

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [editorial.bmjopen@bmj.com](mailto:editorial.bmjopen@bmj.com)

# BMJ Open

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

Journal:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	Paediatrics
Secondary Subject Heading:	Infectious diseases, Paediatrics
Keywords:	pneumonia, missing data, propensity scores, comparative effectiveness

SCHOLARONE™  
Manuscripts

1  
2  
3 1 **Using observational data to compare the effectiveness of antibiotic treatments for**  
4  
5 2 **children hospitalised with pneumonia in Kenya**  
6  
7

8 3 Lucas Malla<sup>1\*</sup> ([lmalla@kemri-wellcome.org](mailto:lmalla@kemri-wellcome.org))  
9

10 4 Rafael Perera-Salazar<sup>2</sup> ([rafael.perera@phc.ox.ac.uk](mailto:rafael.perera@phc.ox.ac.uk))  
11

12 5 Emily McFadden<sup>2</sup> ([emily.mcfadden@phc.ox.ac.uk](mailto:emily.mcfadden@phc.ox.ac.uk))  
13

14 6 Mike English<sup>1,3</sup> ([menglish@kemri-wellcome.org](mailto:menglish@kemri-wellcome.org))  
15  
16

17 7 <sup>1</sup>Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.  
18  
19

20 8 <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United  
21  
22 9 Kingdom.  
23  
24

25 10 <sup>3</sup>Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.  
26  
27

28  
29  
30  
31  
32

33 12 **\*Corresponding author:**  
34

35  
36 13 Lucas Malla  
37  
38

39 14  
40

41 15  
42

43 16  
44

45 17  
46

47 18  
48

49 19  
50

51 20  
52

53 21  
54

55 22  
56

57 23  
58

59 24  
60

1  
2  
3 254  
5 **Abstract**6  
7 **Objectives**

8  
9  
10 28 Kenyan guidelines for antibiotic treatment of pneumonia recommended treatment of  
11  
12 29 pneumonia characterised by indrawing with injectable penicillin alone in inpatient settings  
13  
14 30 until early 2016. At this point, they were revised becoming consistent with WHO guidance  
15  
16 31 after results of a Kenyan trial provided further evidence of equivalence of oral amoxicillin  
17  
18 32 and injectable penicillin. This change also made possible use of oral amoxicillin for  
19  
20 33 outpatient treatment in this patient group. However, given non-trivial mortality in Kenyan  
21  
22 34 children with indrawing pneumonia it remained possible they would benefit from a broader  
23  
24 35 spectrum antibiotic regimen. Therefore, we compared the effectiveness of injectable  
25  
26 36 penicillin monotherapy with a regimen combining penicillin with gentamicin.

27  
28  
29  
30 **Setting**

31  
32  
33 38 We used a large routine observational dataset that captures data on all admissions to 13  
34  
35 39 Kenyan county hospitals.

36  
37  
38 **Participants and measures**

39  
40  
41 41 The analyses included children aged 2 – 59 months. Selection of study population was based  
42  
43 42 on inclusion criteria typical of a prospective trial, primary analysis (experiment 1, n = 4002),  
44  
45 43 but we also explored more pragmatic inclusion criteria (experiment 2, n = 6420) as part of a  
46  
47 44 secondary analysis. To overcome the challenges associated with the non – random allocation  
48  
49 45 of treatments and missing data, we used propensity score(PS) methods and multiple  
50  
51 46 imputation to minimize bias. Further, we estimated mortality risk ratios using log binomial  
52  
53 47 regression and conducted sensitivity analyses using an instrumental variable and PS  
54  
55 48 trimming.

1  
2  
3 49 **Results**  
4  
5

6 50 The estimated risk of dying, in experiment 1, in those receiving penicillin plus gentamicin  
7  
8 51 was 1.46 [0.85, 2.43] compared to the penicillin monotherapy group. In experiment 2, the  
9  
10 52 estimated risk was 1.04 [0.76, 1.40].  
11  
12

13 53 **Conclusion**  
14  
15

16 54 There is no statistical difference in the treatment of indrawing pneumonia with either  
17  
18 55 penicillin or penicillin plus gentamicin. By extension it is unlikely that treatment with  
19  
20 56 penicillin plus gentamicin would offer an advantage to treatment with oral amoxicillin.  
21  
22

23 57 **Strength**  
24

25 58 - This study provides a platform to explore effectiveness of alternative treatments in  
26  
27 59 routine care in a low income setting to improve health outcomes for children.  
28  
29

30 60 **Limitations**  
31  
32

33 61 - The analysis is limited to the variables in the observational dataset – and therefore risk  
34  
35 62 bias due to unmeasured key variables.  
36  
37  
38 63 - The influence of any resulting bias, to alter results, has however been assessed  
39  
40 64 through the use of alternative methods as instrumental variables.  
41  
42

43 65  
44  
45 66  
46  
47 67  
48  
49 68  
50  
51 69  
52  
53 70  
54  
55 71  
56  
57 72  
58  
59 73  
60

1  
2  
3 744  
5 **75 Introduction**

6  
7  
8 76 World Health Organisation (WHO) recommendations guide treatment for millions of children  
9  
10 77 with pneumonia every year across low and middle income countries (1). These guidelines are  
11  
12 78 largely based on moderate certainty in evidence of effects (2-5). However, trials supporting  
13  
14 79 recommendations for hospitalized children have included fewer participants from Africa than  
15  
16 80 other settings (6) and it is suggested that African children with pneumonia have higher  
17  
18 81 mortality (7). Additionally, trial populations may not always include the heterogeneous  
19  
20 82 populations presenting for care, many of whom at hospital level may have co-morbidity (8).  
21  
22 83 Thus despite improving access to recommended treatments and deployment of childhood  
23  
24 84 vaccines at high coverage, including those against *H. influenzae* Type B and pneumococcus,  
25  
26 85 clinically diagnosed pneumonia remains one of the top causes of mortality for children under  
27  
28 86 five in Kenya and other countries (7). According to the Global Health Data exchange  
29  
30 87 (GHDx) website (9), pneumonia caused about 212 under five deaths per 100 000 admission  
31  
32 88 cases in 2015 (which was the highest compared to diarrhoea/dehydration and malaria which  
33  
34 89 are the other top causes of under-five mortality in Kenya). The comparison of mortality rates  
35  
36 90 between 2000 and 2015 for pneumonia, diarrhoea/dehydration and malaria is presented in the  
37  
38 91 supplementary material figure A. The basic and pneumococcal vaccine coverage by 2014 for  
39  
40 92 children aged 12 – 23 months in Kenya was at least 80% (10).

41  
42  
43  
44  
45  
46 93 In a recent change to guidance it is now recommended that pneumonia characterized by lower  
47  
48 94 chest wall indrawing be treated in outpatient settings with oral medication (Box 1) (11, 12).

49  
50 95 Yet it remains associated with non-trivial mortality that may be higher outside trial  
51  
52 96 populations (13). Residual mortality may be associated with causes that are not prevented by  
53  
54 97 currently available conjugate vaccines and organisms, which are not susceptible to the  
55  
56 98 antibiotics currently recommended. Establishing whether there are benefits of alternative  
57  
58  
59  
60

1  
2  
3 99 treatment regimens to help reduce mortality would ideally require large, pragmatic clinical  
4  
5 100 trials (14, 15). However, these remain relatively expensive and time consuming.  
6  
7 101 Observational data may support comparative effectiveness analyses of alternative treatments,  
8  
9 102 may be cheaper and quicker, and may enable evaluation of interventions for which  
10  
11 103 randomization is difficult (16). We use observational data from Kenya to address an  
12  
13 104 important contemporary question for the treatment of pneumonia, a comparison of the  
14  
15 105 effectiveness of gentamicin plus penicillin versus penicillin alone for the treatment of  
16  
17 106 indrawing pneumonia in routine settings. The only previous clinical trial comparing these  
18  
19 107 treatments was a small study of 40 patients in Malaysia (17). In so doing we examine the  
20  
21 108 potential of using data collected by providers as part of their routine practice for comparative  
22  
23 109 effectiveness research in an African setting.  
24  
25  
26  
27

## 28 **Methods**

### 29 *Clinical definitions of pneumonia, primary and secondary analyses.*

30  
31  
32 112 The WHO and Kenyan pneumonia treatment guidelines are implicitly based on risk  
33  
34 113 stratification of illness with children deemed at higher risk of mortality offered broader  
35  
36 114 spectrum antibiotic regimens and those at lower risk narrower spectrum antibiotics (11, 18-  
37  
38 115 20). We present three categories of clinically diagnosed pneumonia in Box 1. This  
39  
40 116 categorization outlines previous and recently revised WHO and Kenyan pneumonia treatment  
41  
42 117 guidelines (11, 19). What we refer to as indrawing pneumonia may be associated with low  
43  
44 118 but clinically significant mortality rates (13, 21). Prior to March 2016 recommended  
45  
46 119 treatment for this group was penicillin monotherapy and our aim is to examine whether there  
47  
48 120 is any advantage of broader spectrum antibiotics in this group. Since March 2016 new  
49  
50 121 guidelines recommend outpatient treatment with oral amoxicillin for this group on the basis  
51  
52 122 of trials suggesting equivalence of amoxicillin and penicillin. However, as indicated above  
53  
54 123 very few patients had been included in studies comparing narrow (amoxicillin or penicillin)  
55  
56  
57  
58  
59  
60

1  
2  
3 124 and broader spectrum antibiotic regimens. As indicated above, beyond the confines of clinical  
4  
5 125 trials amongst all children being treated for indrawing pneumonia, clinical outcomes  
6  
7 126 (including mortality) are worse than seen in the trials (7) and clinicians are often choosing not  
8  
9  
10 127 to use a single drug regime and are in fact often opting to use the combination of gentamicin  
11  
12 128 and penicillin in the group meeting criteria for indrawing pneumonia in real life settings (22).  
13  
14 129 As mortality is higher in real life settings than in trials and as the possibility that broad  
15  
16 130 spectrum antibiotics could have an advantage over monotherapy with penicillin (or  
17  
18 131 amoxicillin) has not been explored in Kenya's previous trials, we feel that examining whether  
19  
20  
21 132 broad spectrum antibiotics confer an advantage is an important question.  
22  
23  
24 133



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

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15
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.*

16  
17  
18  
19

*The previous WHO (41) and pre-2016 Kenyan guidelines (20) named this class as “very severe pneumonia”.*

- 20  
21  
22  
23  
24
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.*

25  
26  
27  
28  
29  
30  
31  
32

*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.*

33  
34  
35  
36  
37  
38  
39

*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.*

40  
41  
42

**Note:** *The term indrawing pneumonia is hereafter used in this analysis to define this category of children to avoid confusion.*

- 43  
44  
45  
46  
47  
48  
49  
50  
51  
52
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.*

134

135

1  
2  
3 136 The ability to use routine data to compare treatment effects requires that patients with similar  
4  
5 137 problems receive different treatments. Previous studies conducted in Kenya and elsewhere  
6  
7 138 have indicated that clinicians often do not follow guideline recommendations in treating  
8  
9 139 pneumonia (22). Variation from the guideline recommended approach can occur at the point  
10  
11 140 of pneumonia severity assignment (clinicians do not follow a nationally approved protocol  
12  
13 141 linking clinical signs and severity category outlined in Box 1) and at the point of treatment  
14  
15 142 assignment (clinicians do not follow this protocol that links treatment and severity). This  
16  
17 143 variability in adherence to protocols provides the opportunity for comparative effectiveness  
18  
19 144 evaluation. More specifically, the adherence and non – adherence to treatment protocols by  
20  
21 145 clinicians allows us to classify indrawing pneumonia admissions in two ways:

- 22  
23  
24  
25  
26 146 1) Those with clinical signs placing them in the group of indrawing pneumonia  
27  
28 147 irrespective of the category or classification assigned to the child by the clinician.  
29  
30  
31 148 2) Those given a clinician classification of indrawing pneumonia irrespective of the  
32  
33 149 actual clinical signs observed by the clinician.  
34  
35

36 150 Based on these two possibilities two experiments were designed (see protocol in press (23))  
37  
38 151 with specific objectives as follows<sup>1</sup>:

- 39  
40  
41 152 1) **Experiment 1:** To compare effectiveness of injectable penicillin versus penicillin  
42  
43 153 plus gentamicin (both injectable) in treatment of indrawing pneumonia; where the  
44  
45 154 child is identified as belonging to a population of children with indrawing pneumonia  
46  
47 155 on the basis of data on their recorded clinical signs. The Experiment 1 population of  
48  
49 156 indrawing pneumonia is therefore consistent with pre-2016 clinical guideline  
50  
51 157 recommendations.  
52  
53  
54

---

55  
56 <sup>1</sup> All children with danger signs were excluded from experiment 1 and in general (both in experiments 1 and 2),  
57 children with the following comorbidities were excluded: HIV, meningitis, tuberculosis and or acute severe  
58 malnutrition.  
59  
60

1  
2  
3 158 2) **Experiment 2:** To compare effectiveness of injectable penicillin versus penicillin  
4  
5 159 plus gentamicin in a population in which we use the clinician assigned categorisation  
6  
7 160 of indrawing pneumonia, which may not be consistent with clinical guideline  
8  
9  
10 161 recommendations.

11  
12 162 We defined Experiment 1 as our primary analysis as we propose it would identify a  
13  
14 163 population similar to that recruited to a randomised trial where the inclusion criteria would be  
15  
16 164 based on specified clinical signs. Experiment 2 offers a scenario that may represent a more  
17  
18 165 pragmatic study design with inclusion criteria based around a clinician led classification.  
19  
20

21  
22 166 ***Data source***

23  
24 167 We use data from the Kenyan Clinical Information Network (CIN) that was initiated to  
25  
26 168 improve inpatient paediatric data availability from county (formerly district) hospitals.  
27  
28 169 Thirteen county referral hospitals were purposively selected with direction from Ministry of  
29  
30 170 Health (MOH) and recruited into the CIN. These hospitals were recruited into the study at  
31  
32 171 different times; four in September 2013, five in October 2013 and four in February 2014.  
33  
34 172 This analysis utilises data up to March 2016. On average, 25 000 paediatric admissions are  
35  
36 173 captured per year. These hospitals typically have one paediatrician leading services  
37  
38 174 predominantly provided by junior clinical teams. Data in these hospitals are collected  
39  
40 175 prospectively post discharge by trained data clerks guided by well-defined standard operating  
41  
42 176 procedures, under supervision by the hospital medical records department and the research  
43  
44 177 team. Clinicians admitting patients fill standardized Paediatric Admission Record (PAR)  
45  
46 178 forms (24) that have been shown to improve documentation of clinical symptoms and signs  
47  
48 179 (25). Together with discharge forms, treatment sheets and laboratory reports these are all part  
49  
50 180 of the patient files that are the primary data source. This data collection system has been  
51  
52 181 described in detail elsewhere (26). Feedback to hospitals as part of the CIN activities has  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 182 helped improve the quality of clinical data (26). The description of hospital selection and  
4  
5 183 their populations of patients is detailed elsewhere (27).  
6  
7

## 8 184 *Statistical analysis*

### 9 185 *i) Defining per protocol and intention to treat populations*

10 186 In typical randomised controlled trials, types of analyses to be conducted are defined  
11  
12 187 beforehand – and this involves defining the type of patient populations that are included in  
13  
14 188 the analyses. Intention to treat and per protocol populations derived from observational  
15  
16 189 datasets have been described in Danaei (2013) (28). We defined per protocol and intention to  
17  
18 190 treat populations based on the dates actual treatments were recorded as prescribed for patients  
19  
20 191 included in our primary and secondary analyses (experiments 1 and 2 respectively). Within  
21  
22 192 each experiment, and after applying inclusion and exclusion criteria, we define the per  
23  
24 193 protocol population as those whose prescription of one of the two study regimens did not  
25  
26 194 change during the admission. The intention to treat population is defined by the original  
27  
28 195 treatment assignment and included children in whom treatment was subsequently changed  
29  
30 196 (see Figure 1 in the Results section).  
31  
32  
33  
34  
35  
36

### 37 197 *ii) Dealing with missing data and propensity score matching*

38  
39 198 As CIN comprises data from routine care settings it faces challenges of non – random  
40  
41 199 treatment allocation and missing data. The missing data and propensity score methods for this  
42  
43 200 analysis have been detailed in the protocol in press linked to this work (23). In brief, after  
44  
45 201 exploring the patient populations, 20 datasets<sup>2</sup> (29) were derived using multiple imputation  
46  
47 202 (with chained equations) for each experiment (all the variables in both the experiments had  
48  
49 203 missing data less than 30% – see table 2 b of the supplementary material). Clinical signs and  
50  
51 204 symptoms data considered were those recorded by clinicians before patients were admitted.  
52  
53  
54  
55  
56

---

57 <sup>2</sup> The current literature (29) recommends the use of more than 5 imputed datasets and therefore 20 should be  
58 sufficient.  
59  
60

1  
2  
3 205 The multiple imputation excluded outcome data as guidance on the use of observational  
4  
5 206 datasets for comparative effectiveness analysis recommends exclusion of outcome data in the  
6  
7 207 design phase (30). Following this, those with missing outcome data were excluded from the  
8  
9 208 analysis (missingness in the outcome data were 0.5% and 0.8% for experiments 1 and 2). For  
10  
11 209 each imputed dataset, patients in the alternative treatment groups (penicillin monotherapy  
12  
13 210 versus penicillin plus gentamicin) were then matched using propensity score (PS) methods to  
14  
15 211 overcome non – random treatment allocation. Propensity scores define the probability of  
16  
17 212 belonging to or being assigned a given treatment based on signs and symptoms (31). PS is a  
18  
19 213 distance measure (32) which is used as a means to overcome allocation bias as treatment  
20  
21 214 outcomes in children with similar propensity scores can then be compared. In these analyses  
22  
23 215 we compared three approaches to reducing possible bias based on PS – optimal full matching,  
24  
25 216 weighting and sub-classification (31, 32). All are aimed at creating groups of patients that are  
26  
27 217 comparable in terms of the distribution of observed signs and symptoms. For each  
28  
29 218 experiment, in order to select the optimum PS implementation method, absolute standardised  
30  
31 219 mean differences (ASMD) were used as diagnostic checks for covariate balance and overlap  
32  
33 220 (33, 34) between the alternative treatment groups. PS methods that resulted in the minimum  
34  
35 221 average absolute standardised mean differences for the majority of the variables while  
36  
37 222 retaining the largest number of patients in the analysis were considered the most appropriate  
38  
39 223 (32).

40  
41  
42  
43  
44  
45  
46 224 *iii) Analytic modelling and sensitivity analyses*

47  
48 225 In sample size calculations conducted prior to the experiments (presented in greater detail  
49  
50 226 elsewhere (protocol)), it was estimated that a sample size of at least 4000 would be sufficient  
51  
52 227 for the planned experiments to detect a minimum difference of 1.5% in mortality between the  
53  
54 228 two treatment groups. The sample size for experiment one was 4002 and experiment two  
55  
56 229 6420 (including 3312 of those that were also in experiment 1). In other words, experiment 2  
57  
58  
59  
60

1  
2  
3 230 largely included those in the experiment 1 population but also children not meeting eligibility  
4  
5 231 criteria for experiment 1. For each of the experiments, after multiple imputation,  
6  
7 232 multivariable log-binomial regression models were fitted to PS weighted datasets and  
8  
9 233 adjusting for all the variables also used in the PS models (also as a form of sensitivity  
10  
11 234 analyses, treatment effects were estimated on PS unweighted datasets). Only pooled  
12  
13 235 treatment effect estimates are reported.

14  
15  
16  
17 236 One possibility is that clinicians' treatment assignment is skewed such that patients who  
18  
19 237 appear sicker (having a greater number of clinical signs of more severe illness) are assigned  
20  
21 238 'stronger' or broad spectrum treatment. In this situation as mentioned by Sturmer (2010),  
22  
23 239 specific types of treatment allocation may be more likely associated with increased mortality  
24  
25 240 (35). In theory, the use of propensity scores is supposed to account for such skewed  
26  
27 241 assignment by comparing only outcomes of those with similar propensity scores assumed to  
28  
29 242 suggest they have similar clinical profiles and thus similar risks. PS trimming attempts to  
30  
31 243 tackle this problem further by excluding patients who are at the extremes of the PS  
32  
33 244 distribution to create a population with clinical characteristics that are as homogeneous as  
34  
35 245 possible. We use PS trimming to define a population between the 5% - 95% PS percentiles in  
36  
37 246 a sensitivity analysis.

