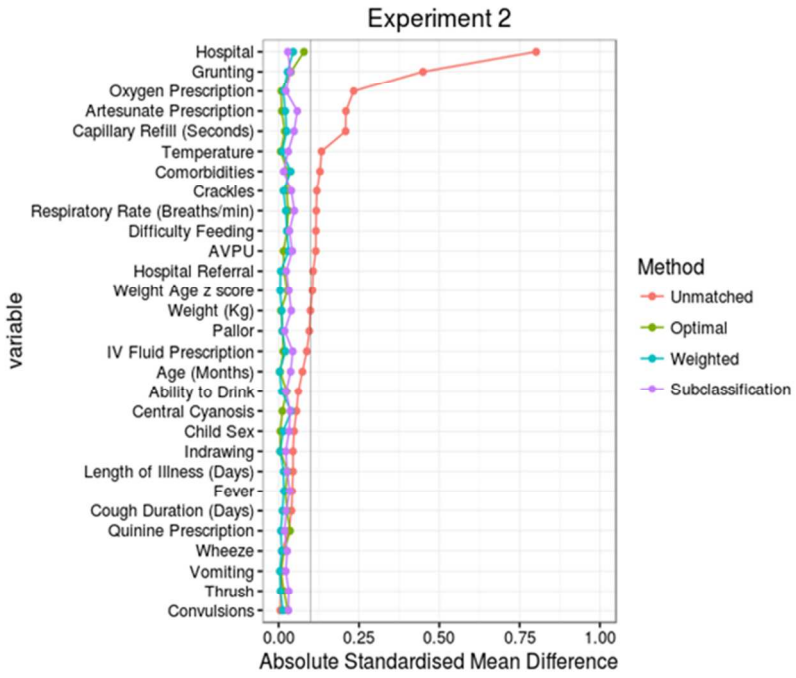


Figure 3: Comparing performance of the three PS implementation methods in experiment 2

63x45mm (600 x 600 DPI)

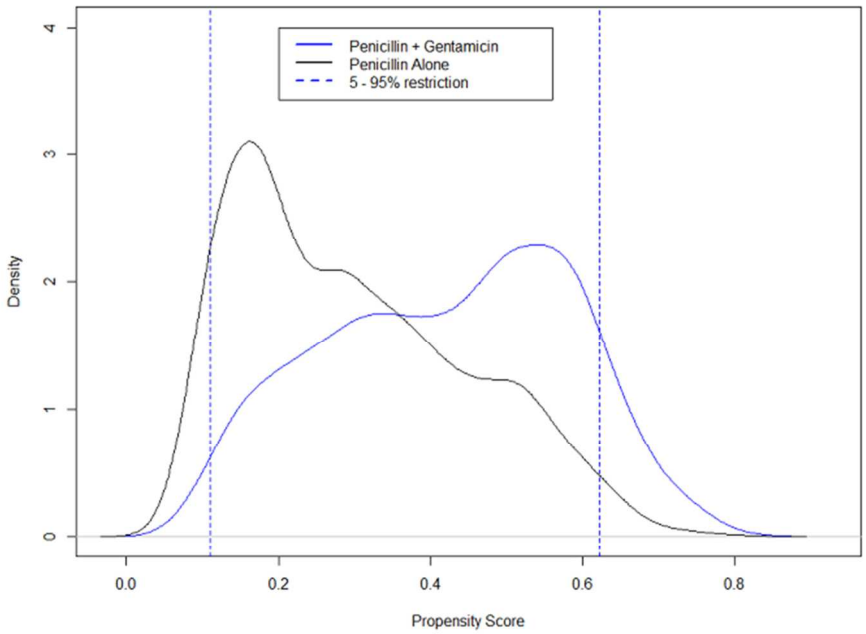


Figure 4: Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those who were given penicillin plus gentamicin. While the continuous black line shows the PS distribution for those who received penicillin alone.

63x45mm (600 x 600 DPI)

Using observational data to compare the effectiveness of antibiotic treatments for children hospitalised with pneumonia in Kenya

Lucas Malla[1*]

Rafael Perera-Salazar[2]

Emily McFadden[2]

Mike English[1, 3]

[1]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom.

[3] Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.

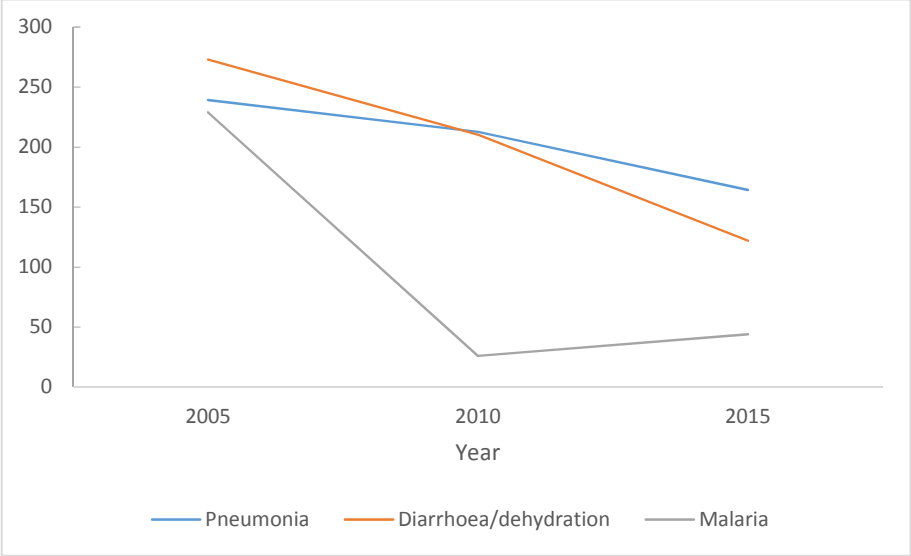
The supplementary material is organised into the following subsections:

- Under-five mortality incident rates in Kenya
- Summary of analysis variables
- Analysis using PS sub classification
- Analysis using optimal full matching
- Experiment 2 trimming
- Analysis using instrumental variables
- Analysis using per protocol population
- Overlap and correctness of penicillin and gentamicin dosing
- Definition of PS methods and how they were used

a) Under-five mortality incident rates in Kenya

The mortality data summarised in figure A were extracted from the Global Health Data Exchange (GHDx) website (accessible through this link: http://ghdx.healthdata.org/gbd-results-tool?params=querytool-permalink/ee043b99b22a223f41b3e9d38c5c596a)

Figure A: Deaths, rate per 100,000



b) Summary of analysis variables

Table 1 a: Summary of key and auxiliary independent variables for experiments 1 and 2[1]

Experiment 1 key variables	Experiment 2 key variables	Auxiliary variables for experiments 1 and 2
Age (2 – 59 months)	Age (2 – 59 months)	Gender (male/female)
Indrawing (present/absent)	Indrawing (present/absent)	Cough duration (days)
Level of consciousness – AVPU (alert/verbal response/pain response/unresponsive)	History of cough (yes/no)	Crackles (present/absent)
	Difficulty breathing (present/absent)	Weight (Kg)
	Level of consciousness – AVPU	Pallor (0, +, +++)
	Central cyanosis	Capillary refill (immediate, 1 – 2 secs, 3 – 6 sec, > 6 secs)
	Grunting	Fever (present/absent)
	Ability to drink	Convulsions (present/absent)

[1] Comorbidities and WAZ variables were derived after multiple imputation

		Vomiting (yes/no)
		Referral (yes/no)
		Length of illness (days)
		Thrush (present/absent)
		Quinine/artesunate (prescribed/not prescribed)
		Weight for age z – score
		Wheeze (present/absent)
		Comorbidities (Malaria and or diarrhoea)

Table 1 b: Percentage of documentation of analysis variables (experiments 1 and 2)

Variable	Experiment 1 (%)	Experiment 2 (%)
Age (2 – 59 months)	99.7	99.5
Indrawing (present/absent)	100.0	96.3
Level of consciousness – AVPU	_	95.5
Central cyanosis	_	95.9
Grunting	_	94.2
Ability to drink	_	91.2
Gender (male/female)	99.6	99.0
Cough duration (days)	84.9	83.4
Crackles (present/absent)	97.4	94.7
Weight (Kg)	96.3	96.0
Pallor (0, +, +++)	96.7	94.5
Capillary refill	83.3	78.0
Fever (present/absent)	98.2	97.6
Temperature	94.1	92.6
Convulsions (present/absent)	96.3	94.3
Vomiting (yes/no)	97.1	95.2
Referral (yes/no)	83.3	73.6
Length of illness (days)	98.4	98.0
Thrush (present/absent)	90.4	83.9
Quinine/artesunate (prescribed/not prescribed)	100.0	100.0
Wheeze (present/absent)	97.1	94.5
Respiratory rate	87.4	85.4
IV fluid prescription	100.0	100.0
Outcome (died/alive)	99.5	99.2

c) Analysis using PS sub - classification

PS should classify children in groups where they share clinical features, as these features are also related to outcomes then in this case they are also grouped by severity. The average proportion of children who died increased consistently from PS subclass one to five for the two experiments. As PS was used as a proxy for disease severity in sub-classification,

children in subclass 1 were likely to have less severe pneumonia (fewer variables with a positive value that may be associated with possible risk) and children in subclass 5 were likely to have more severe pneumonia (more variables with a positive value that may be associated with possible risk) (table 2). Therefore, this relationship of PS subclass with mortality is expected.

Table 2: Severe Pneumonia Deaths in Experiment 1 and 2 (ITT)

	Experiment 1		Experiment 2	
PS Subgroup	Penicillin plus Gentamicin	Penicillin	Penicillin plus Gentamicin	Penicillin
1	7/273 (2.56%)	8/1269 (0.63%)	9/459 (1.96%)	14/2333 (0.60%)
2	3/272 (1.10%)	1/591 (0.17%)	12/459 (2.61%)	10/822 (1.22%)
3	6/273 (2.20%)	8/380 (2.11%)	16/459 (3.49%)	8/467 (1.71%)
4	8/272 (2.94%)	4/266 (1.50%)	17/458 (3.71%)	11/341 (3.23%)
5	9/273 (3.30%)	5/133 (3.76%)	33/460 (7.17%)	7/153 (4.58%)
Total	33/1363 (2.46%)	26/2639 (0.99%)	87/2296 (3.79%)	50/4124 (1.21%)

In PS sub-classification (for experiment 2 – figure 1) the log risk ratios consistently decreased from subclass 1 to 5 though this pattern was not completely clear in experiment 1 (figure 2). In order to obtain pooled treatment effect, estimates were weighted by the number of patients who received penicillin plus gentamicin per subclass. However, the number of patients who received penicillin plus gentamicin were distributed equally (which would imply equal weighting) – and additional weighting was based on how precise the log risk ratios were. This implied that the subclasses were treated as different trials and log RR estimates pooled in the form of a meta-analysis. The pooled estimates across the subclasses for experiments 1 and 2 were not statistically significant though had wider credible intervals as subclassification did not completely achieve balance on some of the variables at the subclass level.

Figure 1: Experiment 1 – ITT

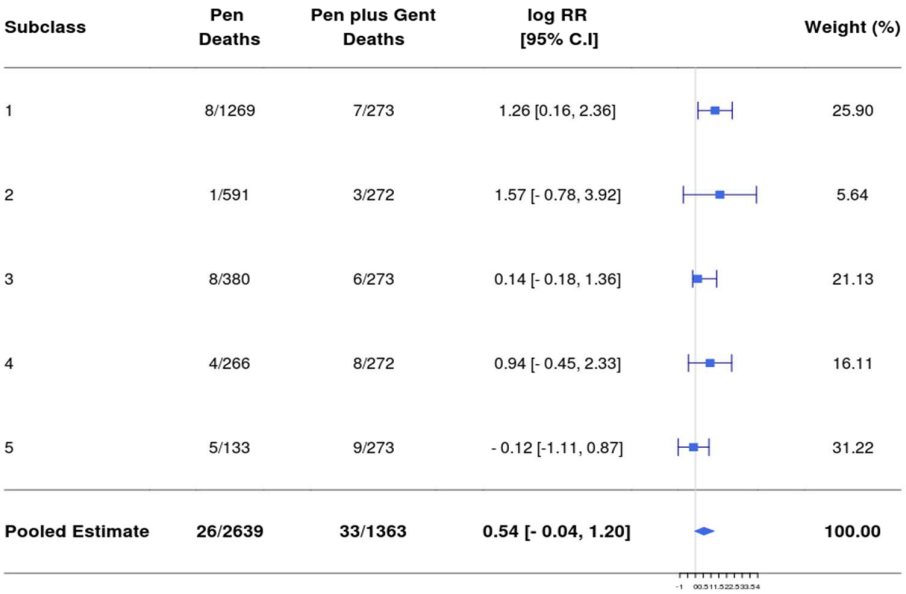
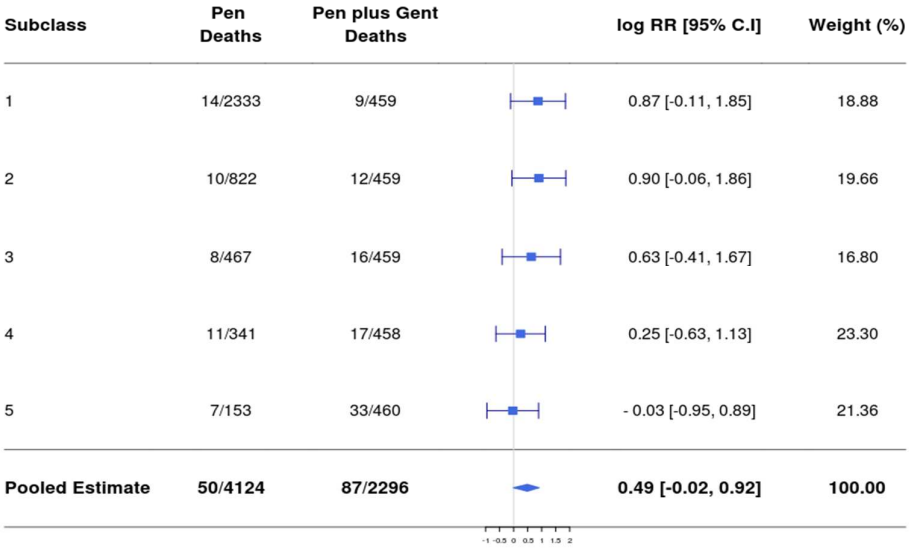


Figure 2: Experiment 2 – ITT



d) Analysis using PS optimal full matching

Also analysis using PS optimal full matching showed no statistical significance in treatment of indrawing pneumonia using either penicillin or penicillin plus gentamicin (table 3).

Table 3: Treatment effect estimates

	log RR (95% C.I)

Experiment 1	
Regression without PS adjustment	0.56 [-0.06, 1.02]
PS optimal matching	0.27 [-0.22, 0.65]
Experiment 2	
Regression without PS adjustment	0.52 [0.14, 0.86]
PS optimal matching	-0.08 [-0.37, 0.18]

e) Trimming in experiment 2 (ITT population)

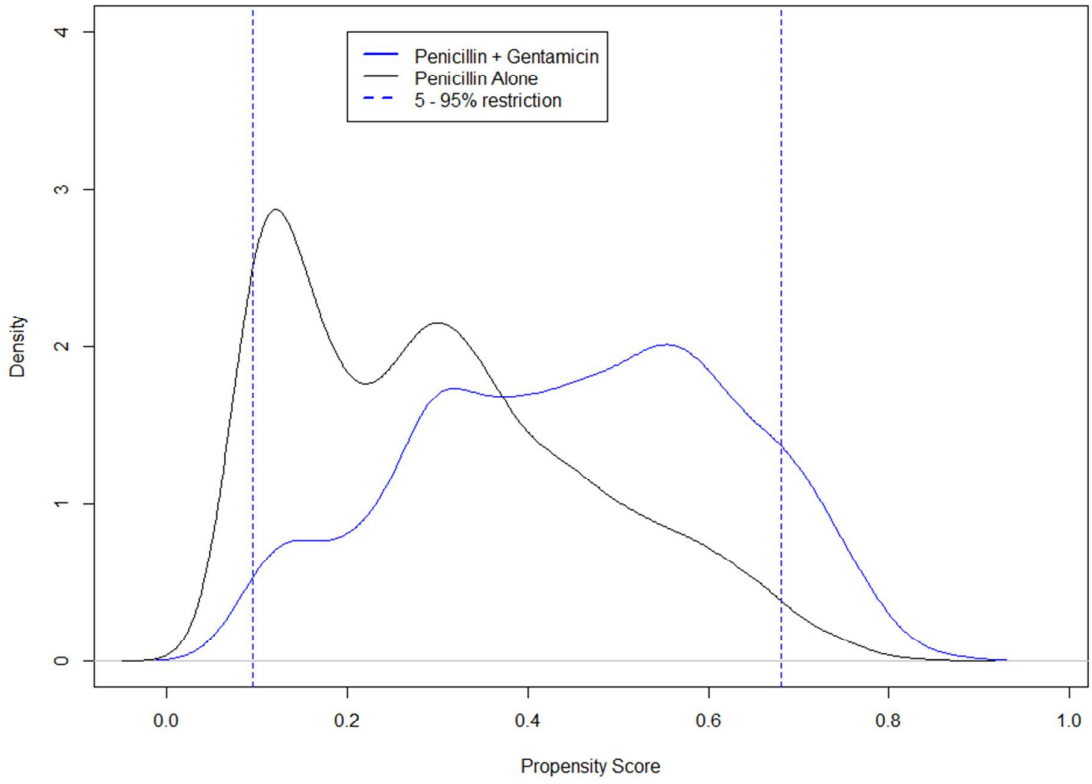


Figure 3: PS trimming in experiment 2

f) Analysis using Instrument variables

Since a valid instrumental variable should be: (i) usable as a variable for randomly and effectively assigning patients into alternative groups, distribution of patients was examined across the levels of the IV as the distribution should be approximately similar between the IV levels; (ii) related with the treatment, a likelihood ratio test was conducted to examine the

treatment – IV relationship. The process of fitting the instrumental variable models has been described in the supplementary material.

