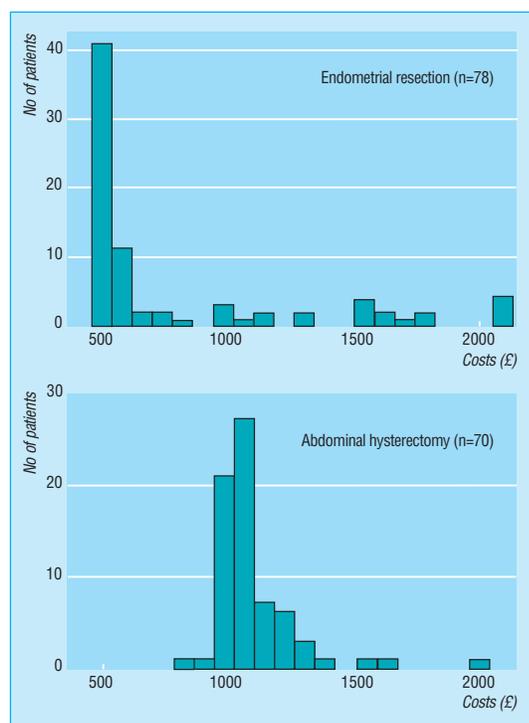


How should cost data in pragmatic randomised trials be analysed?

Simon G Thompson, Julie A Barber

Pragmatic randomised trials are usually large scale multicentre studies in which interventions or medical policies are compared in a realistic setting.¹ The intention is that conclusions from these trials, if accepted, can be adopted directly into medical practice.² Economic evaluations carried out alongside these trials are increasingly common because it is often important to assess costs and cost effectiveness as well as clinical outcomes.³ Costs are usually derived from information about the quantity of healthcare resources used by each patient in the trial. The quantities of each resource used are multiplied by fixed unit cost values and are then summed over the separate types of resource to give a total cost per patient.⁴

This information leads to a range of different costs across participants in the trial. As an example, the figure shows the distribution of costs in women with



Distribution of costs from a trial comparing endometrial resection with hysterectomy in women with menorrhagia. Costs are based on health resource use from randomisation to two years; they include preoperative, operative, hospital stay, complications, retreatment, and primary care components⁵

Summary points

Health economic evaluations are now commonly included in pragmatic clinical trials that inform policy decisions

Despite the usual skewness in the distribution of costs, it is the arithmetic mean that is the most informative measure

Measures other than the arithmetic mean do not provide information about the cost of treating all patients, which is needed as the basis for healthcare policy decisions

Statistical analysis based on transforming cost data or comparing medians using standard non-parametric methods may provide misleading conclusions

menorrhagia randomised to treatment with endometrial resection or abdominal hysterectomy.⁵ Such highly skewed distributions are typical of cost data; the long right hand tail reflects the fact that some patients incur high costs because of factors such as medical complications, reoperation, or extended hospital stay.

What aspect of cost data is important?

When information about the costs of alternative treatments is to be used to guide healthcare policy decision making, it is the total budget needed to treat patients with the disease that is relevant. For example, healthcare planners may need information about the total annual budget required to provide a treatment at a particular hospital. An estimate of this total cost is obtained from data in a trial by multiplying the arithmetic mean (average) cost in a particular treatment group by the total number of patients to be treated. It is therefore the arithmetic mean that is the informative measure for cost data in pragmatic clinical trials.

Other measures, however, are often reported when describing cost data. For example, the median cost is the value below and above which the costs of half the patients lie. Another measure, the geometric mean cost, can be derived by transforming the costs onto a logarithmic scale, calculating the average, and

Department of Medical Statistics and Evaluation, Imperial College School of Medicine, London W12 0NN

Simon G Thompson
professor

Julie A Barber
research fellow

Correspondence to: S G Thompson, MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR

simon.thompson@mrc-bsu.cam.ac.uk

BMJ 2000;320:1197-200

transforming this back. For positively skewed data such as those in the figure, the median and geometric mean are always less than the arithmetic mean. For example, in the endometrial resection group, the median cost was £523, the geometric mean was £683, and the arithmetic mean was £790. The extent of differences between these quantities depends on the shape and skewness of the distribution. Hence, in the hysterectomy group, in which cost data are less skewed, the median of £1053 and the geometric mean of £1100 are closer to the arithmetic mean of £1110.

