How sensitive are cost-effectiveness analyses to choice of parametric distributions?

Simon G Thompson, Richard M Nixon

MRC Biostatistics Unit, Institute of Public Health,

Robinson Way, Cambridge CB2 2SR, UK

January 2004

Correspondence to: Professor S G Thompson

MRC Biostatistics Unit, Institute of Public Health,

Robinson Way, Cambridge CB2 2SR, UK

Tel: +44 1223 330366, Fax: +44 1223 330388

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

How sensitive are cost-effectiveness analyses to choice of parametric distributions?

Summary

Background: Cost-effectiveness analyses of clinical trial data are based on assumptions about the distributions of costs and effects. Cost data usually have very skew distributions, and can be difficult to model. We investigate whether choice of distribution can make a difference to the conclusions drawn.

Methods: We compare three distributions for costs data – normal, gamma and lognormal – using similar parametric models for the cost-effectiveness analyses. Inferences on the cost-effectiveness plane are derived, together with cost-effectiveness acceptability curves. These methods are applied to data from a trial of rapid magnetic resonance imaging (rMRI) investigation in patients with low back pain.

Results: The gamma and lognormal distributions fitted the cost data much better than the normal distribution. However, in terms of inferences about cost-effectiveness, it was the normal and gamma distributions that gave similar results. Using the lognormal distribution led to the conclusion that rMRI was cost-effective for a range of willingness-to-pay values where assuming a gamma or normal distribution did not.

Conclusions: Conclusions from cost-effectiveness analyses are sensitive to choice of distribution and, in particular, to how the upper tail of the cost distribution beyond the observed data is modelled. How well a distribution fits the data is an insufficient guide to model choice, and a sensitivity analysis is necessary. The conclusions from existing cost-effectiveness analyses, which have not explored this issue, should be regarded with caution.

Introduction

Undertaking cost-effectiveness analyses using individual patient data from randomised trials is now common practice. Simple comparisons of average costs, and separately of average effects, between randomised groups need to be supplemented by cost-effectiveness analyses that formally offset, for example, health gains against increased costs [1]. The inferences from such analyses can be presented on the cost-effectiveness plane (CE-plane) [2], to depict the joint uncertainty around the mean cost difference and the mean effect difference [3]. Using the societal or purchaser's 'willingness-to-pay', the economic value put on a unit of health gain, the incremental monetary net benefit (INB) is calculated taking into account the mean differences in both costs and effects between randomised groups [4]. The probability, given the data, that one intervention is cost-effective, for a range of willingness-to-pay values, can be shown as a cost-effectiveness acceptability curve (CEAC) [5]. These are now the accepted tools for cost-effectiveness analysis of clinical trial data.

Such analyses rely on statistical models, sometimes explicit and sometimes implicit, which describe the distribution of costs and effects and their interrelation across individuals in the trial [6,7]. The choice of models used in practice is sometimes rather arbitrarily made, for example on the basis of familiarity or ease of computation. Here we investigate whether the results of cost-effectiveness analyses depend on how the data are modelled. Our context is two-arm randomised trials, say of a new intervention against standard treatment, where continuously distributed effects and costs are assessed for each individual over a fixed period of time.

Often the assumption is made that the data on costs and effects for individuals in each arm of the trial follow normal distributions and that they are linearly correlated [6,8]. This is equivalent to assuming that costs and effects have a bivariate normal distribution in each arm of the trial. While such an assumption may be convenient for

computational purposes, it is rarely realistic. In particular, cost data usually have extremely skew distributions [9], not at all like a symmetric normal distribution. Other methods have therefore been proposed, including non-parametric bootstrapping [10] and parametric modelling [11]. Non-parametric bootstrapping is usually based on the distribution of average costs and effects across repeated samples, drawn from the original data with replacement [7]. While this procedure goes some way to accommodating the skewed nature of cost data, reliance on using simple averages usually gives similar results to assuming normal distributions [12]. This is especially the case for larger sample sizes, through the action of the central limit theorem [7]. We therefore confine our attention to parametric modelling in this paper.