Imbalance of covariates between weekday and weekend admissions were explored (table 4).

Table 4: Imbalance of covariates between weekday and weekend admissions

Variable	Experiment 1			Experiment 2		
	Weekdays (n = 3014)	Weekends (n = 988)	ASMD	Weekdays (n = 4881)	Weekends (n = 1539)	ASMD
Child Sex						
Female	45%	46%	0.03	44%	45%	0.01
Male	55%	54%		56%	55%	
Pallor						
Mild/moderate	4%	5%	0.02	5%	5%	0.00
None	95%	94%		93%	93%	
Severe	1%	2%		2%	2%	
Capillary refill						
1 sec	68%	71%	0.07	66%	68%	0.04
2 sec	30%	27%		31%	29%	
>2 sec	3%	2%		3%	3%	
Fever						
Absent	21%	18%	0.05	19%	16%	0.07
Present	79%	82%		81%	84%	
Convulsions						
Absent	95%	96%	0.02	94%	94%	0.03
Present	5%	4%		6%	6%	
Vomiting						
No	65%	62%	0.06	63%	63%	0.00
Yes	35%	38%		37%	37%	
Referral						
No	82%	86%	0.10	81%	84%	0.09
Yes	18%	14%		19%	16%	
Thrush						
Absent	98%	98%	0.00	98%	98%	0.03
Present	2%	2%		2%	2%	
Comorbidities						
None	84%	83%	0.02	82%	80%	0.03
Malaria	9%	10%		10%	13%	
Diarrhoea	3%	2%		3%	2%	
Malaria and diarrhoea	4%	5%		5%	5%	
Crackles						
Absent	47%	47%	0.01	48%	47%	0.02
Present	53%	53%		52%	53%	
Wheeze						
Absent	85%	84%	0.02	85%	84%	0.02
Present	15%	16%		15%	16%	
IV prescription						
No	97%	96%	0.05	95%	95%	0.01
Yes	3%	4%		5%	5%	
Quinine Prescription						
No	97%	97%	0.02	95%	94%	0.04
Yes	3%	3%		5%	6%	
Artesunate Prescription						
No	92%	92%	0.01	92%	90%	0.05
Yes	8%	8%		8%	10%	
Mean WAZ	0.00	-0.01	0.01	0.01	-0.03	0.03
Mean age (months)	19.59	20.47	0.04	20.29	21.05	0.04
Mean weight (Kg)	9.56	9.61	0.01	9.7	9.89	0.05
Mean resp rate (breaths/min)	52.61	51.65	0.08	51.82	51.34	0.04
Mean temp (degrees C)	37.73	37.79	0.06	37.78	37.85	0.06

Mean cough duration (days)	3.40	3.20	0.07	3.45	3.35	0.04
Mean length of illness (days)	3.70	3.46	0.08	3.73	3.56	0.05

Also mortality between weekend and weekday admissions was explored for experiments 1 and 2 (table 5). The weekend mortalities, in the raw datasets, seemed to be higher than weekday mortalities.

Table 5: Summary of deaths by weekend/weekday admissions

Experiment	Weekend	Weekday
1	17/988 (1.7%)	45/3014 (1.5%)
2	47/1539 (3.1%)	49/4881 (1.0%)

In the next step, the treatment and outcome (mortality) probit models were fitted, with covariates in the treatment model being the same as those used in the corresponding propensity score models – though with the addition of admission timing variable as an IV. On the other hand, the outcome model used the same covariates as the treatment model with the exclusion of the admission timing variable both in experiments 1 and 2. Here, the parameter estimates were only presented for the treatment variable (mainly for comparison with individual treatment effect estimates obtained using propensity score weighting method.

Interpreting individual coefficients (like for treatment here) is less straightforward in probit models compared to linear regression and logit models where estimates are individually interpretable (1). This is because change in probability due to a unit change in a predictor is jointly dependent on other predictor values and their starting values. However, there are limited ways through which probit model parameters may be interpreted individually: (i) without considering the magnitude, the direction of effect may be inferred based on whether the parameter estimate is either positive or negative; (ii) if both the magnitude and direction are of interest (as is the case here), then a set of approximations may be conducted. Amemiya (1981) suggested multiplying the individual estimate from probit model by 1.6 to obtain the result in terms of log odds ratio (2). As the estimates obtained using PS methods were

expressed in terms of log relative risk, the estimated odds ratios are further converted to log risk ratio using the modified relationship documented in (3):

$$log\ RR = \log\left(\frac{OR}{(1 - p_0) + (p_0 \times OR)}\right)$$

Where RR – is the risk ratio; OR – odds ratio and; p_0 – is the proportion of children who died in the penicillin monotherapy treatment group. Results have been presented in table 2 in the main manuscript.

g) Analysis using per protocol population

Analysis using propensity score methods with per protocol population also demonstrated no significance in treatment with either penicillin or penicillin plus gentamicin (table 6).

Table 6 : Per protocol treatment effect estimates

	log RR (95% C.I)
Experiment 1	
Unmatched (regression only)	0.71 [0.03, 1.42]
Optimal Full Matching	0.61 [0.05, 1.29]
Weighting	0.45 [-0.14, 1.09]
Sub-classification (pooled)	0.64 [-0.03, 1.32]
Experiment 2	
Unmatched (regression only)	0.54 [0.09, 0.98]
Optimal Full Matching	-0.33 [-0.66, 0.01]
Weighting	-0.13 [-0.48, 0.21]
Sub-classification (pooled)	0.47-0.08, 0.89]

h) Overlap and correctness of penicillin and gentamicin dosing

A total of 3312 patients were both common to experiments 1 and 2. We also examined if the patients received correct dosages of penicillin and gentamicin: For penicillin, a dose of 40,000 – 60,000 I.U/Kg was considered normal and for gentamicin, a dose of 6 – 9 mg/Kg. These were +/- 20% of recommended dosages in the Kenyan paediatric protocols. Majority of the patients were prescribed normal dosages of penicillin and gentamicin (see table 7).

Table 7: Correctness of penicillin and gentamicin prescription

	Experiment 1		Experiment 2	
	Penicillin	Gentamicin	Penicillin	Gentamicin
Under dose	3%	10%	3%	12%
Normal	93%	87%	92%	85%
Over dose	4%	3%	5%	3%

i) Definition of PS methods and how they were used

We implemented three PS methods and these are briefly introduced:

Optimal full matching

PS matching aims to obtain treatment and (active) control patients who have approximately equivalent propensity score values (4). In optimal full matching, an optimal algorithm is used to obtain subsets of matched patients with the least global distance between them. Distance, here, is defined as the absolute difference in the propensity scores between a treated and control patient with global distance the sum of all distances between matched treated and control patients (5). This is the only form of matching that happens without replacement.

PS Weighting

There are two types of weights that may be estimated using PS. The first is inverse probability of treatment weights (IPTW) such that treated individuals are assigned weights of 1/PS while those in the (active) control group are assigned weights of 1/(1 – PS). The second is weighting by odds such that those treated are assigned a weight of 1 and those in the (active) control are assigned weights of PS/(1 – PS). These weights are used to estimate different treatment quantities. In this analysis we used weighting by odds to estimate what effect would be obtained suppose those who received gentamicin plus penicillin were denied this treatment.

PS sub – classification

Sub-classification divides patients into mutually exclusive groups based on their propensity scores. A standard practice, though not supported by specific recommendations, has been subdividing patients into five subclasses (6). One approach for creating patient subclasses would be to first conduct one on one nearest neighbour matching and then split the population into subclasses (7), alternatively one may use PS quintiles (4). The number of subclasses will usually depend on the sample size, and for large datasets, more classes with reasonable sample sizes would be desirable. This analysis used PS quintiles with five subclasses.

References

1. Probit regression [Internet].; 2017 [cited 22-03-2017]. Available from: http://stats.idre.ucla.edu/stata/output/probit-regression/.

2. Amemiya T. Qualitative Response Models: A Survey. Journal of Economic Literature. 1981;XIX:1483-536.

3. Zhang J, Yu KF. What's the Relative Risk? A method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. JAMA - Journal of the American Medical Association. 1998.

4. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011.

5. Gu X, Rosenbaum PR. Comparison of multivariate matching methods: structures, distances and algorithms. Journal of Computational and Graphical Statistics. 1993;2:405-20.

6. Yang S, Imbens GW, Cui Z, Faries DE, Kadziola Z. Propensity Score Matching and Subclassification in observational Studies with Multilevel Treatments. Journal of the International Biometric Society. 2016.

7. Ho DE, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. 2011;42(8):1-28.

BMJ Open

Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in Kenya: A retrospective observational study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019478.R1
Article Type:	Research
Date Submitted by the Author:	20-Oct-2017
Complete List of Authors:	Malla, Lucas; Kenya Medical Research Institute - Wellcome Trust; University of Oxford Perera, Rafael; University of Oxford, Primary Health Care McFadden, Emily; University of Oxford, Nuffield Department of Primary Care Health Sciences English, Mike; University of Oxford, Nuffield Department of Medicine and Department of Paediatrics
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Infectious diseases, Paediatrics
Keywords:	pneumonia, missing data, propensity scores, comparative effectiveness

Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in Kenya: A retrospective observational study

Lucas Malla[1*] (lmalla@kemri-wellcome.org)

Rafael Perera-Salazar[2] (rafael.perera@phc.ox.ac.uk)

Emily McFadden[2] (emily.mcfadden@phc.ox.ac.uk)

Mike English[1, 3] (menglish@kemri-wellcome.org)

[1]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom.

[3] Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.

***Corresponding author:**

Lucas Malla

Abstract

Objectives

Kenyan guidelines for antibiotic treatment of pneumonia recommended treatment of pneumonia characterised by indrawing with injectable penicillin alone in inpatient settings until early 2016. At this point, they were revised becoming consistent with WHO guidance after results of a Kenyan trial provided further evidence of equivalence of oral amoxicillin and injectable penicillin. This change also made possible use of oral amoxicillin for outpatient treatment in this patient group. However, given non-trivial mortality in Kenyan children with indrawing pneumonia it remained possible they would benefit from a broader spectrum antibiotic regimen. Therefore, we compared the effectiveness of injectable penicillin monotherapy with a regimen combining penicillin with gentamicin.

Setting

We used a large routine observational dataset that captures data on all admissions to 13 Kenyan county hospitals.

Participants and measures

The analyses included children aged 2 – 59 months. Selection of study population was based on inclusion criteria typical of a prospective trial, primary analysis (experiment 1, n = 4002), but we also explored more pragmatic inclusion criteria (experiment 2, n = 6420) as part of a secondary analysis. To overcome the challenges associated with the non – random allocation of treatments and missing data, we used propensity score(PS) methods and multiple imputation to minimize bias. Further, we estimated mortality risk ratios using log binomial regression and conducted sensitivity analyses using an instrumental variable and PS trimming.

Results

The estimated risk of dying, in experiment 1, in those receiving penicillin plus gentamicin was 1.46 [0.85, 2.43] compared to the penicillin monotherapy group. In experiment 2, the estimated risk was 1.04 [0.76, 1.40].

Conclusion

There is no statistical difference in the treatment of indrawing pneumonia with either penicillin or penicillin plus gentamicin. By extension it is unlikely that treatment with penicillin plus gentamicin would offer an advantage to treatment with oral amoxicillin.

Strength

- This study provides a platform to explore effectiveness of alternative treatments in routine care in a low income setting to improve health outcomes for children.

Limitations

- The analysis is limited to the variables in the observational dataset – and therefore risk bias due to unmeasured key variables.
- The influence of any resulting bias, to alter results, has however been assessed through the use of alternative methods as instrumental variables.

Introduction

World Health Organisation (WHO) recommendations guide treatment for millions of children with pneumonia every year across low and middle income countries (1). These guidelines are largely based on moderate certainty in evidence of effects (2-5). However, trials supporting recommendations for hospitalized children have included fewer participants from Africa than other settings (6) and it is suggested that African children with pneumonia have higher mortality (7). Additionally, trial populations may not always include the heterogeneous populations presenting for care, many of whom at hospital level may have co-morbidity (8). Thus despite improving access to recommended treatments and deployment of childhood vaccines at high coverage, including those against *H. influenzae* Type B and pneumococcus, clinically diagnosed pneumonia remains one of the top causes of mortality for children under five in Kenya and other countries (7). According to the mortality data derived from the Global Health Observatory (GHO) Data – published in the WHO website (9), pneumonia caused about 5.4 under five deaths per 1000 children in 2015 (which was the highest compared to diarrhoea/dehydration and malaria which are the other top causes of under-five mortality in Kenya). The comparison of mortality rates between 2000 and 2015 for pneumonia, diarrhoea/dehydration and malaria is presented in the additional file 1: supplementary data

figure A. The basic and pneumococcal vaccine coverage by 2014 for children aged 12 – 23 months in Kenya was at least 80% (10).

In a recent change to guidance it is now recommended that pneumonia characterized by lower chest wall indrawing be treated in outpatient settings with oral medication (Box 1) (11, 12). Yet it remains associated with non-trivial mortality that may be higher outside trial populations (13). Residual mortality may be associated with causes that are not prevented by

currently available conjugate vaccines and organisms, which are not susceptible to the antibiotics currently recommended. Establishing whether there are benefits of alternative treatment regimens to help reduce mortality would ideally require large, pragmatic clinical trials (14, 15). However, these remain relatively expensive and time consuming. Observational data may support comparative effectiveness analyses of alternative treatments, may be cheaper and quicker, and may enable evaluation of interventions for which randomization is difficult (16). We use observational data from Kenya to address an important contemporary question for the treatment of pneumonia, a comparison of the effectiveness of gentamicin plus penicillin versus penicillin alone for the treatment of indrawing pneumonia in routine settings. The only previous clinical trial comparing these treatments was a small study of 40 patients in Malaysia (17). In so doing we examine the potential of using data collected by providers as part of their routine practice for comparative effectiveness research in an African setting.

Methods

Clinical definitions of pneumonia, primary and secondary analyses.

The WHO and Kenyan pneumonia treatment guidelines are implicitly based on risk stratification of illness with children deemed at higher risk of mortality offered broader spectrum antibiotic regimens and those at lower risk narrower spectrum antibiotics (11, 18-20). We present three categories of clinically diagnosed pneumonia in Box 1. This categorization outlines previous and recently revised WHO and Kenyan pneumonia treatment guidelines (11, 19). What we refer to as indrawing pneumonia may be associated with low but clinically significant mortality rates (13, 21). Prior to March 2016 recommended treatment for this group was penicillin monotherapy and our aim is to examine whether there is any advantage of broader spectrum antibiotics in this group. Since March 2016 new guidelines recommend outpatient treatment with oral amoxicillin for this group on the basis

of trials suggesting equivalence of amoxicillin and penicillin. However, as indicated above very few patients had been included in studies comparing narrow (amoxicillin or penicillin) and broader spectrum antibiotic regimens. As indicated above, beyond the confines of clinical trials amongst all children being treated for indrawing pneumonia, clinical outcomes (including mortality) are worse than seen in the trials (7) and clinicians are often choosing not to use a single drug regime and are in fact often opting to use the combination of gentamicin and penicillin in the group meeting criteria for indrawing pneumonia in real life settings (22). As mortality is higher in real life settings than in trials and as the possibility that broad spectrum antibiotics could have an advantage over monotherapy with penicillin (or amoxicillin) has not been explored in Kenya's previous trials, we feel that examining whether broad spectrum antibiotics confer an advantage is an important question.

Box 1: Clinical Pneumonia Classifications and Treatments in use in Kenya

1. ***Severe pneumonia:*** *If a child has either oxygen saturation less than 90% or central cyanosis or is grunting or unable to drink or not alert, then s/he is classified as having severe pneumonia and is put on oxygen and treated with a combination of gentamicin and penicillin.*

   The previous WHO (23) and pre-2016 Kenyan guidelines (20) named this class as "very severe pneumonia".
2. ***Indrawing pneumonia:*** *If a child has lower chest wall indrawing (but does not have any of the qualifying signs for severe pneumonia above) and is alert then s/he is classified as having indrawing pneumonia.*

   In previous WHO (23) and pre-2016 Kenyan guidelines (20) guidelines, this class was named as "severe pneumonia" and treatment recommended was inpatient penicillin monotherapy. Our analyses are based on data from the period before March 2016 when inpatient penicillin monotherapy was recommended for this population.

   Since March 2016 in Kenya, and reflecting updated WHO guidance and results of a local trial (24), it has been recommended that this group be treated in outpatient settings with oral amoxicillin as part of an expanded group of non-severe pneumonia.

   Note: *The term indrawing pneumonia is hereafter used in this analysis to define this category of children to avoid confusion.*
3. ***Non – severe pneumonia:*** *If a child has none of the clinical signs in the 2 categories above but has cough or difficulty breathing and a respiratory rate greater than or equal to 50 breaths/minute (for age between 2 and 11 months) or respiratory rate greater than or equal to 40 breaths/minute (for age above 12 months) then Kenyan guidelines in the period pre and post March 2016 recommend s/he is classified as having non severe pneumonia and treated with oral amoxicillin as an outpatient.*

The ability to use routine data to compare treatment effects requires that patients with similar problems receive different treatments. Previous studies conducted in Kenya and elsewhere have indicated that clinicians often do not follow guideline recommendations in treating pneumonia (22). Variation from the guideline recommended approach can occur at the point of pneumonia severity assignment (clinicians do not follow a nationally approved protocol linking clinical signs and severity category outlined in Box 1) and at the point of treatment assignment (clinicians do not follow this protocol that links treatment and severity). This variability in adherence to protocols provides the opportunity for comparative effectiveness evaluation. More specifically, the adherence and non – adherence to treatment protocols by clinicians allows us to classify indrawing pneumonia admissions in two ways:

1) Those with clinical signs placing them in the group of indrawing pneumonia irrespective of the category or classification assigned to the child by the clinician.