Measures other than arithmetic means may be useful for some purposes. For example, the median cost may be used to describe a “typical” cost for an individual. Knowledge of the probability of incurring a particularly extreme cost may be useful to a medical insurance company. Measures other than the arithmetic mean, however, do not provide information about the total cost that will be incurred by treating all patients, which is needed as the basis for healthcare policy decisions.

How should costs be compared?

Many commonly used statistical methods require that data approximate a symmetrical bell shaped—or normal—distribution. Researchers have therefore chosen statistical techniques which try to deal with the skewness in the distribution of cost data. At first sight this is reasonable, given the advice in statistical guidelines and textbooks. For example, the *BMJ*'s statistical guidelines state that “data which have a highly skewed (asymmetrical) distribution ... may require either some transformation before analysis or the use of alternative ‘distribution free’ methods.”⁶ A transformation of the data, such as a logarithmic transformation, might be used to achieve a more normal distribution, for which “parametric” methods such as a *t* test are appropriate. Alternatively, “non-parametric” or distribution free methods, which are appropriate for any shape of distribution, could be used.

This conventional advice implies that the method of analysis should be chosen on the basis of the shape of the distribution of the data. However, the method of analysis used also has important implications for the interpretation of results, since different methods compare different aspects of the distributions. A *t* test on untransformed data compares arithmetic means, while a *t* test on log transformed data compares geometric means. The Mann-Whitney U test, a non-parametric method, is often interpreted as a comparison of medians, although it is in fact an overall comparison of distributions in terms of both shape and location.⁷ Out of these three tests, only the *t* test on untransformed data can be appropriate for costs, since it is the only one that addresses a comparison of arithmetic means. A legitimate concern, and the basis of the conventional

statistical guidelines, is that methods based on the *t* test are strictly valid only if the cost data are normally distributed.⁸ However, a *t* test, and the confidence interval derived from it, will be reliable if either the skewness is not too extreme or the sample sizes are moderately large—an issue to which we return later.

Examples from three recent publications

In a pragmatic randomised trial comparing hospital at home with inpatient hospital care, the strategy for statistical analysis was as follows: “When appropriate, data with non-normal distributions was log transformed before further parametric analysis was done. The Mann-Whitney U test was used for continuous data that did not approximate a normal distribution after log transformation.”⁹ The table shows the result of this strategy for the group of hip replacement patients included in the trial. Arithmetic mean hospital costs were compared by using a *t* test, general practitioner costs were presented as medians and compared with a Mann-Whitney U test, and, although total costs were presented as arithmetic means, geometric means were compared statistically by using an analysis based on log transformed values. The confusion over methods of analysis and their resulting presentation is obvious. It stems, however, from following the conventional guidelines for the statistical analysis of continuous data. In addition, presenting arithmetic means while comparing geometric means statistically (which was, it seems, recommended recently¹⁰) can only encourage misinterpretation.

In a second example, a pragmatic randomised trial was carried out to assess the cost effectiveness over one year of day hospital compared with inpatient treatment for patients with acute psychiatric illness.¹¹ Because the cost data were skewed, the authors used medians to summarise the distributions and the Mann-Whitney U test to make comparisons between groups. This analysis showed that total patient costs were statistically significantly lower in the day hospital group. It does not follow, however, that the arithmetic mean costs were also significantly lower. So the authors' conclusion that day hospital treatment was cheaper overall, which has direct policy implications, is not justified by the statistical analysis presented.

A similar example is provided by a pragmatic randomised trial evaluating care for discharged psychiatric patients. In this study, community multidisciplinary teams and hospital based care over one year were compared.¹² Arithmetic mean, median, and geometric mean costs were presented, but only the geometric mean costs were compared statistically, using a *t* test on log transformed values “to correct for skewed distribution.” As for medians in the previous example, the non-significant difference in geometric mean costs cannot be taken to imply a similar result for arithmetic mean costs.