38  
39  
40  
41  
42 247 In a further sensitivity analysis, we used an instrumental variable to examine the potential  
43  
44 248 influence of any unmeasured variables (36). An instrumental variable method aims to find a  
45  
46 249 proxy randomised experiment in a routine or observational dataset (37). We used  
47  
48 250 weekend/weekday admission as an instrumental variable as it was demonstrated in a study  
49  
50 251 conducted by Berkley (2004) (38) in a Kenyan hospital that children who were admitted  
51  
52 252 during the weekend experienced higher mortality compared to those admitted during the  
53  
54 253 weekdays. This, in theory, implies that the type of treatment and care received depends on the  
55  
56 254 day of admission – and this later determines the type of health outcome of the patient. The  
57  
58  
59  
60

1  
2  
3 255 process of fitting the instrumental variable models has been described in the supplementary  
4  
5 256 material. The two sensitivity approaches described above were done for both primary and  
6  
7 257 secondary analyses.  
8  
9

## 10 258 **Results**

### 11 12 13 259 *a) Creating per protocol and intention to treat populations*

14  
15 260 Examining the dates treatments were given, five treatment arms (per experimental scenario)  
16  
17 261 were defined – specifically those who received: (1) penicillin alone without changes, (2) a  
18  
19 262 combination of penicillin plus gentamicin without changes, (3) penicillin but switched to a  
20  
21 263 combination of penicillin plus gentamicin, (4) penicillin but switched to ceftriaxone, and (5) a  
22  
23 264 combination of penicillin plus gentamicin but switched to ceftriaxone (ceftriaxone is the  
24  
25 265 recommended second line treatment for severe pneumonia). Therefore, per protocol analyses  
26  
27 266 would compare patients in treatment arm 1 versus 2, while intention to treat analyses would  
28  
29 267 compare patients in treatment arms 1, 3, and 4 versus 2 and 5 (figure 1).  
30  
31

32  
33 268 [Insert figure 1]  
34  
35

### 36 269 **Figure 1:** Summary of patients per treatment arm in experiments 1 – 2

37  
38  
39 270 In this analysis, intention to treat populations were considered primary and are reported in  
40  
41 271 experiments 1 and 2 in keeping with clinical trial reporting guidelines. These analyses include  
42  
43 272 a relatively larger number of patients compared to per protocol analyses. The recommended  
44  
45 273 doses of penicillin and gentamicin in these hospitals are 50000 iu/Kg and 7.5 mg/Kg given 4  
46  
47 274 and once daily respectively. Additional data suggest most clinicians prescribed these doses  
48  
49 275 correctly (see table 7 of the supplementary material).  
50  
51

52  
53 276  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 277 *b) Comparing performance of optimal full matching, weighting and PS sub-*  
4  
5 278 *classification in experiments 1 and 2 respectively*  
6  
7

8 279 For each experiment, the three PS implementation methods were compared to determine the  
9  
10 280 one which would result in the least absolute standardised mean differences for most of the  
11  
12 281 variables in the analysis (even though all the three methods resulted in variables with  
13  
14 282 ASMD $\leq$ 10%). For experiment 1, PS weighting performed better than PS optimal full  
15  
16 283 matching and sub-classification and for experiment 2, the performance of weighting was  
17  
18 284 comparable to that of optimal full matching (see figures 2 and 3). In both experiments, PS  
19  
20 285 sub-classification reduced covariate imbalance the least. Thus, in the subsequent sections,  
21  
22 286 outcome analyses are based on PS weighted datasets for both experiments.  
23  
24  
25  
26  
27

28  
29 288 [Insert Figure 2]  
30  
31

32 289 **Figure 2:** Comparing performance of the three PS implementation methods in experiment 1:

33  
34 290 The y – axis contains all the variables used in the PS models. While x – axis shows absolute standardised mean  
35 291 difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value  
36 292 of  $\leq$  10% indicates the method has performed well in creating comparable groups.  
37

38 293

39 294 [Insert figure 3]  
40  
41

42 295 **Figure 3:** Comparing performance of the three PS implementation methods in experiment 2

43 296 *c) Outcome Analysis Results*

44 297 *i) Exploring mortality in raw datasets*  
45  
46 298

47  
48 299 Examining the raw datasets without PS adjustments in experiment 1, the average number of  
49  
50 300 pneumonia deaths (across the 20 imputed datasets) in penicillin plus gentamicin group was  
51  
52 301 33/1363 (2.42%) and in penicillin monotherapy was 26/2639 (0.99%). And for experiment 2,  
53  
54 302 the average number of deaths were 87/2296 (3.79%) and 50/4124 (1.21%) in penicillin plus  
55  
56 gentamicin and penicillin monotherapy groups respectively. Overall, the average number of  
57  
58  
59  
60



1  
2  
3 303 pneumonia deaths in the penicillin plus gentamicin group was approximately two and a half  
4  
5 304 to three times the number of mortality events in the penicillin monotherapy group in  
6  
7 305 experiments 1 and 2 respectively.  
8  
9

10 306

11  
12 307

13  
14  
15  
16 308 ***ii) Modelling mortality risk ratios***

17  
18  
19 309 The analysis considered penicillin monotherapy as the reference group and mortality as the  
20  
21 310 outcome – and therefore a RR greater than one would be interpreted to favour penicillin over  
22  
23 311 penicillin plus gentamicin. For both experiments, the treatment risk ratios (RR) estimated on  
24  
25 312 the unmatched datasets were larger than the RR estimated on datasets obtained through PS  
26  
27 313 weighting (see table 1 for all results). In experiment 2, the PS unadjusted analysis showed  
28  
29 314 that penicillin monotherapy was significantly more effective than penicillin plus gentamicin  
30  
31 315 (1.68 [1.15, 2.36]). However, the PS weighted effect estimate (1.04 [0.76, 1.40]) was much  
32  
33 316 reduced and suggested that use of PS had corrected (to a degree) for allocation bias indicating  
34  
35 317 that there was no statistical difference in mortality outcomes between penicillin plus  
36  
37 318 gentamicin and penicillin monotherapy treatments. We also observed that the adjusted point  
38  
39 319 estimate for any effect difference in experiment 2 (1.04 [0.76, 1.40]) was less than that in  
40  
41 320 experiment 1 (1.46 [0.85, 2.43]). This may be due to an increase in the number of covariables  
42  
43 321 available for PS weighting that could be used in Experiment 2 resulting in closer matching  
44  
45 322 (see table 1 a of the supplementary material).  
46  
47  
48  
49

50  
51 323 ***d) Sensitivity analysis through trimming using 5 – 95% PS population restriction***

52  
53 324 After excluding 10% of the populations as a result of PS trimming in sensitivity analyses for  
54  
55 325 experiments 1 and 2, the resulting sample sizes were 3583 and 5778. The skewed assignment  
56  
57 326 of children to treatment with gentamicin and penicillin is demonstrated by their higher PS  
58  
59  
60

1  
2  
3 327 scores in figure 4 for experiment 1 (and figure 3 in supplementary material for experiment 2).  
4  
5 328 As higher PS scores are associated with the presence of a greater number of clinical signs of  
6  
7 329 illness this also suggests an association between more severe illness and treatment with  
8  
9  
10 330 gentamicin and penicillin. For experiment 1, the estimated average mortality events (on PS  
11  
12 331 unadjusted datasets) were 26/1201 (2.16%) and 24/2382 (1.01%) for penicillin plus  
13  
14 332 gentamicin and penicillin monotherapy groups. While the estimated events in experiment 2  
15  
16 333 were 62/2026 (3.06%) and 46/3752 (1.22%). Thus in sensitivity analyses for both  
17  
18 334 experiments, trimming excluded more mortality events in the penicillin plus gentamicin  
19  
20  
21 335 group compared with the penicillin monotherapy group. The treatment effects estimated  
22  
23 336 using PS weighted models for the restricted populations as a result of PS trimming showed no  
24  
25 337 statistical difference between the two treatments (table 1).  
26  
27

28 338 [Insert figure 4]  
29  
30

31 339 **Figure 4:** Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity  
32  
33 340 scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those  
34  
35 341 who were given penicillin plus gentamicin. While the continuous black line shows the PS distribution for those  
36  
37 342 who received penicillin alone.  
38  
39

40 343  
41

42 344 *e) Sensitivity Analysis through the use of weekend/weekday as an instrumental*  
43 345 *variable*  
44

45 346 In order to assess whether a timing of admission variable would form a natural and random  
46  
47 347 experiment, the distributions of covariates were examined across the levels of the  
48  
49 348 instrumental variable (weekend/weekday) in experiments 1 and 2. The distribution of each of  
50  
51 349 the patient characteristics between weekend and weekday admissions was approximately  
52  
53 350 similar (table 4 in supplementary material) suggesting that weekend/weekday admission  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 351 satisfactorily satisfies one of the criteria as a valid IV (also see supplementary material for the  
4  
5 352 set of criteria for a valid IV).  
6  
7  
8 353 The estimated treatment effects, both in experiments 1 and 2, suggest there is no statistical  
9  
10 354 difference in treating indrawing pneumonia with either penicillin alone or penicillin plus  
11  
12 355 gentamicin. The effect estimates obtained using our IV in both experiments are less than one  
13  
14 356 as compared to those obtained with PS weighting which are greater than one. Biologically,  
15  
16  
17 357 the effectiveness of gentamicin plus penicillin (when administered in correct doses) is  
18  
19 358 expected to be the same or greater than that of penicillin monotherapy. Based on the  
20  
21 359 magnitude and direction of effects, the use of the IV seems to demonstrate that the effects  
22  
23 360 obtained through PS weighting may have had some residual bias. However, it is important to  
24  
25 361 highlight that for all analyses the 95%CI obtained are consistent with the Null Hypothesis of  
26  
27 362 no different effect for the treatments.  
28  
29  
30

31 **Table 1:** Treatment effect estimates (RR (95% C.I))  
32  
33  
34

	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]

35  
36  
37  
38  
39 365  
40

## 41 366 Discussion

42  
43 367 We compared penicillin alone with penicillin plus gentamicin in treatment of indrawing  
44  
45 368 pneumonia in populations with overall mortality of 1.5% and 2% in experiments 1 and 2  
46  
47 369 respectively. There were more fatal events in the penicillin plus gentamicin group than the  
48  
49 370 penicillin group (approximately 2.5 times) and unadjusted analyses pointed, therefore, to a  
50  
51 371 protective effect of penicillin treatment. However, adjusted analyses, both in experiments 1  
52  
53 372 and 2, that aim to account for allocation bias that can result from non-random treatment  
54  
55 373 allocation suggest that there is no appreciable difference in outcomes between penicillin and  
56  
57  
58  
59  
60

1  
2  
3 374 gentamicin plus penicillin treatment of indrawing pneumonia. Such adjusted analyses were  
4  
5 375 conducted with multiple propensity score methods (PS weighting, optimal full matching and  
6  
7 376 sub-classification) and both intention to treat and per protocol populations, all of which  
8  
9  
10 377 showed similar findings (see the provided supplementary material). Propensity score  
11  
12 378 trimming and instrumental variable analyses further support the suggestion that poor outcome  
13  
14 379 in this population is not associated with the antibiotic regimen received.

15  
16  
17 380 Our analyses were conducted using data from over 4,000 children, one hundred times more  
18  
19 381 participants than were included in the only prior randomised controlled trial of penicillin  
20  
21 382 monotherapy and penicillin plus gentamicin in treatment of pneumonia in an Asian  
22  
23 383 population (17). There are continuing concerns of clinically important mortality in children  
24  
25 384 with indrawing pneumonia in Africa (21). This has led to hesitation to adopt new WHO and  
26  
27  
28 385 Kenyan guidelines that now recommend the treatment of indrawing pneumonia as an  
29  
30 386 outpatient using amoxicillin (11, 19). Our results suggest that there are likely to be two  
31  
32 387 distinct issues. Firstly, they suggest that offering broader spectrum injectable antibiotic  
33  
34 388 treatment to children with indrawing pneumonia may not improve outcomes compared to  
35  
36 389 treatment with penicillin monotherapy. As other studies have suggested equivalence between  
37  
38  
39 390 oral (high dose) amoxicillin therapy and injectable penicillin therapy (2-5, 39) it seems likely  
40  
41 391 therefore that oral amoxicillin and penicillin plus gentamicin combination therapy would  
42  
43 392 result in similar outcomes when used to treat indrawing pneumonia. Clinicians should  
44  
45 393 therefore carefully adhere to guidelines for treatment of indrawing pneumonia and avoid  
46  
47 394 using gentamicin helping to prevent any possible toxicity.

48  
49  
50  
51 395 Secondly, however, our results suggest that children fulfilling a definition of indrawing  
52  
53 396 pneumonia based on clinical signs, and having excluded serious co-morbidities, may still  
54  
55 397 have an appreciable risk of mortality irrespective of their antibiotic treatment (1.5% in all  
56  
57 398 children in experiment 1). When clinicians categorise children with indrawing pneumonia

1  
2  
3 399 and imperfectly adhere to clinical sign based guidance mortality tends to be higher (2% in all  
4  
5 400 children in experiment 2). These findings point to as yet uncharacterised risk factors that  
6  
7 401 could be important in determining which children need admission to hospital as current  
8  
9 402 guidance indicates that all those with indrawing pneumonia can be treated as an outpatient.  
10  
11 403 While offering an alternative antibiotic to amoxicillin to this group may not improve  
12  
13 404 outcomes it is possible that closer and continuing observation in hospital may help identify  
14  
15 405 co-morbid or alternative conditions that are contributing to this mortality and that may be  
16  
17 406 treated.

18  
19  
20  
21 407 The trials that informed the basis for the revised WHO guidelines (2-5) showed extremely  
22  
23 408 low mortality (0 – 0.2%) suggesting that the populations included in such trials may not be  
24  
25 409 directly representative of all those to whom guidelines are applied in routine settings. In the  
26  
27 410 trial by Agweyu (2015) conducted in Kenya (which compared penicillin versus oral  
28  
29 411 amoxicillin for indrawing pneumonia) overall mortality was 0.8% (39). In a parallel  
30  
31 412 observational cohort providing data from the same hospitals over the same time period for  
32  
33 413 children treated with penicillin alone but not included in the Kenyan trial mortality was not  
34  
35 414 significantly different but marginally higher at 1.2% (Agweyu (2017), submitted) perhaps  
36  
37 415 suggesting that even the limited exclusion criteria in this pragmatic trial might result in  
38  
39 416 exclusion of some sicker children. Taken together with data from the analyses presented here  
40  
41 417 it does appear there is a need to explore whether guidelines might be modified to  
42  
43 418 accommodate additional clinical risk factors for possible life-threatening illness that should  
44  
45 419 prompt admission. In a population with high coverage with conjugate vaccines this may more  
46  
47 420 usefully be for more rigorous evaluation to identify alternative diagnoses or for improved  
48  
49 421 supportive care than for different antibiotics.

50  
51  
52  
53  
54  
55 422 **Strengths and limitations**  
56  
57  
58  
59  
60

1  
2  
3 423 Conducting comparative effectiveness analyses using observational datasets can offer the  
4  
5 424 advantage of larger sample sizes at lower cost than randomised controlled clinical trials. They  
6  
7 425 also include patients that may not qualify for enrolment in a typical explanatory randomised  
8  
9 426 controlled trial – and therefore perhaps provide more true to life estimates of treatment effects  
10  
11 427 similar to those observed in highly pragmatic trials (15). However, as most observational  
12  
13 428 datasets are not meant for research, they have challenges of non-random treatment allocation  
14  
15 429 and missing data. We employed a rigorous ‘experimental design’ strategy as is recommended  
16  
17 430 when using observational data (30). We used PS and multiple imputation methods in an effort  
18  
19 431 to minimise bias due to non-random treatment allocation and missing data and analyses  
20  
21 432 suggested no appreciable difference in outcomes of indrawing pneumonia treated with  
22  
23 433 penicillin alone compared with penicillin plus gentamicin. This was in contrast to unadjusted  
24  
25 434 regression analyses that pointed towards better outcomes with penicillin alone suggesting  
26  
27 435 presence of allocation bias. As most observational datasets are limited to observed variables,  
28  
29 436 it is important to conduct sensitivity analysis to explore if the estimated effects are potentially  
30  
31 437 sensitive to unmeasured variables. We used an instrumental variable and PS trimming, both  
32  
33 438 supported the idea of no appreciable difference regimens when treating indrawing  
34  
35 439 pneumonia. While there are differences (in terms of magnitude) in the mortality observed in  
36  
37 440 the different groups that suggest some residual bias in treatment allocation, these mortality  
38  
39 441 differences are no greater than might occur by chance after PS adjustment (with the type 1  
40  
41 442 and 2 errors specified in the protocol). In that sense the PS approach may still have  
42  
43 443 limitations but it does allow us to conclude no statistical difference in mortality outcomes  
44  
45 444 between the two treatment arms.  
46  
47  
48  
49  
50

51  
52 445 The WHO recommended guidelines for treating pneumonia have considerable influence on  
53  
54 446 policy and practice in low and middle income countries. While the evidence base and rigour  
55  
56 447 of guideline development have improved considerably there remain few data on their  
57  
58  
59  
60

1  
2  
3 448 effectiveness when implemented in non-trial settings. Even though well-designed, large  
4  
5 449 pragmatic trials would be preferred, we demonstrate that carefully collected routine data may  
6  
7 450 be useful for assessing the effectiveness of alternative treatments (15). Such analyses may  
8  
9 451 become increasingly possible as electronic medical records are deployed in low and middle  
10  
11 452 income countries (40) but it is important that such studies are carefully designed to limit as  
12  
13 453 far as possible the biases that arise from non-random treatment allocation (30). Our results  
14  
15 454 suggest that children with indrawing pneumonia may gain little benefit from treatment with  
16  
17 455 broader spectrum antibiotic regimens. However, they also suggest that further work is needed  
18  
19 456 to identify those who are at higher risk of death who might be prioritised for an inpatient  
20  
21 457 diagnostic work up and improved supportive care rather than treated as outpatients.  
22  
23  
24

#### 25 26 458 **Additional files**

27  
28  
29 459 **Additional file 1:** Analysis protocol (published in BMJ Open)

30  
31 460 **Additional file 2:** Supplementary material

#### 32 33 34 461 **Acknowledgement**

35  
36  
37 462 We would like to thank Ambrose Agweyu for his comments that helped improve this  
38  
39 463 manuscript. We also appreciate the valuable contribution offered by the CIN team: Lydia  
40  
41 464 Thurania & Grace Ochieng' (Kiambu County Hospital), Barnabas Kigen (Busia County  
42  
43 465 Hospital), Melab Musabi & Rachel Ingina (Kitale County Hospital), Anne Kamunya & Sam  
44  
45 466 Otido (Embu County Hospital), Margaret Kuria (Kisumu East County Hospital), Agnes  
46  
47 467 Mithamo & Francis Kanyingi (Nyeri County Hospital), Celia Muturi, Caren Emadau &  
48  
49 468 Cecilia Mutiso (Mama Lucy Kibaki County Hospital), David Kimutai & Loice Mutai  
50  
51 469 (Mbagathi County Hospital), Nick Aduro (Kakamega County Hospital), Samuel Ng'arng'ar  
52  
53 470 (Vihiga County Hospital). Fred Were & David Githanga (Kenya Paediatric Association).  
54  
55 471 Rachel Nyamai (Ministry of Health).  
56  
57  
58  
59  
60



1  
2  
3 472  
4  
5  
6

7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

473 **Funding**

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

479 **Availability of Data**

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

482 **Authors Contributions**

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

486 **Competing Interests**

487 The authors declare they have no competing interests.

488 **Ethics approval**

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

491

492

493

494

495

496

497

498



1  
2  
3 499  
4 500  
5 501  
6 502  
7 503  
8 504  
9 505  
10 506  
11 507

12  
13  
14 **References**

- 15  
16 509 1. Olayemi E, Asare EV, Benneh-Akwasi Kuma AA. Guidelines in lower-middle income  
17 510 countries. *Br J Haematol*. 2017.
- 18  
19  
20 511 2. Atkinson M, et al. A multicentre randomised controlled equivalence trial comparing oral  
21 512 amoxicillin and intravenous benzyl penicillin for community acquired pneumonia in children  
22 513 PIVOT Trial. *Thorax*. 2007;61:1102-6.
- 23  
24 514 3. Hazir T, et al. Ambulatory short-course high-dose oral amoxicillin for treatment of severe  
25 515 pneumonia in children: a randomised equivalency trial. *Lancet*. 2008;371:49-56.
- 26  
27  
28 516 4. Addo-Yobo E, Chisaka N, Hassan M, Hibberd P, Lozano JM, Jeena P, et al. Oral  
29 517 amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59  
30 518 months: a randomised multicentre equivalency study. *Lancet*. 2004;364:1141-8.
- 31  
32 519 5. Addo-Yobo E, et al. Outpatient treatment of children with severe pneumonia with oral  
33 520 amoxicillin in four countries: the MASS study. *Tropical Medicine and International Health*.  
34 521 2011;16:995-1006.
- 35  
36 522 6. Mulholland K, Carlin JB, Duke T, Weber M. Challenges of trials of antibiotics for  
37 523 pneumonia in low-income countries. *The Lancet Respiratory Medicine*. 2014;2(12):952-4.
- 38  
39  
40 524 7. IVAC. Pneumonia and Diarrhoea Progress Report 2014. John Hopkins Bloomberg School  
41 525 of Public Health; 2014.
- 42  
43 526 8. Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, et al. Bacteremia  
44 527 among Children Admitted to a Rural Hospital in Kenya. *The new england journal of*  
45 528 *medicine*. 2005;352:39-47.
- 46  
47 529 9. Global Health Data Exchange [Internet].; 2015 []. Available from:  
48 530 [http://ghdx.healthdata.org/gbd-results-tool?params=querytool-](http://ghdx.healthdata.org/gbd-results-tool?params=querytool-permalink/ee043b99b22a223f41b3e9d38c5c596a)  
49 531 [permalink/ee043b99b22a223f41b3e9d38c5c596a](http://ghdx.healthdata.org/gbd-results-tool?params=querytool-permalink/ee043b99b22a223f41b3e9d38c5c596a).
- 50  
51  
52 532 10. Kenya Demographic and Health Survey [Internet].; 2014 []. Available from:  
53 533 <http://dhsprogram.com/pubs/pdf/FR308/FR308.pdf>.
- 54  
55 534 11. Basic paediatric protocols [Internet].: Kenyan Ministry of Health; 2016 []. Available  
56 535 from: [https://www.tropicalmedicine.ox.ac.uk/\\_asset/file/basic-paediatric-protocols-2016.pdf](https://www.tropicalmedicine.ox.ac.uk/_asset/file/basic-paediatric-protocols-2016.pdf).
- 57  
58  
59  
60