2) Those given a clinician classification of indrawing pneumonia irrespective of the actual clinical signs observed by the clinician.

Based on these two possibilities two experiments were designed (see additional file 2: analysis protocol (25)) with specific objectives as follows[1]:

1) **Experiment 1:** To compare effectiveness of injectable penicillin versus penicillin plus gentamicin (both injectable) in treatment of indrawing pneumonia; where the child is identified as belonging to a population of children with indrawing pneumonia on the basis of data on their recorded clinical signs. The Experiment 1 population of

[1] All children with danger signs were excluded from experiment 1 and in general (both in experiments 1 and 2), children with the following comorbidities were excluded: HIV, meningitis, tuberculosis and or acute severe malnutrition.

indrawing pneumonia is therefore consistent with pre-2016 clinical guideline recommendations.

2) **Experiment 2:** To compare effectiveness of injectable penicillin versus penicillin plus gentamicin in a population in which we use the clinician assigned categorisation of indrawing pneumonia, which may not be consistent with clinical guideline recommendations.

We defined Experiment 1 as our primary analysis as we propose it would identify a population similar to that recruited to a randomised trial where the inclusion criteria would be based on specified clinical signs. Experiment 2 offers a scenario that may represent a more pragmatic study design with inclusion criteria based around a clinician led classification.

Data source

We use data from the Kenyan Clinical Information Network (CIN) that was initiated to improve inpatient paediatric data availability from county (formerly district) hospitals. Thirteen county referral hospitals were purposively selected with direction from Ministry of Health (MOH) and recruited into the CIN. These hospitals were recruited into the study at different times; four in September 2013, five in October 2013 and four in February 2014. This analysis utilises data up to March 2016. On average, 25 000 paediatric admissions are captured per year. These hospitals typically have one paediatrician leading services predominantly provided by junior clinical teams. Data systems and standardised clinical forms were specifically implemented in all hospitals at the start of this work to optimise the quality of routine data. Patient data in these hospitals are collected post discharge by trained data clerks guided by well-defined standard operating procedures, under supervision by the hospital medical records department and the research team. Clinicians admitting patients fill standardized Paediatric Admission Record (PAR) forms (26) that have been shown to improve documentation of clinical symptoms and signs (27). Together with discharge forms,

treatment sheets and laboratory reports these are all part of the patient files that are the primary data source. This data collection system has been described in detail elsewhere (28). Feedback to hospitals as part of the CIN activities has helped improve the quality of clinical data (28). The description of hospital selection and their populations of patients is detailed elsewhere (29).

Statistical analysis

i) Defining per protocol and intention to treat populations

In typical randomised controlled trials, types of analyses to be conducted are defined beforehand – and this involves defining the type of patient populations that are included in the analyses. Intention to treat and per protocol populations derived from observational datasets have been described in Danaei (2013) (30). We defined per protocol and intention to treat populations based on the dates actual treatments were recorded as prescribed for patients included in our primary and secondary analyses (experiments 1 and 2 respectively). Within each experiment, and after applying inclusion and exclusion criteria, we define the per protocol population as those whose prescription of one of the two study regimens did not change during the admission. The intention to treat population is defined by the original treatment assignment and included children in whom treatment was subsequently changed (see Figure 1 in the Results section).

ii) Dealing with missing data and propensity score matching

As CIN comprises data from routine care settings it faces challenges of non – random treatment allocation and missing data. The missing data and propensity score methods for this analysis have been detailed in the additional file 2: analysis protocol linked to this work (25). In brief, after exploring the patient populations, 20 datasets[2] (31) were derived using multiple

[2] The current literature (31) recommends the use of more than 5 imputed datasets and therefore 20 should be sufficient.

imputation (with chained equations) for each experiment (all the variables in both the experiments had missing data less than 30% – see table A in the additional file 1: supplementary data). Clinical signs and symptoms data considered were those recorded by clinicians before patients were admitted. The multiple imputation excluded outcome data as guidance on the use of observational datasets for comparative effectiveness analysis recommends exclusion of outcome data in the design phase (32). Following this, those with missing outcome data were excluded from the analysis (missingness in the outcome data were 0.5% and 0.8% for experiments 1 and 2). For each imputed dataset, patients in the alternative treatment groups (penicillin monotherapy versus penicillin plus gentamicin) were then matched using propensity score (PS) methods to overcome non – random treatment allocation. Propensity scores define the probability of belonging to or being assigned a given treatment based on signs and symptoms (33). PS is a distance measure (34) which is used as a means to overcome allocation bias as treatment outcomes in children with similar propensity scores can then be compared. In these analyses we compared three approaches to reducing possible bias based on PS – optimal full matching, weighting and sub-classification (33, 34). All are aimed at creating groups of patients that are comparable in terms of the distribution of observed signs and symptoms. For each experiment, in order to select the optimum PS implementation method, absolute standardised mean differences (ASMD) were used as diagnostic checks for covariate balance and overlap (35, 36) between the alternative treatment groups. PS methods that resulted in the minimum average absolute standardised mean differences for the majority of the variables while retaining the largest number of patients in the analysis were considered the most appropriate (34).

iii) Analytic modelling and sensitivity analyses

In sample size calculations conducted prior to the experiments (presented in greater detail elsewhere (see additional file 2: analysis protocol)), it was estimated that a sample size of at

least 4000 would be sufficient for the planned experiments to detect a minimum difference of 1.5% in mortality between the two treatment groups. The sample size for experiment one was 4002 and experiment two 6420 (including 3312 of those that were also in experiment 1). In other words, experiment 2 largely included those in the experiment 1 population but also children not meeting eligibility criteria for experiment 1. For each of the experiments, after multiple imputation, multivariable log-binomial regression models were fitted to PS weighted datasets and adjusting for all the variables also used in the PS models (also as a form of sensitivity analyses, treatment effects were estimated on PS unweighted datasets). Only pooled treatment effect estimates are reported.

One possibility is that clinicians' treatment assignment is skewed such that patients who appear sicker (having a greater number of clinical signs of more severe illness) are assigned 'stronger' or broad spectrum treatment. In this situation as mentioned by Stürmer (2010), specific types of treatment allocation may be more likely associated with increased mortality (37). In theory, the use of propensity scores is supposed to account for such skewed assignment by comparing only outcomes of those with similar propensity scores assumed to suggest they have similar clinical profiles and thus similar risks. PS trimming attempts to tackle this problem further by excluding patients who are at the extremes of the PS distribution to create a population with clinical characteristics that are as homogeneous as possible. We use PS trimming to define a population between the 5% - 95% PS percentiles in a sensitivity analysis.

In a further sensitivity analysis, we used an instrumental variable to examine the potential influence of any unmeasured variables (38). An instrumental variable method aims to find a proxy randomised experiment in a routine or observational dataset (39). We used weekend/weekday admission as an instrumental variable as it was demonstrated in a study conducted by Berkley (2004) (40) in a Kenyan hospital that children who were admitted

during the weekend experienced higher mortality compared to those admitted during the weekdays. This, in theory, implies that the type of treatment and care received depends on the day of admission – and this later determines the type of health outcome of the patient. The process of fitting the instrumental variable models has been described in the additional file 1: supplementary data. The two sensitivity approaches described above were done for both primary and secondary analyses.

Results

a) Creating per protocol and intention to treat populations

Examining the dates treatments were given, five treatment arms (per experimental scenario) were defined – specifically those who received: (1) penicillin alone without changes, (2) a combination of penicillin plus gentamicin without changes, (3) penicillin but switched to a combination of penicillin plus gentamicin, (4) penicillin but switched to ceftriaxone, and (5) a combination of penicillin plus gentamicin but switched to ceftriaxone (ceftriaxone is the recommended second line treatment for severe pneumonia). Therefore, per protocol analyses would compare patients in treatment arm 1 versus 2, while intention to treat analyses would compare patients in treatment arms 1, 3, and 4 versus 2 and 5 (figure 1).

[Insert figure 1]

Figure 1: Summary of patients per treatment arm in experiments 1 – 2

In this analysis, intention to treat populations were considered primary and are reported in experiments 1 and 2 in keeping with clinical trial reporting guidelines. These analyses include a relatively larger number of patients compared to per protocol analyses. The recommended doses of penicillin and gentamicin in these hospitals are 50000 iu/Kg and 7.5 mg/Kg given 4 and once daily respectively. Additional data suggest most clinicians prescribed these doses correctly (see table B in the additional file 1: supplementary data).

b) Comparing performance of optimal full matching, weighting and PS sub-classification in experiments 1 and 2 respectively

For each experiment, the three PS implementation methods were compared to determine the one which would result in the least absolute standardised mean differences for most of the variables in the analysis (even though all the three methods resulted in variables with ASMD<=10%). For experiment 1, PS weighting performed better than PS optimal full matching and sub-classification and for experiment 2, the performance of weighting was comparable to that of optimal full matching (see figures 2 and 3). In both experiments, PS sub-classification reduced covariate imbalance the least. Thus, in the subsequent sections, outcome analyses are based on PS weighted datasets for both experiments.

[Insert Figure 2]

Figure 2: Comparing performance of the three PS implementation methods in experiment 1:

The y – axis contains all the variables used in the PS models. While x – axis shows absolute standardised mean difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value of <= 10% indicates the method has performed well in creating comparable groups.

[Insert figure 3]

Figure 3: Comparing performance of the three PS implementation methods in experiment 2

c) Outcome Analysis Results

i) Exploring mortality in raw datasets

Examining the raw datasets without PS adjustments in experiment 1, the average number of pneumonia deaths (across the 20 imputed datasets) in penicillin plus gentamicin group was 33/1363 (2.42%) and in penicillin monotherapy was 26/2639 (0.99%). And for experiment 2,

the average number of deaths were 87/2296 (3.79%) and 50/4124 (1.21%) in penicillin plus gentamicin and penicillin monotherapy groups respectively. Overall, the average number of pneumonia deaths in the penicillin plus gentamicin group was approximately two and a half to three times the number of mortality events in the penicillin monotherapy group in experiments 1 and 2 respectively.

ii) Modelling mortality risk ratios

The analysis considered penicillin monotherapy as the reference group and mortality as the outcome – and therefore a risk ratio (RR) greater than one would be interpreted to favour penicillin over penicillin plus gentamicin. For both experiments, the treatment risk ratios estimated on the unmatched datasets were larger than the RR estimated on datasets obtained through PS weighting (see table 1 for all results). In experiment 2, the PS unadjusted analysis showed that penicillin monotherapy was significantly more effective than penicillin plus gentamicin (1.68 [1.15, 2.36]). However, the PS weighted effect estimate (1.04 [0.76, 1.40]) was much reduced and suggested that use of PS had corrected (to a degree) for allocation bias indicating that there was no statistical difference in mortality outcomes between penicillin plus gentamicin and penicillin monotherapy treatments. We also observed that the adjusted point estimate for any effect difference in experiment 2 (1.04 [0.76, 1.40]) was less than that in experiment 1 (1.46 [0.85, 2.43]). This may be due to an increase in the number of covariables available for PS weighting that could be used in Experiment 2 resulting in closer matching (see table C in the additional file 1: supplementary data).

d) Sensitivity analysis through trimming using 5 – 95% PS population restriction

After excluding 10% of the populations as a result of PS trimming in sensitivity analyses for experiments 1 and 2, the resulting sample sizes were 3583 and 5778. The skewed assignment of children to treatment with gentamicin and penicillin is demonstrated by their higher PS scores in figure 4 for experiment 1 (and figure B for experiment 2 in the additional file 1: supplementary data). As higher PS scores are associated with the presence of a greater number of clinical signs of illness this also suggests an association between more severe illness and treatment with gentamicin and penicillin. For experiment 1, the estimated average mortality events (on PS unadjusted datasets) were 26/1201 (2.16%) and 24/2382 (1.01%) for penicillin plus gentamicin and penicillin monotherapy groups. While the estimated events in experiment 2 were 62/2026 (3.06%) and 46/3752 (1.22%). Thus in sensitivity analyses for both experiments, trimming excluded more mortality events in the penicillin plus gentamicin group compared with the penicillin monotherapy group. The treatment effects estimated using PS weighted models for the restricted populations as a result of PS trimming showed no statistical difference between the two treatments (table 1).

[Insert figure 4]

Figure 4: Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those who were given penicillin plus gentamicin. While the continuous black line shows the PS distribution for those who received penicillin alone.

e) Sensitivity Analysis through the use of weekend/weekday as an instrumental variable

In order to assess whether a timing of admission variable would form a natural and random experiment, the distributions of covariates were examined across the levels of the instrumental variable (weekend/weekday) in experiments 1 and 2. The distribution of each of

the patient characteristics between weekend and weekday admissions was approximately similar (table D in the additional file 1: supplementary data) suggesting that weekend/weekday admission satisfactorily satisfies one of the criteria as a valid IV (also see the additional file 1: supplementary data for the set of criteria for a valid IV). The weekend mortalities, in the raw datasets, seemed to be higher than weekday mortalities (Table E – additional file 1: supplementary data).

The estimated treatment effects, both in experiments 1 and 2, suggest there is no statistical difference in treating indrawing pneumonia with either penicillin alone or penicillin plus gentamicin. The effect estimates obtained using our IV in both experiments are less than one as compared to those obtained with PS weighting which are greater than one. Biologically, the effectiveness of gentamicin plus penicillin (when administered in correct doses) is expected to be the same or greater than that of penicillin monotherapy. Based on the magnitude and direction of effects, the use of the IV seems to demonstrate that the effects obtained through PS weighting may have had some residual bias. However, it is important to highlight that for all analyses the 95% CI obtained are consistent with the Null Hypothesis of no different effect for the treatments.

Table 1: Treatment effect estimates (RR (95% C.I))

	Experiment 1	**Experiment 2**
Regression without PS adjustment	1.75 [0.94, 2.77]	1.68 [1.15, 2.36]
PS Weighting	1.46 [0.85, 2.43]	1.04 [0.76, 1.40]
PS trimming (5% – 95% restriction)	1.39 [0.74, 2.15]	1.05 [0.74, 1.41]
Instrumental variable	0.91 [0.41, 2.20]	0.44 [0.34, 1.32]

Discussion

We compared penicillin alone with penicillin plus gentamicin in treatment of indrawing pneumonia in populations with overall mortality of 1.5% and 2% in experiments 1 and 2 respectively. There were more fatal events in the penicillin plus gentamicin group than the

penicillin group (approximately 2.5 times) and unadjusted analyses pointed, therefore, to a protective effect of penicillin treatment. However, adjusted analyses, both in experiments 1 and 2, that aim to account for allocation bias using PS weighting that can result from non-random treatment allocation suggest that there is no appreciable difference in outcomes between penicillin and gentamicin plus penicillin treatment of indrawing pneumonia. In addition, we conducted analyses using alternative PS methods – sub-classification (results presented in additional file 1: supplementary data – figures C and D, and table F) and optimal full matching (results presented in additional file 1: supplementary data – table G) and analyses of both intention to treat and per protocol populations. All analyses showed similar findings (see the provided additional file 1: supplementary data – table H). We undertook two formal approaches to sensitivity analysis. First, we employed PS trimming to exclude 10% of the analysis populations in experiments 1 and 2. Effect estimates in this case are based on analyses of 90% of cases that PS suggest are best matched. Second, we used an instrumental variable. These techniques employ different approaches to account for possible confounding that might contribute to estimated treatment effects. Both these forms of analysis provided results that support the suggestion that poor outcome in this population is not associated with the antibiotic regimen received.

Our analyses were conducted using data from over 4,000 children, one hundred times more participants than were included in the only prior randomised controlled trial of penicillin monotherapy and penicillin plus gentamicin in treatment of pneumonia in an Asian population (17). There are continuing concerns of clinically important mortality in children with indrawing pneumonia in Africa (21). This has led to hesitation to adopt new WHO and Kenyan guidelines that now recommend the treatment of indrawing pneumonia as an outpatient using amoxicillin (11, 19). Our results suggest that there are likely to be two distinct issues. Firstly, they suggest that offering broader spectrum injectable antibiotic

treatment to children with indrawing pneumonia may not improve outcomes compared to treatment with penicillin monotherapy. As other studies have suggested equivalence between oral (high dose) amoxicillin therapy and injectable penicillin therapy (2-5, 24) it seems likely therefore that oral amoxicillin and penicillin plus gentamicin combination therapy would result in similar outcomes when used to treat indrawing pneumonia. Clinicians should therefore carefully adhere to guidelines for treatment of indrawing pneumonia and avoid using gentamicin helping to prevent any possible toxicity.

Secondly, however, our results suggest that children fulfilling a definition of indrawing pneumonia based on clinical signs, and having excluded serious co-morbidities, may still have an appreciable risk of mortality irrespective of their antibiotic treatment (1.5% in all children in experiment 1). When clinicians categorise children with indrawing pneumonia and imperfectly adhere to clinical sign based guidance mortality tends to be higher (2% in all children in experiment 2). These findings point to as yet uncharacterised risk factors that could be important in determining which children need admission to hospital as current guidance indicates that all those with indrawing pneumonia can be treated as an outpatient. While offering an alternative antibiotic to amoxicillin to this group may not improve outcomes it is possible that closer and continuing observation in hospital may help identify co-morbid or alternative conditions that are contributing to this mortality and that may be treated.