Example of published comparison of cost data: costs (£) over 3 months after hospital admission for hip replacement patients allocated to hospital at home or inpatient care⁹

	Hospital at home (n=36)	Inpatient (n=49)	Difference (95% CI)	P value
Mean (SD) hospital costs	515.42 (473.20)	776.30 (364.53)	Arithmetic mean: -260.87 (-441.56 to -80.19)	<0.01
Mean (SD) hospital at home costs	351.24 (240.58)	—	—	—
Median (interquartile range) GP costs	42.84 (0-64.61)	15.49 (0-45.19)	Mann-Whitney U test	0.06
Mean (SD) total health service costs	911.39 (563.76)	815.70 (347.99)	Ratio of geometric means: 1.05 (0.87 to 1.27)	0.59

Does the choice of method matter?

In these examples, it is not clear whether using a comparison of arithmetic means would have changed the conclusions. The reader cannot be sure and cannot therefore draw reliable conclusions from the analyses presented. As the necessary analyses can readily be performed when original data are available, it is easy to find examples to show that the choice of method of analysis can make a difference to the conclusions. In a trial comparing a community based exercise programme and usual general practitioner care for patients with low back pain, the arithmetic mean costs over 12 months were £360 and £508 respectively.¹³ Using *t* test based methods to assess the mean difference of £148 gave a 95% confidence interval of –£146 to £442 and a non-significant P value of 0.32, thus providing no evidence of a difference. However, a Mann-Whitney U test applied to the same data gave a significant P value of 0.02, which would be interpreted as substantial evidence of a cost difference. Clearly, these two methods lead to very different interpretations for the cost evaluation, and if the Mann-Whitney U test had been used it would have been extremely misleading.

Another example is provided by the subgroup of hysterectomy patients included in the hospital at home trial described above.⁹ It was stated that in this case “health service costs were significantly higher for those allocated to hospital at home care.” The conclusion was based on a comparison of geometric means, the cited P value being <0.01. However, using the arithmetic means and standard deviations reported in the paper to carry out a standard *t* test gives a less significant P value of 0.1. Again, these two analyses lead to different interpretations.

How common are these problems?

A recent review of 45 randomised trials that included economic evaluations and were published in 1995 showed serious inadequacies in the use of statistical methods for costs.¹⁴ Among the papers that reported statistical comparisons, only half used methods that addressed differences in arithmetic means, and others used inappropriate non-parametric approaches (for example, Mann-Whitney U test) or log transformation approaches. The situation is made worse by recent articles giving incorrect or misleading advice about the statistical analysis of cost data. Although it has been mentioned that standard non-parametric methods are inappropriate, several authors have (wrongly) recommended carrying out analyses on log transformed cost data.^{15–18} These recommendations have influenced methods of analysis used in subsequent studies.¹⁹ In the context of cost data, the unthinking application of conventional statistical guidelines for analysing skewed data leads to inappropriate analyses and potentially misleading conclusions.

Appropriate methods of analysis

Given the need to compare treatment groups in terms of arithmetic mean costs, standard approaches such as *t* tests seem to be appropriate. Indeed, in the review of published economic evaluations, *t* tests were used for all the comparisons of arithmetic means reported.¹⁴

Their validity, however, relies on assumptions of “normality” and so is questionable for skewed cost data.⁸ Although these methods are known to be fairly robust to non-normality, especially if the sample size is large, robustness for a particular data set is difficult to judge.⁷ Standard methods for comparing arithmetic mean costs therefore may have to be used with caution.