- 1  
2  
3 536 12. WHO. Hospital Care for Children. 2013.  
4  
5 537 13. UNICEF DATA: Monitoring the situation of women and children [Internet].; 2017 [].  
6 538 Available from: <https://data.unicef.org/topic/child-health/pneumonia/>.  
7  
8  
9 539 14. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised  
10 540 controlled trials with non-compliance: a simulation study. *BMJ Open Research Methods*.  
11 541 2014.  
12  
13 542 15. Thadhani R. Formal trials versus observational studies. *Oxford PharmaGenesis*. 2006.  
14  
15 543 16. West SG, Duan N, Pequegnat W, Gaist P, Jarlais DCD, Holtgrave D, et al. Alternatives to  
16 544 Randomised Controlled Trials. *American Public Health Association*. 2008;98(8):1359-66.  
17  
18 545 17. Hasali MAA, Ibrahim MIM, Sulaiman SAS, Ahmad Z, Hasali JBA. A clinical and  
19 546 economic study of community-acquired pneumonia between single versus combination  
20 547 therapy. *Pharm World Sci*. 2005;27(3):249-53.  
21  
22  
23 548 18. Ayieko P, English M. Case Management of Childhood Pneumonia in Developing  
24 549 Countries. *Pediatr Infect Dis J*. 2007;26:432-40.  
25  
26 550 19. Hospital Care for Children [Internet].; 2013 []. Available from:  
27 551 [http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf).  
28  
29  
30 552 20. Basic Paediatric Protocols: November 2013 Edition [Internet].: Ministry of Health; 2013  
31 553 []. Available from: [http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-](http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition)  
32 554 [basic-paediatric-protocols-november-2013-edition](http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition).  
33  
34 555 21. Koh JW, Wong JJ, Sultana R, Wong PP, Mok YH, Lee JH.  
35 556 Risk factors for mortality in children with pneumonia admitted to the pediatric intensive care  
36 557 unit. *Pediatric pulmonology*. 2017.  
37  
38  
39 558 22. Agweyu A, Kibore M, Digolo L, Kosgei C, Maina V, Mugane S, et al. Prevalence and  
40 559 correlates of treatment failure among Kenyan children hospitalised with severe community-  
41 560 acquired pneumonia: a prospective study of the clinical effectiveness of WHO pneumonia  
42 561 case management guidelines. *Tropical Medicine & International Health*. 2014.  
43  
44 562 23. Malla L, Perera R, McFadden E, English M.  
45 563 Comparative effectiveness of injectable penicillin versus a combination of penicillin and  
46 564 gentamicin in children with pneumonia characterised by indrawing in Kenya: A protocol for  
47 565 an observational study. *BMJ Open (in press) [Internet]*. 2017.  
48  
49  
50 566 24. Mwakyusa S, Wamae A, Wasunna A, Were F, Esamai F, Ogotu B, et al.  
51 567 Implementation of a structured paediatric admission record for district hospitals in Kenya –  
52 568 results of a pilot study. *BMC International Health and human rights*. 2006.  
53  
54  
55 569 25. Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A Multifaceted  
56 570 Intervention to Implement Guidelines and Improve Admission Paediatric Care in Kenyan  
57 571 District Hospitals: A Cluster Randomised Trial. *PLoS Medicine*. 2011.  
58  
59  
60

- 1  
2  
3 572 26. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation  
4 573 of clinical care within a clinical information network: an essential initial step in efforts to  
5 574 understand and improve care in Kenyan hospitals. *BMJ Global Health*. 2016.
- 7 575 27. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of  
8 576 admissions and variations in the use of basic investigations, treatments and outcomes in  
9 577 Kenyan hospitals within a new Clinical Information Network. *Archives of Disease in*  
10 578 *Childhood*. 2015.
- 13 579 28. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA.  
14 580 Observational data for comparative effectiveness research: an emulation of randomised trials  
15 581 of statins and primary prevention of coronary heart disease. *Stat.Methods Med.Res.* 2013.
- 17 582 29. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues  
18 583 and guidance for practice. *Stat Med*. 2011;30(4):377-99.
- 21 584 30. Rubin D.B. For Objective Causal Inference, Design Trumps Analysis. *The Annals of*  
22 585 *Applied Statistics*. 2008;2(3):808-40.
- 24 586 31. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of  
25 587 Confounding in Observational Studies. *Multivariate Behavioral Research*. 2011.
- 27 588 32. Stuart EA. Matching methods for causal inference: A review and a look forward. *Statist.*  
28 589 *Sci*. 2010;25(1):1-21.
- 31 590 33. Austin PC. Assessing balance in measured baseline covariates when using many-to-one  
32 591 matching on the propensity-score. *Pharmacoepidemiol Drug Saf*. 2008 2008;17(12):1218-25.
- 34 592 34. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates  
35 593 between treatment groups in propensity-score matched samples. *Stat Med*. 2009 10 Nov  
36 594 2009;28(25):3083-107.
- 38 595 35. Stumer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the presence of  
39 596 Unmeasured Confounding: Dealing with observations in the Tails of the Propensity Score  
40 597 Distribution - A Simulation Study. *American Journal of Epidemiology*. 2010.
- 43 598 36. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat*  
44 599 *Med*. 2014 15 Jun 2014;33(13):2297-340.
- 46 600 37. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods  
47 601 for Causal Inference. *Statistics in Medicine*. 2014;33(13):2297-340.
- 49 602 38. Berkley JA, Brent A, Mwangi I, English M, Maitland K, Marsh K, et al. Mortality among  
50 603 Kenyan children admitted to a rural district hospital on weekends as compared with  
51 604 weekdays . *Pediatrics*. 2004;114:1737-8.
- 54 605 39. Agweyu A, Gathara D, Oliwa J, Muinga N, Edwards T, Allen E, et al. Oral amoxicillin  
55 606 versus benzyl penicillin for severe pneumonia among kenyan children: a pragmatic  
56 607 randomised controlled noninferiority trial. *Clin Infect Dis*. 2015.

1  
2  
3 608 40. English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al.  
4 609 Building Learning Health Systems to Accelerate Research and Improve Outcomes of Clinical  
5 610 Care in Low-and Middle-Income Countries. PLOS Medicine. 2016.  
6

7  
8 611 41. Hospital Care for Children [Internet].; 2005 []. Available from:  
9 612 <http://apps.who.int/iris/bitstream/10665/43206/1/9241546700.pdf>.

10  
11 613  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

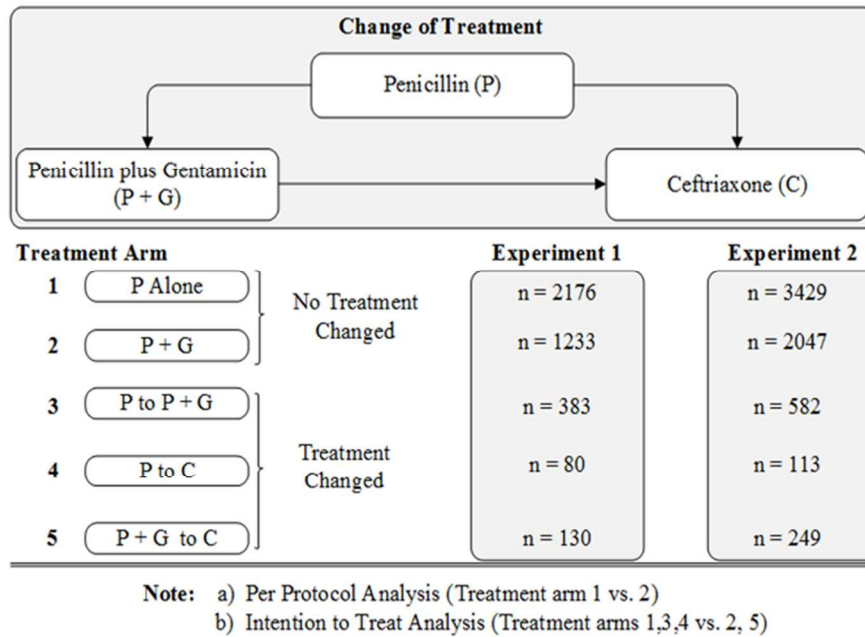


Figure 1: Summary of patients per treatment arm in experiments 1 – 2

63x45mm (600 x 600 DPI)

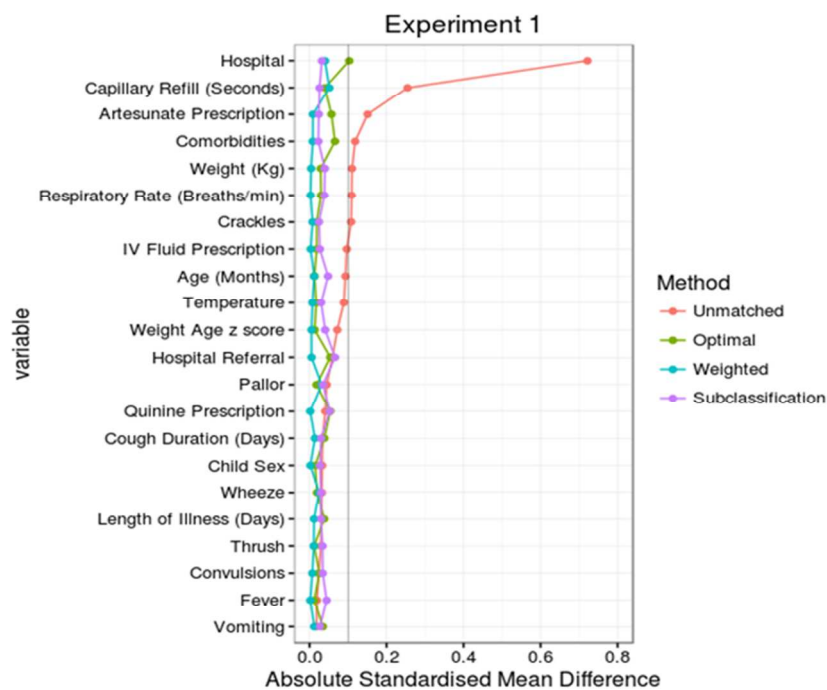


Figure 2: Comparing performance of the three PS implementation methods in experiment 1: The y - axis contains all the variables used in the PS models. While x - axis shows absolute standardised mean difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value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

63x45mm (600 x 600 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

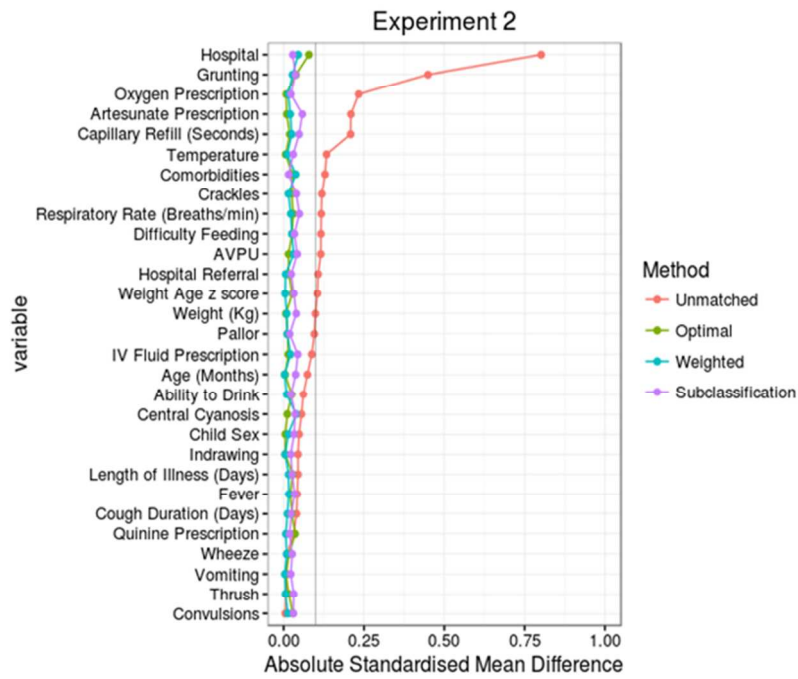


Figure 3: Comparing performance of the three PS implementation methods in experiment 2

63x45mm (600 x 600 DPI)

Review only

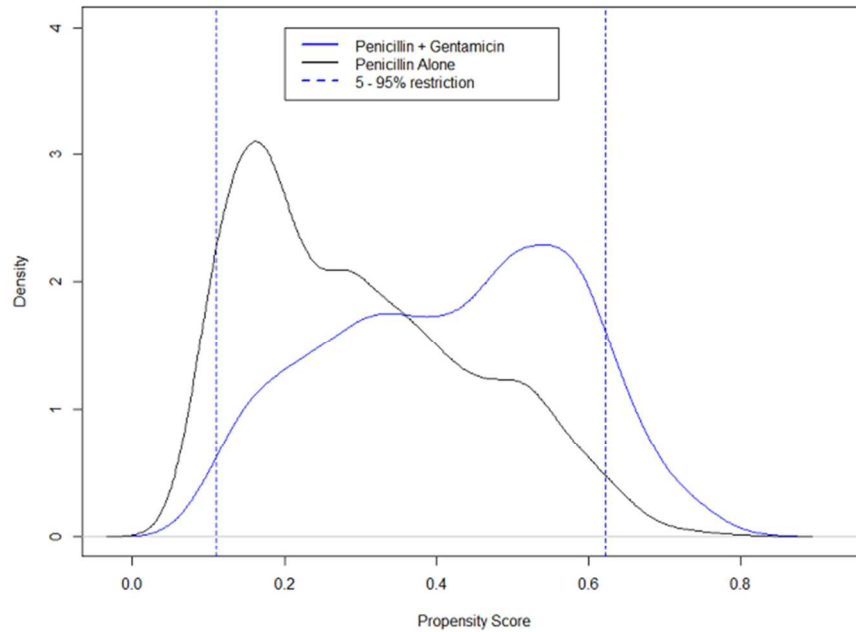


Figure 4: Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those who were given penicillin plus gentamicin. While the continuous black line shows the PS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

63x45mm (600 x 600 DPI)



1  
2  
3 **Using observational data to compare the effectiveness of antibiotic treatments for**  
4 **children hospitalised with pneumonia in Kenya**  
5

6 Lucas Malla<sup>1\*</sup>  
7

8  
9 Rafael Perera-Salazar<sup>2</sup>  
10

11 Emily McFadden<sup>2</sup>  
12

13  
14 Mike English<sup>1,3</sup>  
15

16  
17 <sup>1</sup>Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.  
18

19  
20 <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United  
21  
22 Kingdom.  
23

24  
25 <sup>3</sup>Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

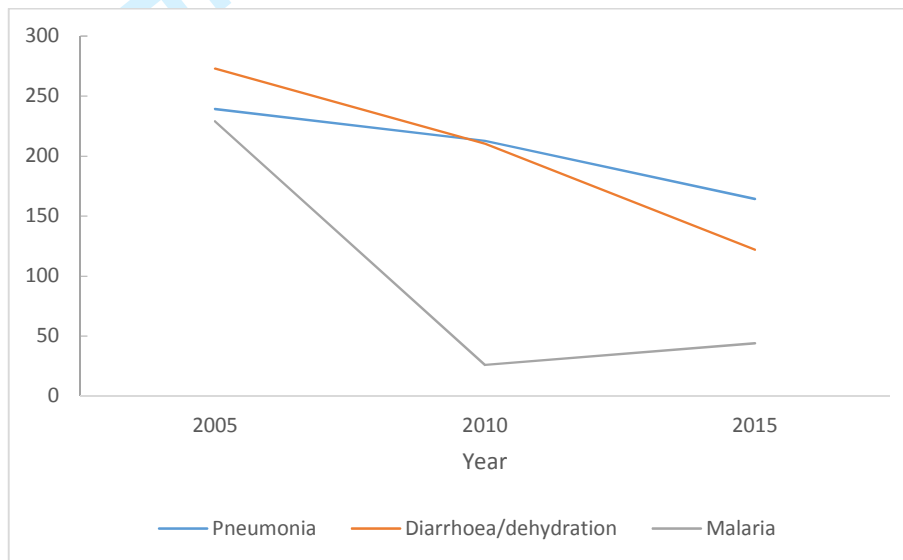
1  
2  
3 The supplementary material is organised into the following subsections:  
4  
5

- 6 • Under-five mortality incident rates in Kenya
- 7
- 8 • Summary of analysis variables
- 9
- 10 • Analysis using PS sub classification
- 11
- 12 • Analysis using optimal full matching
- 13
- 14 • Experiment 2 trimming
- 15
- 16 • Analysis using instrumental variables
- 17
- 18 • Analysis using per protocol population
- 19
- 20 • Overlap and correctness of penicillin and gentamicin dosing
- 21
- 22 • Definition of PS methods and how they were used
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

### 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<sup>1</sup>

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)

<sup>1</sup> 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)

PS Subgroup	Experiment 1		Experiment 2	
	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%)
<b>Total</b>	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

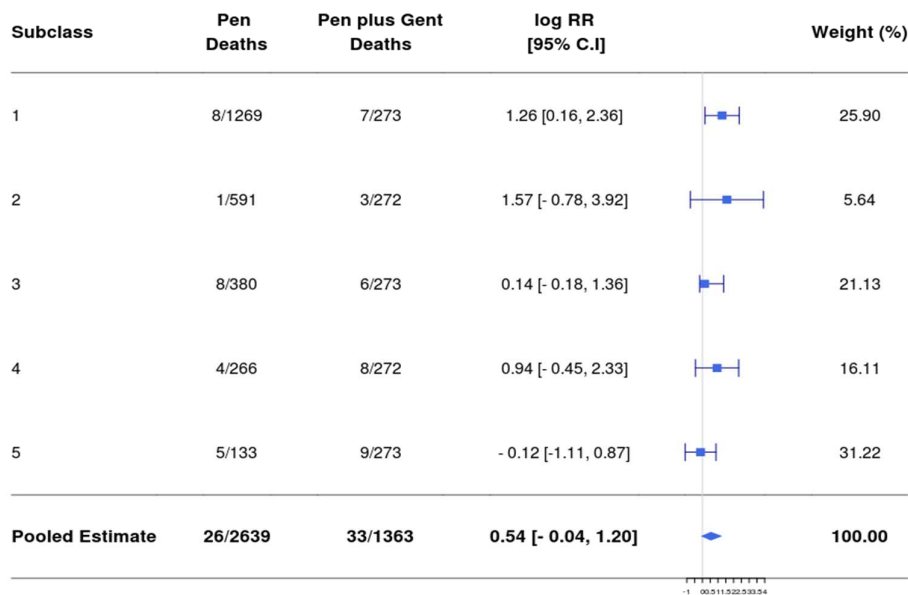
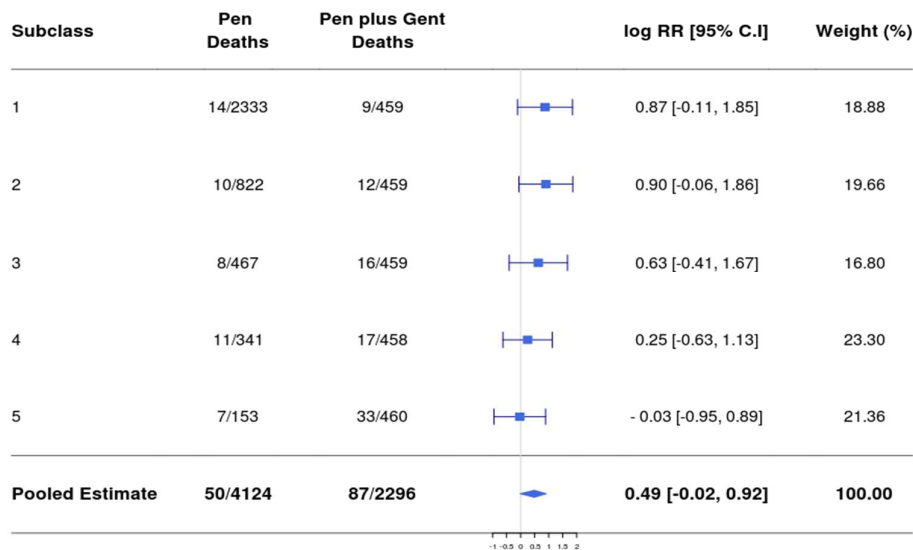


