Methodology

Methods for analyzing cost effectiveness data from cluster randomized trials

Max O Bachmann*[1], Lara Fairall[2], Allan Clark[1] and Miranda Mugford[1]

Address: [1]School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, UK and [2]Lara Fairall, Research Fellow, University of Cape Town Lung Institute, University of Cape Town, Cape Town, South Africa

Email: Max O Bachmann* - m.bachmann@uea.ac.uk; Lara Fairall - lfairall@uctgsh1.uct.ac.za; Allan Clark - allan.clark@uea.ac.uk; Miranda Mugford - m.mugford@uea.ac.uk

* Corresponding author

Abstract

Background: Measurement of individuals' costs and outcomes in randomized trials allows uncertainty about cost effectiveness to be quantified. Uncertainty is expressed as probabilities that an intervention is cost effective, and confidence intervals of incremental cost effectiveness ratios. Randomizing clusters instead of individuals tends to increase uncertainty but such data are often analysed incorrectly in published studies.

Methods: We used data from a cluster randomized trial to demonstrate five appropriate analytic methods: 1) joint modeling of costs and effects with two-stage non-parametric bootstrap sampling of clusters then individuals, 2) joint modeling of costs and effects with Bayesian hierarchical models and 3) linear regression of net benefits at different willingness to pay levels using a) least squares regression with Huber-White robust adjustment of errors, b) a least squares hierarchical model and c) a Bayesian hierarchical model.

Results: All five methods produced similar results, with greater uncertainty than if cluster randomization was not accounted for.

Conclusion: Cost effectiveness analyses alongside cluster randomized trials need to account for study design. Several theoretically coherent methods can be implemented with common statistical software.

Background

Cluster randomized trials are commonly used to evaluate the effectiveness and cost effectiveness of interventions in health care, health promotion and health professional education. Groups of individuals, such as doctors' patients or schools' pupils, are allocated together to receive different interventions or to follow usual practice. One key advantage of randomly allocating groups rather than individuals is that it permits inferences about the intervention's effects on service providers as well as on

users. For example, in a trial of an educational intervention aimed at doctors, allocating doctors together with their patients permits inferences about the intervention's effects on doctors as well as on their patients. But cluster randomization tends to reduce the statistical power and precision of trials because of similarities between individuals within each cluster, compared to individuals in other clusters. This similarity, or non-independence, is expressed as an intra-cluster, or intra-class, correlation coefficient (ICC) [1]. The ICC is the proportion of

response variance that occurs between clusters, as a proportion of the total variance (within and between clusters) [1]. Statistical methods for comparing outcomes in cluster randomized trials are now well developed [1,2]. But there has been little methodological research on appropriate methods for jointly analyzing cost and effectiveness data from cluster randomized trials [3,4].

Cost effectiveness analysis with individual level costs and outcomes is more complex than analysis of effects or costs alone, because differences in costs and differences in outcomes need to be analysed together. Uncertainty about cost effectiveness estimates can be quantified as confidence intervals for incremental cost effectiveness ratios (ICERs), or probabilities that interventions are cost effective, using two general approaches [5]. One approach combines cost and effect data on a two dimensional cost effectiveness plane (for example, Figure 1) [5-7]. It models the cost difference in one dimension and the outcome difference in the other, taking into account their variances and covariance and producing a cost effectiveness ellipse. The other approach combines cost and outcome measures for each individual by calculating net benefits, which are then compared between trial arms [4,8,9]. Both general approaches can be used for cluster randomized trials [3,4].

Clustering of costs and of outcomes is often neglected in economic evaluations alongside cluster randomized trials that have individual cost data. Several economic evaluations alongside cluster randomized trials have used bootstrapping to deal with asymmetrically distributed costs [10-12], but without specifying whether they sampled individuals or clusters. If they simply sampled individuals then cluster randomization design effects would not have been accounted for. Other studies adjusted for clustering using hierarchical regression models to estimate cost differences and outcome differences and their variances [13,14]. They then used these values to plot confidence ellipses. But neither of these two studies estimated ICER confidence intervals or considered correlations between costs and effects. Others have adjusted for clustering when comparing costs and comparing outcomes, but without estimating ICERs or probabilities that the intervention was cost effective [15].

