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

Lucas Malla[1*]

Rafael Perera-Salazar[2]

Emily McFadden[2]

Mike English[1, 3]

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

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

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

The supplementary material is organised into the following subsections:

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

a) Under-five mortality incident rates in Kenya

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

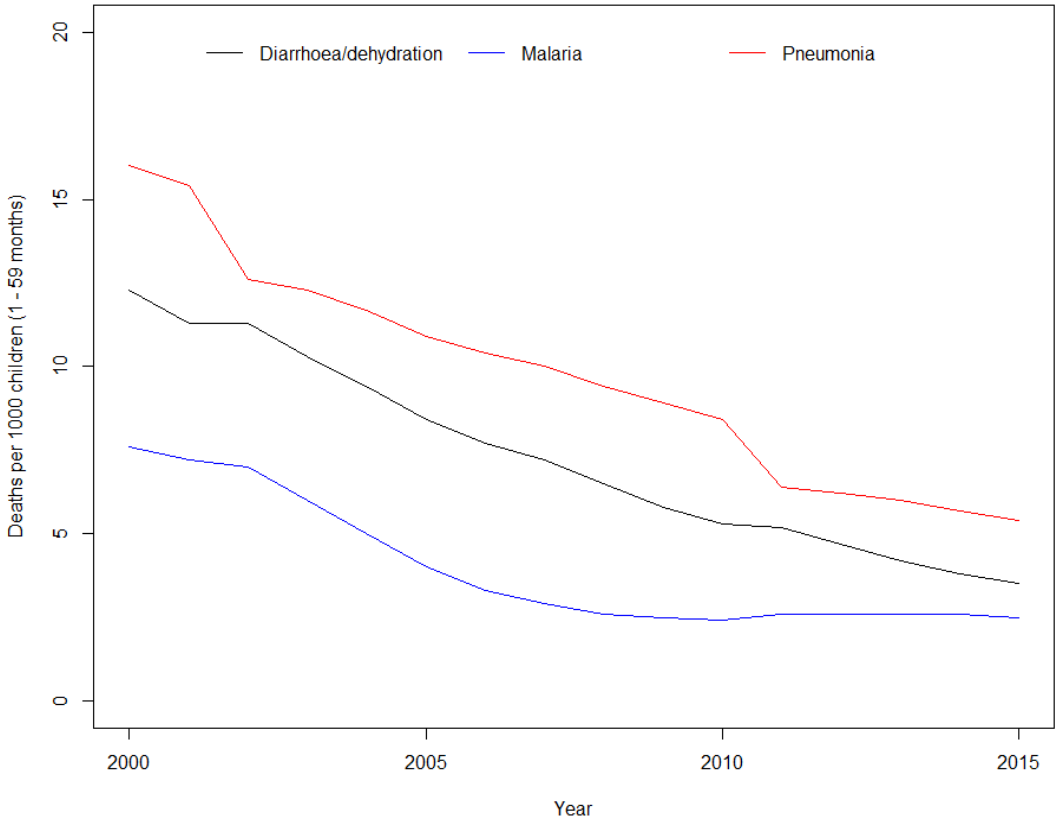

Figure A: Deaths, rate per 1000

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

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

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

c) Overlap and correctness of penicillin and gentamicin dosing

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

Table B: Correctness of penicillin and gentamicin prescription

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

d) Summary of key and auxiliary independent variables

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

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

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

e) Trimming in experiment 2 (ITT population)