Figure 2: Experiment 2 – ITT



d) Analysis using PS optimal full matching

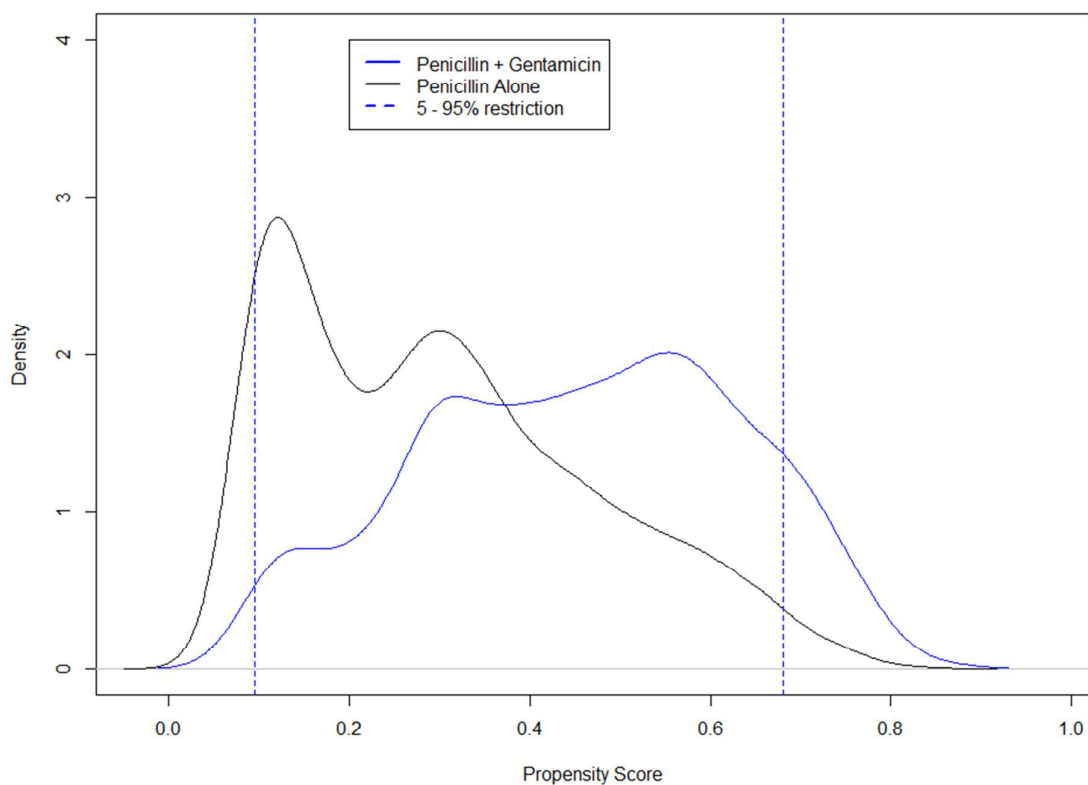
Also analysis using PS optimal full matching showed no statistical significance in treatment of indrawing pneumonia using either penicillin or penicillin plus gentamicin (table 3).

Table 3: Treatment effect estimates

	log RR (95% C.I)
--	------------------

<b>Experiment 1</b>	
Regression without PS adjustment	0.56 [-0.06, 1.02]
PS optimal matching	0.27 [-0.22, 0.65]
<b>Experiment 2</b>	
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
<b>Child Sex</b>						
Female	45%	46%	0.03	44%	45%	0.01
Male	55%	54%		56%	55%	
<b>Pallor</b>						
Mild/moderate	4%	5%	0.02	5%	5%	0.00
None	95%	94%		93%	93%	
Severe	1%	2%		2%	2%	
<b>Capillary refill</b>						
1 sec	68%	71%	0.07	66%	68%	0.04
2 sec	30%	27%		31%	29%	
>2 sec	3%	2%		3%	3%	
<b>Fever</b>						
Absent	21%	18%	0.05	19%	16%	0.07
Present	79%	82%		81%	84%	
<b>Convulsions</b>						
Absent	95%	96%	0.02	94%	94%	0.03
Present	5%	4%		6%	6%	
<b>Vomiting</b>						
No	65%	62%	0.06	63%	63%	0.00
Yes	35%	38%		37%	37%	
<b>Referral</b>						
No	82%	86%	0.10	81%	84%	0.09
Yes	18%	14%		19%	16%	
<b>Thrush</b>						
Absent	98%	98%	0.00	98%	98%	0.03
Present	2%	2%		2%	2%	
<b>Comorbidities</b>						
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%	
<b>Crackles</b>						
Absent	47%	47%	0.01	48%	47%	0.02
Present	53%	53%		52%	53%	
<b>Wheeze</b>						
Absent	85%	84%	0.02	85%	84%	0.02
Present	15%	16%		15%	16%	
<b>IV prescription</b>						
No	97%	96%	0.05	95%	95%	0.01
Yes	3%	4%		5%	5%	
<b>Quinine Prescription</b>						
No	97%	97%	0.02	95%	94%	0.04
Yes	3%	3%		5%	6%	
<b>Artesunate Prescription</b>						
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**

	<b>log RR (95% C.I)</b>
<b>Experiment 1</b>	
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]
<b>Experiment 2</b>	
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

	<b>Experiment 1</b>		<b>Experiment 2</b>	
	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*

1  
2  
3 Sub-classification divides patients into mutually exclusive groups based on their propensity  
4 scores. A standard practice, though not supported by specific recommendations, has been  
5 subdividing patients into five subclasses (6). One approach for creating patient subclasses  
6 would be to first conduct one on one nearest neighbour matching and then split the  
7 population into subclasses (7), alternatively one may use PS quintiles (4). The number of  
8 subclasses will usually depend on the sample size, and for large datasets, more classes with  
9 reasonable sample sizes would be desirable. This analysis used PS quintiles with five  
10 subclasses.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

## 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

### 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:	<i>BMJ Open</i>
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
<b>Primary Subject Heading</b>:	Paediatrics
Secondary Subject Heading:	Infectious diseases, Paediatrics
Keywords:	pneumonia, missing data, propensity scores, comparative effectiveness

SCHOLARONE™  
Manuscripts

1  
2  
3 1 **Comparative effectiveness of injectable penicillin versus a combination of penicillin and**  
4  
5 2 **gentamicin in children with pneumonia characterised by indrawing in Kenya: A**  
6  
7 3 **retrospective observational study**  
8  
9

10 4 Lucas Malla<sup>1\*</sup> ([lmalla@kemri-wellcome.org](mailto:lmalla@kemri-wellcome.org))  
11

12  
13 5 Rafael Perera-Salazar<sup>2</sup> ([rafael.perera@phc.ox.ac.uk](mailto:rafael.perera@phc.ox.ac.uk))  
14

15  
16 6 Emily McFadden<sup>2</sup> ([emily.mcfadden@phc.ox.ac.uk](mailto:emily.mcfadden@phc.ox.ac.uk))  
17

18  
19 7 Mike English<sup>1,3</sup> ([menglish@kemri-wellcome.org](mailto:menglish@kemri-wellcome.org))  
20

21  
22 8 <sup>1</sup>Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.  
23

24  
25 9 <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United  
26  
27 10 Kingdom.  
28

29  
30 11 <sup>3</sup>Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.  
31

32  
33  
34  
35 13 **\*Corresponding author:**  
36

37  
38 14 Lucas Malla  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **25 Abstract**

4  
5 **26 Objectives**

6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
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.

36 **Setting**

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

39 **Participants and measures**

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

48 **Results**

1  
2  
3 49 The estimated risk of dying, in experiment 1, in those receiving penicillin plus gentamicin  
4  
5 50 was 1.46 [0.85, 2.43] compared to the penicillin monotherapy group. In experiment 2, the  
6  
7 51 estimated risk was 1.04 [0.76, 1.40].  
8  
9

## 10 **Conclusion**

11  
12  
13 53 There is no statistical difference in the treatment of indrawing pneumonia with either  
14  
15 54 penicillin or penicillin plus gentamicin. By extension it is unlikely that treatment with  
16  
17 55 penicillin plus gentamicin would offer an advantage to treatment with oral amoxicillin.  
18  
19

## 20 **Strength**

21  
22 57 - This study provides a platform to explore effectiveness of alternative treatments in  
23  
24 58 routine care in a low income setting to improve health outcomes for children.  
25  
26

## 27 **Limitations**

28  
29  
30 60 - The analysis is limited to the variables in the observational dataset – and therefore risk  
31  
32 61 bias due to unmeasured key variables.  
33  
34  
35 62 - The influence of any resulting bias, to alter results, has however been assessed  
36  
37 63 through the use of alternative methods as instrumental variables.  
38  
39

40 64

41 65

42 66

43 67

44 68

45 69

46 70

47 71

48 72

49 73

50

51

52

53



## 74 Introduction

75 World Health Organisation (WHO) recommendations guide treatment for millions of children  
76 with pneumonia every year across low and middle income countries (1). These guidelines are  
77 largely based on moderate certainty in evidence of effects (2-5). However, trials supporting  
78 recommendations for hospitalized children have included fewer participants from Africa than  
79 other settings (6) and it is suggested that African children with pneumonia have higher  
80 mortality (7). Additionally, trial populations may not always include the heterogeneous  
81 populations presenting for care, many of whom at hospital level may have co-morbidity (8).  
82 Thus despite improving access to recommended treatments and deployment of childhood  
83 vaccines at high coverage, including those against *H. influenzae* Type B and pneumococcus,  
84 clinically diagnosed pneumonia remains one of the top causes of mortality for children under  
85 five in Kenya and other countries (7). According to the mortality data derived from the  
86 Global Health Observatory (GHO) Data – published in the WHO website (9), pneumonia  
87 caused about 5.4 under five deaths per 1000 children in 2015 (which was the highest  
88 compared to diarrhoea/dehydration and malaria which are the other top causes of under-five  
89 mortality in Kenya). The comparison of mortality rates between 2000 and 2015 for  
90 pneumonia, diarrhoea/dehydration and malaria is presented in the additional file 1:  
91 supplementary data  
92 figure A. The basic and pneumococcal vaccine coverage by 2014 for children aged 12 – 23  
93 months in Kenya was at least 80% (10).

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

1  
2  
3 98 currently available conjugate vaccines and organisms, which are not susceptible to the  
4  
5 99 antibiotics currently recommended. Establishing whether there are benefits of alternative  
6  
7 100 treatment regimens to help reduce mortality would ideally require large, pragmatic clinical  
8  
9  
10 101 trials (14, 15). However, these remain relatively expensive and time consuming.  
11  
12 102 Observational data may support comparative effectiveness analyses of alternative treatments,  
13  
14 103 may be cheaper and quicker, and may enable evaluation of interventions for which  
15  
16 104 randomization is difficult (16). We use observational data from Kenya to address an  
17  
18 105 important contemporary question for the treatment of pneumonia, a comparison of the  
19  
20 106 effectiveness of gentamicin plus penicillin versus penicillin alone for the treatment of  
21  
22 107 indrawing pneumonia in routine settings. The only previous clinical trial comparing these  
23  
24 108 treatments was a small study of 40 patients in Malaysia (17). In so doing we examine the  
25  
26 109 potential of using data collected by providers as part of their routine practice for comparative  
27  
28 110 effectiveness research in an African setting.

## 111 **Methods**

### 112 *Clinical definitions of pneumonia, primary and secondary analyses.*

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

1  
2  
3 123 of trials suggesting equivalence of amoxicillin and penicillin. However, as indicated above  
4  
5 124 very few patients had been included in studies comparing narrow (amoxicillin or penicillin)  
6  
7 125 and broader spectrum antibiotic regimens. As indicated above, beyond the confines of clinical  
8  
9 126 trials amongst all children being treated for indrawing pneumonia, clinical outcomes  
10  
11 127 (including mortality) are worse than seen in the trials (7) and clinicians are often choosing not  
12  
13 128 to use a single drug regime and are in fact often opting to use the combination of gentamicin  
14  
15 129 and penicillin in the group meeting criteria for indrawing pneumonia in real life settings (22).  
16  
17 130 As mortality is higher in real life settings than in trials and as the possibility that broad  
18  
19 131 spectrum antibiotics could have an advantage over monotherapy with penicillin (or  
20  
21 132 amoxicillin) has not been explored in Kenya's previous trials, we feel that examining whether  
22  
23 133 broad spectrum antibiotics confer an advantage is an important question.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

134

**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.*

135

136

1  
2  
3 137  
4  
5  
6 138  
7  
8 139  
9  
10 140  
11  
12 141  
13  
14 142  
15  
16 143  
17  
18 144  
19  
20 145  
21  
22 146  
23  
24 147  
25  
26  
27  
28  
29 148  
30  
31 149  
32  
33  
34 150  
35  
36 151  
37  
38  
39 152  
40  
41 153  
42  
43  
44 154  
45  
46 155  
47  
48 156  
49  
50 157  
51  
52  
53  
54  
55

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<sup>1</sup>:

- 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

---

<sup>1</sup> 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.

1  
2  
3 158 indrawing pneumonia is therefore consistent with pre-2016 clinical guideline  
4  
5 159 recommendations.  
6

7 160 2) **Experiment 2:** To compare effectiveness of injectable penicillin versus penicillin  
8  
9 161 plus gentamicin in a population in which we use the clinician assigned categorisation  
10  
11 162 of indrawing pneumonia, which may not be consistent with clinical guideline  
12  
13 163 recommendations.  
14

15  
16  
17 164 We defined Experiment 1 as our primary analysis as we propose it would identify a  
18  
19 165 population similar to that recruited to a randomised trial where the inclusion criteria would be  
20  
21 166 based on specified clinical signs. Experiment 2 offers a scenario that may represent a more  
22  
23 167 pragmatic study design with inclusion criteria based around a clinician led classification.  
24

#### 25 26 168 *Data source*

27  
28 169 We use data from the Kenyan Clinical Information Network (CIN) that was initiated to  
29  
30 170 improve inpatient paediatric data availability from county (formerly district) hospitals.  
31

32  
33 171 Thirteen county referral hospitals were purposively selected with direction from Ministry of  
34  
35 172 Health (MOH) and recruited into the CIN. These hospitals were recruited into the study at  
36  
37 173 different times; four in September 2013, five in October 2013 and four in February 2014.  
38

39 174 This analysis utilises data up to March 2016. On average, 25 000 paediatric admissions are  
40  
41 175 captured per year. These hospitals typically have one paediatrician leading services  
42

43 176 predominantly provided by junior clinical teams. Data systems and standardised clinical  
44  
45 177 forms were specifically implemented in all hospitals at the start of this work to optimise the  
46  
47 178 quality of routine data. Patient data in these hospitals are collected post discharge by trained  
48

49 179 data clerks guided by well-defined standard operating procedures, under supervision by the  
50  
51 180 hospital medical records department and the research team. Clinicians admitting patients fill  
52

53 181 standardized Paediatric Admission Record (PAR) forms (26) that have been shown to  
54  
55 182 improve documentation of clinical symptoms and signs (27). Together with discharge forms,  
56

57  
58  
59  
60

1  
2  
3 183 treatment sheets and laboratory reports these are all part of the patient files that are the  
4  
5 184 primary data source. This data collection system has been described in detail elsewhere (28).  
6  
7 185 Feedback to hospitals as part of the CIN activities has helped improve the quality of clinical  
8  
9 186 data (28). The description of hospital selection and their populations of patients is detailed  
10  
11 187 elsewhere (29).  
12  
13

## 14 ***Statistical analysis***

### 15 ***i) Defining per protocol and intention to treat populations***

16  
17  
18 190 In typical randomised controlled trials, types of analyses to be conducted are defined  
19  
20 191 beforehand – and this involves defining the type of patient populations that are included in  
21  
22 192 the analyses. Intention to treat and per protocol populations derived from observational  
23  
24 193 datasets have been described in Danaei (2013) (30). We defined per protocol and intention to  
25  
26 194 treat populations based on the dates actual treatments were recorded as prescribed for patients  
27  
28 195 included in our primary and secondary analyses (experiments 1 and 2 respectively). Within  
29  
30 196 each experiment, and after applying inclusion and exclusion criteria, we define the per  
31  
32 197 protocol population as those whose prescription of one of the two study regimens did not  
33  
34 198 change during the admission. The intention to treat population is defined by the original  
35  
36 199 treatment assignment and included children in whom treatment was subsequently changed  
37  
38 200 (see Figure 1 in the Results section).  
39  
40  
41  
42

### 43 ***ii) Dealing with missing data and propensity score matching***

44  
45  
46 202 As CIN comprises data from routine care settings it faces challenges of non – random  
47  
48 203 treatment allocation and missing data. The missing data and propensity score methods for this  
49  
50 204 analysis have been detailed in the additional file 2: analysis protocol linked to this work (25).  
51  
52 205 In brief, after exploring the patient populations, 20 datasets<sup>2</sup> (31) were derived using multiple  
53  
54  
55

---

56  
57 <sup>2</sup> The current literature (31) recommends the use of more than 5 imputed datasets and therefore 20 should be  
58 sufficient.  
59

1  
2  
3 206 imputation (with chained equations) for each experiment (all the variables in both the  
4  
5 207 experiments had missing data less than 30% – see table A in the additional file 1:  
6  
7 208 supplementary data). Clinical signs and symptoms data considered were those recorded by  
8  
9 209 clinicians before patients were admitted. The multiple imputation excluded outcome data as  
10  
11 210 guidance on the use of observational datasets for comparative effectiveness analysis  
12  
13 211 recommends exclusion of outcome data in the design phase (32). Following this, those with  
14  
15 212 missing outcome data were excluded from the analysis (missingness in the outcome data  
16  
17 213 were 0.5% and 0.8% for experiments 1 and 2). For each imputed dataset, patients in the  
18  
19 214 alternative treatment groups (penicillin monotherapy versus penicillin plus gentamicin) were  
20  
21 215 then matched using propensity score (PS) methods to overcome non – random treatment  
22  
23 216 allocation. Propensity scores define the probability of belonging to or being assigned a given  
24  
25 217 treatment based on signs and symptoms (33). PS is a distance measure (34) which is used as a  
26  
27 218 means to overcome allocation bias as treatment outcomes in children with similar propensity  
28  
29 219 scores can then be compared. In these analyses we compared three approaches to reducing  
30  
31 220 possible bias based on PS – optimal full matching, weighting and sub-classification (33, 34).  
32  
33 221 All are aimed at creating groups of patients that are comparable in terms of the distribution of  
34  
35 222 observed signs and symptoms. For each experiment, in order to select the optimum PS  
36  
37 223 implementation method, absolute standardised mean differences (ASMD) were used as  
38  
39 224 diagnostic checks for covariate balance and overlap (35, 36) between the alternative  
40  
41 225 treatment groups. PS methods that resulted in the minimum average absolute standardised  
42  
43 226 mean differences for the majority of the variables while retaining the largest number of  
44  
45 227 patients in the analysis were considered the most appropriate (34).  
46  
47  
48  
49  
50

51  
52 228 *iii) Analytic modelling and sensitivity analyses*  
53  
54

55 229 In sample size calculations conducted prior to the experiments (presented in greater detail  
56  
57 230 elsewhere (see additional file 2: analysis protocol)), it was estimated that a sample size of at  
58  
59  
60



1  
2  
3 231 least 4000 would be sufficient for the planned experiments to detect a minimum difference of  
4  
5 232 1.5% in mortality between the two treatment groups. The sample size for experiment one was  
6  
7 233 4002 and experiment two 6420 (including 3312 of those that were also in experiment 1). In  
8  
9  
10 234 other words, experiment 2 largely included those in the experiment 1 population but also  
11  
12 235 children not meeting eligibility criteria for experiment 1. For each of the experiments, after  
13  
14 236 multiple imputation, multivariable log-binomial regression models were fitted to PS weighted  
15  
16 237 datasets and adjusting for all the variables also used in the PS models (also as a form of  
17  
18 238 sensitivity analyses, treatment effects were estimated on PS unweighted datasets). Only  
19  
20  
21 239 pooled treatment effect estimates are reported.