In this paper we describe several appropriate methods for analyzing cost effectiveness data from cluster randomized trials. We show how to apply these methods using data from one such trial [16].

Example
This cluster randomized trial evaluated an educational intervention aimed at improving the management of lung disease in adults attending South African primary care
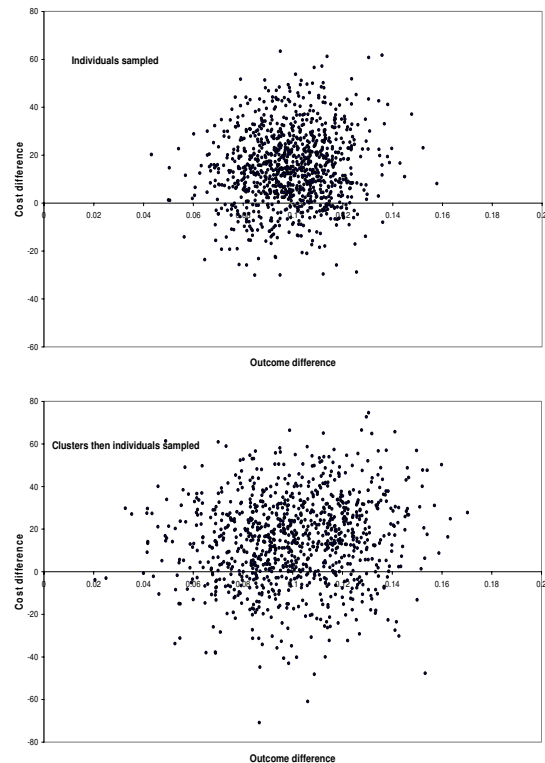

Figure 1
Cost effectiveness plane from bootstrap sampling of individuals, or clusters then individuals.

clinics [16]. Forty clinics were randomized to intervention or control arms. In each clinic 50 patients were interviewed at baseline and 3 months later. The trial outcome, indicating appropriate care, was defined as present if a patient was 1) newly diagnosed as having tuberculosis or 2) treated with inhaled corticosteroids for asthma or 3) referred for higher level care in the presence of defined indicators of severe illness. Health service costs were measured for each subject and included costs of clinic and hospital attendance, investigations, drugs, ambulance transport and the educational intervention. Complete data were available on 1856 patients. The prevalence of desirable outcome was 21% in the intervention arm and 11% in the control arm (odds ratio 2.2, 95% CI adjusted for clustering 1.5–3.2, ICC = 0.061). Costs were positively skewed, with mean costs of 220 South African Rand (ZAR) (median 159, interquartile range 77–266) in the intervention arm and ZAR 205 (median 140, interquartile range 70–218) in the control arm, and ICCs of 0.01 in each arm. The point estimate for the incremental cost effectiveness ratio was thus ZAR150 per unit or effect ((ZAR 220–205)/

(21%–11%)). Net benefits were not normally distributed. At willingness to pay of zero they were, by definition, the negative of costs, that is, with negative values and negatively skewed. At higher willingness to pay levels, net benefits were bimodal, depending on whether subjects had experienced the outcome or not. The ICC for net benefits ranged from 0.01 (at willingness to pay of zero) to 0.036 (at willingness to pay of ZAR2000). These ICCs indicated that clustering of costs and outcomes needed to be accounted for. The remainder of this paper shows how this can be done. For each method we outline the analytic principles, describe how they were implemented with Stata [17] or WinBUGS [18] software, and show the results.