One alternative approach is the non-parametric bootstrap.²⁰ This method avoids the need to make assumptions about the shape of the distribution, such as normality, and uses instead the observed distributions of the cost data in the study being analysed. Statistical analysis is based on repeatedly sampling from the observed data, using a computer program.²¹ Bootstrap methods can be used for hypothesis tests, calculating confidence intervals and regression analyses. The application of the non-parametric bootstrap to test and derive confidence intervals for differences in arithmetic mean costs has recently been described.^{21–22} As yet, bootstrap methods have not often been used for analysing costs in practice, although there are some recent examples.^{13–23–25}

In our experience, the results from standard *t* tests and *t* test based confidence intervals are adequate in most realistic situations for comparing arithmetic mean costs between two groups. For cost data in general, we prefer methods that do not assume that the standard deviations in the two groups are the same.²⁶ For example, in the menorrhagia trial (see figure), the 95% confidence intervals for the difference in arithmetic mean costs between groups (£320) were very similar whether a *t* test based method or bootstrap method was used (£204 to £437 and £192 to £426, respectively). This is despite the skewness of the cost data, especially in the endometrial resection group, and the moderate number of patients in each group (78 and 70). Even with lower sample sizes of about 15–20 patients per group and highly skewed cost data, results can be similar. For example, in a pilot trial of cognitive behavioural therapy for patients with deliberate self harm, P values for the *t* test and a bootstrap test were almost identical (0.20 and 0.21 respectively) and the methods again gave fairly similar confidence intervals.²³

Conclusions

In cost evaluations designed to have an impact on medical policy, it is the total healthcare cost that is important. Thus, despite the usually skewed distribution of cost data, it is analyses of arithmetic means that are informative. A simple *t* test of untransformed costs may be sufficient, but the validity of these results, especially for small samples or extremely skewed data, should be checked by using bootstrap techniques. There is a need for economic and statistical guidelines to be revised to emphasise these issues, since basing important policy decisions on studies that use inappropriate methods of analysis for costs may do more harm than good.

We thank Mark Sculpher and colleagues and Jennifer Klaber Moffett and colleagues for permission to use data from their studies.

Funding: ST was funded by HEFC London University, JB by Thames NHS Executive.

Competing interests: None declared.

- 1 Roland M, Torgerson D. What are pragmatic trials? *BMJ* 1998;316:285.
- 2 Fayers PM, Hand DJ. Generalisation from phase III clinical trials: survival, quality of life and health economics. *Lancet* 1997;350:1025-7.
- 3 Drummond MF, Stoddart GL. Economic analysis and clinical trials. *Control Clin Trials* 1984;5:115-28.
- 4 Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programs*. 2nd ed. Oxford: Oxford Medical Publications, 1997.
- 5 Sculpher MJ, Dwyer N, Byford S, Stirrat GM. Randomised trial comparing hysterectomy and transcervical endometrial resection: effect on health related quality of life and costs two years after surgery. *Br J Obstet Gynaecol* 1996;103:142-9.
- 6 Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *BMJ* 1983;286:1489-93.
- 7 Bradley JV. *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall, 1968.
- 8 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
- 9 Sheppard S, Harwood D, Gray A, Vessey M, Morgan P. Randomised controlled trial comparing hospital at home care with inpatient hospital care. II: cost minimisation analysis. *BMJ* 1998;316:1791-6.
- 10 Briggs AH, Gray AM. Handling uncertainty in economic evaluations of healthcare interventions. *BMJ* 1999;319:635-8.
- 11 Creed F, Mbaya P, Lancashire S, Tomenson B, Williams B, Holmes S. Cost effectiveness of day and inpatient psychiatric treatment: results of a randomised controlled trial. *BMJ* 1997;314:1381-5.
- 12 Tyrer P, Evans K, Gandhi N, Lamont A, Harrison-Read P, Johnson T. A randomised controlled trial of two models of care for discharged psychiatric patients. *BMJ* 1998;316:106-9.
- 13 Klaber Moffett J, Torgerson D, Bell-Syer S, Jackson D, Llewlyn-Phillips H, Farrin A, et al. Randomised controlled trial of exercise for low back pain: clinical outcomes, costs, and preferences. *BMJ* 1999;319:279-83.
- 14 Barber JA, Thompson SG. Analysis and interpretation of cost data in randomised controlled trials: review of published studies. *BMJ* 1998;317:1195-200.
- 15 Coyle D. Statistical analysis in pharmacoeconomic studies: a review of current issues and standards. *Pharmacoeconomics* 1996;9:506-16.
- 16 Rutten-van Molken MP, van Doorslaer EK, van Vliet RC. Statistical analysis of cost outcomes in a randomized controlled clinical trial. *Health Econ* 1994;3:333-45.
- 17 Gray AM, Marshall M, Lockwood A, Morris J. Problems in conducting economic evaluations alongside clinical trials. Lessons from a study of case management for people with mental disorders. *Br J Psychiatry* 1997;170:47-52.
- 18 Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy* 1998;3:233-45.
- 19 Knapp M, Marks I, Wolstenholme J, Beecham J, Astin J, Audini B, et al. Home-based versus hospital-based care for serious mental illness: controlled cost-effectiveness study over four years. *Br J Psychiatry* 1998;172:506-12.
- 20 Efron B, Tibshirani RJ. *An introduction to the bootstrap*. New York: Chapman and Hall, 1993.
- 21 Barber JA, Thompson SG. Analysis of cost data in randomised trials: an application of the non-parametric bootstrap. *Stat Med* (in press).
- 22 Desgagne A, Castilloux A, Angers J, LeLorier J. The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *Pharmacoeconomics* 1998;13:487-97.
- 23 Evans K, Tyrer P, Catalan J, Schmidt U, Davidson K, Dent J, et al. Manual-assisted cognitive-behaviour therapy (MACT): a randomised controlled trial of a brief intervention with bibliotherapy in the treatment of recurrent deliberate self-harm. *Psychol Med* 1999;29:19-25.
- 24 Lambert CM, Hurst NP, Forbes JF, Lochhead A, Macleod M, Nuki G. Is day care equivalent to inpatient care for active rheumatoid arthritis? Randomised controlled clinical and economic evaluation. *BMJ* 1998;316:965-9.
- 25 UK Small Aneurysm Trial Participants. Health service costs and quality of life for early elective surgery or ultrasonographic surveillance for small aortic aneurysms. *Lancet* 1998;352:1656-60.
- 26 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell Scientific, 1987.