22  
23  
24 240 One possibility is that clinicians' treatment assignment is skewed such that patients who  
25  
26 241 appear sicker (having a greater number of clinical signs of more severe illness) are assigned  
27  
28 242 'stronger' or broad spectrum treatment. In this situation as mentioned by Stürmer (2010),  
29  
30 243 specific types of treatment allocation may be more likely associated with increased mortality  
31  
32 244 (37). In theory, the use of propensity scores is supposed to account for such skewed  
33  
34 245 assignment by comparing only outcomes of those with similar propensity scores assumed to  
35  
36 246 suggest they have similar clinical profiles and thus similar risks. PS trimming attempts to  
37  
38 247 tackle this problem further by excluding patients who are at the extremes of the PS  
39  
40 248 distribution to create a population with clinical characteristics that are as homogeneous as  
41  
42 249 possible. We use PS trimming to define a population between the 5% - 95% PS percentiles in  
43  
44 250 a sensitivity analysis.

45  
46  
47  
48  
49 251 In a further sensitivity analysis, we used an instrumental variable to examine the potential  
50  
51 252 influence of any unmeasured variables (38). An instrumental variable method aims to find a  
52  
53 253 proxy randomised experiment in a routine or observational dataset (39). We used  
54  
55 254 weekend/weekday admission as an instrumental variable as it was demonstrated in a study  
56  
57 255 conducted by Berkley (2004) (40) in a Kenyan hospital that children who were admitted

1  
2  
3 256 during the weekend experienced higher mortality compared to those admitted during the  
4  
5 257 weekdays. This, in theory, implies that the type of treatment and care received depends on the  
6  
7 258 day of admission – and this later determines the type of health outcome of the patient. The  
8  
9  
10 259 process of fitting the instrumental variable models has been described in the additional file 1:  
11  
12 260 supplementary data. The two sensitivity approaches described above were done for both  
13  
14 261 primary and secondary analyses.

## 17 262 **Results**

### 20 263 *a) Creating per protocol and intention to treat populations*

21 264 Examining the dates treatments were given, five treatment arms (per experimental scenario)  
22  
23 265 were defined – specifically those who received: (1) penicillin alone without changes, (2) a  
24  
25 266 combination of penicillin plus gentamicin without changes, (3) penicillin but switched to a  
26  
27 267 combination of penicillin plus gentamicin, (4) penicillin but switched to ceftriaxone, and (5) a  
28  
29 268 combination of penicillin plus gentamicin but switched to ceftriaxone (ceftriaxone is the  
30  
31 269 recommended second line treatment for severe pneumonia). Therefore, per protocol analyses  
32  
33 270 would compare patients in treatment arm 1 versus 2, while intention to treat analyses would  
34  
35 271 compare patients in treatment arms 1, 3, and 4 versus 2 and 5 (figure 1).

36  
37  
38  
39  
40 272 [Insert figure 1]

### 43 273 **Figure 1:** Summary of patients per treatment arm in experiments 1 – 2

44  
45  
46 274 In this analysis, intention to treat populations were considered primary and are reported in  
47  
48 275 experiments 1 and 2 in keeping with clinical trial reporting guidelines. These analyses include  
49  
50 276 a relatively larger number of patients compared to per protocol analyses. The recommended  
51  
52 277 doses of penicillin and gentamicin in these hospitals are 50000 iu/Kg and 7.5 mg/Kg given 4  
53  
54 278 and once daily respectively. Additional data suggest most clinicians prescribed these doses  
55  
56  
57 279 correctly (see table B in the additional file 1: supplementary data).

1  
2  
3 280  
4  
5

6 281 ***b) Comparing performance of optimal full matching, weighting and PS sub-***  
7  
8 282 ***classification in experiments 1 and 2 respectively***

9  
10  
11 283 For each experiment, the three PS implementation methods were compared to determine the  
12  
13 284 one which would result in the least absolute standardised mean differences for most of the  
14  
15 285 variables in the analysis (even though all the three methods resulted in variables with  
16  
17 286 ASMD $\leq$ 10%). For experiment 1, PS weighting performed better than PS optimal full  
18  
19 287 matching and sub-classification and for experiment 2, the performance of weighting was  
20  
21 288 comparable to that of optimal full matching (see figures 2 and 3). In both experiments, PS  
22  
23 289 sub-classification reduced covariate imbalance the least. Thus, in the subsequent sections,  
24  
25 290 outcome analyses are based on PS weighted datasets for both experiments.  
26  
27  
28

29 291  
30  
3132 292 [Insert Figure 2]  
33  
3435 293 **Figure 2:** Comparing performance of the three PS implementation methods in experiment 1:

36  
37 294 The y – axis contains all the variables used in the PS models. While x – axis shows absolute standardised mean  
38 295 difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value  
39 296 of  $\leq$  10% indicates the method has performed well in creating comparable groups.  
40

41 297

42 298 [Insert figure 3]  
43  
4445 299 **Figure 3:** Comparing performance of the three PS implementation methods in experiment 246 300 ***c) Outcome Analysis Results***47 301 ***i) Exploring mortality in raw datasets***  
48  
49

50  
51  
52 302 Examining the raw datasets without PS adjustments in experiment 1, the average number of  
53  
54 303 pneumonia deaths (across the 20 imputed datasets) in penicillin plus gentamicin group was  
55  
56 304 33/1363 (2.42%) and in penicillin monotherapy was 26/2639 (0.99%). And for experiment 2,  
57  
58  
59  
60

1  
2  
3 305 the average number of deaths were 87/2296 (3.79%) and 50/4124 (1.21%) in penicillin plus  
4  
5 306 gentamicin and penicillin monotherapy groups respectively. Overall, the average number of  
6  
7 307 pneumonia deaths in the penicillin plus gentamicin group was approximately two and a half  
8  
9  
10 308 to three times the number of mortality events in the penicillin monotherapy group in  
11  
12 309 experiments 1 and 2 respectively.

310

311

312 ***ii) Modelling mortality risk ratios***

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

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

1  
2  
3 328 After excluding 10% of the populations as a result of PS trimming in sensitivity analyses for  
4  
5 329 experiments 1 and 2, the resulting sample sizes were 3583 and 5778. The skewed assignment  
6  
7 330 of children to treatment with gentamicin and penicillin is demonstrated by their higher PS  
8  
9 331 scores in figure 4 for experiment 1 (and figure B for experiment 2 in the additional file 1:  
10  
11 332 supplementary data). As higher PS scores are associated with the presence of a greater  
12  
13 333 number of clinical signs of illness this also suggests an association between more severe  
14  
15 334 illness and treatment with gentamicin and penicillin. For experiment 1, the estimated average  
16  
17 335 mortality events (on PS unadjusted datasets) were 26/1201 (2.16%) and 24/2382 (1.01%) for  
18  
19 336 penicillin plus gentamicin and penicillin monotherapy groups. While the estimated events in  
20  
21 337 experiment 2 were 62/2026 (3.06%) and 46/3752 (1.22%). Thus in sensitivity analyses for  
22  
23 338 both experiments, trimming excluded more mortality events in the penicillin plus gentamicin  
24  
25 339 group compared with the penicillin monotherapy group. The treatment effects estimated  
26  
27 340 using PS weighted models for the restricted populations as a result of PS trimming showed no  
28  
29 341 statistical difference between the two treatments (table 1).  
30  
31  
32  
33

34  
35 342 [Insert figure 4]  
36

37  
38 343 **Figure 4:** Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity  
39  
40 344 scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those  
41  
42 345 who were given penicillin plus gentamicin. While the continuous black line shows the PS distribution for those  
43  
44 346 who received penicillin alone.  
45

46  
47 347

48  
49 348 *e) Sensitivity Analysis through the use of weekend/weekday as an instrumental*  
50 349 *variable*  
51

52 350 In order to assess whether a timing of admission variable would form a natural and random  
53  
54 351 experiment, the distributions of covariates were examined across the levels of the  
55  
56 352 instrumental variable (weekend/weekday) in experiments 1 and 2. The distribution of each of  
57  
58  
59  
60

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

1  
2  
3 376 penicillin group (approximately 2.5 times) and unadjusted analyses pointed, therefore, to a  
4  
5 377 protective effect of penicillin treatment. However, adjusted analyses, both in experiments 1  
6  
7 378 and 2, that aim to account for allocation bias using PS weighting that can result from non-  
8  
9  
10 379 random treatment allocation suggest that there is no appreciable difference in outcomes  
11  
12 380 between penicillin and gentamicin plus penicillin treatment of indrawing pneumonia. In  
13  
14 381 addition, we conducted analyses using alternative PS methods – sub-classification (results  
15  
16 382 presented in additional file 1: supplementary data – figures C and D, and table F) and  
17  
18 383 optimal full matching (results presented in additional file 1: supplementary data – table G)  
19  
20  
21 384 and analyses of both intention to treat and per protocol populations. All analyses showed  
22  
23 385 similar findings (see the provided additional file 1: supplementary data – table H). We  
24  
25 386 undertook two formal approaches to sensitivity analysis. First, we employed PS trimming to  
26  
27 387 exclude 10% of the analysis populations in experiments 1 and 2. Effect estimates in this case  
28  
29 388 are based on analyses of 90% of cases that PS suggest are best matched. Second, we used an  
30  
31 389 instrumental variable. These techniques employ different approaches to account for possible  
32  
33  
34 390 confounding that might contribute to estimated treatment effects. Both these forms of  
35  
36 391 analysis provided results that support the suggestion that poor outcome in this population is  
37  
38 392 not associated with the antibiotic regimen received.

39  
40  
41 393 Our analyses were conducted using data from over 4,000 children, one hundred times more  
42  
43 394 participants than were included in the only prior randomised controlled trial of penicillin  
44  
45 395 monotherapy and penicillin plus gentamicin in treatment of pneumonia in an Asian  
46  
47 396 population (17). There are continuing concerns of clinically important mortality in children  
48  
49 397 with indrawing pneumonia in Africa (21). This has led to hesitation to adopt new WHO and  
50  
51 398 Kenyan guidelines that now recommend the treatment of indrawing pneumonia as an  
52  
53 399 outpatient using amoxicillin (11, 19). Our results suggest that there are likely to be two  
54  
55 400 distinct issues. Firstly, they suggest that offering broader spectrum injectable antibiotic  
56  
57  
58  
59  
60



1  
2  
3 401 treatment to children with indrawing pneumonia may not improve outcomes compared to  
4  
5 402 treatment with penicillin monotherapy. As other studies have suggested equivalence between  
6  
7 403 oral (high dose) amoxicillin therapy and injectable penicillin therapy (2-5, 24) it seems likely  
8  
9 404 therefore that oral amoxicillin and penicillin plus gentamicin combination therapy would  
10  
11 405 result in similar outcomes when used to treat indrawing pneumonia. Clinicians should  
12  
13 406 therefore carefully adhere to guidelines for treatment of indrawing pneumonia and avoid  
14  
15 407 using gentamicin helping to prevent any possible toxicity.

16  
17  
18  
19 408 Secondly, however, our results suggest that children fulfilling a definition of indrawing  
20  
21 409 pneumonia based on clinical signs, and having excluded serious co-morbidities, may still  
22  
23 410 have an appreciable risk of mortality irrespective of their antibiotic treatment (1.5% in all  
24  
25 411 children in experiment 1). When clinicians categorise children with indrawing pneumonia  
26  
27 412 and imperfectly adhere to clinical sign based guidance mortality tends to be higher (2% in all  
28  
29 413 children in experiment 2). These findings point to as yet uncharacterised risk factors that  
30  
31 414 could be important in determining which children need admission to hospital as current  
32  
33 415 guidance indicates that all those with indrawing pneumonia can be treated as an outpatient.  
34  
35 416 While offering an alternative antibiotic to amoxicillin to this group may not improve  
36  
37 417 outcomes it is possible that closer and continuing observation in hospital may help identify  
38  
39 418 co-morbid or alternative conditions that are contributing to this mortality and that may be  
40  
41 419 treated.

42  
43  
44  
45  
46 420 The trials that informed the basis for the revised WHO guidelines (2-5) showed extremely  
47  
48 421 low mortality (0 – 0.2%) suggesting that the populations included in such trials may not be  
49  
50 422 directly representative of all those to whom guidelines are applied in routine settings. In the  
51  
52 423 trial by Agweyu (2015) conducted in Kenya (which compared penicillin versus oral  
53  
54 424 amoxicillin for indrawing pneumonia) overall mortality was 0.8% (24). In a parallel  
55  
56 425 observational cohort providing data from the same hospitals over the same time period for  
57  
58  
59  
60



1  
2  
3 426 children treated with penicillin alone but not included in the Kenyan trial mortality was not  
4  
5 427 significantly different but marginally higher at 1.2% (Agweyu (2017), submitted) perhaps  
6  
7 428 suggesting that even the limited exclusion criteria in this pragmatic trial might result in  
8  
9 429 exclusion of some sicker children. Taken together with data from the analyses presented here  
10  
11 430 it does appear there is a need to explore whether guidelines might be modified to  
12  
13 431 accommodate additional clinical risk factors for possible life-threatening illness that should  
14  
15 432 prompt admission. In a population with high coverage with conjugate vaccines this may more  
16  
17 433 usefully be for more rigorous evaluation to identify alternative diagnoses or for improved  
18  
19 434 supportive care than for different antibiotics.  
20  
21  
22

### 23 435 **Strengths and limitations**

24  
25  
26 436 Conducting comparative effectiveness analyses using observational datasets can offer the  
27  
28 437 advantage of larger sample sizes at lower cost than randomised controlled clinical trials. They  
29  
30 438 also include patients that may not qualify for enrolment in a typical explanatory randomised  
31  
32 439 controlled trial – and therefore perhaps provide more true to life estimates of treatment effects  
33  
34 440 similar to those observed in highly pragmatic trials (15). However, as most observational  
35  
36 441 datasets are not meant for research, they have challenges of non-random treatment allocation  
37  
38 442 and missing data. We employed a rigorous ‘experimental design’ strategy as is recommended  
39  
40 443 when using observational data (32). We used PS and multiple imputation methods in an effort  
41  
42 444 to minimise bias due to non-random treatment allocation and missing data and analyses  
43  
44 445 suggested no appreciable difference in outcomes of indrawing pneumonia treated with  
45  
46 446 penicillin alone compared with penicillin plus gentamicin. This was in contrast to unadjusted  
47  
48 447 regression analyses that pointed towards better outcomes with penicillin alone suggesting  
49  
50 448 presence of allocation bias. As most observational datasets are limited to observed variables,  
51  
52 449 it is important to conduct sensitivity analysis to explore if the estimated effects are potentially  
53  
54 450 sensitive to unmeasured variables. We used an instrumental variable and PS trimming, both  
55  
56  
57  
58  
59  
60

1  
2  
3 451 supported the idea of no appreciable difference regimens when treating indrawing  
4  
5 452 pneumonia. While there are differences (in terms of magnitude) in the mortality observed in  
6  
7 453 the different groups that suggest some residual bias in treatment allocation, these mortality  
8  
9 454 differences are no greater than might occur by chance after PS adjustment (with the type 1  
10  
11 455 and 2 errors specified in the additional file 2: analysis protocol). In that sense the PS  
12  
13 456 approach may still have limitations but it does allow us to conclude no statistical difference in  
14  
15 457 mortality outcomes between the two treatment arms.  
16  
17  
18

19 458 The WHO recommended guidelines for treating pneumonia have considerable influence on  
20  
21 459 policy and practice in low and middle income countries. While the evidence base and rigour  
22  
23 460 of guideline development have improved considerably there remain few data on their  
24  
25 461 effectiveness when implemented in non-trial settings. Even though well-designed, large  
26  
27 462 pragmatic trials would be preferred, we demonstrate that carefully collected routine data may  
28  
29 463 be useful for assessing the effectiveness of alternative treatments (15). Such analyses may  
30  
31 464 become increasingly possible as electronic medical records are deployed in low and middle  
32  
33 465 income countries (41) but it is important that such studies are carefully designed to limit as  
34  
35 466 far as possible the biases that arise from non-random treatment allocation (32). Our results  
36  
37 467 suggest that children with indrawing pneumonia may gain little benefit from treatment with  
38  
39 468 broader spectrum antibiotic regimens. However, they also suggest that further work is needed  
40  
41 469 to identify those who are at higher risk of death who might be prioritised for an inpatient  
42  
43 470 diagnostic work up and improved supportive care rather than treated as outpatients.  
44  
45  
46  
47

48  
49 471 **Additional files**

50  
51 472 **Additional file 1:** supplementary data

52  
53 473 **Additional file 2:** analysis protocol

54  
55 474 **Additional file 3:** STROBE checklist  
56  
57  
58  
59  
60

1  
2  
3 475 **Additional file 4:** Main manuscript with changes highlighted  
4  
5

6 476 **Acknowledgement**  
7

8  
9 477 We would like to thank Ambrose Agweyu for his comments that helped improve this  
10  
11 478 manuscript. We also appreciate the valuable contribution offered by the CIN team: Lydia  
12  
13 479 Thurania & Grace Ochieng' (Kiambu County Hospital), Barnabas Kigen (Busia County  
14  
15 480 Hospital), Melab Musabi & Rachel Ingina (Kitale County Hospital), Anne Kamunya & Sam  
16  
17 481 Otido (Embu County Hospital), Margaret Kuria (Kisumu East County Hospital), Agnes  
18  
19 482 Mithamo & Francis Kanyingi (Nyeri County Hospital), Celia Muturi, Caren Emadau &  
20  
21 483 Cecilia Mutiso (Mama Lucy Kibaki County Hospital), David Kimutai & Loice Mutai  
22  
23 484 (Mbagathi County Hospital), Nick Aduro (Kakamega County Hospital), Samuel Ng'arng'ar  
24  
25 485 (Vihiga County Hospital). Fred Were & David Githanga (Kenya Paediatric Association).  
26  
27 486 Rachel Nyamai (Ministry of Health).  
28  
29  
30

31 487  
32  
33

34 488 **Funding**  
35

36 489 We are grateful for the funds from the Wellcome Trust (#097170) that support ME through a  
37  
38 490 fellowship and additional funds from a Wellcome Trust core grant awarded to the KEMRI-  
39  
40 491 Wellcome Trust Research Programme (#092654) that supported this work. LM is supported  
41  
42 492 by a Nuffield Department of Medicine Prize DPhil Studentship and Clarendon Scholarship  
43  
44 493 (Oxford University). The funders had no role in drafting or submitting this manuscript.  
45  
46

47 494 **Availability of Data**  
48

49 495 The hospital specific datasets are in custody of the hospitals participating in CIN (these  
50  
51 496 datasets have been de-identified).  
52  
53

54 497 **Authors Contributions**  
55  
56  
57  
58  
59  
60

1  
2  
3 498 The contributions of the authors were as follows: LM did an initial draft of this manuscript  
4  
5 499 with the support of RP, EM and ME. Thereafter, all authors edited subsequent versions and  
6  
7 500 approved the final copy.  
8

### 9 501 **Competing Interests**

10  
11 502 The authors declare they have no competing interests.  
12

### 13 503 **Ethics approval**

14  
15  
16  
17 504 This analysis is based on a larger project (CIN) which was cleared by the Kenya Medical  
18  
19 505 Research Institute ethics and review board (Protocol number: 2465).  
20

21  
22 506  
23

24  
25 507  
26

27  
28 508  
29

### 30 509 **References**

31  
32 510 1. Olayemi E, Asare EV, Benneh-Akwasi Kuma AA. Guidelines in lower-middle income  
33 511 countries. *Br J Haematol*. 2017;177(6):846-54.  
34

35 512 2. Atkinson M, Lakhanpaul M, Smyth A, Vyas H, Weston V, Sithole J, et al. Comparison of  
36 513 oral amoxicillin and intravenous benzyl penicillin for community acquired pneumonia in  
37 514 children (PIVOT trial): a multicentre pragmatic randomised controlled equivalence trial.  
38 515 *Thorax*. 2007;61(12):1102-6.  
39

40 516 3. Hazir T, Fox LM, Nisar YB, Fox MP, Ashraf YP, MacLeod WB, et al. Ambulatory short-  
41 517 course high-dose oral amoxicillin for treatment of severe pneumonia in children: a  
42 518 randomised equivalency trial. *Lancet*. 2008;371:49-56.  
43

44  
45 519 4. Addo-Yobo E, Chisaka N, Hassan M, Hibberd P, Lozano JM, Jeena P, et al. Oral  
46 520 amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59  
47 521 months: a randomised multicentre equivalency study. *Lancet*. 2004;364(9440):1141-8.  
48