The trials that informed the basis for the revised WHO guidelines (2-5) showed extremely low mortality (0 – 0.2%) suggesting that the populations included in such trials may not be directly representative of all those to whom guidelines are applied in routine settings. In the trial by Agweyu (2015) conducted in Kenya (which compared penicillin versus oral amoxicillin for indrawing pneumonia) overall mortality was 0.8% (24). In a parallel observational cohort providing data from the same hospitals over the same time period for

children treated with penicillin alone but not included in the Kenyan trial mortality was not significantly different but marginally higher at 1.2% (Agweyu (2017), submitted) perhaps suggesting that even the limited exclusion criteria in this pragmatic trial might result in exclusion of some sicker children. Taken together with data from the analyses presented here it does appear there is a need to explore whether guidelines might be modified to accommodate additional clinical risk factors for possible life-threatening illness that should prompt admission. In a population with high coverage with conjugate vaccines this may more usefully be for more rigorous evaluation to identify alternative diagnoses or for improved supportive care than for different antibiotics.

Strengths and limitations

Conducting comparative effectiveness analyses using observational datasets can offer the advantage of larger sample sizes at lower cost than randomised controlled clinical trials. They also include patients that may not qualify for enrolment in a typical explanatory randomised controlled trial – and therefore perhaps provide more true to life estimates of treatment effects similar to those observed in highly pragmatic trials (15). However, as most observational datasets are not meant for research, they have challenges of non-random treatment allocation and missing data. We employed a rigorous 'experimental design' strategy as is recommended when using observational data (32). We used PS and multiple imputation methods in an effort to minimise bias due to non-random treatment allocation and missing data and analyses suggested no appreciable difference in outcomes of indrawing pneumonia treated with penicillin alone compared with penicillin plus gentamicin. This was in contrast to unadjusted regression analyses that pointed towards better outcomes with penicillin alone suggesting presence of allocation bias. As most observational datasets are limited to observed variables, it is important to conduct sensitivity analysis to explore if the estimated effects are potentially sensitive to unmeasured variables. We used an instrumental variable and PS trimming, both

supported the idea of no appreciable difference regimens when treating indrawing pneumonia. While there are differences (in terms of magnitude) in the mortality observed in the different groups that suggest some residual bias in treatment allocation, these mortality differences are no greater than might occur by chance after PS adjustment (with the type 1 and 2 errors specified in the additional file 2: analysis protocol). In that sense the PS approach may still have limitations but it does allow us to conclude no statistical difference in mortality outcomes between the two treatment arms.

The WHO recommended guidelines for treating pneumonia have considerable influence on policy and practice in low and middle income countries. While the evidence base and rigour of guideline development have improved considerably there remain few data on their effectiveness when implemented in non-trial settings. Even though well-designed, large pragmatic trials would be preferred, we demonstrate that carefully collected routine data may be useful for assessing the effectiveness of alternative treatments (15). Such analyses may become increasingly possible as electronic medical records are deployed in low and middle income countries (41) but it is important that such studies are carefully designed to limit as far as possible the biases that arise from non-random treatment allocation (32). Our results suggest that children with indrawing pneumonia may gain little benefit from treatment with broader spectrum antibiotic regimens. However, they also suggest that further work is needed to identify those who are at higher risk of death who might be prioritised for an inpatient diagnostic work up and improved supportive care rather than treated as outpatients.

Additional files

Additional file 1: supplementary data

Additional file 2: analysis protocol

Additional file 3: STROBE checklist

Additional file 4: Main manuscript with changes highlighted

Acknowledgement

We would like to thank Ambrose Agweyu for his comments that helped improve this manuscript. We also appreciate the valuable contribution offered by the CIN team: Lydia Thuranira & Grace Ochieng' (Kiambu County Hospital), Barnabas Kigen (Busia County Hospital), Melab Musabi & Rachel Inginia (Kitale County Hospital), Anne Kamunya & Sam Otido (Embu County Hospital), Margaret Kuria (Kisumu East County Hospital), Agnes Mithamo & Francis Kanyingi (Nyeri County Hospital), Celia Muturi, Caren Emadau & Cecilia Mutiso (Mama Lucy Kibaki County Hospital), David Kimutai & Loice Mutai (Mbagathi County Hospital), Nick Aduro (Kakamega County Hospital), Samuel Ng'arng'ar (Vihiga County Hospital). Fred Were & David Githanga (Kenya Paediatric Association). Rachel Nyamai (Ministry of Health).

Funding

We are grateful for the funds from the Wellcome Trust (#097170) that support ME through a fellowship and additional funds from a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (#092654) that supported this work. LM is supported by a Nuffield Department of Medicine Prize DPhil Studentship and Clarendon Scholarship (Oxford University). The funders had no role in drafting or submitting this manuscript.

Availability of Data

The hospital specific datasets are in custody of the hospitals participating in CIN (these datasets have been de-identified).

Authors Contributions

The contributions of the authors were as follows: LM did an initial draft of this manuscript with the support of RP, EM and ME. Thereafter, all authors edited subsequent versions and approved the final copy.

Competing Interests

The authors declare they have no competing interests.

Ethics approval

This analysis is based on a larger project (CIN) which was cleared by the Kenya Medical Research Institute ethics and review board (Protocol number: 2465).

References

1. Olayemi E, Asare EV, Benneh-Akwasi Kuma AA. Guidelines in lower-middle income countries. Br J Haematol. 2017;177(6):846-54.

2. Atkinson M, Lakhanpaul M, Smyth A, Vyas H, Weston V, Sithole J, et al. Comparison of oral amoxicillin and intravenous benzyl penicillin for community acquired pneumonia in children (PIVOT trial): a multicentre pragmatic randomised controlled equivalence trial. Thorax. 2007;61(12):1102-6.

3. Hazir T, Fox LM, Nisar YB, Fox MP, Ashraf YP, MacLeod WB, et al. Ambulatory short-course high-dose oral amoxicillin for treatment of severe pneumonia in children: a randomised equivalency trial. Lancet. 2008;371:49-56.

4. Addo-Yobo E, Chisaka N, Hassan M, Hibberd P, Lozano JM, Jeena P, et al. Oral amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59 months: a randomised multicentre equivalency study. Lancet. 2004;364(9440):1141-8.

5. Addo-Yobo E, Anh DD, El-Sayed HF, Fox LM, Fox MP, MacLeod W, et al. Outpatient treatment of children with severe pneumonia with oral amoxicillin in four countries: the MASS study . Tropical Medicine and International Health. 2011;16(8):995-1006.

6. Mulholland K, Carlin JB, Duke T, Weber M. Challenges of trials of antibiotics for pneumonia in low-income countries. The Lancet Respiratory Medicine. 2014;2(12):952-4.

7. Pneumonia and Diarrhoea Progress Report 2014 [Internet].: John Hopkins Bloomberg School of Public Health; 2014 []. Available from: http://www.jhsph.edu/research/centers-and-institutes/ivac/resources/IVAC-2014-Pneumonia-Diarrhea-Progress-Report.pdf;.

8. Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, et al. Bacteremia among Children Admitted to a Rural Hospital in Kenya. The new england journal of medicine. 2005;352(1):39-47.

9. Global Health Observatory data repository [Internet].; 2016 []. Available from: http://apps.who.int/gho/data/node.main.686?lang=en.

10. Kenya Demographic and Health Survey [Internet].; 2014 []. Available from: http://dhsprogram.com/pubs/pdf/FR308/FR308.pdf.

11. Basic paediatric protocols [Internet].: Kenyan Ministry of Health; 2016 []. Available from: https://www.tropicalmedicine.ox.ac.uk/_asset/file/basic-paediatric-protocols-2016.pdf.

12. Hospital Care for Children [Internet].; 2013 []. Available from: http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf.

13. UNICEF DATA: Monioring the situation of women and children [Internet].; 2017 []. Available from: https://data.unicef.org/topic/child-health/pneumonia/.

14. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. BMJ Open Research Methods. 2014;4(6):e005362.

15. Thadhani R. Formal trials versus observational studies. In: Mehta A, Beck M, Sunder-Plassmann G, editors. Fabry Disease: Perspectives from 5 Years of FOS. Oxford: Oxford PharmaGenesis; 2006. Chapter 14. Available from: https://www.ncbi.nlm.nih.gov/books/NBK11597/.

16. West SG, Duan N, Pequegnat W, Gaist P, Jarlais DCD, Holtgrave D, et al. Alternatives to Randomised Controlled Trials. American Public Health Association. 2008;98(8):1359-66.

17. Hasali MAA, Ibrahim MIM, Sulaiman SAS, Ahmad Z, Hasali JBA. A clinical and economic study of community-acquired pneumonia between single versus combination therapy . Pharm World Sci. 2005;27(3):249-53.

18. Ayieko P, English M. Case Management of Childhood Pneumonia in Developing Countries . Pediatr Infect Dis J. 2007;26:432-40.

19. Hospital Care for Children [Internet].; 2013 []. Available from: http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf.

20. Basic Paediatric Protocols: November 2013 Edition [Internet].: Ministry of Health; 2013 []. Available from: http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition.

21. Koh JW, Wong JJ, Sultana R, Wong PP, Mok YH, Lee JH.
Risk factors for mortality in children with pneumonia admitted to the pediatric intensive care unit. Pediatric pulmonology. 2017;52(8):1076-84.

22. Agweyu A, Kibore M, Digolo L, Kosgei C, Maina V, Mugane S, et al. Prevalence and correlates of treatment failure among Kenyan children hospitalised with severe community-acquired pneumonia: a prospective study of the clinical effectiveness of WHO pneumonia case management guidelines. Tropical Medicine & International Health. 2014;19(11):1310-20.

23. Hospital Care for Children [Internet].; 2005 []. Available from: http://apps.who.int/iris/bitstream/10665/43206/1/9241546700.pdf.

24. Agweyu A, Gathara D, Oliwa J, Muinga N, Edwards T, Allen E, et al. Oral amoxicillin versus benzyl penicillin for severe pneumonia among kenyan children: a pragmatic randomised controlled noninferiority trial. Clin Infect Dis. 2015;60(8):1216-24.

25. Malla L, Perera R, McFadden E, English M.
Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in Kenya: A protocol for an observational study. BMJ Open. 2017;7(9):e016784.

26. Mwakyusa S, Wamae A, Wasunna A, Were F, Esamai F, Ogutu B, et al.
Implementation of a structured paediatric admission record for district hospitals in Kenya – results of a pilot study. BMC International Health and human rights. 2006;6:9.

27. Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A Multifaceted Intervention to Implement Guidelines and Improve Admission Paediatric Care in Kenyan District Hospitals: A Cluster Randomised Trial. PLoS Medicine. 2011;8(4).

28. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation of clinical care within a clinical information network: an essential initial step in efforts to understand and improve care in Kenyan hospitals. BMJ Global Health. 2016;1(1):e000028.

29. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in Kenyan hospitals within a new Clinical Information Network. Archives of Disease in Childhood. 2015;101(3):223-9.

30. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA.
Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. Stat.Methods Med.Res. 2011;22(1):70-96.

31. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.

32. Rubin D.B. For Objective Causal Inference, Design Trumps Analysis. The Annals of Applied Statistics. 2008;2(3):808-40.

33. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011;46(3):399-424.

34. Stuart EA. Matching methods for causal inference:A review and a look forward. Statist. Sci. 2010;25(1):1-21.

35. Austin PC. Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. Pharmacoepidemiol Drug Saf. 2008 2008;17(12):1218-25.

36. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009 10 Nov 2009;28(25):3083-107.

37. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the presence of Unmeasured Confounding: Dealing with observations in the Tails of the Propensity Score Distribution - A Simulation Study. American Journal of Epidemiology. 2010;172(7):843-54.

38. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014 15 Jun 2014;33(13):2297-340.

39. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference. Statistics in Medicine. 2014;33(13):2297-340.

40. Berkley JA, Brent A, Mwangi I, English M, Maitland K, Marsh K, et al. Mortality among Kenyan children admitted to a rural district hospital on weekends as compared with weekdays . Pediatrics. 2004;114:1737-8.

41. English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al. Building Learning Health Systems to Accelerate Research and Improve Outcomes of Clinical Care in Low-and Middle-Income Countries. PLOS Medicine. 2016;13(4):e1001991.

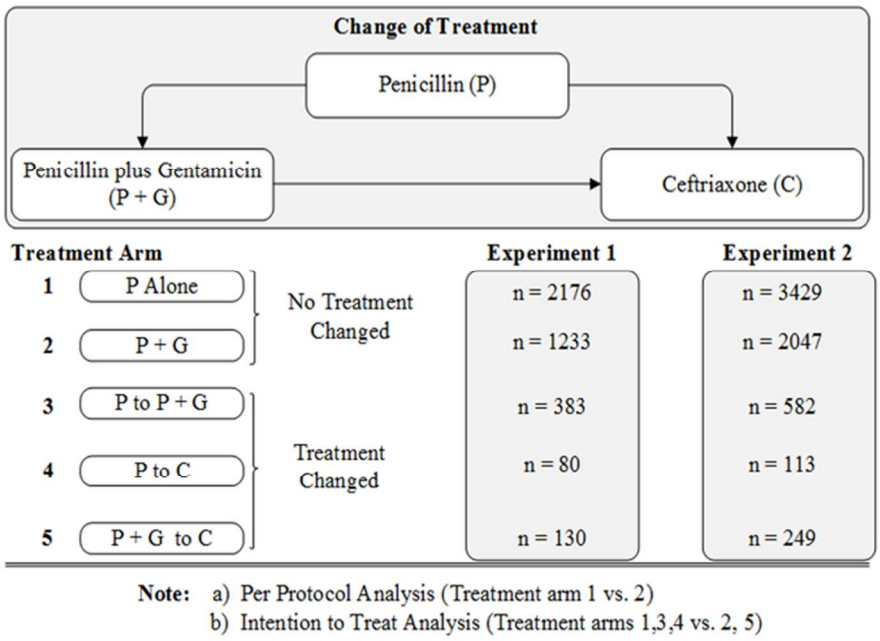


Figure 1: Summary of patients per treatment arm in experiments 1 – 2

63x45mm (600 x 600 DPI)

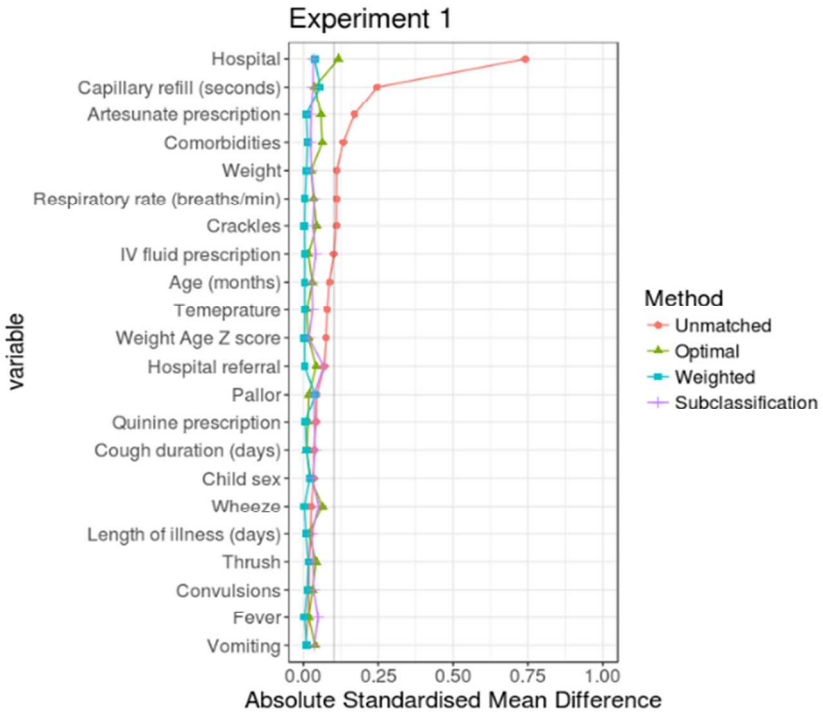


Figure 2: Comparing performance of the three PS implementation methods in experiment 1: The y – axis contains all the variables used in the PS models. While x – axis shows absolute standardised mean difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value of <= 10% indicates the method has performed well in creating comparable groups.

63x45mm (600 x 600 DPI)

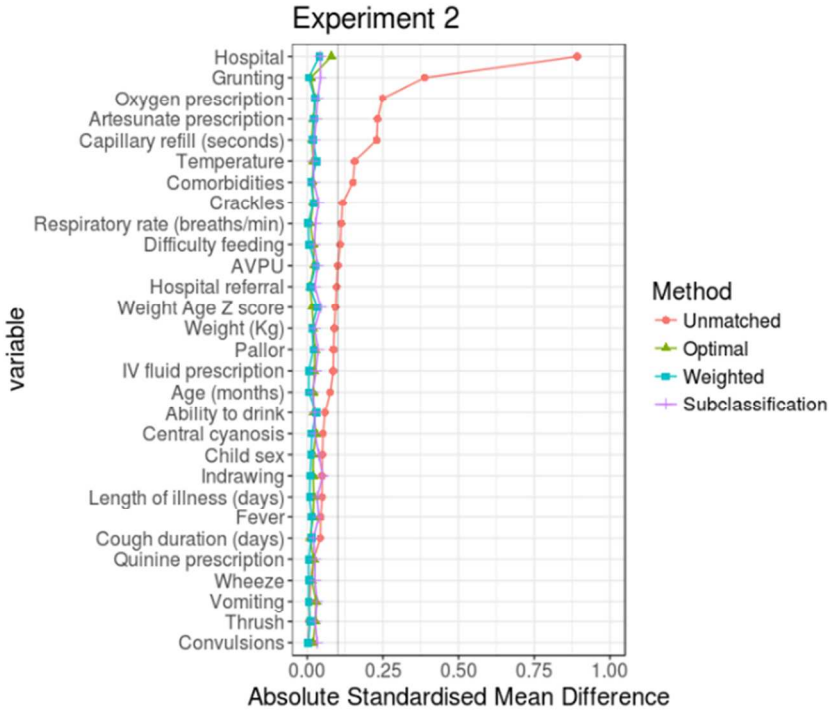


Figure 3: Comparing performance of the three PS implementation methods in experiment 2.