(Accepted 13 December 1999)

Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions

John W Lynch, George Davey Smith, George A Kaplan, James S House

Department of Epidemiology, School of Public Health, University of Michigan, 109 Observatory Street, Ann Arbor, MI 48109-2029, USA

John W Lynch
assistant professor
George A Kaplan
chair

Survey Research Center, Institute for Social Research, University of Michigan

James S House
director

Department of Social Medicine, University of Bristol, Bristol BSS 2PR

George Davey Smith
chair

Correspondence to:
JW Lynch
jwlynch@sph.umich.edu

BMJ 2000;320:1200-4

Studies on the health effects of income inequality have generated great interest. The evidence on this association between countries is mixed,¹⁻⁴ but income inequality and health have been linked within the United States,⁵⁻¹¹ Britain,¹² and Brazil.¹³ Questions remain over how to interpret these findings and the mechanisms involved. We discuss three interpretations of the association between income inequality and health: the individual income interpretation, the psychosocial environment interpretation, and the neo-material interpretation.

Methods

We reviewed the literature through traditional and electronic means and supplemented this with correlational analyses of gross domestic product and life expectancy and of income inequality and mortality trends based on data from the World Bank,¹⁴ the World Health Organization,¹⁵ and two British sources.^{16, 17}

The individual income interpretation

According to the individual income interpretation, aggregate level associations between income inequality and health reflect only the individual level association between income and health. The curvilinear relation between income and health at the individual level^{18, 19} is

Summary points

Income inequality has generally been associated with differences in health

A psychosocial interpretation of health inequalities, in terms of perceptions of relative disadvantage and the psychological consequences of inequality, raises several conceptual and empirical problems

Income inequality is accompanied by many differences in conditions of life at the individual and population levels, which may adversely influence health

Interpretation of links between income inequality and health must begin with the structural causes of inequalities, and not just focus on perceptions of that inequality

Reducing health inequalities and improving public health in the 21st century requires strategic investment in neo-material conditions via more equitable distribution of public and private resources