49 522 5. Addo-Yobo E, Anh DD, El-Sayed HF, Fox LM, Fox MP, MacLeod W, et al. Outpatient  
50 523 treatment of children with severe pneumonia with oral amoxicillin in four countries: the  
51 524 MASS study. *Tropical Medicine and International Health*. 2011;16(8):995-1006.  
52

53  
54 525 6. Mulholland K, Carlin JB, Duke T, Weber M. Challenges of trials of antibiotics for  
55 526 pneumonia in low-income countries. *The Lancet Respiratory Medicine*. 2014;2(12):952-4.  
56  
57  
58  
59  
60

- 1  
2  
3 527 7. Pneumonia and Diarrhoea Progress Report 2014 [Internet].: John Hopkins Bloomberg  
4 528 School of Public Health; 2014 []. Available from: [http://www.jhsph.edu/research/centers-](http://www.jhsph.edu/research/centers-and-institutes/ivac/resources/IVAC-2014-Pneumonia-Diarrhea-Progress-Report.pdf)  
5 529 [and-institutes/ivac/resources/IVAC-2014-Pneumonia-Diarrhea-Progress-Report.pdf](http://www.jhsph.edu/research/centers-and-institutes/ivac/resources/IVAC-2014-Pneumonia-Diarrhea-Progress-Report.pdf).  
6  
7  
8 530 8. Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, et al. Bacteremia  
9 531 among Children Admitted to a Rural Hospital in Kenya. *The new england journal of*  
10 532 *medicine*. 2005;352(1):39-47.  
11  
12 533 9. Global Health Observatory data repository [Internet].; 2016 []. Available from:  
13 534 <http://apps.who.int/gho/data/node.main.686?lang=en>.  
14  
15 535 10. Kenya Demographic and Health Survey [Internet].; 2014 []. Available from:  
16 536 <http://dhsprogram.com/pubs/pdf/FR308/FR308.pdf>.  
17  
18 537 11. Basic paediatric protocols [Internet].: Kenyan Ministry of Health; 2016 []. Available  
19 538 from: [https://www.tropicalmedicine.ox.ac.uk/\\_asset/file/basic-paediatric-protocols-2016.pdf](https://www.tropicalmedicine.ox.ac.uk/_asset/file/basic-paediatric-protocols-2016.pdf).  
20  
21 539 12. Hospital Care for Children [Internet].; 2013 []. Available from:  
22 540 [http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf).  
23  
24 541 13. UNICEF DATA: Monitoring the situation of women and children [Internet].; 2017 [].  
25 542 Available from: <https://data.unicef.org/topic/child-health/pneumonia/>.  
26  
27 543 14. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised  
28 544 controlled trials with non-compliance: a simulation study. *BMJ Open Research Methods*.  
29 545 2014;4(6):e005362.  
30  
31 546 15. Thadhani R. Formal trials versus observational studies. In: Mehta A, Beck M, Sunder-  
32 547 Plassmann G, editors. *Fabry Disease: Perspectives from 5 Years of FOS*. Oxford: Oxford  
33 548 PharmaGenesis; 2006. Chapter 14. Available from:  
34 549 <https://www.ncbi.nlm.nih.gov/books/NBK11597/>.  
35  
36 550 16. West SG, Duan N, Pequegnat W, Gaist P, Jarlais DCD, Holtgrave D, et al. Alternatives to  
37 551 Randomised Controlled Trials. *American Public Health Association*. 2008;98(8):1359-66.  
38  
39 552 17. Hasali MAA, Ibrahim MIM, Sulaiman SAS, Ahmad Z, Hasali JBA. A clinical and  
40 553 economic study of community-acquired pneumonia between single versus combination  
41 554 therapy . *Pharm World Sci*. 2005;27(3):249-53.  
42  
43 555 18. Ayieko P, English M. Case Management of Childhood Pneumonia in Developing  
44 556 Countries . *Pediatr Infect Dis J*. 2007;26:432-40.  
45  
46 557 19. Hospital Care for Children [Internet].; 2013 []. Available from:  
47 558 [http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf).  
48  
49 559 20. Basic Paediatric Protocols: November 2013 Edition [Internet].: Ministry of Health; 2013  
50 560 []. Available from: [http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-](http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition)  
51 561 [basic-paediatric-protocols-november-2013-edition](http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition).  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 562 21. Koh JW, Wong JJ, Sultana R, Wong PP, Mok YH, Lee JH.  
4 563 Risk factors for mortality in children with pneumonia admitted to the pediatric intensive care  
5 564 unit. *Pediatric pulmonology*. 2017;52(8):1076-84.
- 7 565 22. Agweyu A, Kibore M, Digolo L, Kosgei C, Maina V, Mugane S, et al. Prevalence and  
8 566 correlates of treatment failure among Kenyan children hospitalised with severe community-  
9 567 acquired pneumonia: a prospective study of the clinical effectiveness of WHO pneumonia  
10 568 case management guidelines. *Tropical Medicine & International Health*. 2014;19(11):1310-  
11 569 20.
- 14 570 23. Hospital Care for Children [Internet].; 2005 []. Available from:  
15 571 <http://apps.who.int/iris/bitstream/10665/43206/1/9241546700.pdf>.
- 17 572 24. Agweyu A, Gathara D, Oliwa J, Muinga N, Edwards T, Allen E, et al. Oral amoxicillin  
18 573 versus benzyl penicillin for severe pneumonia among kenyan children: a pragmatic  
19 574 randomised controlled noninferiority trial. *Clin Infect Dis*. 2015;60(8):1216-24.
- 22 575 25. Malla L, Perera R, McFadden E, English M.  
23 576 Comparative effectiveness of injectable penicillin versus a combination of penicillin and  
24 577 gentamicin in children with pneumonia characterised by indrawing in Kenya: A protocol for  
25 578 an observational study. *BMJ Open*. 2017;7(9):e016784.
- 27 579 26. Mwakyusa S, Wamae A, Wasunna A, Were F, Esamai F, Ogotu B, et al.  
28 580 Implementation of a structured paediatric admission record for district hospitals in Kenya –  
29 581 results of a pilot study. *BMC International Health and human rights*. 2006;6:9.
- 32 582 27. Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, et al. A Multifaceted  
33 583 Intervention to Implement Guidelines and Improve Admission Paediatric Care in Kenyan  
34 584 District Hospitals: A Cluster Randomised Trial. *PLoS Medicine*. 2011;8(4).
- 36 585 28. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation  
37 586 of clinical care within a clinical information network: an essential initial step in efforts to  
38 587 understand and improve care in Kenyan hospitals. *BMJ Global Health*. 2016;1(1):e000028.
- 40 588 29. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of  
41 589 admissions and variations in the use of basic investigations, treatments and outcomes in  
42 590 Kenyan hospitals within a new Clinical Information Network. *Archives of Disease in*  
43 591 *Childhood*. 2015;101(3):223-9.
- 46 592 30. Danaei G, Rodríguez LA, Cantero OF, Logan R, Hernán MA.  
47 593 Observational data for comparative effectiveness research: an emulation of randomised trials  
48 594 of statins and primary prevention of coronary heart disease. *Stat.Methods Med.Res.*  
49 595 2011;22(1):70-96.
- 52 596 31. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues  
53 597 and guidance for practice. *Stat Med*. 2011;30(4):377-99.
- 55 598 32. Rubin D.B. For Objective Causal Inference, Design Trumps Analysis. *The Annals of*  
56 599 *Applied Statistics*. 2008;2(3):808-40.

- 1  
2  
3 600 33. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of  
4 601 Confounding in Observational Studies. *Multivariate Behavioral Research*. 2011;46(3):399-  
5 602 424.
- 6  
7 603 34. Stuart EA. Matching methods for causal inference:A review and a look forward. *Statist.*  
8 604 *Sci*. 2010;25(1):1-21.
- 9  
10 605 35. Austin PC. Assessing balance in measured baseline covariates when using many-to-one  
11 606 matching on the propensity-score. *Pharmacoepidemiol Drug Saf*. 2008 2008;17(12):1218-25.
- 12  
13 607 36. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates  
14 608 between treatment groups in propensity-score matched samples. *Stat Med*. 2009 10 Nov  
15 609 2009;28(25):3083-107.
- 16  
17 610 37. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the presence of  
18 611 Unmeasured Confounding: Dealing with observations in the Tails of the Propensity Score  
19 612 Distribution - A Simulation Study. *American Journal of Epidemiology*. 2010;172(7):843-54.
- 20  
21 613 38. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat*  
22 614 *Med*. 2014 15 Jun 2014;33(13):2297-340.
- 23  
24 615 39. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods  
25 616 for Causal Inference. *Statistics in Medicine*. 2014;33(13):2297-340.
- 26  
27 617 40. Berkley JA, Brent A, Mwangi I, English M, Maitland K, Marsh K, et al. Mortality among  
28 618 Kenyan children admitted to a rural district hospital on weekends as compared with  
29 619 weekdays . *Pediatrics*. 2004;114:1737-8.
- 30  
31 620 41. English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al.  
32 621 Building Learning Health Systems to Accelerate Research and Improve Outcomes of Clinical  
33 622 Care in Low-and Middle-Income Countries. *PLOS Medicine*. 2016;13(4):e1001991.

34  
35  
36  
37  
38  
39 623  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



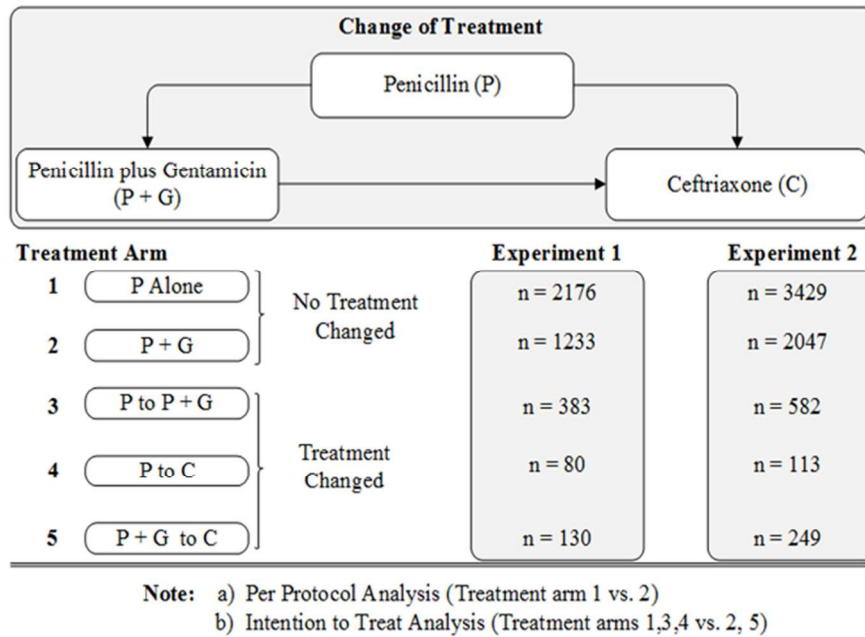


Figure 1: Summary of patients per treatment arm in experiments 1 – 2

63x45mm (600 x 600 DPI)



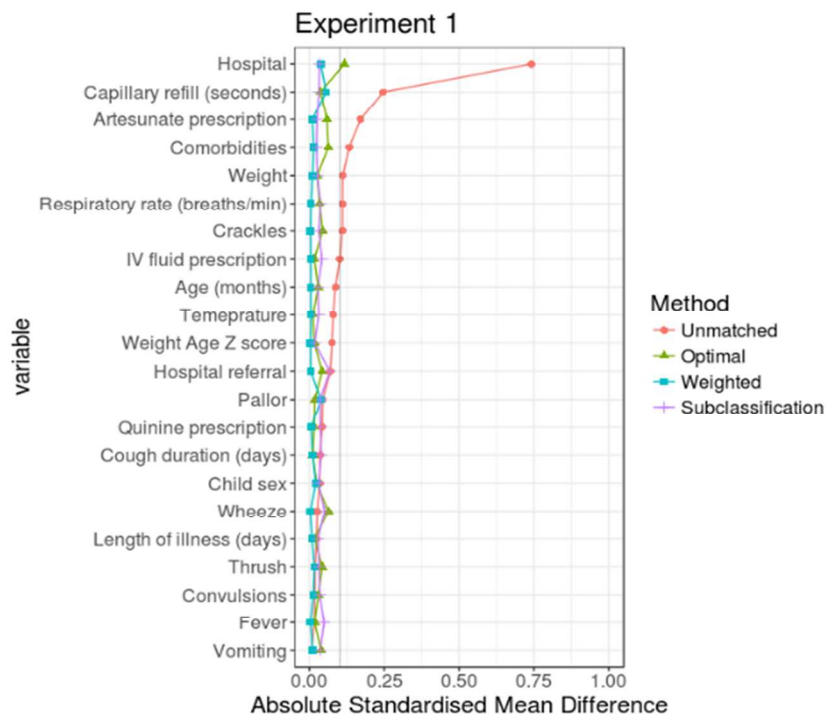


Figure 2: Comparing performance of the three PS implementation methods in experiment 1: The y - axis contains all the variables used in the PS models. While x - axis shows absolute standardised mean difference (ASMD) which is a measure of covariate balance between the two treatment groups. An ASMD value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

63x45mm (600 x 600 DPI)

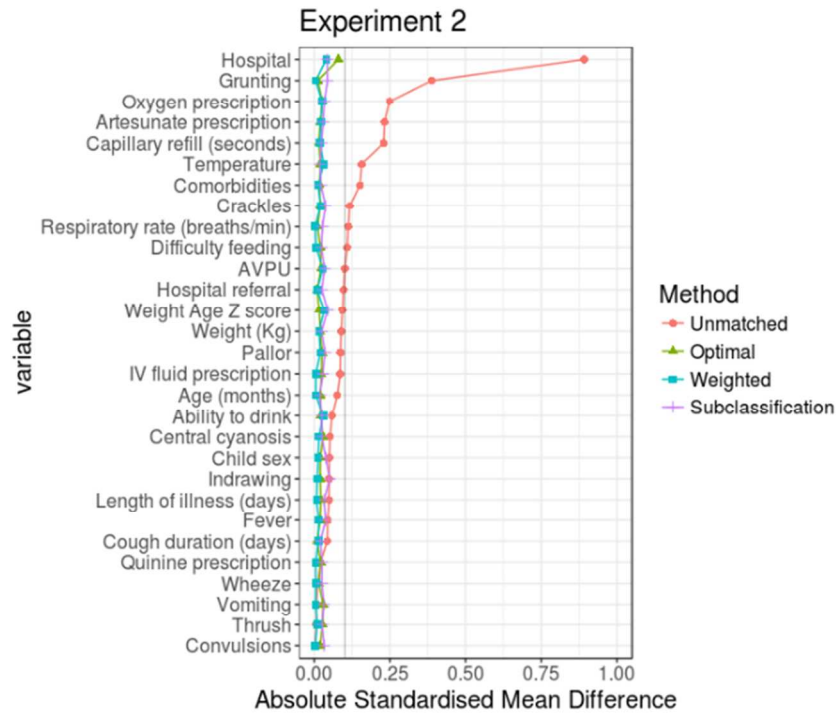


Figure 3: Comparing performance of the three PS implementation methods in experiment 2.

63x45mm (600 x 600 DPI)

Review only

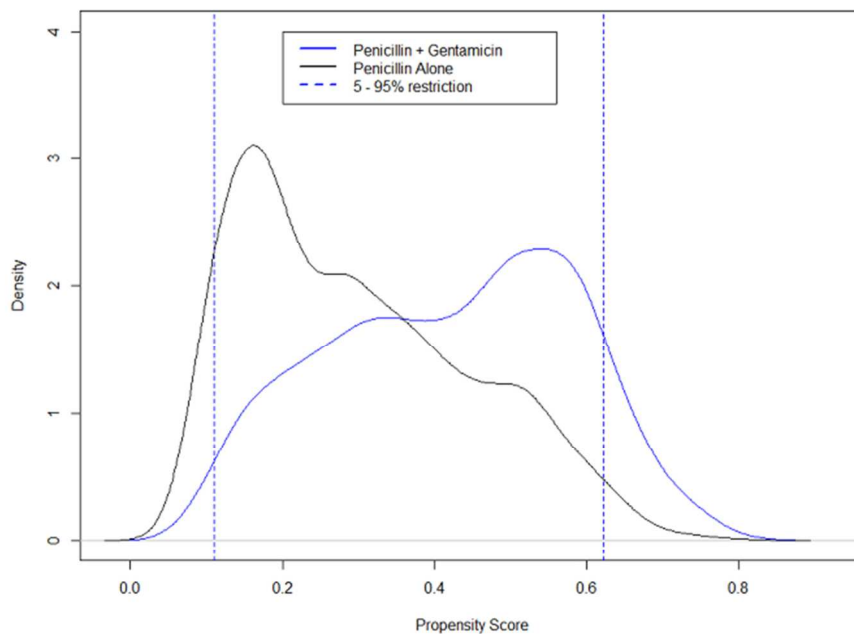


Figure 4: Experiment 1 PS distribution curves: The dotted lines show the distribution of propensity scores for patients in the 5 – 95%. The continuous blue line shows the distribution of propensity scores for those who were given penicillin plus gentamicin. While the continuous black line shows the PS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

63x45mm (600 x 600 DPI)

1  
2  
3 **Comparative effectiveness of injectable penicillin versus a combination of penicillin and**  
4 **gentamicin in children with pneumonia characterised by indrawing in Kenya: A**  
5 **retrospective observational study**  
6  
7  
8  
9

10 Lucas Malla<sup>1\*</sup>

11 Rafael Perera-Salazar<sup>2</sup>

12 Emily McFadden<sup>2</sup>

13 Mike English<sup>1,3</sup>

14  
15  
16  
17  
18  
19  
20  
21  
22  
23 <sup>1</sup>Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

24  
25  
26 <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Kingdom.

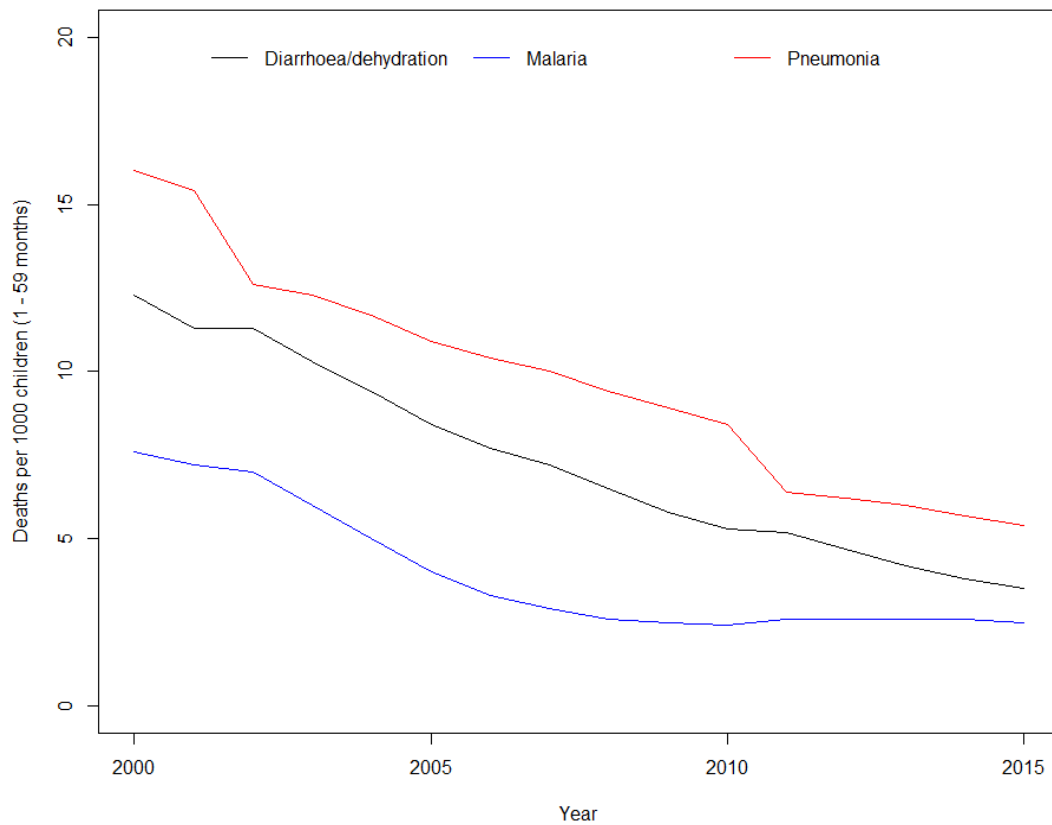
<sup>3</sup> Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.