Methods
1. Joint modelling of costs and effects
1.1 Bayesian parametric model
Nixon and Thompson have shown how to model costs and effects jointly for individually randomized trial data [7]. We have adapted the model to cluster randomized trials. We assume that we have two treatments which we wish to compare. In general, we assume that cost (C) and effectiveness (E) have distributions which can be characterised by their means and variances:

$$c_{ijk} \sim \text{Dist}(\mu_{ijk}^C, \sigma_{ijk}^C)$$
$$e_{ijk} \sim \text{Dist}(\mu_{ijk}^E, \sigma_{ijk}^E)$$

where μ_{ijk}^C is the mean cost for the ith the individual in the jth cluster in arm k of the trial and σ_{ijk}^C is the standard deviation of the cost. Similarly μ_{ijk}^E is the mean effectiveness for the ith the individual in the jth cluster in arm k of the trial and σ_{ijk}^E is the standard deviation of the effectiveness. We assume that we can model the mean cost and effectiveness as a linear combination:

$$\mu_{ijk}^C = \alpha^C + \tau_k^C + \eta_j^C,$$

$$\mu_{ijk}^E = \alpha^E + \tau_k^E + \eta_j^E + (\beta_k + \phi_j)(c_{ijk} - \mu_{ijk}^C),$$

In equation (2), α^C is the average cost for the control treatment and τ_k^C is the additional cost for treatment k (by default $\tau_1^C = 0$); η_j^C is the deviation from the average cost of centre j. It is possible to extend this model to allow for covariates [7]. Similarly, in equation (3) α^E is the average effect, or outcome, for the control treatment and τ_k^E is the

additional effect of treatment k (by default $\tau_1^C = 0$); η_j^C is the deviation from the average effectiveness of centre j. β_k is a parameter which allows for the relationship between costs and effects, ϕ_j and allows this to vary between clusters. From this model we can define the ICER as:

$$ICER = \frac{\tau_2^C}{\tau_2^E}.$$

We also define the probability that the intervention is cost effective (for a given willingness to pay, say λ) as $\Pr(e\lambda - c > 0)$, where e is the effect and c is the difference in cost.

In our example we shall compare two models, one assuming a normal distribution for the costs and another assuming a gamma distributions for the costs. Both models will assume a Bernoulli distribution for the effectiveness since this was a binary outcome measure. In both the normal and gamma distribution models we shall assume a linear link function (although traditionally a log link function would be used for the gamma distribution in order to ensure that the mean was estimated to be above zero). For the Bernoulli model we shall assume a logit link function since this is the standard model for effectiveness in clinical trials with binary outcome data. None of these models allow for individual (or cluster) level covariates.

In order to complete the Bayesian specification of the model we must assign prior distributions to all the unknown parameters. The particular priors that we shall use are:

$$\eta_j^E \sim N(0, \sigma_{E\eta}^2) \quad \eta_j^C \sim N(0, \sigma_{C\eta}^2)$$
$$\beta_k \sim N(0, \sigma_\beta^2) \quad \phi_j \sim N(0, \sigma_\phi^2)$$

All other parameters are assumed to follow a normal distribution with large variance.

1.2 Non-parametric bootstrapping
We can apply the bootstrap to the model defined by equation (1) whilst allowing both for the relationship between the cost and the effectiveness and for the non-independence of the cost and the effectiveness due to clustering of the data. This method has the advantage of not assuming a specific distribution for either the cost or the effectiveness [19].

The following algorithm will construct a bootstrap sample of K replications with m clusters each of size w. However this particular algorithm is only applicable in the situation where each cluster is of the same size; alternative algorithms are appropriate when this assumption is not

true [19,24]. The relationship between costs and effectiveness is retained due to their joint sampling.

Bootstrap algorithm
1. For the observed data set estimate the ICER, say $ICER^0$;

2. For i in 1 to K

a. For j in 1 to m

i. Randomly select (with replacement) a cluster centre, say k_j

ii. Within that cluster randomly select (with replacement) w sets of costs and effectiveness, these must be selected together in order to preserve the relationship between costs and effectivess.

b. Estimate the ICER on the basis of the bootstrap sample constructed in part *a* (say $ICER^i$), as the difference between treatment and control groups in mean costs, divided by the difference in mean outcomes.

Confidence intervals can then be constructed for the ICER in various ways [22]. We shall use the bias corrected accelerated percentile method [17]. We also estimated the probability that the intervention was cost effective. This was the proportion of iterations in which the effect, multiplied by the corresponding willingness to pay per effect, was greater than the cost difference.