63x45mm (600 x 600 DPI)

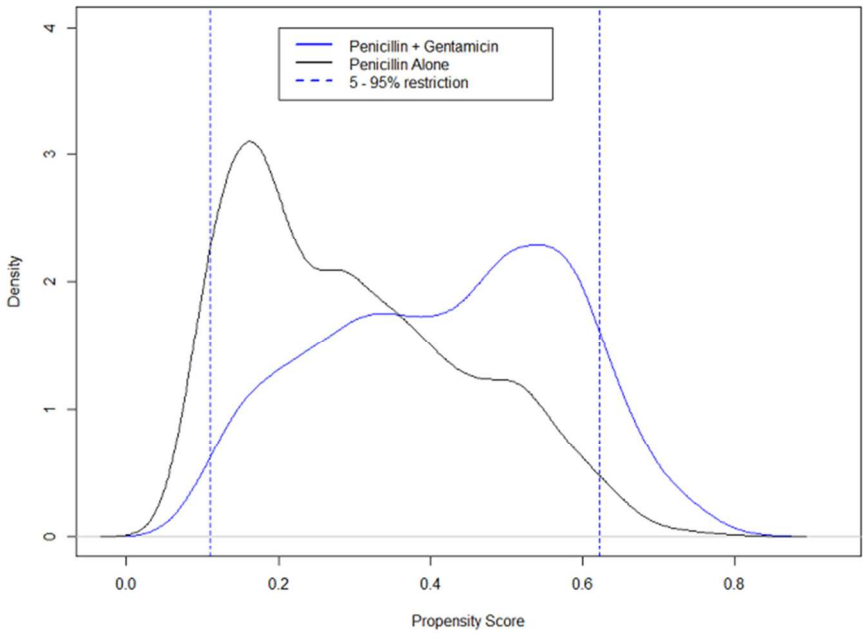


Figure 4: Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those who were given penicillin plus gentamicin. While the continuous black line shows the PS distribution for those who received penicillin alone.

63x45mm (600 x 600 DPI)

Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in Kenya: A retrospective observational study

Lucas Malla[1*]

Rafael Perera-Salazar[2]

Emily McFadden[2]

Mike English[1, 3]

[1]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom.

[3] Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.

The supplementary material is organised into the following subsections:

- Under-five mortality incident rates in Kenya
- Percentage of completeness of variables (experiments 1 and 2)
- Overlap and correctness of penicillin and gentamicin dosing
- Summary of key and auxiliary independent variables
- Trimming in experiment 2 (ITT population)
- Analysis using Instrument variables
- Analysis using PS sub - classification
- Analysis using PS optimal full matching
- Analysis using per protocol population
- Definition of PS methods and how they were used

a) Under-five mortality incident rates in Kenya

The mortality data summarised in figure A were extracted from the Global Health Observatory (GHO) Data – published in the WHO website (1).

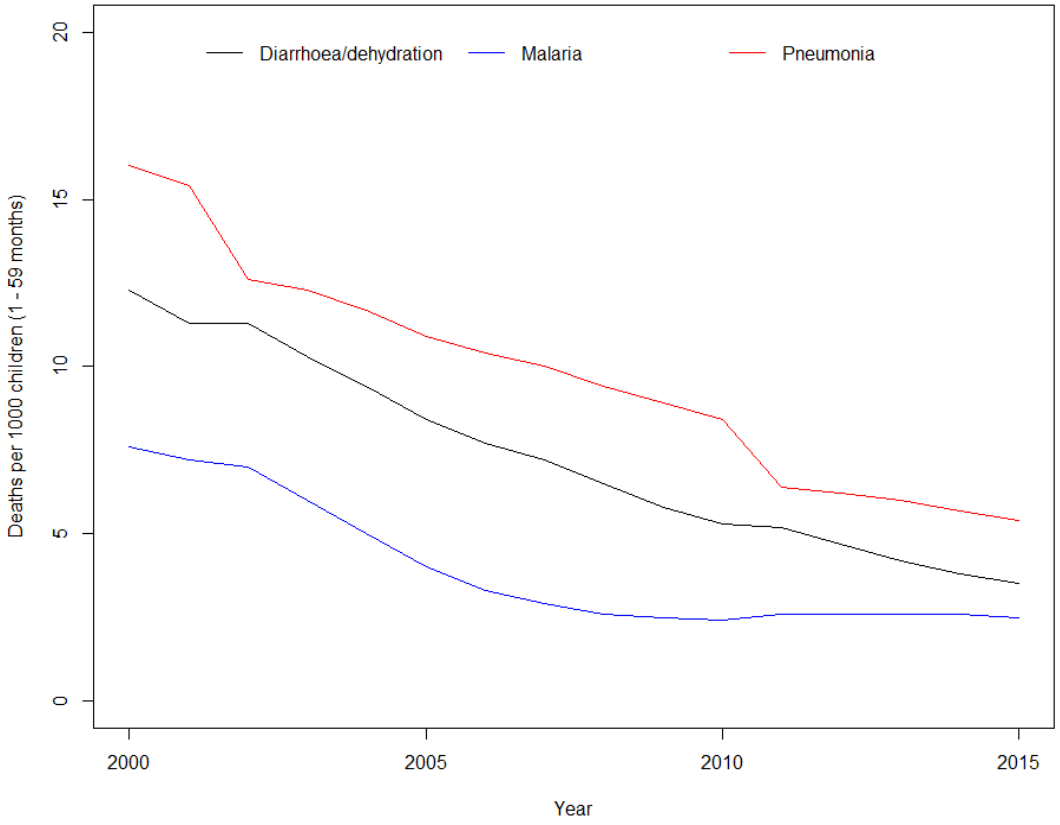


Figure A: Deaths, rate per 1000

b) Percentage of completeness of variables (experiments 1 and 2)

Table A: Percentage of completeness of variables (experiments 1 and 2)

Variable	Experiment 1 (%)	Experiment 2 (%)
Age (2 – 59 months)	99.7	99.5
Indrawing (present/absent)	100.0	96.3
Level of consciousness – AVPU	_	95.5
Central cyanosis	_	95.9
Grunting	_	94.2
Ability to drink	_	91.2
Gender (male/female)	99.6	99.0
Cough duration (days)	84.9	83.4
Crackles (present/absent)	97.4	94.7
Weight (Kg)	96.3	96.0
Pallor (0, +, +++)	96.7	94.5
Capillary refill	83.3	78.0
Fever (present/absent)	98.2	97.6
Temperature	94.1	92.6
Convulsions (present/absent)	96.3	94.3
Vomiting (yes/no)	97.1	95.2
Referral (yes/no)	83.3	73.6
Length of illness (days)	98.4	98.0
Thrush (present/absent)	90.4	83.9
Quinine/artesunate (prescribed/not prescribed)	100.0	100.0
Wheeze (present/absent)	97.1	94.5
Respiratory rate	87.4	85.4
IV fluid prescription	100.0	100.0
Outcome (died/alive)	99.5	99.2

c) Overlap and correctness of penicillin and gentamicin dosing

A total of 3312 patients were both common to experiments 1 and 2. We also examined if the patients received correct dosages of penicillin and gentamicin: For penicillin, a dose of 40,000 – 60,000 I.U/Kg was considered normal and for gentamicin, a dose of 6 – 9 mg/Kg. These were +/- 20% of recommended dosages in the Kenyan paediatric protocols. Majority of the patients were prescribed normal dosages of penicillin and gentamicin (see table B).

Table B: Correctness of penicillin and gentamicin prescription

	Experiment 1		**Experiment 2**	
	Penicillin	Gentamicin	Penicillin	Gentamicin
Under dose	3%	10%	3%	12%
Normal	93%	87%	92%	85%
Over dose	4%	3%	5%	3%

d) Summary of key and auxiliary independent variables

Table C: Summary of key and auxiliary independent variables for experiments 1 and 2[1]

Experiment 1 key variables	**Experiment 2 key variables**	**Auxiliary variables for experiments 1 and 2**
Age (2 – 59 months)	Age (2 – 59 months)	Gender (male/female)
Indrawing (present/absent)	Indrawing (present/absent)	Cough duration (days)
Level of consciousness – AVPU (alert/verbal response/pain response/unresponsive)	History of cough (yes/no)	Crackles (present/absent)
	Difficulty breathing (present/absent)	Weight (Kg)
	Level of consciousness – AVPU	Pallor (0, +, +++)
	Central cyanosis	Capillary refill (immediate, 1 – 2 secs, 3 – 6 sec, > 6 secs)
	Grunting	Fever (present/absent)
	Ability to drink	Convulsions (present/absent)
		Vomiting (yes/no)
		Referral (yes/no)
		Length of illness (days)
		Thrush (present/absent)
		Quinine/artesunate (prescribed/not prescribed)
		Weight for age z – score
		Wheeze (present/absent)
		Comorbidities (Malaria and or diarrhoea)

[1] Comorbidities and WAZ variables were derived after multiple imputation

e) Trimming in experiment 2 (ITT population)

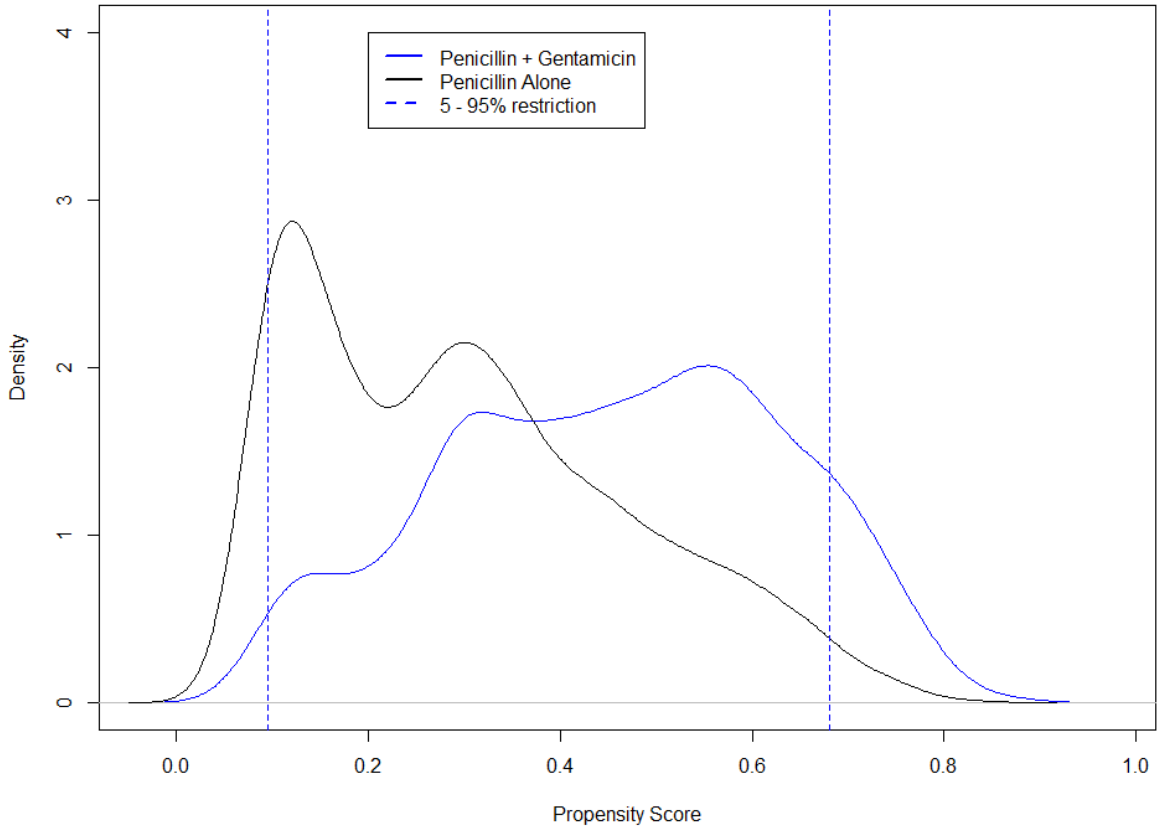


Figure B: PS trimming in experiment 2

f) Analysis using Instrument variables

Since a valid instrumental variable should be: (i) usable as a variable for randomly and effectively assigning patients into alternative groups, distribution of patients was examined across the levels of the IV as the distribution should be approximately similar between the IV levels; (ii) related with the treatment, a likelihood ratio test was conducted to examine the treatment – IV relationship. The process of fitting the instrumental variable models has been described in the supplementary material.

Imbalance of covariates between weekday and weekend admissions were explored (table D).

Table D: Imbalance of covariates between weekday and weekend admissions

Variable	Experiment 1			Experiment 2		
	Weekdays (n = 3014)	Weekends (n = 988)	ASMD	Weekdays (n = 4881)	Weekends (n = 1539)	ASMD
Child Sex						
Female	45%	46%	0.03	44%	45%	0.01
Male	55%	54%		56%	55%	
Pallor						
Mild/moderate	4%	5%	0.02	5%	5%	0.00
None	95%	94%		93%	93%	
Severe	1%	2%		2%	2%	
Capillary refill						
1 sec	68%	71%	0.07	66%	68%	0.04
2 sec	30%	27%		31%	29%	
>2 sec	3%	2%		3%	3%	
Fever						
Absent	21%	18%	0.05	19%	16%	0.07
Present	79%	82%		81%	84%	
Convulsions						
Absent	95%	96%	0.02	94%	94%	0.03
Present	5%	4%		6%	6%	
Vomiting						
No	65%	62%	0.06	63%	63%	0.00
Yes	35%	38%		37%	37%	
Referral						
No	82%	86%	0.10	81%	84%	0.09
Yes	18%	14%		19%	16%	
Thrush						
Absent	98%	98%	0.00	98%	98%	0.03
Present	2%	2%		2%	2%	
Comorbidities						
None	84%	83%	0.02	82%	80%	0.03
Malaria	9%	10%		10%	13%	
Diarrhoea	3%	2%		3%	2%	
Malaria and diarrhoea	4%	5%		5%	5%	
Crackles						
Absent	47%	47%	0.01	48%	47%	0.02
Present	53%	53%		52%	53%	
Wheeze						
Absent	85%	84%	0.02	85%	84%	0.02
Present	15%	16%		15%	16%	
IV prescription						
No	97%	96%	0.05	95%	95%	0.01
Yes	3%	4%		5%	5%	
Quinine Prescription						
No	97%	97%	0.02	95%	94%	0.04
Yes	3%	3%		5%	6%	
Artesunate Prescription						
No	92%	92%	0.01	92%	90%	0.05
Yes	8%	8%		8%	10%	
Mean WAZ	0.00	-0.01	0.01	0.01	-0.03	0.03
Mean age (months)	19.59	20.47	0.04	20.29	21.05	0.04
Mean weight (Kg)	9.56	9.61	0.01	9.7	9.89	0.05
Mean resp rate (breaths/min)	52.61	51.65	0.08	51.82	51.34	0.04
Mean temp (degrees C)	37.73	37.79	0.06	37.78	37.85	0.06
Mean cough duration (days)	3.40	3.20	0.07	3.45	3.35	0.04
Mean length of illness (days)	3.70	3.46	0.08	3.73	3.56	0.05

Also mortality between weekend and weekday admissions was explored for experiments 1 and 2 (table E). The weekend mortalities, in the raw datasets, seemed to be higher than weekday mortalities.

Table E: Summary of deaths by weekend/weekday admissions

Experiment	Weekend	Weekday
1	17/988 (1.7%)	45/3014 (1.5%)
2	47/1539 (3.1%)	49/4881 (1.0%)

In the next step, the treatment and outcome (mortality) probit models were fitted, with covariates in the treatment model being the same as those used in the corresponding propensity score models – though with the addition of admission timing variable as an IV. On the other hand, the outcome model used the same covariates as the treatment model with the exclusion of the admission timing variable both in experiments 1 and 2. Here, the parameter estimates were only presented for the treatment variable (mainly for comparison with individual treatment effect estimates obtained using propensity score weighting method.

Interpreting individual coefficients (like for treatment here) is less straightforward in probit models compared to linear regression and logit models where estimates are individually interpretable (2). This is because change in probability due to a unit change in a predictor is jointly dependent on other predictor values and their starting values. However, there are limited ways through which probit model parameters may be interpreted individually: (i) without considering the magnitude, the direction of effect may be inferred based on whether the parameter estimate is either positive or negative; (ii) if both the magnitude and direction are of interest (as is the case here), then a set of approximations may be conducted. Amemiya (1981) suggested multiplying the individual estimate from probit model by 1.6 to obtain the result in terms of log odds ratio (3). As the estimates obtained using PS methods were expressed in terms of log relative risk, the estimated odds ratios are further converted to log risk ratio using the modified relationship documented in (4):

$$log\ RR = \log\left(\frac{OR}{(1 - p_0) + (p_0 \times OR)}\right)$$

Where RR – is the risk ratio; OR – odds ratio and; p_0 – is the proportion of children who died in the penicillin monotherapy treatment group. Results have been presented in table 2 in the main manuscript.

g) Analysis using PS sub - classification

PS should classify children in groups where they share clinical features, as these features are also related to outcomes then in this case they are also grouped by severity. The average proportion of children who died increased consistently from PS subclass one to five for the two experiments. As PS was used as a proxy for disease severity in sub-classification, children in subclass 1 were likely to have less severe pneumonia (fewer variables with a positive value that may be associated with possible risk) and children in subclass 5 were likely to have more severe pneumonia (more variables with a positive value that may be associated with possible risk) (table F). Therefore, this relationship of PS subclass with mortality is expected.