1  
2  
3  
4 The supplementary material is organised into the following subsections:  
5  
6

- 7 • Under-five mortality incident rates in Kenya
- 8
- 9 • Percentage of completeness of variables (experiments 1 and 2)
- 10
- 11 • Overlap and correctness of penicillin and gentamicin dosing
- 12
- 13 • Summary of key and auxiliary independent variables
- 14
- 15
- 16 • Trimming in experiment 2 (ITT population)
- 17
- 18 • Analysis using Instrument variables
- 19
- 20
- 21 • Analysis using PS sub - classification
- 22
- 23 • Analysis using PS optimal full matching
- 24
- 25
- 26 • Analysis using per protocol population
- 27
- 28 • Definition of PS methods and how they were used
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

### 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

only

## 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

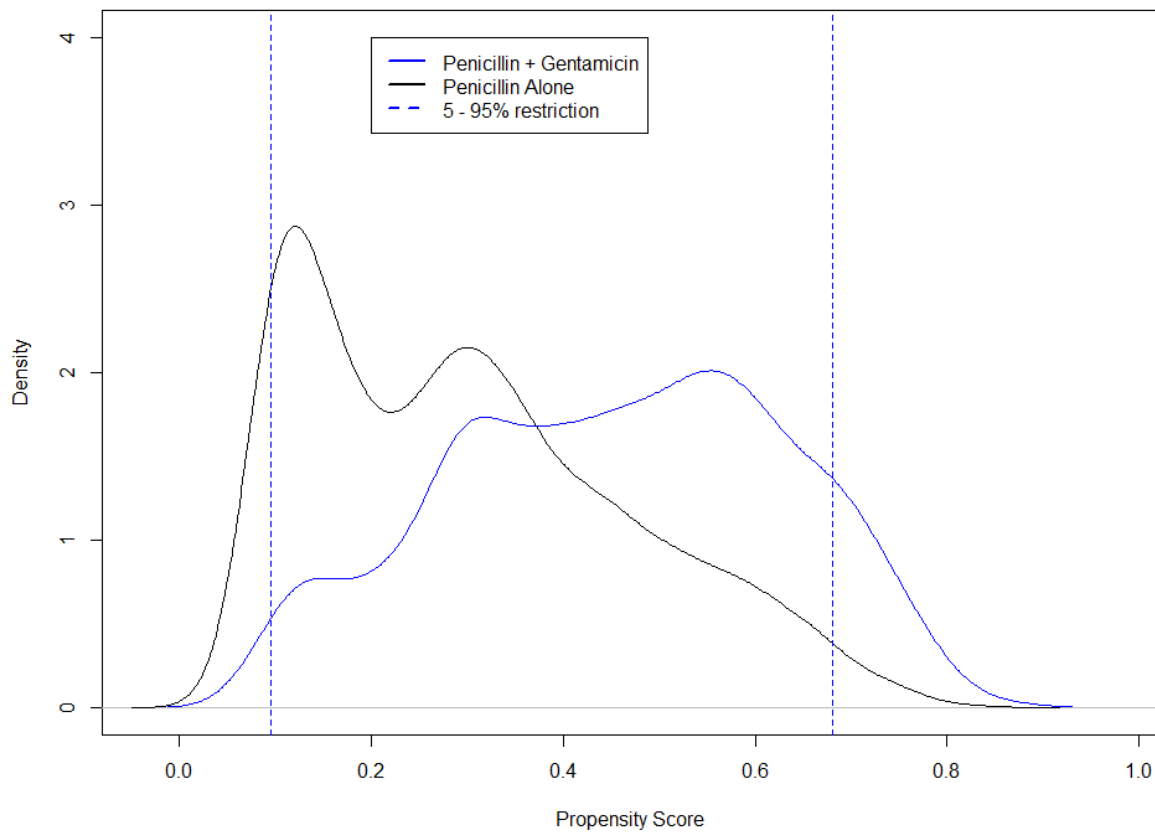
**Table C:** Summary of key and auxiliary independent variables for experiments 1 and 2<sup>1</sup>

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)

<sup>1</sup> 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
<b>Child Sex</b>						
Female	45%	46%	0.03	44%	45%	0.01
Male	55%	54%		56%	55%	
<b>Pallor</b>						
Mild/moderate	4%	5%	0.02	5%	5%	0.00
None	95%	94%		93%	93%	
Severe	1%	2%		2%	2%	
<b>Capillary refill</b>						
1 sec	68%	71%	0.07	66%	68%	0.04
2 sec	30%	27%		31%	29%	
>2 sec	3%	2%		3%	3%	
<b>Fever</b>						
Absent	21%	18%	0.05	19%	16%	0.07
Present	79%	82%		81%	84%	
<b>Convulsions</b>						
Absent	95%	96%	0.02	94%	94%	0.03
Present	5%	4%		6%	6%	
<b>Vomiting</b>						
No	65%	62%	0.06	63%	63%	0.00
Yes	35%	38%		37%	37%	
<b>Referral</b>						
No	82%	86%	0.10	81%	84%	0.09
Yes	18%	14%		19%	16%	
<b>Thrush</b>						
Absent	98%	98%	0.00	98%	98%	0.03
Present	2%	2%		2%	2%	
<b>Comorbidities</b>						
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%	
<b>Crackles</b>						
Absent	47%	47%	0.01	48%	47%	0.02
Present	53%	53%		52%	53%	
<b>Wheeze</b>						
Absent	85%	84%	0.02	85%	84%	0.02
Present	15%	16%		15%	16%	
<b>IV prescription</b>						
No	97%	96%	0.05	95%	95%	0.01
Yes	3%	4%		5%	5%	
<b>Quinine Prescription</b>						
No	97%	97%	0.02	95%	94%	0.04
Yes	3%	3%		5%	6%	
<b>Artesunate Prescription</b>						
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)

PS Subgroup	Experiment 1		Experiment 2	
	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%)
<b>Total</b>	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.

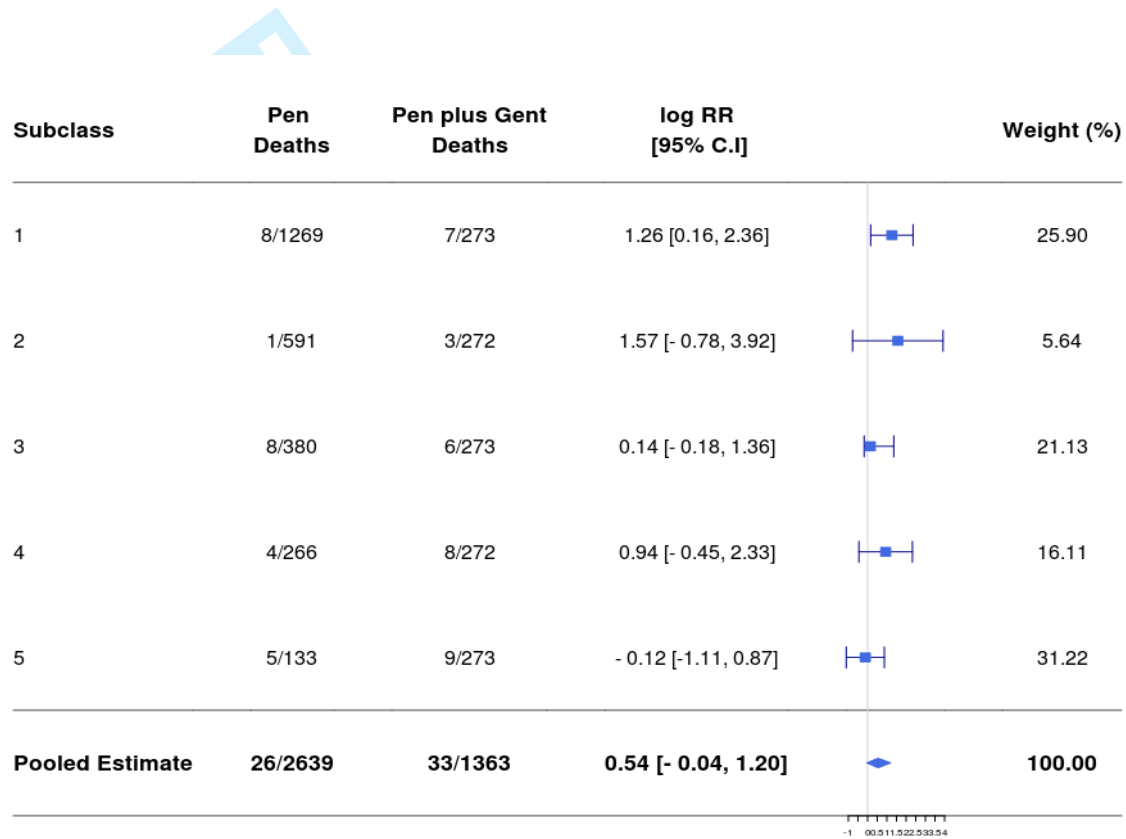
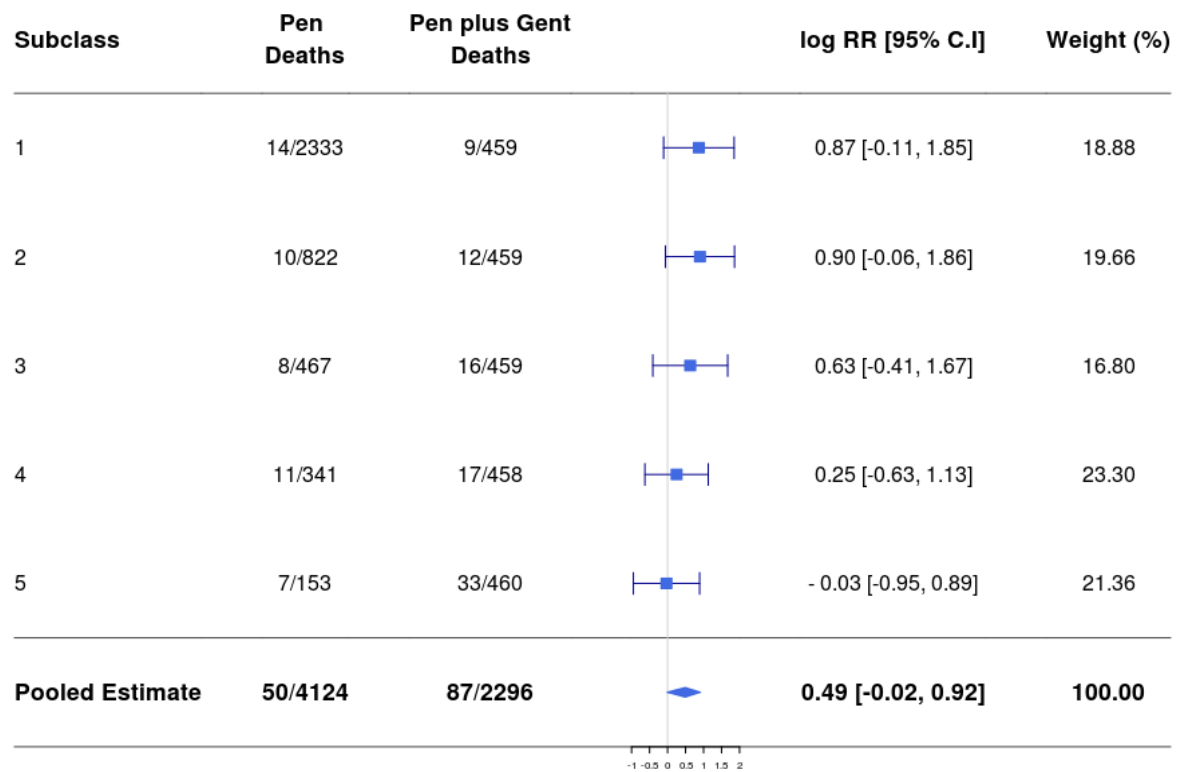


Figure C: Experiment 1 – ITT



**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)
<b>Experiment 1</b>	
Regression without PS adjustment	0.56 [-0.06, 1.02]
PS optimal matching	0.27 [-0.22, 0.65]
<b>Experiment 2</b>	
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)
<b>Experiment 1</b>	
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]
<b>Experiment 2</b>	
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.

1  
2  
3 **Comparative effectiveness of injectable penicillin versus a combination of penicillin and**  
4 **gentamicin in children with pneumonia characterised by indrawing in Kenya: A protocol**  
5 **for an observational study**  
6  
7  
8  
9

10  
11 Lucas Malla<sup>1\*</sup> ([lucas.malla@some.ox.ac.uk](mailto:lucas.malla@some.ox.ac.uk))  
12

13  
14 Rafael Perera-Salazar<sup>2</sup> ([rafael.perera@phc.ox.ac.uk](mailto:rafael.perera@phc.ox.ac.uk))  
15

16  
17 Emily McFadden<sup>2</sup> ([emily.mcfadden@phc.ox.ac.uk](mailto:emily.mcfadden@phc.ox.ac.uk))  
18

19  
20 Mike English<sup>1,3</sup> ([menglish@kemri-wellcome.org](mailto:menglish@kemri-wellcome.org))  
21  
22

23 <sup>1</sup>Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.  
24

25  
26 <sup>2</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United  
27  
28 Kingdom.  
29

30  
31 <sup>3</sup>Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.  
32  
33  
34  
35  
36  
37

38 **\*Corresponding author:**  
39

40  
41 Lucas Malla  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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

1  
2  
3 rigorous approach to study design, propensity score methods and multiple imputation to  
4 minimize bias.  
5  
6  
7

### 8 9 **Ethics and dissemination**

10  
11 The primary data are held by hospitals participating in the Kenyan Clinical Information  
12 Network (CIN) project with de-identified data shared with the KEMRI-Wellcome Trust  
13 Research Programme for agreed analyses. The use of data for the analysis described received  
14 ethical clearance from the Kenya Medical Research Institute Scientific and Ethical Review  
15 Committee. The findings of this analysis will be published.  
16  
17  
18  
19  
20  
21  
22

### 23 24 **Strength**

- 25  
26  
27 - This study will be used as a platform to explore effectiveness of alternative treatments  
28 in routine care in a low income setting to improve health outcomes for children.  
29  
30  
31

### 32 33 **Limitation**

- 34  
35  
36 - The analysis will be limited to the variables in the observational dataset – and therefore  
37 risk bias due to unmeasured key variables.  
38  
39  
40  
41 - The influence of any resulting bias, to alter results, will however be assessed through  
42 the use of alternative methods as instrumental variables.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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

1  
2  
3 opportunity for comparative effectiveness evaluation if similar populations of children with  
4 pneumonia are prescribed different treatments. Clinicians may create such a situation by not  
5 following recommendations because they have inadequate knowledge or if they believe  
6 (potentially contrary to the evidence) that certain treatments result in better health outcomes.  
7  
8  
9  
10  
11

12  
13 **Box 1: Pneumonia treatment algorithm**

14  
15 *The pneumonia severity classification that was recommended by Kenyan guidelines up to*  
16 *March 2016 (9) (and previously by WHO guidelines (1)) defined the following three severity*  
17 *classes:*  
18

- 19  
20  
21 1. **Very severe pneumonia:** *If a child had either oxygen saturation less than 90% or*  
22 *central cyanosis or was grunting or unable to drink or not alert, then s/he was*  
23 *classified as having very severe pneumonia, put on oxygen and treated with a*  
24 *combination of gentamicin and penicillin.*

25  
26  
27 *{The new WHO (2) and Kenyan guidelines (9) renamed this class as “severe*  
28 *pneumonia” – and currently recommend treatment with a combination of ampicillin*  
29 *(or penicillin) with gentamicin plus oxygen}.*  
30  
31

- 32  
33 2. **Severe pneumonia:** *If a child had lower chest wall indrawing (but did not have any*  
34 *of qualifying signs for very severe pneumonia above) and was alert then s/he was to*  
35 *be classified as having severe pneumonia and be treated with benzyl penicillin only.*

36  
37  
38 **Note:** *The term indrawing pneumonia is hereafter used in this protocol to define this*  
39 *category of children to avoid confusion.*  
40

- 41  
42 3. **(Non – severe) Pneumonia:** *If a child had none of the mentioned signs but had cough*  
43 *or difficulty breathing and a respiratory rate greater than or equal to 50*  
44 *breaths/minute (for age between 2 and 11 months) or respiratory rate greater than or*  
45 *equal to 40 breaths/minute (for age above 12 months) then s/he was classified as*  
46 *having non severe pneumonia and treated with cotrimoxazole or amoxicillin if*  
47 *previously treated with cotrimoxazole.*

48  
49  
50 *{The current WHO and Kenyan guidelines collapsed severity classes 2 and 3 into one*  
51 *category referred to as “non –severe pneumonia”. This group of patients are*  
52 *currently treated with oral amoxicillin – partly informed by a local trial (18)}.*  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 In particular, a previous study showed that clinicians over-prescribed gentamicin, adding this  
4  
5 to penicillin for the treatment of pneumonia characterized by lower chest wall indrawing but  
6  
7 no other signs of severe illness instead of penicillin alone as was recommended (4)<sup>1</sup>. Therefore,  
8  
9 this protocol is for a study that seeks to explore whether there is any benefit from adding  
10  
11 gentamicin to penicillin in treating children with indrawing pneumonia. Such a benefit could  
12  
13 accrue if bacterial causes of pneumonia that were previously (prior to introduction of new  
14  
15 vaccines) proportionately less common (eg. *S. aureus* and gram negative bacteria) are now  
16  
17 accounting for an increased proportion of pneumonia deaths – as in such cases, the addition of  
18  
19 gentamicin might provide effective treatment for a broader spectrum of pathogens. Tackling  
20  
21 this question is of importance as WHO have recently changed indrawing pneumonia treatment  
22  
23 guidance based on trials that suggest equivalence of oral amoxicillin and injectable penicillin  
24  
25 (12-15). New guidance recommends outpatient oral treatment for a population of children  
26  
27 previously admitted to hospital (10). However, mortality from pneumonia has been reported to  
28  
29 be higher in African settings (16, 17) despite the increasing use of multiple vaccines spanning:  
30  
31 measles, pertussis, HiB and pneumococcal conjugate vaccines. It remains possible therefore  
32  
33 that for a small number of children a broader spectrum antibiotic regimen might be of benefit.  
34  
35 This study addresses this question that has not been the subject of prior community and  
36  
37 pragmatic clinical trials.  
38  
39  
40  
41  
42  
43  
44  
45

## 46 Objectives

### 47 Primary

- 48  
49  
50  
51  
52  
53 1) Experiment 1: To compare the effectiveness of injectable penicillin versus penicillin  
54  
55 plus gentamicin (both injectable) in treatment of indrawing pneumonia; where severity  
56  
57  
58  
59

---

60 <sup>1</sup> 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).

1  
2  
3 level is constructed (imputed) using data recorded on each child's clinical signs  
4  
5 (hospitals use a structured record form that supports recording of signs highlighted in  
6  
7 guidelines) such that severity classification is consistent with guideline  
8  
9 recommendations.  
10  
11

### 12 13 *Secondary*

- 14  
15  
16 2) Experiment 2: To compare effectiveness of injectable penicillin versus penicillin plus  
17  
18 gentamicin in treatment of indrawing pneumonia; where we use clinician assigned  
19  
20 severity level.  
21  
22  
23 3) Experiment 3: To compare effectiveness of injectable penicillin versus penicillin plus  
24  
25 gentamicin in treatment of all cases of pneumonia admitted to hospital.  
26  
27  
28

29 Experiment 1 will be primary as it most approximates a typical randomised trial where  
30  
31 recruitment would be based on specified clinical signs. This scenario will provide an evaluation  
32  
33 of alternative therapies within a guideline class (where children have very similar clinical signs)  
34  
35 and thus is the best mimic of a prospectively designed comparative evaluation in which  
36  
37 clinicians stick to the rules of severity classification (see (18) for an example of a RCT in Kenya  
38  
39 that this would be similar to – where classification is based on clinical signs). Recommended  
40  
41 treatment for this disease classification was penicillin alone, treatment with combination  
42  
43 therapy may therefore represent over-treatment. Alternatively, the combination treatment that  
44  
45 provides broader antimicrobial cover could provide an advantage in a small proportion of cases  
46  
47 that would only be detected in moderately large studies – where the addition of gentamicin  
48  
49 offers improved treatment for specific organisms not susceptible to penicillin alone.  
50  
51  
52

53 Experiment 2 will provide a test of alternative therapies amongst those where clinicians used  
54  
55 their own judgement (possibly including gut feeling) to classify and treat (19) and have on  
56  
57 occasions (potentially) over-ridden or ignored the guideline recommendations. In this case  
58  
59  
60



1  
2  
3 although the same label of indrawing pneumonia is given to all, the treatment selected may be  
4  
5 an indicator of perceived severity and there may be a potential bias as a result – and the  
6  
7 propensity score distributions (see below) may help demonstrate this and in theory may  
8  
9 overcome this potential bias. Here if there is no clinically relevant difference between  
10  
11 treatments within a group of patients that reflects clinicians' actual classification decisions this  
12  
13 could reassure them that monotherapy with penicillin (or amoxicillin) would be acceptable.  
14  
15 Lastly, experiment 3 is an extension of the logic of experiment two. To date there have been  
16  
17 no pragmatic trials of penicillin alone compared with alternative combination therapies for all  
18  
19 forms of inpatient pneumonia, and addressing this question may be relevant for two reasons.  
20  
21 First, the population of children admitted with severe forms of pneumonia is now largely one  
22  
23 that has received *H. influenzae* Type B and pneumococcal conjugate vaccines that have likely  
24  
25 changed the aetiology of this illness. Second, if clinicians are poorly trained and unable to  
26  
27 classify illness severity – resulting in non-adherence to guidelines - it would be useful to  
28  
29 explore the potential impact of this across all levels of severity of pneumonia. This analysis has  
30  
31 the largest numbers of subjects.  
32  
33  
34  
35  
36  
37  
38

### 39 **Methods and analysis**

40  
41  
42 To answer these three questions, we will use the Kenyan Clinical Information Network (CIN)  
43  
44 dataset that provides observational data on all admissions to 13 Kenyan County hospitals (Box  
45  
46 2).  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**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.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21
- 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.