2. Regression-based models of net benefits

The calculation of net benefits reduces costs and effectiveness to a single variable which can be used in standard regression analyses [5]. We define the net benefit (*nb*) as,

$$nb_{ijk} = e_{ijk}\lambda - c_{ijk},$$

Where, as before, e_{ijk} is the effectiveness on the ith person in the jth cluster in arm k, λ is the money society would be willing to pay for a unit of effectiveness, and c_{ijk} is the cost. Net benefit is expressed in monetary terms and so is also called net monetary benefit.

In a standard simple linear regression model with net benefit as the outcome variable and trial arm as the explanatory variable, the regression coefficient for the treatment term represents the incremental net benefit attributable to the intervention, for that level of willingness to pay. Willingness to pay is explored for a range of levels because it is usually not known. To estimate the corresponding incremental net benefit, 95% confidence limits and P values, a separate regression analysis is done at each willingness to pay level.

The intervention is defined to be cost effective, at a given willingness to pay level, if the corresponding incremental net benefit is greater than zero. Therefore the probability that the intervention is cost effective at a given willingness to pay level is the probability that the incremental net benefit is greater than zero. If the coefficient is greater than zero, then the probability that the intervention is cost effective is one minus half the one sided P value for the treatment term. If the coefficient is less than zero, then the probability that the intervention is cost effective is half the one sided P value. In Bayesian models the probability that the intervention is cost effective is the predictive probability that the net benefit is greater than zero. If the model is estimated using the Markov chain Monte Carlo method this is simply estimated as the proportion of iterations for which the incremental net benefit is greater than zero.

The ICER and its confidence limits can also be estimated from these regression results because the values of λ at which the 95% confidence intervals of the incremental net benefit estimate are equal to zero are thus the 95% confidence intervals of the ICER [5]. These can be estimated from the estimated incremental net benefits, and their confidence limits.

Net benefit regression has previously been used to analyse cost effectiveness data from multi-centred trials [4,9], and can be adapted to account for cluster randomization in various ways. We detail three possible methods.

2.1 Least squares regression of net benefits with robust estimates of standard errors
The standard error from the traditional regression model will be inaccurate and lead to an underestimation of the standard errors of the parameter estimates, but it is possible to compensate for this by using the Huber-White sandwich estimator [17]. The Huber-White sandwich estimator accounts for the non-independence of observations within each cluster. However, because it adjusts the standard errors post estimation the likelihood-ratio test statistics are not applicable.

2.2 Least squares regression of net benefits with a hierarchical model
Hierarchical (multi-level) linear regression models account for clustering of net benefits by modelling individuals at the first level and clusters at the second. We used Stata's xtmixed procedure [17] to specify a hierarchical model with net benefit as outcome, with trial arm as explanatory variable and with coefficients and intercepts varying randomly between clusters. In particular, we consider a model of the form:

$$nb_{ijk} = \mu_{ijk} + \varepsilon_{ijk}$$

$$\mu_{ijk} = \mu + \tau_k + \theta_j + \upsilon_{jk}$$

$$\theta_j \sim N(0,\sigma_\theta^2)$$

$$v_{jk} \sim N(0,\sigma_{jk}^2)$$

Where μ is the overall mean net benefit, τ_2 is the additional net benefit associated with the treatment arm (by default $\tau_2 = 0$), θ_j is the deviation in net benefit due to the jth clusters centre and $v_{j,k}$ is the deviation in net benefit in centre j (by default all $v_{1,k}$ are zero).

2.3 Net benefit regression with a Bayesian hierarchical model
Spiegelhalter [24] described a Bayesian two level linear regression model for comparing continuous outcomes in cluster randomized trials, defined as follows:

$$nb_{ijk} \sim N\ (\mu_{ijk},\ \sigma^2),$$

$$\mu_{ijk} = \mu + \tau_k + \theta_j,$$

Where, μ is the mean net benefit for the control arm; τ_2 is the additional net benefit associated with the treatment arm (by default $\tau_1 = 0$); and θ_j is the deviation in net benefit due to the jth clusters centre. In order to complete the model description we shall use the following prior distributions:

$$\theta_j \sim N(0,\sigma_\theta^2)$$

$$\tau_2 \sim N(0,100000)$$

$$\mu \sim N(0,100000)$$

$$\sigma_\theta^2 \sim U(0,100000)$$

This model must be estimated separately for each level of willingness to pay. It could be extended by including an additional random effect of incremental net benefit by centre.