Table F: Severe Pneumonia Deaths in Experiment 1 and 2 (ITT)

	Experiment 1		**Experiment 2**	
PS Subgroup	**Penicillin plus Gentamicin**	**Penicillin**	**Penicillin plus Gentamicin**	**Penicillin**
1	7/273 (2.56%)	8/1269 (0.63%)	9/459 (1.96%)	14/2333 (0.60%)
2	3/272 (1.10%)	1/591 (0.17%)	12/459 (2.61%)	10/822 (1.22%)
3	6/273 (2.20%)	8/380 (2.11%)	16/459 (3.49%)	8/467 (1.71%)
4	8/272 (2.94%)	4/266 (1.50%)	17/458 (3.71%)	11/341 (3.23%)
5	9/273 (3.30%)	5/133 (3.76%)	33/460 (7.17%)	7/153 (4.58%)
Total	33/1363 (2.46%)	26/2639 (0.99%)	87/2296 (3.79%)	50/4124 (1.21%)

In PS sub-classification (for experiment 2 – figure B) the log risk ratios consistently decreased from subclass 1 to 5 though this pattern was not completely clear in experiment 1 (figure C). In order to obtain pooled treatment effect, estimates were weighted by the number of patients who received penicillin plus gentamicin per subclass. However, the number of patients who received penicillin plus gentamicin were distributed equally (which would imply equal

weighting) – and additional weighting was based on how precise the log risk ratios were. This implied that the subclasses were treated as different trials and log RR estimates pooled in the form of a meta-analysis. The pooled estimates across the subclasses for experiments 1 and 2 were not statistically significant though had wider credible intervals as subclassification did not completely achieve balance on some of the variables at the subclass level.

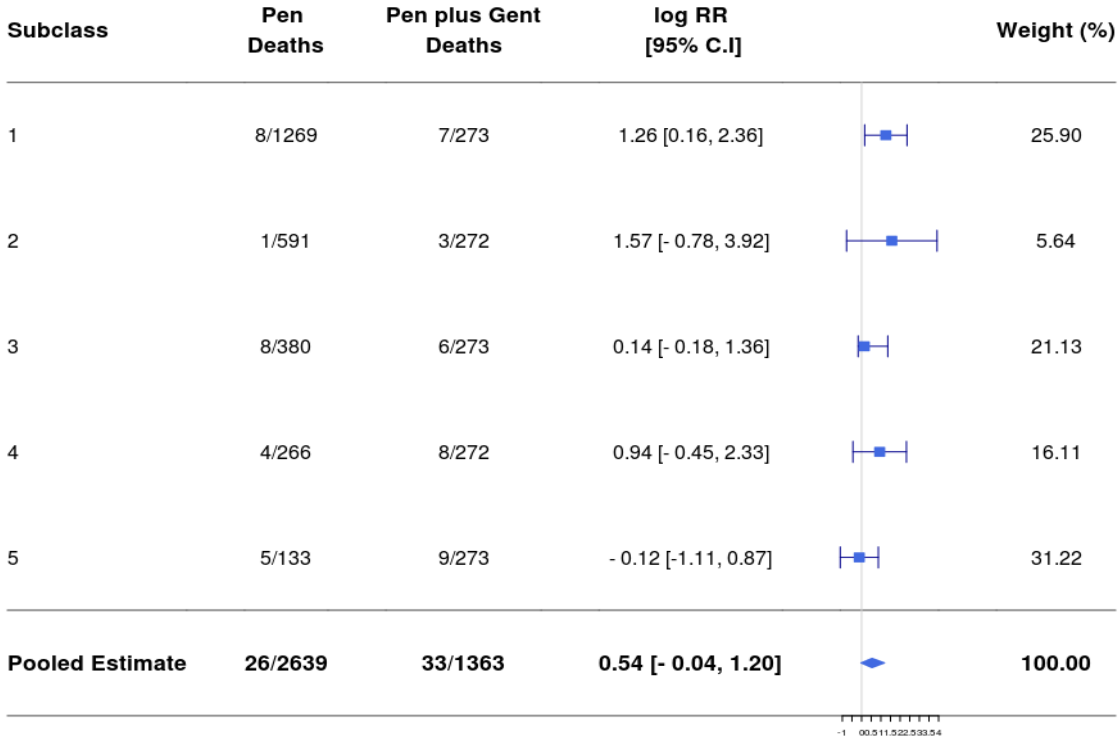


Figure C: Experiment 1 – ITT

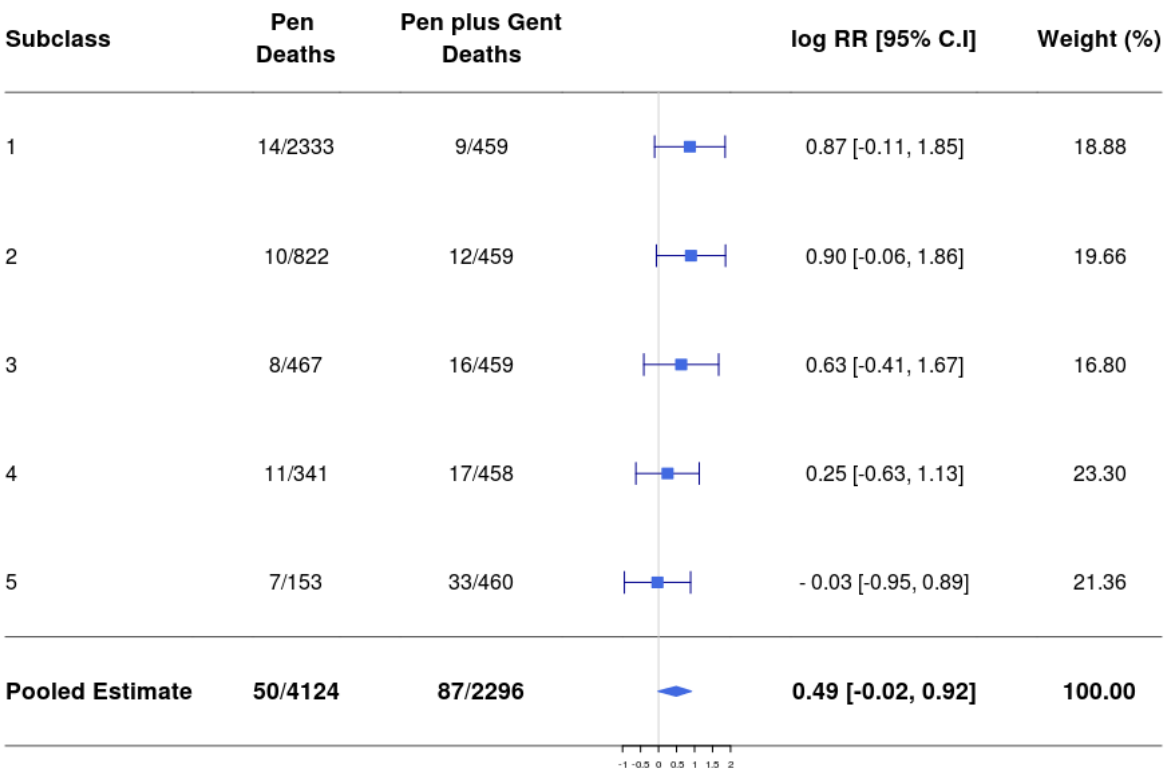


Figure D: Experiment 2 – ITT

h) Analysis using PS optimal full matching

Also analysis using PS optimal full matching showed no statistical significance in treatment of indrawing pneumonia using either penicillin or penicillin plus gentamicin (table G).

Table G: Treatment effect estimates

	log RR (95% C.I)
Experiment 1	
Regression without PS adjustment	0.56 [-0.06, 1.02]
PS optimal matching	0.27 [-0.22, 0.65]
Experiment 2	
Regression without PS adjustment	0.52 [0.14, 0.86]
PS optimal matching	-0.08 [-0.37, 0.18]

i) Analysis using per protocol population

Analysis using propensity score methods with per protocol population also demonstrated no significance in treatment with either penicillin or penicillin plus gentamicin (table H).

Table H: Per protocol treatment effect estimates

	log RR (95% C.I)
Experiment 1	
Unmatched (regression only)	0.71 [0.03, 1.42]
Optimal Full Matching	0.61 [0.05, 1.29]
Weighting	0.45 [-0.14, 1.09]
Sub-classification (pooled)	0.64 [-0.03, 1.32]
Experiment 2	
Unmatched (regression only)	0.54 [0.09, 0.98]
Optimal Full Matching	-0.33 [-0.66, 0.01]
Weighting	-0.13 [-0.48, 0.21]
Sub-classification (pooled)	0.47-0.08, 0.89]

j) Definition of PS methods and how they were used

We implemented three PS methods and these are briefly introduced:

Optimal full matching

PS matching aims to obtain treatment and (active) control patients who have approximately equivalent propensity score values (5). In optimal full matching, an optimal algorithm is used to obtain subsets of matched patients with the least global distance between them. Distance, here, is defined as the absolute difference in the propensity scores between a treated and control patient with global distance the sum of all distances between matched treated and control patients (6). This is the only form of matching that happens without replacement.

PS Weighting

There are two types of weights that may be estimated using PS. The first is inverse probability of treatment weights (IPTW) such that treated individuals are assigned weights of 1/PS while those in the (active) control group are assigned weights of 1/(1 – PS). The second is weighting by odds such that those treated are assigned a weight of 1 and those in the (active) control are assigned weights of PS/(1 – PS). These weights are used to estimate different treatment quantities. In this analysis we used weighting by odds to estimate what effect would be obtained suppose those who received gentamicin plus penicillin were denied this treatment.

PS sub – classification

Sub-classification divides patients into mutually exclusive groups based on their propensity scores. A standard practice, though not supported by specific recommendations, has been subdividing patients into five subclasses (7). One approach for creating patient subclasses would be to first conduct one on one nearest neighbour matching and then split the population into subclasses (8), alternatively one may use PS quintiles (5). The number of subclasses will usually depend on the sample size, and for large datasets, more classes with reasonable sample sizes would be desirable. This analysis used PS quintiles with five subclasses.

References

1. Global Health Observatory data repository [Internet].; 2016 []. Available from: http://apps.who.int/gho/data/node.main.686?lang=en.

2. Probit regression [Internet].; 2017 [cited 22-03-2017]. Available from: http://stats.idre.ucla.edu/stata/output/probit-regression/.

3. Amemiya T. Qualitative Response Models: A Survey. Journal of Economic Literature. 1981;XIX:1483-536.

4. Zhang J, Yu KF. What's the Relative Risk? A method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. JAMA - Journal of the American Medical Association. 1998.

5. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011;46(3):399-424.

6. Gu X, Rosenbaum PR. Comparison of multivariate matching methods: structures, distances and algorithms. Journal of Computational and Graphical Statistics. 1993;2:405-20.

7. Yang S, Imbens GW, Cui Z, Faries DE, Kadziola Z. Propensity Score Matching and Subclassification in observational Studies with Multilevel Treatments. Journal of the International Biometric Society. 2016.

8. Ho DE, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. 2011;42(8):1-28.

Comparative effectiveness of injectable penicillin versus a combination of penicillin and gentamicin in children with pneumonia characterised by indrawing in Kenya: A protocol for an observational study

Lucas Malla[1*] (lucas.malla@some.ox.ac.uk)

Rafael Perera-Salazar[2] (rafael.perera@phc.ox.ac.uk)

Emily McFadden[2] (emily.mcfadden@phc.ox.ac.uk)

Mike English[1, 3] (menglish@kemri-wellcome.org)

[1]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

[2]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom.

[3] Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.

***Corresponding author:**

Lucas Malla

Abstract

Introduction

WHO treatment guidelines are widely recommended for guiding treatment for millions of children with pneumonia every year across multiple low and middle income countries. Guidelines are based on synthesis of available evidence that provides moderate certainty in evidence of effects for forms of pneumonia that can result in hospitalisation. However, trials have included fewer children from Africa than other settings and it is suggested that African children with pneumonia have higher mortality. Thus despite improving access to recommended treatments and deployment with high coverage of childhood vaccines, pneumonia remains one of the top causes of mortality for children in Kenya. Establishing whether there are benefits of alternative treatment regimens to help reduce mortality would utilize pragmatic clinical trials. However, these remain relatively expensive and time consuming. This protocol describes an approach to using secondary analysis of a new, large observational dataset as a potentially cheaper and quicker way to examine the comparative effectiveness of penicillin versus penicillin plus gentamicin in treatment of indrawing pneumonia. Addressing this question is important as although it is now recommended that this form of pneumonia is treated with oral medication as an outpatient it remains associated with non-trivial mortality that may be higher outside trial populations.

Methods and analysis

We will use a large observational dataset that captures data on all admissions to 13 Kenyan county hospitals. These data represent the findings of clinicians in practice and, because the system was developed for large observational research, pose challenges of non-random treatment allocation and missing data. To overcome these challenges this analysis will use a

rigorous approach to study design, propensity score methods and multiple imputation to minimize bias.

Ethics and dissemination

The primary data are held by hospitals participating in the Kenyan Clinical Information Network (CIN) project with de-identifed data shared with the KEMRI-Wellcome Trust Research Programme for agreed analyses. The use of data for the analysis described received ethical clearance from the Kenya Medical Research Institute Scientific and Ethical Review Committee. The findings of this analysis will be published.

Strength

- This study will be used as a platform to explore effectiveness of alternative treatments in routine care in a low income setting to improve health outcomes for children.

Limitation

- The analysis will be limited to the variables in the observational dataset – and therefore risk bias due to unmeasured key variables.
- The influence of any resulting bias, to alter results, will however be assessed through the use of alternative methods as instrumental variables.

Introduction

Kenya has developed and disseminated national treatment guidelines largely drawing on those of WHO for a number of childhood diseases including pneumonia (1, 2). These pneumonia guideline recommendations are based on synthesis of available evidence that provides moderate certainty in evidence of effects of treatments for forms of pneumonia that can result in hospitalization (1, 2). Such guidelines have been shown to be effective in reducing pneumonia related mortality and thus Kenyan clinicians are supposed to use them in routine practice to treat pneumonia (and other diseases) (3, 4). However, although the guidelines are based on the best available evidence, the evidence available from trials conducted in Africa remains limited (5). There has also been little thorough investigation of the effectiveness of treatments in non-trial populations in routine settings that may often differ from those enrolled in formal clinical trials. For example many children admitted with pneumonia may have co-morbidity that might exclude them from trials (6). These issues can prove problematic when making national guidelines where study generalisability can be contested (7).

The WHO and Kenyan pneumonia treatment guidelines are implicitly based on risk stratification of illness with children deemed at higher risk of severe illness and mortality offered broad spectrum antibiotic regimens and those at lower risk narrow spectrum antibiotics (2, 8-10). This risk stratification approach is operationalized by requiring clinicians to look for specific features in the clinical history and examination that are used to define illness severity and therefore recommended treatment (Box 1). Previous studies conducted in Kenya have, however, indicated that clinicians do not always follow guideline recommendations in treating pneumonia (4). Variation from the guideline recommended approach can occur at the point of pneumonia severity assignment (clinicians do not follow the rules linking clinical signs and severity category) and at the point of treatment assignment (clinicians do not follow the rules linking treatment and severity). This variability in treatment assignment provides the

opportunity for comparative effectiveness evaluation if similar populations of children with pneumonia are prescribed different treatments. Clinicians may create such a situation by not following recommendations because they have inadequate knowledge or if they believe (potentially contrary to the evidence) that certain treatments result in better health outcomes.

Box 1: Pneumonia treatment algorithm

The pneumonia severity classification that was recommended by Kenyan guidelines up to March 2016 (9) (and previously by WHO guidelines (1)) defined the following three severity classes:

1. ***Very severe pneumonia:*** *If a child had either oxygen saturation less than 90% or central cyanosis or was grunting or unable to drink or not alert, then s/he was classified as having very severe pneumonia, put on oxygen and treated with a combination of gentamicin and penicillin.*

   {The new WHO (2) and Kenyan guidelines (9) renamed this class as "severe pneumonia" – and currently recommend treatment with a combination of ampicillin (or penicillin) with gentamicin plus oxygen}.
2. ***Severe pneumonia:*** *If a child had lower chest wall indrawing (but did not have any of qualifying signs for very severe pneumonia above) and was alert then s/he was to be classified as having severe pneumonia and be treated with benzyl penicillin only.*

   Note: *The term indrawing pneumonia is hereafter used in this protocol to define this category of children to avoid confusion.*
3. ***(Non – severe) Pneumonia:*** *If a child had none of the mentioned signs but had cough or difficulty breathing and a respiratory rate greater than or equal to 50 breaths/minute (for age between 2 and 11 months) or respiratory rate greater than or equal to 40 breaths/minute (for age above 12 months) then s/he was classified as having non severe pneumonia and treated with cotrimoxazole or amoxicillin if previously treated with cotrimoxazole.*

   {*The current WHO and Kenyan guidelines collapsed severity classes 2 and 3 into one category referred to as "non –severe pneumonia". This group of patients are currently treated with oral amoxicillin – partly informed by a local trial (18)}.*

In particular, a previous study showed that clinicians over-prescribed gentamicin, adding this to penicillin for the treatment of pneumonia characterized by lower chest wall indrawing but no other signs of severe illness instead of penicillin alone as was recommended (4)[1]. Therefore, this protocol is for a study that seeks to explore whether there is any benefit from adding gentamicin to penicillin in treating children with indrawing pneumonia. Such a benefit could accrue if bacterial causes of pneumonia that were previously (prior to introduction of new vaccines) proportionately less common (eg. *S. aureus* and gram negative bacteria) are now accounting for an increased proportion of pneumonia deaths – as in such cases, the addition of gentamicin might provide effective treatment for a broader spectrum of pathogens. Tackling this question is of importance as WHO have recently changed indrawing pneumonia treatment guidance based on trials that suggest equivalence of oral amoxicillin and injectable penicillin (12-15). New guidance recommends outpatient oral treatment for a population of children previously admitted to hospital (10). However, mortality from pneumonia has been reported to be higher in African settings (16, 17) despite the increasing use of multiple vaccines spanning: measles, pertussis, HiB and pneumococcal conjugate vaccines. It remains possible therefore that for a small number of children a broader spectrum antibiotic regimen might be of benefit. This study addresses this question that has not been the subject of prior community and pragmatic clinical trials.