22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

**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.

43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**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,

1  
2  
3 blood culture or tests for inflammatory markers are not routinely available (6). As indicated  
4  
5 above clinicians may fail to adhere to guideline recommendations by making errors or over-  
6  
7 riding recommendations at any of the three steps of assessment, severity classification and  
8  
9 treatment assignment. However, based on the clinical symptoms and signs recorded it is  
10  
11 possible to assign a severity classification (and thus expected treatment) based on the data. It  
12  
13 is a data informed and investigator assigned classification as indrawing pneumonia that is used  
14  
15 in the primary analysis (experiment 1).  
16  
17  
18  
19

### 20 *c) Analysis Variables*

#### 21 *Outcome variable*

22  
23  
24 Mortality will be used as the outcome variable in all the three experiments.  
25  
26  
27  
28

#### 29 *Independent variables*

30  
31  
32 These variables are grouped into key and auxiliary. Key variables are defined as those that  
33  
34 should influence pneumonia severity classification and hence treatment based on the treatment  
35  
36 protocol (9) (Box 1 above). Auxiliary variables are defined as those that might, *a priori*, be  
37  
38 expected to influence treatment assignment based on clinical reasoning (for example they  
39  
40 might make a clinician concerned for severe illness), although according to the formal rules  
41  
42 (the guidelines) they are not considered reasons to alter treatment assignment. Such auxiliary  
43  
44 variables were identified from those clinical symptoms and signs that are routinely collected  
45  
46 within CIN. See table 1 for a summary of key and auxiliary variables that will be used in the  
47  
48 analyses.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1:** Summary of key and auxiliary independent variables for experiments 1, 2 and 3<sup>2</sup>

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):

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

<sup>2</sup> 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.

1  
2  
3 ns = size of smaller group.

4 k = ratio of larger group to smaller group.

5  $p_1 - p_2$  = clinical difference in proportions of the outcome.

6  $Z_b$  = corresponds to power of 80%

7  $Z_{1-\alpha/2}$  = corresponds to two-tailed significance level (1.96 for  $\alpha = .05$ ).

8  $\bar{p}$  = corresponds to average of outcome proportions in two groups.

9  
10  
11  
12  
13  
14  
15 The value for  $\bar{p}$  is estimated from studies – two of which formed evidence for earlier WHO  
16  
17 indrawing pneumonia treatment guidelines. See table 2 that shows the number of deaths per  
18  
19 treatment arm reported in these studies.  
20  
21

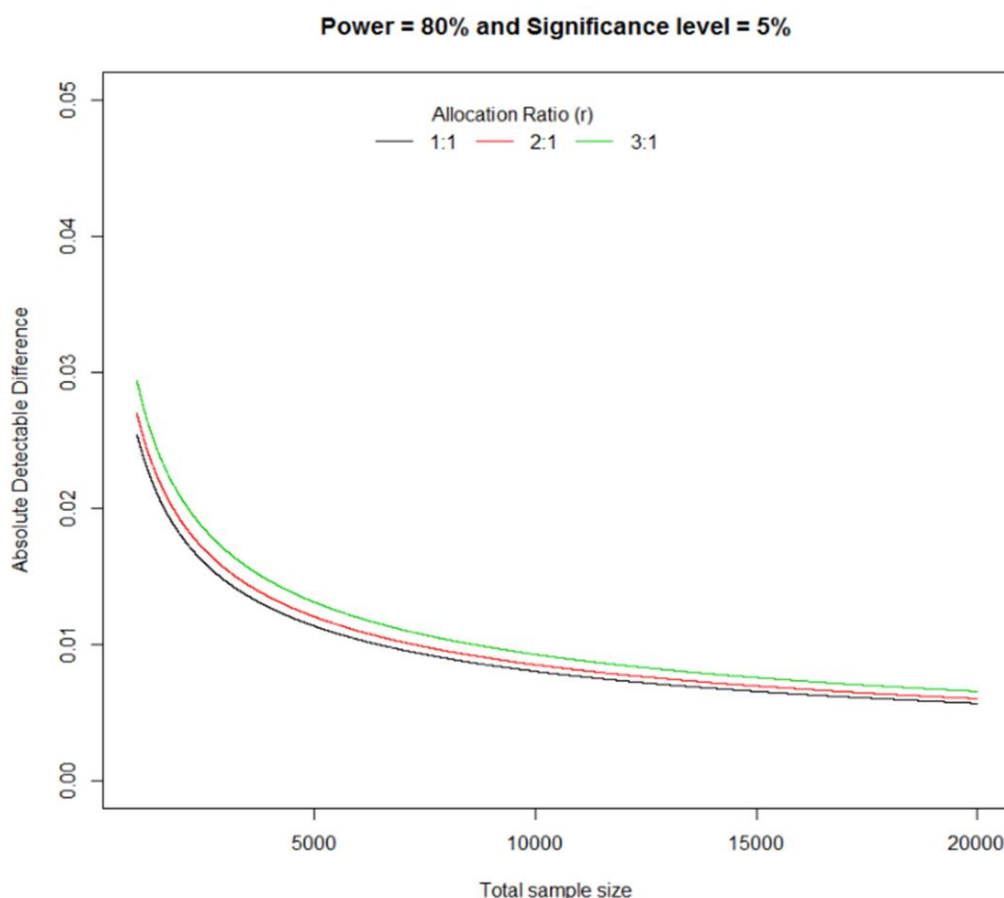
22  
23 **Table 2:** Summary of some of pneumonia studies that informed previous WHO guidelines  
24

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	

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37 For assessment of sample size for indrawing pneumonia experiments, a weighted<sup>3</sup>  $\bar{p}$  of 0.041  
38  
39 from these studies is used. The ratio r is varied between 1 and 3. Figure 1 was generated by  
40  
41 fixing power and significance level at 80% and 5% respectively. Estimates of  $\bar{p}(1 - \bar{p})$  derived  
42  
43 from WHO studies were substituted in the sample size formula and data simulated in order to  
44  
45 see what detectable differences would be achieved by different sample sizes. A total sample  
46  
47 size of about 4000 would be sufficient to detect a minimum difference of 1.5% (absolute  
48  
49 difference e.g. a reduction of mortality from X% to X – 1.5%) in any of these experiments<sup>4</sup>.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<sup>3</sup> Weighting was done using the total sample sizes per experiment.

<sup>4</sup> 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<sup>5</sup> (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.

<sup>5</sup> 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<sup>6</sup>. A hospital variable will be modelled as a fixed effect

---

<sup>6</sup> 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.



1  
2  
3 in the log binomial regression that measures treatment effects on pooled data. These models  
4  
5 will be fitted on each imputed dataset (adjusting for other variables used in PS models) and  
6  
7 results pooled using Rubin rules (31).  
8  
9

10  
11 **Step 4:** sensitivity analysis will be conducted to investigate effects of unmeasured confounders  
12  
13 and validity of estimates obtained through multiple imputation. Propensity score methods  
14  
15 generate matched treated and (active) control patients whose distribution of measured  
16  
17 covariates are as similar as possible. However, two patients with similar covariate distribution  
18  
19 may differ in terms of unmeasured variables – and this may introduce bias in estimated  
20  
21 treatment effects (32). On the other hand, if outcome and explanatory variables have missing  
22  
23 data, then inclusion of outcome data in multiple imputation may contribute minor information  
24  
25 in the substantive (outcome) model (33).  
26  
27  
28  
29

#### 30 31 *Exploring effects of unmeasured confounders*

32  
33 Sensitivity analysis for unmeasured confounders will involve the use of an instrumental  
34  
35 variable (IV) (34) – weekend admission and PS trimming (35). A few IV sources in health  
36  
37 studies have been described in Baiocchi (2014) (36). These include: distance to specialty,  
38  
39 genes, insurance plan, timing of admission, calendar time and preference based IVs. Of  
40  
41 relevance to this analysis would be timing of admission IVs. A study conducted by Berkley  
42  
43 (2004) (37) in a Kenyan hospital demonstrated that children who were admitted during the  
44  
45 weekend experienced higher mortality compared to those admitted during the weekdays –  
46  
47 which is a possible indication of poor quality of care and treatment during the weekend. In  
48  
49 other words, it is anticipated that children admitted during the weekdays would have better  
50  
51 health outcomes. This, in theory, implies that the type of treatment and care received depend  
52  
53 on the day of admission – and which later determines the type of health outcome of the patient.  
54  
55  
56  
57  
58  
59  
60

#### *Examining validity of multiple imputation*

1  
2  
3 The analysis steps 1 – 3 above will exclude outcome data in the imputation model – however  
4  
5 sensitivity analysis will include models in which the outcome variable is included in the  
6  
7 imputation approach. This will aim to investigate if including outcome data in the imputation  
8  
9 model has an influence<sup>7</sup>.  
10  
11

### 12 13 **Ethics and dissemination**

14  
15  
16 The primary data are held by hospitals participating in the Kenyan Clinical Information  
17  
18 Network (CIN) project with de-identified data shared with the KEMRI-Wellcome Trust  
19  
20 Research Programme for agreed analyses. The analyses described in this protocol are part of  
21  
22 this larger project (CIN) which was approved by the Kenya Medical Research Institute  
23  
24 Scientific and Ethical Review Committee (Protocol number: 2465). This committee agreed the  
25  
26 use of de-identified patient data derived from retrospective case record review without gaining  
27  
28 individual patient consent as is common practice in service evaluation research. The findings  
29  
30 will be useful in understanding the external validity of current treatments – and will provide a  
31  
32 platform on which to do more similar analyses for different (combinations of) treatments. The  
33  
34 results of this analysis will be shared with the Kenyan Ministry of Health and will inform  
35  
36 discussions on national pneumonia treatment guidelines to which the research team have made  
37  
38 major prior contributions. The work will also be submitted for publication.  
39  
40  
41  
42  
43  
44

### 45 **Competing Interests**

46  
47 The authors declare they have no competing interests.  
48  
49

### 50 **Authors Contributions**

51  
52  
53  
54  
55  
56  
57  
58  
59  
60  

---

<sup>7</sup> 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).

1  
2  
3 The contributions of the authors were as follows: LM did an initial draft of this manuscript  
4  
5 with the support of RP, EM and ME. Thereafter, all authors edited subsequent versions and  
6  
7 approved the final copy.  
8  
9

## 10 Funding

11  
12 We are grateful for the funds from the Wellcome Trust (#097170) that support ME through a  
13  
14 fellowship and additional funds from a Wellcome Trust core grant awarded to the KEMRI-  
15  
16 Wellcome Trust Research Programme (#092654) that supported this work. LM is supported by  
17  
18 a Nuffield Department of Medicine Prize DPhil Studentship and Clarendon Scholarship  
19  
20 (Oxford University). The funders had no role in drafting or submitting this manuscript.  
21  
22  
23  
24  
25  
26  
27  
28

## 29 References

- 30  
31  
32 1. Hospital Care for Children [Internet].; 2005 []. Available from:  
33 <http://apps.who.int/iris/bitstream/10665/43206/1/9241546700.pdf>.  
34  
35 2. Hospital Care for Children [Internet].; 2013 []. Available from:  
36 [http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/81170/1/9789241548373_eng.pdf).  
37  
38 3. Sazawal S, Black R. Pneumonia Case Management Trials Group. Lancet Infect Dis.  
39 2003;3(9):547-56.  
40  
41 4. Agweyu A, Kibore M, Digolo L, Kosgei C, Maina V, Mugane S, et al. Prevalence and  
42 correlates of treatment failure among Kenyan children hospitalised with severe community-  
43 acquired pneumonia: a prospective study of the clinical effectiveness of WHO pneumonia  
44 case management guidelines. Tropical Medicine & International Health. 2014.  
45  
46 5. Mulholland K, Carlin JB, Duke T, Weber M. Challenges of trials of antibiotics for  
47 pneumonia in low-income countries. The Lancet Respiratory Medicine. 2014;2(12):952-4.  
48  
49 6. Ayieko P, Ogero M, Makone B, Julius T, Mbevi G, Nyachiro W, et al. Characteristics of  
50 admissions and variations in the use of basic investigations, treatments and outcomes in  
51 Kenyan hospitals within a new Clinical Information Network. Archives of Disease in  
52 Childhood. 2015.  
53  
54 7. Agweyu A, Opiyo N, English M. Experience developing national evidence-based clinical  
55 guidelines for childhood pneumonia in a low-income setting - making the GRADE? BMC  
56 Pediatric. 2012.  
57  
58  
59  
60

- 1  
2  
3 8. Ayieko P, English M. Case Management of Childhood Pneumonia in Developing  
4 Countries . *Pediatr Infect Dis J*. 2007;26:432-40.  
5
- 6  
7 9. Basic Paediatric Protocols: November 2013 Edition [Internet].: Ministry of Health; 2013  
8 []. Available from: [http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-](http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition)  
9 [basic-paediatric-protocols-november-2013-edition](http://www.idoc-africa.org/index.php/86-clinical-guide/ken-guide/134-basic-paediatric-protocols-november-2013-edition).  
10
- 11  
12 10. Basic paediatric protocols [Internet].: Kenyan Ministry of Health; 2016 []. Available  
13 from: [https://www.tropicalmedicine.ox.ac.uk/\\_asset/file/basic-paediatric-protocols-2016.pdf](https://www.tropicalmedicine.ox.ac.uk/_asset/file/basic-paediatric-protocols-2016.pdf).  
14
- 15  
16 11. Graham SM, English M, Hazir T, Enarson P, Duke T. Challenges to improving case  
17 management of childhood pneumonia at health facilities in resource limited settings. *Bulletin*  
18 *of the WHO*. 2015.  
19
- 20  
21 12. Atkinson M, et al. A multicentre randomised controlled equivalence trial comparing oral  
22 amoxicillin and intravenous benzyl penicillin for community acquired pneumonia in children  
23 PIVOT Trial. *Thorax*. 2007;61:1102-6.  
24
- 25  
26 13. Hazir T, et al. Ambulatory short-course high-dose oral amoxicillin for treatment of severe  
27 pneumonia in children: a randomised equivalency trial. *Lancet*. 2008;371:49-56.  
28
- 29  
30 14. Addo-Yobo E, Chisaka N, Hassan M, Hibberd P, Lozano JM, Jeena P, et al. Oral  
31 amoxicillin versus injectable penicillin for severe pneumonia in children aged 3 to 59  
32 months: a randomised multicentre equivalency study. *Lancet*. 2004;364:1141-8.  
33
- 34  
35 15. Addo-Yobo E, et al. Outpatient treatment of children with severe pneumonia with oral  
36 amoxicillin in four countries: the MASS study . *Tropical Medicine and International Health*.  
37 2011;16:995-1006.  
38
- 39  
40 16. IVAC. Pneumonia and Diarrhoea Progress Report 2014. John Hopkins Bloomberg School  
41 of Public Health; 2014.  
42
- 43  
44 17. Onyango D, Kikuvi G, Amukoye E, Omolo J. Risk factors of severe pneumonia among  
45 children aged 2-59 months in western Kenya: a case control study. *The Pan African Medical*  
46 *Journal*. 2012(13):45.  
47
- 48  
49 18. Agweyu A, Gathara D, Oliwa J, Muinga N, Edwards T, Allen E, et al. Oral amoxicillin  
50 versus benzyl penicillin for severe pneumonia among kenyan children: a pragmatic  
51 randomised controlled noninferiority trial. *Clin Infect Dis*. 2015.  
52
- 53  
54 19. Stopler E, Van Roeyen P, Van de Wiel M, Van Bokhoven M, Houben P, Van der Weijde  
55 T, et al. Consensus on gut feelings in general practice. *BMC Family Practice*. 2009.  
56
- 57  
58 20. Tuti T, Bitok M, Malla L, Paton C, Muinga N, Gathara D, et al. Improving documentation  
59 of clinical care within a clinical information network: an essential initial step in efforts to  
60 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.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
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.

- 1  
2  
3 35. Lee BK, Lessler J, Stuart EA. Weight Trimming and Propensity Score Weighting. PLoS  
4 ONE [Electronic Resource]. 2011.  
5  
6  
7 36. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods  
8 for Causal Inference. *Statistics in Medicine*. 2014;33(13):2297-340.  
9  
10 37. Berkley JA, Brent A, Mwangi I, English M, Maitland K, Marsh K, et al. Mortality among  
11 Kenyan children admitted to a rural district hospital on weekends as compared with  
12 weekdays . *Pediatrics*. 2004;114:1737-8.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

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

	Item No	Recommendation	Page No.
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract  <i>A retrospective observational study</i>	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found <b>Participants and measures</b> <i>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.</i> <b>Results</b> <i>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].</i> <b>Conclusion</b> <i>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.</i>	2 – 3
<b>Introduction</b>			
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported  <i>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</i>	2, 4 – 5
Objectives	3	State specific objectives, including any prespecified hypotheses: <i>Objectives:</i> <b>1) Experiment 1:</b> <i>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.</i>	8



		<i>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.</i>	
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper: <i>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)</i>	5 – 8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection <i>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</i>	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: <i>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.</i>	8 – 9
		(b) For matched studies, give matching criteria and the number of controls per case <i>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</i>	10 – 13
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable <i>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).</i>	
Data sources/ measurements	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  <i>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.</i>	9
Bias	9	Describe any efforts to address potential sources of bias <i>As the routine dataset used was limited to only observed variables,</i>	10 – 12



		<i>sensitivity analyses were conducted using propensity score trimming and instrumental variables to examine any potential influence of unmeasured variables</i>	
Study size	10	Explain how the study size was arrived at: <i>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).</i>	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why <i>All quantitative variables were used as continuous variables in the propensity score and outcome models</i>	<i>See the additional file 1: analysis protocol l on variables selected</i>
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding <i>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.</i>	10 – 12
		(b) Describe any methods used to examine subgroups and interactions <i>No sub – group analysis conducted</i>	
		(c) Explain how missing data were addressed <i>Missing data were handled using multiple imputation</i>	10
		(d) If applicable, explain how matching of cases and controls was addressed <i>Matching was conducted using propensity score methods – and has been outlined in the methodology section</i>	10
		(e) Describe any sensitivity analyses <i>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</i>	11 – 12
<b>Results</b>			
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 <i>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.</i>	13
		(b) Give reasons for non-participation at each stage <i>As we used routine dataset in analysis – the exclusion and inclusion were already defined in the analysis protocol (additional file 1: analysis protocol)</i>	
		(c) Consider use of a flow diagram <i>The analysis populations have been defined in figure 1</i>	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders <i>See figure 1, also see figure D of the additional file 2: supplementary</i>	

		<i>data</i>	
		(b) Indicate number of participants with missing data for each variable of interest <i>This has been referred to on page 10 and full results presented in the supplementary material provided (additional file 2: supplementary data)</i>	
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure <i>See figure 1</i>	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 <i>Treatment effect estimates are summarised in table 1</i>	See table 1, page 17
		(b) Report category boundaries when continuous variables were categorized <i>Continuous variables were not categorised</i>	
		(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 <i>No subgroup analyses were conducted</i>	
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives <i>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).</i>	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 <i>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.</i>	20 – 21

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
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.

<p>15 Interpretation 20</p>	<p>Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence <i>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.</i></p>	<p>20</p>
<p>26 Generalisability 21</p>	<p>Discuss the generalisability (external validity) of the study results <i>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.</i></p>	<p>21</p>
<p>44 <b>Other information</b></p>		
<p>45 Funding 22</p>	<p>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 <i>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.</i></p>	<p>22</p>

56 \*Give information separately for cases and controls.

57  
58  
59 **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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

For peer review only