Results
Joint models of cost and effect
The joint models were fitted using WinBUGS. In particular we ran the Bernoulli-Normal and Bernoulli-Gamma models with random intercept defined by equations 2 and 3. We ran the models for 10,000 iterations, using the first 5,000 as burn-in and the remaining 5,000 were thinned at an interval of 5 in order to give an approximately independent sample. We found that the best fitting model, as indicated by the smallest deviance information criterion, was the Bernoulli-Gamma model including ϕ_j. This model gave an ICER of 189 (95% CI -194 to 647) (Table 1) with probabilities of being cost effective as shown in Figure 2.

Table 1: Incremental cost effectiveness ratios (ZAR).
Incremental cost effectiveness ratios (in South African Rand)

Analytic method	ICER	95% confidence limits
Cost effectiveness plane		
Bootstrapping individuals	150	-143, 489
Bootstrapping clusters then individuals	150	-918, 217
Bayesian hierarchical model	189	-194, 647
Linear regression of net benefits		
Least squares model without adjustment	154	-162, 481
Least squares model with robust adjustment	154	-257, 575
Least squares hierarchical model	155	-244, 568
Bayesian hierarchical model	157	-282, 600

ICER incremental cost effectiveness ratio.

All of these models gave similar probabilities of being cost effective.

The non-parametric bootstrap was carried out with 1,000 bootstrap replications and gave an ICER of 150 (95% CI -918 to 217), the results are shown in Table 1 and in Figure 2. Applying the bootstrap while ignoring the clusters and simply randomly selecting individuals and costs resulted in the same estimated ICER but with much narrower confidence intervals. These results are summarised in Table 1 and Figure 2. The long left tail of the ICER distribution was due to bootstrap samples that produced moderate (negative) differences in costs accompanying small effects, resulting in large negative ICERs.
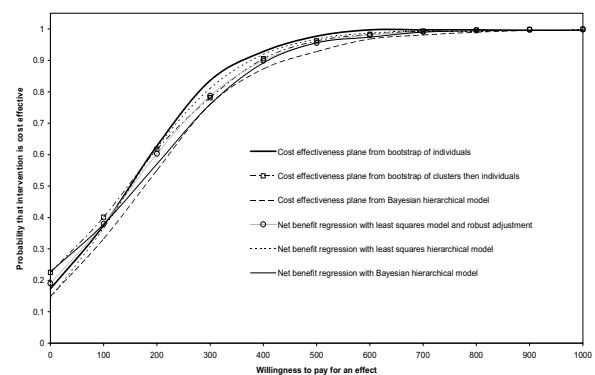

Figure 2
Cost effectiveness acceptability curve: probability that the intervention was cost effective for different levels of willingness to pay and with different analytic methods.
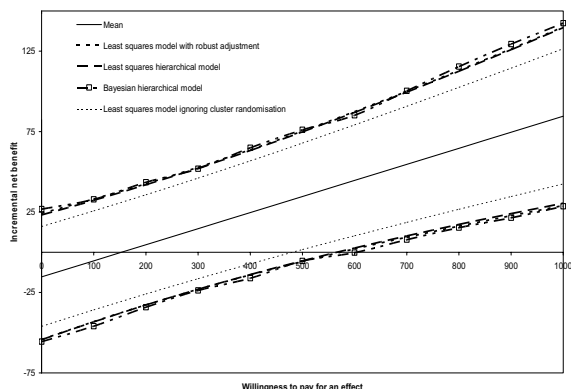
Figure 3
Incremental net benefit at different levels of willingness to pay per unit of effect: 95% confidence limits estimated with different regression models.