Objectives

Primary

1) Experiment 1: To compare the effectiveness of injectable penicillin versus penicillin plus gentamicin (both injectable) in treatment of indrawing pneumonia; where severity

[1] The fact that inadequate knowledge in handling childhood pneumonia may result in inconsistent treatment allocation is supported by a survey conducted in seven developing countries showing that 56% of nurses and doctors had inadequate knowledge in managing pneumonia in children (11).

level is constructed (imputed) using data recorded on each child's clinical signs (hospitals use a structured record form that supports recording of signs highlighted in guidelines) such that severity classification is consistent with guideline recommendations.

Secondary

2) Experiment 2: To compare effectiveness of injectable penicillin versus penicillin plus gentamicin in treatment of indrawing pneumonia; where we use clinician assigned severity level.
3) Experiment 3: To compare effectiveness of injectable penicillin versus penicillin plus gentamicin in treatment of all cases of pneumonia admitted to hospital.

Experiment 1 will be primary as it most approximates a typical randomised trial where recruitment would be based on specified clinical signs. This scenario will provide an evaluation of alternative therapies within a guideline class (where children have very similar clinical signs) and thus is the best mimic of a prospectively designed comparative evaluation in which clinicians stick to the rules of severity classification (see (18) for an example of a RCT in Kenya that this would be similar to – where classification is based on clinical signs). Recommended treatment for this disease classification was penicillin alone, treatment with combination therapy may therefore represent over-treatment. Alternatively, the combination treatment that provides broader antimicrobial cover could provide an advantage in a small proportion of cases that would only be detected in moderately large studies – where the addition of gentamicin offers improved treatment for specific organisms not susceptible to penicillin alone.

Experiment 2 will provide a test of alternative therapies amongst those where clinicians used their own judgement (possibly including gut feeling) to classify and treat (19) and have on occasions (potentially) over-ridden or ignored the guideline recommendations. In this case

although the same label of indrawing pneumonia is given to all, the treatment selected may be an indicator of perceived severity and there may be a potential bias as a result – and the propensity score distributions (see below) may help demonstrate this and in theory may overcome this potential bias. Here if there is no clinically relevant difference between treatments within a group of patients that reflects clinicians' actual classification decisions this could reassure them that monotherapy with penicillin (or amoxicillin) would be acceptable. Lastly, experiment 3 is an extension of the logic of experiment two. To date there have been no pragmatic trials of penicillin alone compared with alternative combination therapies for all forms of inpatient pneumonia, and addressing this question may be relevant for two reasons. First, the population of children admitted with severe forms of pneumonia is now largely one that has received *H. influenzae* Type B and pneumococcal conjugate vaccines that have likely changed the aetiology of this illness. Second, if clinicians are poorly trained and unable to classify illness severity – resulting in non-adherence to guidelines - it would be useful to explore the potential impact of this across all levels of severity of pneumonia. This analysis has the largest numbers of subjects.

Methods and analysis

To answer these three questions, we will use the Kenyan Clinical Information Network (CIN) dataset that provides observational data on all admissions to 13 Kenyan County hospitals (Box 2).

Box 2: Clinical Information Network

CIN was initiated to improve data availability from secondary care in paediatrics and as a model for demonstrating the value of routine data in improving quality of care in the county (formerly district) hospitals. These hospitals typically have a single paediatrician leading services predominantly provided by junior clinical teams. Data in these hospitals are collected prospectively post discharge by trained data clerks, guided by well-defined standard operating procedures, under close supervision by the hospital medical records department and the research team. It is worth noting that the research team has no personnel checking quality of clinical process and whether clinicians correctly document what they do. However, the patient record is the formal (and legal) document describing the clinical condition and management. These documents are used for data abstraction and they include patient files with standardized Paediatric Admission Record (PAR) forms, treatment sheets, discharge summary forms, laboratory reports and clinician notes. The collected data are used to assess documentation of history, physical examination, diagnosis, laboratory investigations, treatment and discharge plans. Feedback to hospitals as part of the CIN activities has helped improve the quality of clinical data (20). The description of hospital selection and their populations of patients is detailed in Ayieko (2015) (6).

The analysis will proceed in two stages – design and outcome analysis as suggested by Rubin (2008) (21) as an objective way for analysing observational datasets.

Study Design

This will be an observational study conducting secondary analyses of data routinely collected from hospital paediatric wards in Kenya's CIN. The design process for the three experimental scenarios will be similar and broadly consists of the following steps suggested in Rubin (2008) (21):

a) Definition of inclusion and exclusion criteria.

b) Understanding the pneumonia diagnosis and treatment assignment processes. This is to help understand key and auxiliary variables required for analysis.

c) Verification of sample size if sufficient for any meaningful analyses.

d) Creation of comparable treatment arms – which will be addressed analytically aiming to overcome non – random treatment assignment and deal with missing data.

e) Outcome analysis follows after conceptualisation of design in steps a – d.

a) Inclusion and exclusion

This analysis will include all children aged 2 – 59 months and will exclude children with any co-morbidity of HIV, meningitis, tuberculosis and or acute malnutrition as there are specific antibiotic treatment rules for these children that supersede those for pneumonia. Specifically Kenyan guidelines for the inpatient treatment of pneumonia in children that are HIV infected recommend only combination therapy. Importantly therefore children with other co-morbidities such as mild anaemia, diarrhoea and malaria are not necessarily excluded from the analysis.

b) Understanding the diagnosis and treatment assignment rules for pneumonia paediatric patients

Clinicians are supposed to use guidelines widely disseminated as the 'Basic Paediatric Protocols' in Kenya (9) that are adapted from WHO guidance, based on available evidence and developed by consensus by a national guideline panel (see (22-24)). In standard practice, the process of treatment assignment happens in three steps; first, there is assessment and documentation of each clinical sign. Step two involves integration of clinical information into severity classification, and in step three severity classification is translated into a treatment assignment (see Box 1 above). In Kenya, as in many low and middle income countries these recommendations reflect the absence of access to further diagnostic tests. Thus pulse oximetry,

blood culture or tests for inflammatory markers are not routinely available (6). As indicated above clinicians may fail to adhere to guideline recommendations by making errors or over-riding recommendations at any of the three steps of assessment, severity classification and treatment assignment. However, based on the clinical symptoms and signs recorded it is possible to assign a severity classification (and thus expected treatment) based on the data. It is a data informed and investigator assigned classification as indrawing pneumonia that is used in the primary analysis (experiment 1).

c) Analysis Variables

Outcome variable

Mortality will be used as the outcome variable in all the three experiments.

Independent variables

These variables are grouped into key and auxiliary. Key variables are defined as those that should influence pneumonia severity classification and hence treatment based on the treatment protocol (9) (Box 1 above). Auxiliary variables are defined as those that might, *a priori*, be expected to influence treatment assignment based on clinical reasoning (for example they might make a clinician concerned for severe illness), although according to the formal rules (the guidelines) they are not considered reasons to alter treatment assignment. Such auxiliary variables were identified from those clinical symptoms and signs that are routinely collected within CIN. See table 1 for a summary of key and auxiliary variables that will be used in the analyses.

Table 1: Summary of key and auxiliary independent variables for experiments 1, 2 and 3[2]

Experiment 1 and 2 key variables	**Experiment 3 key variables**	**Auxiliary variables for experiments 1, 2 and 3**
Age (2 – 59 months)	Age (2 – 59 months)	Gender (male/female)
Indrawing (present/absent)	Indrawing (present/absent)	Cough duration (days)
History of cough (yes/no)	History of cough (yes/no)	Crackles (present/absent)
Difficulty breathing (present/absent)	Difficulty breathing (present/absent)	Weight (Kg)
Level of consciousness – AVPU (alert/verbal response/pain response/unresponsive)	Level of consciousness – AVPU (alert/verbal response/pain response/unresponsive)	Pallor (0, +, +++)
	Oxygen ordered (yes/no)	Capillary refill (immediate, 1 – 2 secs, 3 – 6 sec, > 6 secs)
	Cyanosis (present/absent)	Fever (present/absent)
	Inability to drink/breastfeed (yes/no)	Diarrhoea (present/absent)
	Grunting (present/absent)	Convulsions (present/absent)
	Respiratory rate (breaths/min)	Vomiting (yes/no)
		Referral (yes/no)
		Length of illness (days)
		Number of fits
		Thrush (present/absent)
		Quinine/artesunate (prescribed/not prescribed)
		Weight for age z – score
		Wheeze (present/absent)
		Comorbidities (Malaria and or diarrhoea)

d) Sample size verification

Here, sample size verification uses the formula cited in (25):

$$ns = \frac{k+1}{k}\frac{\bar{p}(1-\bar{p})(Z_{\beta}+Z_{1-\alpha/2})^2}{(p_1-p_2)^2}$$, where:

[2] Experiment 3 has more key variables than experiment 2 as it considers patient populations with "very severe, severe and non –severe pneumonia" – as classified in the previous WHO and Kenyan treatment guidelines. Therefore, in addition to variables used to classify severe pneumonia, other variables used to classify very severe and non-severe pneumonia are considered.

ns = size of smaller group.
k = ratio of larger group to smaller group.
p_1 - p_2 = clinical difference in proportions of the outcome.
Z_b = corresponds to power of 80%
$Z_{1-a/2}$ = corresponds to two-tailed significance level (1.96 for a = .05).
$\bar{p}$ = corresponds to average of outcome proportions in two groups.

The value for $\bar{p}$ is estimated from studies – two of which formed evidence for earlier WHO indrawing pneumonia treatment guidelines. See table 2 that shows the number of deaths per treatment arm reported in these studies.

Table 2: Summary of some of pneumonia studies that informed previous WHO guidelines

Study	Treatment arms	Mortality	$\bar{p}$
Shann et. al(1985)	Chloramphenicol alone	48/377	0.1470
	Chloraphenicol+Penicillin	62/371	
Addo – Yobo et. al (2002)	Injectable penicillin	7/845	0.0050
	Oral amoxicillin	2/857	
Agweyu (2015)	Injectable penicillin	3/264	0.008
	Oral amoxicillin	1/263	

For assessment of sample size for indrawing pneumonia experiments, a weighted[3] $\bar{p}$ of 0.041 from these studies is used. The ratio r is varied between 1 and 3. Figure 1 was generated by fixing power and significance level at 80% and 5% respectively. Estimates of $\bar{p}(1-\bar{p})$ derived from WHO studies were substituted in the sample size formula and data simulated in order to see what detectable differences would be achieved by different sample sizes. A total sample size of about 4000 would be sufficient to detect a minimum difference of 1.5% (absolute difference e.g. a reduction of mortality from X% to X – 1.5%) in any of these experiments[4].

[3] Weighting was done using the total sample sizes per experiment.
[4] A sample size of at least 4000 would be required for experiment 3 as this is the minimum sample for experiments 1 and 2 which are nested in experiment 3.

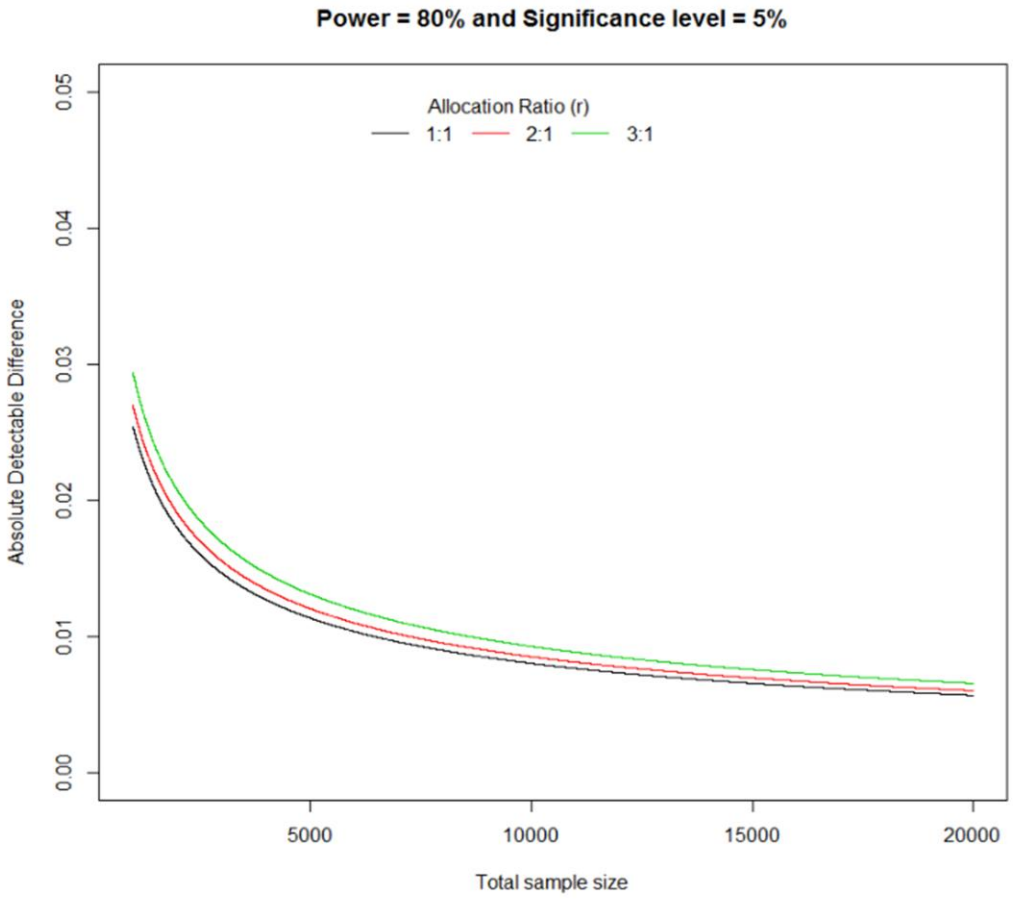


Figure 1: Sample size verification.

Statistical and outcome Analysis

Statistical analysis will proceed in the following four steps:

Step 1 – subset of patients of interest for the experiments will be obtained.

- Experiment 1: First, missing clinical signs data will be multiply imputed[5] (excluding outcome data) – and then key clinical signs data used to impute (construct) a pneumonia severity level for all patients based on the algorithms in the pneumonia treatment protocol (9). Thereafter, a subset of patients with guideline-defined indrawing pneumonia (for each of the imputed datasets) will be obtained for further analyses.

[5] For the three experiments, 20 datasets will be multiply imputed using chained equations (26).

- Experiment 2: A subset of indrawing pneumonia patients (with severity as indicated by the clinicians) will be obtained from the raw dataset – and clinical signs data imputed using multiple imputation (without the outcome data).
- Experiment 3: The raw dataset containing all the patients with all forms of pneumonia severity will be used –and clinical signs data imputed using multiple imputation (without the outcome data).

Step 2 – patients in the alternative treatment arms will be matched using propensity score (PS) methods to overcome non – random treatment allocation. Standardised mean differences (and where necessary density plots) will be used as diagnostic checks for covariate balance and overlap (27, 28) between penicillin and penicillin plus gentamicin treatment groups. PS methods that utilise all the data (PS optimal full matching, weighting and sub classification) will be examined in experiments 1 and 2 (on each imputed dataset) and the method that results in the minimum average absolute standardised mean differences for the majority of the variables and retains the largest number of patients in the analysis will be considered appropriate (29). While only PS sub-classification will be used for experiment 3. As experiment 3 aims to investigate comparative effectiveness in all cases of pneumonia, propensity score will be used as a proxy for disease severity thus patients with lower propensity scores will be considered less ill while those with higher propensity scores will be considered more ill (grouped in propensity score subclasses for analysis).

Step 3: conducting outcome analysis.

For each imputed dataset (per experiment), outcome analysis will aim to investigate treatment causal effects across all the hospitals. Bayesian log binomial regression models (30) will be used to estimate overall treatment effects[6]. A hospital variable will be modelled as a fixed effect

[6] Bayesian models will be used to overcome any bias due to sparsity of data as PS sub-classification in itself reduces the effective sample size.

in the log binomial regression that measures treatment effects on pooled data. These models will be fitted on each imputed dataset (adjusting for other variables used in PS models) and results pooled using Rubin rules (31).

Step 4: sensitivity analysis will be conducted to investigate effects of unmeasured confounders and validity of estimates obtained through multiple imputation. Propensity score methods generate matched treated and (active) control patients whose distribution of measured covariates are as similar as possible. However, two patients with similar covariate distribution may differ in terms of unmeasured variables – and this may introduce bias in estimated treatment effects (32). On the other hand, if outcome and explanatory variables have missing data, then inclusion of outcome data in multiple imputation may contribute minor information in the substantive (outcome) model (33).

Exploring effects of unmeasured confounders

Sensitivity analysis for unmeasured confounders will involve the use of an instrumental variable (IV) (34) – weekend admission and PS trimming (35). A few IV sources in health studies have been described in Baiocchi (2014) (36). These include: distance to specialty, genes, insurance plan, timing of admission, calendar time and preference based IVs. Of relevance to this analysis would be timing of admission IVs. A study conducted by Berkley (2004) (37) in a Kenyan hospital demonstrated that children who were admitted during the weekend experienced higher mortality compared to those admitted during the weekdays – which is a possible indication of poor quality of care and treatment during the weekend. In other words, it is anticipated that children admitted during the weekdays would have better health outcomes. This, in theory, implies that the type of treatment and care received depend on the day of admission – and which later determines the type of health outcome of the patient.