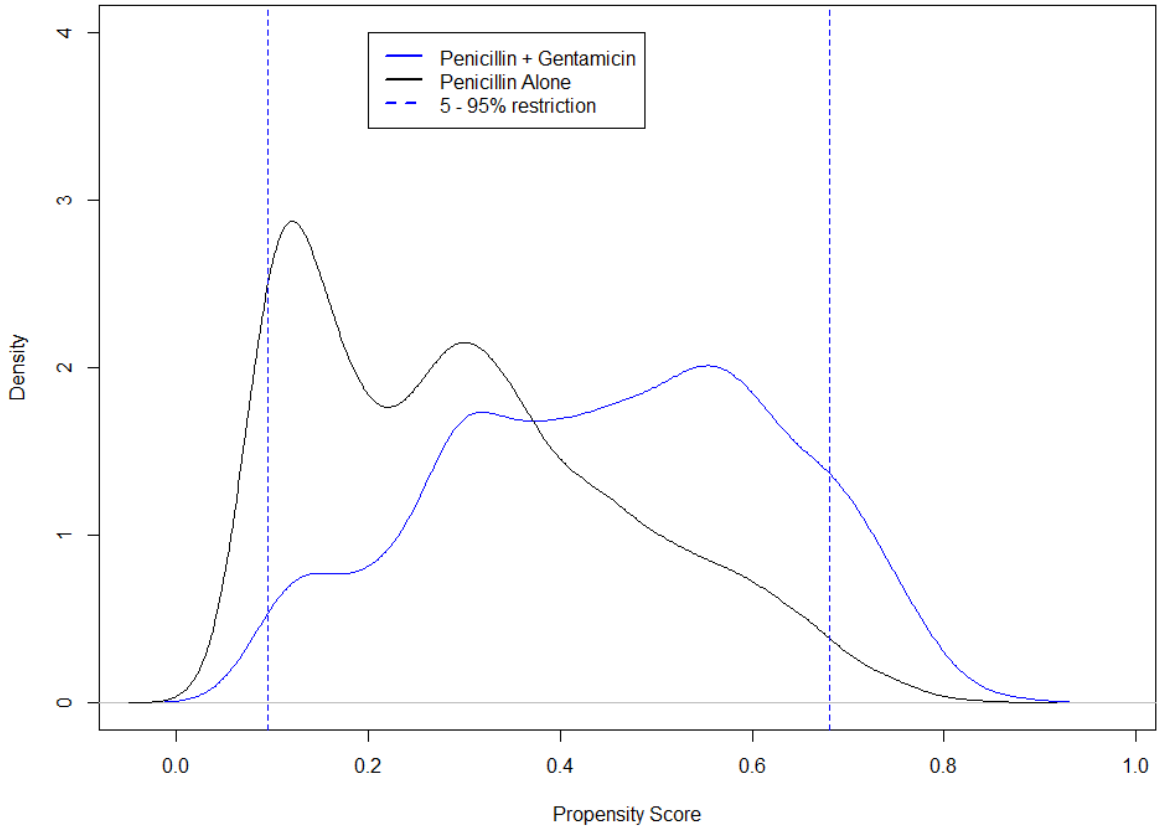

Figure B: PS trimming in experiment 2

f) Analysis using Instrument variables

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

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

Table D: Imbalance of covariates between weekday and weekend admissions

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

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

Table E: Summary of deaths by weekend/weekday admissions

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

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

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

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

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

g) Analysis using PS sub - classification

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

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

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

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

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

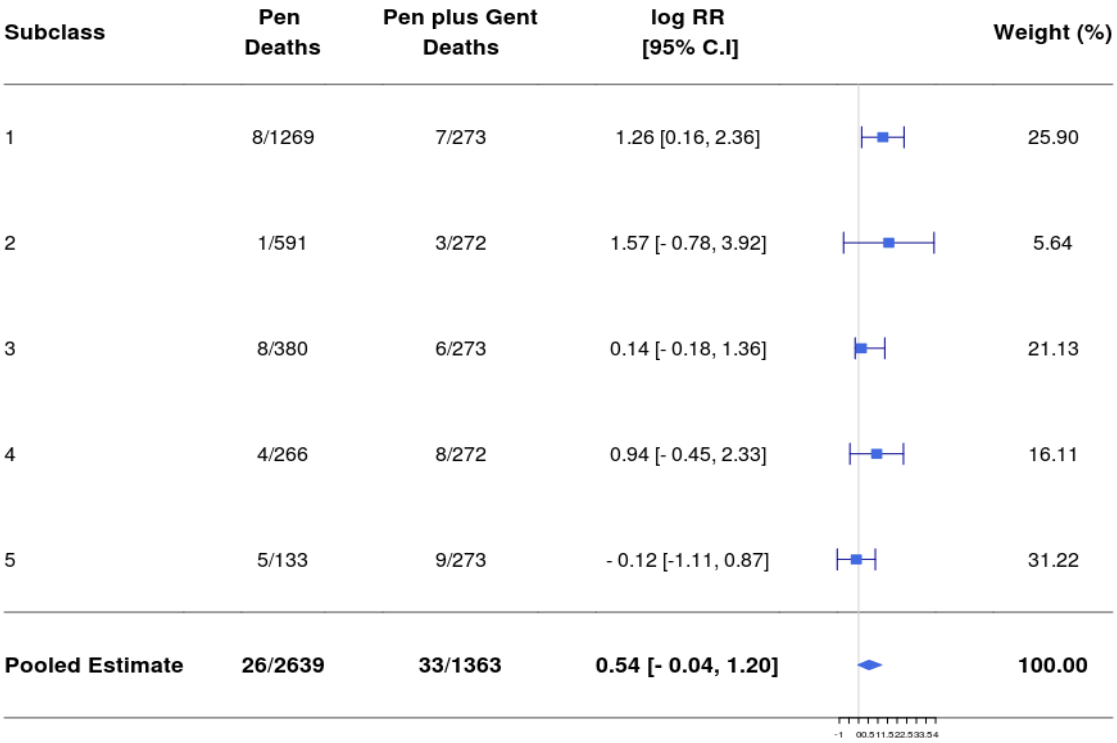

Figure C: Experiment 1 – ITT

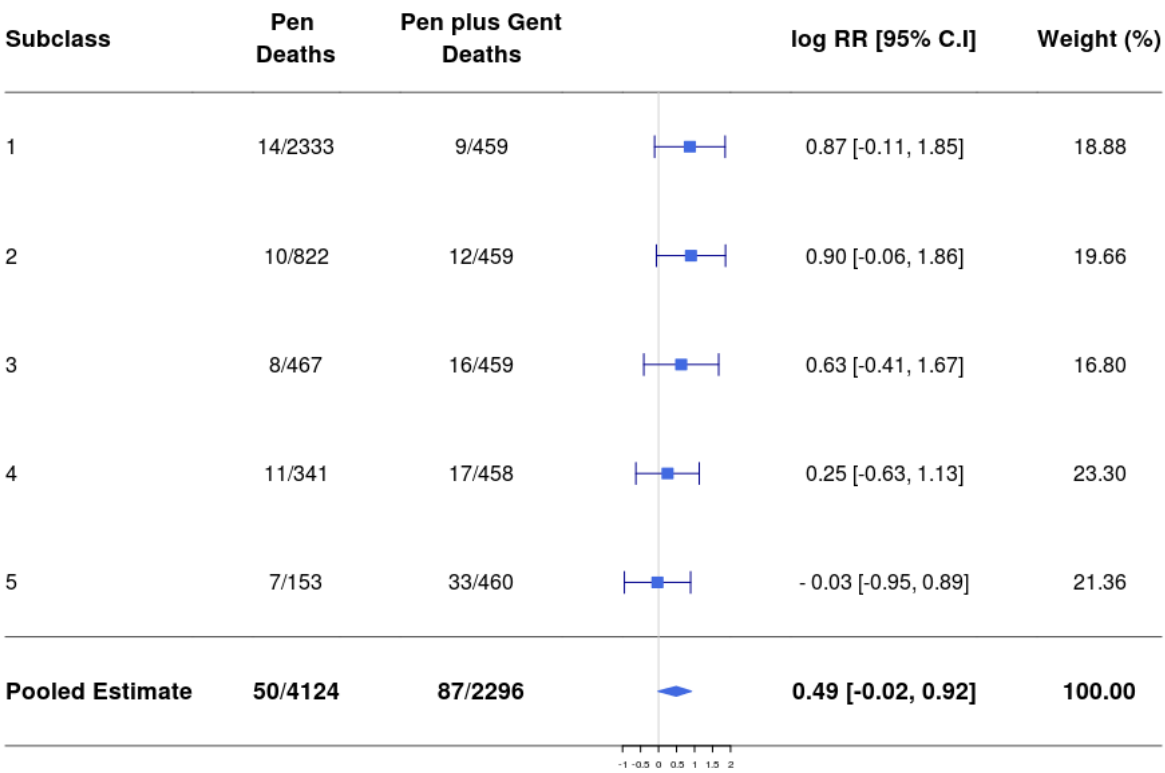

Figure D: Experiment 2 – ITT

h) Analysis using PS optimal full matching

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

Table G: Treatment effect estimates

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

i) Analysis using per protocol population

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

Table H: Per protocol treatment effect estimates

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

j) Definition of PS methods and how they were used

We implemented three PS methods and these are briefly introduced:

Optimal full matching

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

PS Weighting

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

PS sub – classification

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

References

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

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

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

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

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

6. Gu X, Rosenbaum PR. Comparison of multivariate matching methods: structures, distances and algorthms. 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.