Net benefit models

The standard linear regression model gave an estimated ICER of 154 (95% CI -162 to 481); once the robust standard errors were taken into account the confidence interval for the ICER increased to (-257 to 575). The least squares hierarchical model resulted in an estimated ICER of 155 (95% CI -244 to 568) and the Bayesian hierarchical model had an estimated ICER of 157 (95% CI -282,600). These results are contained in Table 1 and Figures 2 and 3.

Discussion

We have demonstrated how individual cost and outcome data from cluster randomized trials can be analysed in various ways with widely available software. In our example different appropriate methods produced similar results. Predictably, adjusting for clustering resulted in greater uncertainty about ICERs, and lower probabilities that the intervention was cost effective, compared to methods that ignored clustering. Of all our results, the ICER confidence intervals from 2 stage bootstrapping differed most from confidence intervals from other methods. ICER point estimates were similar (150–157) except for the Bayesian hierarchical joint model of costs and effects (189) because the latter assumed a gamma distribution of costs.

To be able to interpret ICER confidence intervals that include zero, one needs to plot uncertainties about differences in costs and effects on a cost effectiveness plane (Figure 1). This is because, if the cost effectiveness ellipse extends to non-adjacent quadrants, a negative ICER confidence interval is uninterpretable. This is because it combines information about dominant situations with greater costs and worse outcomes, and situations with lower costs and better outcomes [5]. But in this example the negative ICER represented the latter situation only (Figure 1). That is, even if society was willing to pay up to these amounts (the lower confidence limits in Table 1) to *avoid* one unit of effect, the intervention would still be cost effective because of cost saving. So the lower confidence limit, which is reassuring, is here of less interest than the upper confidence limit, which shows how much might have to be paid for an effect.

Our example of one trial has limited generalizability, which would be enhanced by comparing these methods using different data from other trials and from simulations. Problems could potentially occur with fewer clusters or with varying numbers of individuals per cluster. For example, Flynn and Peters used simulated data to show that, with 24 or fewer clusters per arm, Stata's bootstrap ICER estimates may be spuriously precise [3]. They also found that Stata's robust adjustment performed better than its bootstrap procedures in estimating cost differences [23]. Net benefit regression may be invalid with small samples if net benefits are not normally distributed. In this example, however, estimates from net benefit regression were similar to nonparametric bootstrap estimates, as predicted by the central limit theorem. The bootstrap methods we describe may be inappropriate if cluster sizes vary [3,20], in which case more sophisticated methods might be needed [24]. Net benefit regression models and Nixon's and Thompson's two dimensional model [7] need not assume equal cluster size. Nixon's and Thompson's model has several other advantages. It can accommodate various cost distributions, does not need to be repeated at different willingness to pay levels, does not need separate regression analyses to estimate ICER confidence intervals, and can produce the cost effectiveness ellipse needed to interpret a negative ICER. Two stage bootstrapping has similar advantages, but we found its results to be unreliable over repeated analyses, even with 10000 iterations. A pragmatic approach is to check the robustness of the primary analysis by also using another method, especially if there are few clusters or if their sizes vary.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

MOB conceived of the paper, chose the statistical methods, carried out the Stata analyses, and was lead writer of the paper and co-investigator of the randomised trial and economic evaluation. LF carried out Stata analyses and was principal investigator of the trial and the economic evaluation. AC helped choose the Bayesian models and

carried out and described the WinBUGS analyses. MM was a co-investigator on the economic evaluation. All authors contributed to writing the paper and read and approved the final manuscript.