Examining validity of multiple imputation

The analysis steps 1 – 3 above will exclude outcome data in the imputation model – however sensitivity analysis will include models in which the outcome variable is included in the imputation approach. This will aim to investigate if including outcome data in the imputation model has an influence[7].

Ethics and dissemination

The primary data are held by hospitals participating in the Kenyan Clinical Information Network (CIN) project with de-identifed data shared with the KEMRI-Wellcome Trust Research Programme for agreed analyses. The analyses described in this protocol are part of this larger project (CIN) which was approved by the Kenya Medical Research Institute Scientific and Ethical Review Committee (Protocol number: 2465). This committee agreed the use of de-identified patient data derived from retrospective case record review without gaining individual patient consent as is common practice in service evaluation research. The findings will be useful in understanding the external validity of current treatments – and will provide a platform on which to do more similar analyses for different (combinations of) treatments. The results of this analysis will be shared with the Kenyan Ministry of Health and will inform discussions on national pneumonia treatment guidelines to which the research team have made major prior contributions. The work will also be submitted for publication.

Competing Interests

The authors declare they have no competing interests.

Authors Contributions

[7] The primary interpretations will consider results of multiple imputations without outcome if results differ from those of MI with outcome – as is the standard recommendation to analysis of observational datasets in Rubin (2008) (21).

The contributions of the authors were as follows: LM did an initial draft of this manuscript with the support of RP, EM and ME. Thereafter, all authors edited subsequent versions and approved the final copy.

Funding

We are grateful for the funds from the Wellcome Trust (#097170) that support ME through a fellowship and additional funds from a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (#092654) that supported this work. LM is supported by a Nuffield Department of Medicine Prize DPhil Studentship and Clarendon Scholarship (Oxford University). The funders had no role in drafting or submitting this manuscript.

References

1. Hospital Care for Children [Internet].; 2005 []. Available from: http://apps.who.int/iris/bitstream/10665/43206/1/9241546700.pdf.

2. Hospital Care for Children [Internet].; 2013 []. Available from: http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf.

3. Sazawal S, Black R. Pneumonia Case Management Trials Group. Lancet Infect Dis. 2003;3(9):547-56.

4. Agweyu A, Kibore M, Digolo L, Kosgei C, Maina V, Mugane S, et al. Prevalence and correlates of treatment failure among Kenyan children hospitalised with severe community-acquired pneumonia: a prospective study of the clinical effectiveness of WHO pneumonia case management guidelines. Tropical Medicine & International Health. 2014.

5. Mulholland K, Carlin JB, Duke T, Weber M. Challenges of trials of antibiotics for pneumonia in low-income countries. The Lancet Respiratory Medicine. 2014;2(12):952-4.

6. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of admissions and variations in the use of basic investigations, treatments and outcomes in Kenyan hospitals within a new Clinical Information Network. Archives of Disease in Childhood. 2015.

7. Agweyu A, Opiyo N, English M. Experience developing national evidence-based clinical guidelines for childhood pneumonia in a low-income setting - making the GRADE? BMC Pediatric. 2012.

8. Ayieko P, English M. Case Management of Childhood Pneumonia in Developing Countries . Pediatr Infect Dis J. 2007;26:432-40.

9. Basic Paediatric Protocols: November 2013 Edition [Internet].: Ministry of Health; 2013 []. Available from: http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition.

10. Basic paediatric protocols [Internet].: Kenyan Ministry of Health; 2016 []. Available from: https://www.tropicalmedicine.ox.ac.uk/_asset/file/basic-paediatric-protocols-2016.pdf.

11. Graham SM, English M, Hazir T, Enarson P, Duke T. Challenges to improving case management of childhood pneumonia at health facilities in resource limited settings. Bulletin of the WHO. 2015.

12. Atkinson M, et al. A multicentre randomised controlled equivalence trial comparing oral amoxicillin and intravenous benzyl penicillin for community acquired pneumonia in children PIVOT Trial. Thorax. 2007;61:1102-6.

13. Hazir T, et al. Ambulatory short-course high-dose oral amoxicillin for treatment of severe pneumonia in children: a randomised equivalency trial. Lancet. 2008;371:49-56.

14. Addo-Yobo E, Chisaka N, Hassan M, Hibberd P, Lozano JM, Jeena P, et al. Oral amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59 months: a randomised multicentre equivalency study. Lancet. 2004;364:1141-8.

15. Addo-Yobo E, et al. Outpatient treatment of children with severe pneumonia with oral amoxicillin in four countries: the MASS study . Tropical Medicine and International Health. 2011;16:995-1006.

16. IVAC. Pneumonia and Diarrhoea Progress Report 2014. John Hopkins Bloomberg School of Public Health; 2014.

17. Onyango D, Kikuvi G, Amukoye E, Omolo J. Risk factors of severe pneumonia among children aged 2-59 months in western Kenya: a case control study. The Pan African Medical Journal. 2012(13):45.

18. Agweyu A, Gathara D, Oliwa J, Muinga N, Edwards T, Allen E, et al. Oral amoxicillin versus benzyl penicillin for severe pneumonia among kenyan children: a pragmatic randomised controlled noninferiority trial. Clin Infect Dis. 2015.

19. Stopler E, Van Roeyn P, Van de Wiel M, Van Bokhoven M, Houben P, Van der Weijde T, et al. Consensus on gut feelings in general practice. BMC Family Practice. 2009.

20. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation of clinical care within a clinical information network: an essential initial step in efforts to understand and improve care in Kenyan hospitals. BMJ Global Health. 2016.

21. Rubin DB. For Objective Causal Inference, Design Trumps Analysis. . 2008;2(3):808-40.

22. Draft recommendation for management of children with severe febrile illness and impaired circulation without signs of severely impaired circulation: Guideline Panel Meeting [Internet].: Health Services, Implementation Research and Clinical Excellence Collaboration; 2013 []. Available from: http://www.idoc-africa.org/images/documents/guidelines/Panel%20recommendation%20on%20fluid%20resuscitation%20I%20final.pdf.

23. Draft recommendations for care of umbilical cord in newborns: Guideline Panel Meeting. [Internet].: Health Services, Implementation Research and Clinical Excellence Collaboration; 2013 []. Available from: http://www.idoc-africa.org/images/documents/guidelines/Panel%20recommendations%20on%20cord%20care%20final.pdf.

24. Draft recommendations for hospital management of children (less than 5 years) with sickle cell disease: Guideline Panel Meeting 2013. [Internet].: Health Services, Implementation Research and Clinical Excellence Collaboration; 2013 []. Available from: http://www.idoc-africa.org/images/documents/guidelines/Panel%20recommendations%20on%20sickle%20cell%20anaemia%20final.pdf.

25. Wittes J. Sample size calculations for randomised controlled trials. Epidemiologic Reviews. 2002;24(1).

26. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? International Journal of Methods in Psychiatric Research. 2011;20(1):40-9.

27. Austin PC. Assessing balance in measured baseline covariates when using many-to-one matching on the propensity-score. Pharmacoepidemiol Drug Saf. 2008 2008;17(12):1218-25.

28. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med. 2009 10 Nov 2009;28(25):3083-107.

29. Stuart EA. Matching methods for causal inference:A review and a look forward. Statist. Sci. 2010;25(1):1-21.

30. Fine J, Gray R. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496-509.

31. Rubin DB. An overview of Multiple Imputation. . 1988.

32. Rosenbaum PR. Design sensitivity and efficiency in observational studies. Journal of the American Statistical Association. 2010;105:692-702.

33. Roderick JAL. Regression With Missing X's: A Review. Journal of the American Statistical Association. 1992;87(420):1227-37.

34. Chiba Y. Bias analysis of the instrumental variable estimator as an estimator of the average causal effect. Contemporary Clinical Trials. 2010 January 2010;31(1):12-7.

35. Lee BK, Lessler J, Stuart EA. Weight Trimming and Propensity Score Weighting. PLoS ONE [Electronic Resource]. 2011.

36. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference. Statistics in Medicine. 2014;33(13):2297-340.

37. Berkley JA, Brent A, Mwangi I, English M, Maitland K, Marsh K, et al. Mortality among Kenyan children admitted to a rural district hospital on weekends as compared with weekdays . Pediatrics. 2004;114:1737-8.

STROBE Statement—Checklist of items that should be included in reports of ***case-control studies***

	Item No	Recommendation	Page No.
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract *A retrospective observational study*	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found ***Participants and measures*** *The analyses included children aged 2 – 59 months. Selection of study population was based on inclusion criteria typical of a prospective trial, primary analysis (experiment 1, n = 4002), but we also explored more pragmatic inclusion criteria (experiment 2, n = 6420) as part of a secondary analysis. To overcome the challenges associated with the non – random allocation of treatments and missing data, we used propensity score(PS) methods and multiple imputation to minimize bias. Further, we estimated mortality risk ratios using log binomial regression and conducted sensitivity analyses using an instrumental variable and PS trimming.* ***Results*** *The estimated risk of dying, in experiment 1, in those receiving penicillin plus gentamicin was 1.46 [0.85, 2.43] compared to the penicillin monotherapy group. In experiment 2, the estimated risk was 1.04 [0.76, 1.40].* ***Conclusion*** *There is no statistical difference in the treatment of indrawing pneumonia with either penicillin or penicillin plus gentamicin. By extension it is unlikely that treatment with penicillin plus gentamicin would offer an advantage to treatment with oral amoxicillin.*	2 – 3
Introduction			
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported *Kenyan guidelines for antibiotic treatment of pneumonia recommended treatment of pneumonia characterised by indrawing with injectable penicillin alone in inpatient settings until early 2016. At this point, they were revised becoming consistent with WHO guidance after results of a Kenyan trial provided further evidence of equivalence of oral amoxicillin and injectable penicillin. This change also made possible use of oral amoxicillin for outpatient treatment in this patient group. However, given non-trivial mortality in Kenyan children with indrawing pneumonia it remained possible they would benefit from a broader spectrum antibiotic regimen. Therefore, we compared the effectiveness of injectable penicillin monotherapy with a regimen combining penicillin with gentamicin. This has been explained on pages 2, 4 – 5*	2, 4 – 5
Objectives	3	State specific objectives, including any prespecified hypotheses: *Objectives:* ***1) Experiment 1:*** *To compare effectiveness of injectable penicillin versus penicillin plus gentamicin (both injectable) in treatment of indrawing pneumonia; where the child is identified as belonging to a population of children with indrawing pneumonia on the basis of data on their recorded clinical signs. The Experiment 1 population of indrawing pneumonia is therefore consistent with pre-2016 clinical guideline recommendations.*	8

		2) Experiment 2: *To compare effectiveness of injectable penicillin versus penicillin plus gentamicin in a population in which we use the clinician assigned categorisation of indrawing pneumonia, which may not be consistent with clinical guideline recommendations.*	
Methods			
Study design	4	Present key elements of study design early in the paper: *Key elements have been included in the clinical definitions of pneumonia, primary (experiment 1) and secondary (experiment 2) analyses subsection. Also see definitions of intention to treat and per protocol analyses population definitions (page 10)*	5 – 8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *The setting of this study is Kenyan hospitals and this has been defined in the data source subsection of the methodology. This subsection also describes how data collection happens in this setting*	9 – 10
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls: *The eligibility criteria have been outlined in the analysis protocol. And the question to be addressed was justified in the introduction (see pages, 2, 4 – 5). Also comparison groups discussed in pages 8 – 9 selected/adjusted using propensity score methods.*	8 – 9
		(*b*) For matched studies, give matching criteria and the number of controls per case *The analysis compared various propensity score methods. The best performing method was propensity score weighting in which outcome analysis was adjusted using weights generated from the estimated propensity scores. The primary analysis used propensity score methods that retained all the participants*	10 – 13
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable *All the variables used in the analyses were defined in the protocol alongside the justification for their selection. The protocol has been submitted alongside this paper (additional file 1: analysis protocol).*	
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *The variables are captured in forms filled by clinicians in the hospitals. Clinicians admitting patients fill standardized Paediatric Admission Record (PAR) forms that have been shown to improve documentation of clinical symptoms and signs. Together with discharge forms, treatment sheets and laboratory reports these are all part of the patient files that are the primary data source.*	9
Bias	9	Describe any efforts to address potential sources of bias *As the routine dataset used was limited to only observed variables,*	10 – 12

		sensitivity analyses were conducted using propensity score trimming and instrumental variables to examine any potential influence of unmeasured variables	
Study size	10	Explain how the study size was arrived at: *We used routine dataset already collected and therefore verified the sample sizes if adequate for meaningful analyses. The sample size verification is presented in detail in the published protocol submitted alongside this paper (additional file 1: analysis protocol).*	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why *All quantitative variables were used as continuous variables in the propensity score and outcome models*	*See the additional file 1: analysis protocol l on variables selected*
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding *All the statistical methods have been described in detail in pages 10 – 12. These include propensity score methods, instrumental variable analysis and log – binomial regression models for the outcome analysis.*	10 – 12
		(*b*) Describe any methods used to examine subgroups and interactions *No sub – group analysis conducted*	
		(*c*) Explain how missing data were addressed *Missing data were handled using multiple imputation*	10
		(*d*) If applicable, explain how matching of cases and controls was addressed *Matching was conducted using propensity score methods – and has been outlined in the methodology section*	10
		(*e*) Describe any sensitivity analyses *As the routine dataset used was limited to only observed variables, sensitivity analyses were conducted using propensity score trimming and instrumental variables to examine any potential influence of unmeasured variables*	11 – 12
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed *The number of patients included have been defined in figure 1 by treatment group. Experiment 1 included a total of 4002 while experiment 2 included 6420 as intention to treat populations.*	13
		(b) Give reasons for non-participation at each stage *As we used routine dataset in analysis – the exclusion and inclusion were already defined in the analysis protocol (additional file 1: analysis protocol)*	
		(c) Consider use of a flow diagram *The analysis populations have been defined in figure 1*	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders *See figure 1, also see figure D of the additional file 2: supplementary*	

		data	
		(b) Indicate number of participants with missing data for each variable of interest *This has been referred to on page 10 and full results presented in the supplementary material provided (additional file 2: supplementary data)*	
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure *See figure 1*	See figure 1
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included *Treatment effect estimates are summarised in table 1*	See table 1, page 17
		(*b*) Report category boundaries when continuous variables were categorized *Continuous variables were not categorised*	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses *No subgroup analyses were conducted*	
Discussion			
Key results	18	Summarise key results with reference to study objectives *We compared penicillin alone with penicillin plus gentamicin in treatment of indrawing pneumonia in populations with overall mortality of 1.5% and 2% in experiments 1 and 2 respectively. There were more fatal events in the penicillin plus gentamicin group than the penicillin group (approximately 2.5 times) and unadjusted analyses pointed, therefore, to a protective effect of penicillin treatment. However, adjusted analyses, both in experiments 1 and 2, that aim to account for allocation bias using PS weighting that can result from non-random treatment allocation suggest that there is no appreciable difference in outcomes between penicillin and gentamicin plus penicillin treatment of indrawing pneumonia. In addition, we conducted analyses using alternative PS methods (optimal full matching and sub-classification) and analyses of both intention to treat and per protocol populations. All analyses showed similar findings (see the provided supplementary material).*	17 – 18
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias *Conducting comparative effectiveness analyses using observational datasets can offer the advantage of larger sample sizes at lower cost than randomised controlled clinical trials. They also include patients that may not qualify for enrolment in a typical explanatory randomised controlled trial – and therefore perhaps provide more true to life estimates of treatment effects similar to those observed in highly pragmatic trials.*	20 – 21

		However, as most observational datasets are not meant for research, they have challenges of non-random treatment allocation and missing data. We employed a rigorous 'experimental design' strategy as is recommended when using observational data. We used PS and multiple imputation methods in an effort to minimise bias due to non-random treatment allocation and missing data and analyses suggested no appreciable difference in outcomes of indrawing pneumonia treated with penicillin alone compared with penicillin plus gentamicin. This was in contrast to unadjusted regression analyses that pointed towards better outcomes with penicillin alone suggesting presence of allocation bias.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence *While there are differences (in terms of magnitude) in the mortality observed in the different groups that suggest some residual bias in treatment allocation, these mortality differences are no greater than might occur by chance after PS adjustment (with the type 1 and 2 errors specified in the protocol). In that sense the PS approach may still have limitations but it does allow us to conclude no statistical difference in mortality outcomes between the two treatment arms.*	20
Generalisability	21	*Discuss the generalisability (external validity) of the study results* *The WHO recommended guidelines for treating pneumonia have considerable influence on policy and practice in low and middle income countries. While the evidence base and rigour of guideline development have improved considerably there remain few data on their effectiveness when implemented in non-trial settings. Even though well-designed, large pragmatic trials would be preferred, we demonstrate that carefully collected routine data may be useful for assessing the effectiveness of alternative treatments. Such analyses may become increasingly possible as electronic medical records are deployed in low and middle income countries but it is important that such studies are carefully designed to limit as far as possible the biases that arise from non-random treatment allocation. Our results suggest that children with indrawing pneumonia may gain little benefit from treatment with broader spectrum antibiotic regimens. However, they also suggest that further work is needed to identify those who are at higher risk of death who might be prioritised for an inpatient diagnostic work up and improved supportive care rather than treated as outpatients.*	21
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based *We are grateful for the funds from the Wellcome Trust (#097170) that support ME through a fellowship and additional funds from a Wellcome Trust core grant awarded to the KEMRI-Wellcome Trust Research Programme (#092654) that supported this work. LM is supported by a Nuffield Department of Medicine Prize DPhil Studentship and Clarendon Scholarship (Oxford University). The funders had no role in drafting or submitting this manuscript.*	22

*Give information separately for cases and controls.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at

http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.