Acknowledgements

References

1. Donner A, Klar N: *Design and Analysis of Cluster Randomization Trials in Health Research* London: Hodder Arnold; 2000.
2. Special Issue: **Design and Analysis of Cluster Randomized Trials.** *Stat Med* 2001, **20**:329-496.
3. Flynn TN, Peters TJ: **Cluster randomized trials: another problem for cost-effectiveness ratios.** *Int J Technol Assess Health Care* 2005, **21**:403-9.
4. Manca A, Rice N, Sculpher MJ, Briggs AH: **Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models.** *Health Econ* 2005, **14**:471-85.
5. Drummond MF, Sculpher MJ, Torrance GW, O'Brien B, Stoddart : *Methods for the Evaluation of Health Care Programmes* 3rd edition. Oxford: Oxford University Press; 2005:247-268.
6. Briggs A: **Handling uncertainty in economic evaluation and presenting the results.** In *Economic Evaluation in Health Care. Merging Theory with Practice* Edited by: Drummond M, McGuire A. Oxford: Oxford University Press; 2001:172-214.
7. Nixon RM, Thompson SG: **Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations.** *Health Econ* 2005, **14**:1217-1229.
8. Hoch JS, Briggs AH, Willan AR: **Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis.** *Health Econ* 2002, **11**:415-430.
9. Willan AR, Briggs AH, Hoch JS: **Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectivenessdata.** *Health Econ* 2004, **13**:461-475.
10. Sullivan SD, Lee TA, Blough DK, Finkelstein JA, Lozano P, Inui TS, Fuhlbrigge AL, Carey VJ, Wagner E, Weiss KB: **A multisite randomized trial of the effects of physician education and organizational change in chronic asthma care: cost-effectiveness analysis of the Pediatric Asthma Care Patient Outcomes Research Team II (PAC-PORT II).** *Arch Pediatr Adolesc Med* 2005, **159**:428-34.
11. Goodacre S, Nicholl J, Dixon S, Cross E, Angelini K, Arnold J, Revill S, Locker T, Capewell SJ, Quinney D, Campbell S, Morris F: **Randomized controlled trial and economic evaluation of a chest pain observation unit compared with routine care.** *BMJ* 2004, **328**:254.
12. Gilbert FJ, Grant AM, Gillan MG, Vale LD, Campbell MK, Scott NW, Knight DJ, Wardlaw D, Scottish Back Trial Group: **Low back pain: influence of early MR imaging or CT on treatment and outcome-multicenter randomized trial.** *Radiology* 2004, **231**:343-51.
13. Johnston K, Gray A, Moher M, Yudkin P, Wright L, Mant D: **Reporting the cost-effectiveness of interventions with nonsignificant effect differences: example from a trial of secondary prevention of coronary heart disease.** *Int J Technol Assess Health Care* 2003, **19**:476-89.
14. Morgan K, Dixon S, Mathers N, Thompson J, Tomeny M: **Psychological treatment for insomnia in the regulation of long-term hypnotic drug use.** *Health Technol Assess* 2004:8.
15. Venning P, Durie A, Roland M, Roberts C, Leese B: **Randomized controlled trial comparing cost effectiveness of general practitioners and nurse practitioners in primary care.** *BMJ* 2000, **320**:1048-53.
16. Fairall L, Zwarenstein M, Bateman EB, Bachmann MO, Lombard C, Majara B, Joubert G, English R, Bheekie A, Mayers P, Peters A, Chapman R: **Educational outreach to nurses improves tuberculosis case detection and primary care of respiratory illness: a pragmatic cluster randomized controlled trial.** *BMJ* 2005, **331**:750-754.
17. StataCorp: *Stata Statistical SoFtware: Release 9* College Station, TX; 2005.
18. Spiegelhalter D, Thomas A, Best N, Lunn D: *WinBUGS User Manual. Version 1.4.1* [http://www.mrc-bsu.cam.ac.uk/bugs].
19. Davison AC, Hinkley DV: *Bootstrap Methods and Their Application* Cambridge: Cambridge University Press; 1997:100-102.
20. Barber JA, Thompson SG: **Analysis of cost data in randomized trials: an application of the non-parametric bootstrap.** *Stat Med* 2000, **19**:3219-36.
21. Carpenter J, Bithell J: **Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians.** *Stat Med* 2000, **19**:1141-64.
22. Spiegelhalter DJ: **Bayesian methods for cluster randomized trials with continuous responses.** *Stat Med* 2001, **20**:435-452.
23. Flynn TN, Peters TJ: **Use of the bootstrap in analysing cost data from cluster randomized trials: some simulation results.** *BMC Health Services Res* 2004, **4**:33.
24. Rao JNK, Wu CHJ: **Resampling inference with complex survey data.** *J Am Stat Assoc* 1998, **83**:231-241.