We compare cost-effectiveness analyses for three commonly used distributions for cost data – normal, gamma and lognormal distributions. We propose statistical models for each of these cases, and compare the results as applied to a trial in patients with low back pain. In our example, the effects data are apparently adequately represented by a normal distribution.

Methods

Low back pain trial

The data we use to exemplify the methods were collected in a randomised trial of patients with low back pain [13]. In brief, 380 patients were recruited when they were referred by their physician to one of four imaging centres in Washington State, USA. 190 were randomised to investigation by rapid magnetic resonance imaging (rMRI), and 190 to standard X-ray investigation. The issue being addressed by the trial was whether rMRI would allow better diagnosis and treatment, or simply lead to unnecessary treatment without improvement in symptoms. Follow-up of patients continued for 12 months. The effectiveness data considered here is the change from

baseline of each patient's modified Roland back pain score [14]. Although this is not a utility measure, it is adequate for exemplifying the methods being proposed. The total health care costs (in US $) for each patient were calculated from resource use data (investigations, interventions, medical services) and fixed unit costs [13]. Resource use unrelated to back pain was excluded. Data on costs or effects were available for a total of 365 of the 380 patients randomised.

Normal distribution for costs

We first model both costs and effects using normal distributions. In each arm of the trial, the effect (E_i) and cost (C_i) for patient i are assumed to follow correlated normal distributions:

$$\begin{aligned} E_i &\sim Normal(\mu_{Ei}, \sigma_E^2) \\ C_i &\sim Normal(\mu_C, \sigma_C^2) \\ \mu_{Ei} &= \mu_E + \beta\,(C_i - \mu_C) \end{aligned} \qquad (1)$$

Here the costs have a normal distribution with mean μ_C and standard deviation σ_C. Effects have a standard deviation σ_E and a mean that depends, through the parameter β, on how much the cost C_i is above the mean cost μ_C. The subtraction of μ_C in the third line of (1) ensures that μ_E remains interpretable as the overall mean effect in the arm of the trial. Any bivariate normal distribution can be represented by the equations in (1), the correlation being provided for by the regression coefficient β. However the formulation in (1) is convenient for making extension to other distributions for costs, as described later.

The above model, for a two-arm trial, has 10 parameters that have to be estimated from the data. These are the mean and standard deviation of effects and costs, and the regression coefficient β, in each group. The model allows the correlation between

costs and effects to be different in the two arms of the trial, through the separate respective β parameters.

Gamma distribution for costs

To represent the usual skewness in cost data, we modify the above by using a gamma rather than a normal distribution for costs, while maintaining the normal distribution for effects. In each arm of the trial, we assume:

$$\begin{aligned} E_i &\sim Normal(\mu_{Ei}, \sigma_E^2) \\ C_i &\sim Gamma(\mu_C, \rho_C) \\ \mu_{Ei} &= \mu_E + \beta\,(C_i - \mu_C) \end{aligned} \qquad (2)$$

The gamma distribution for the costs above is parameterised by its mean μ_C and 'shape' ρ_C [15]; the standard deviation is then the mean divided by the square-root of the shape, that is $\mu_C/\sqrt{\rho_C}$. Otherwise the formulation of model (2) is the same as model (1), and it also has 10 parameters for a two-arm trial.

Lognormal distribution for costs

As an alternative to the gamma distribution for modelling skewed cost data, we also use a lognormal distribution. This is equivalent to assuming that the log costs are normally distributed. In each arm of the trial:

$$\begin{aligned} E_i &\sim Normal(\mu_{Ei}, \sigma_E^2) \\ C_i &\sim Lognormal(\nu_C, \tau_C^2) \\ \mu_{Ei} &= \mu_E + \beta\,(C_i - \mu_C) \\ \text{where} \quad \mu_C &= exp(\nu_C + \tau_C^2/2) \end{aligned} \qquad (3)$$

Here ν_C is the mean and τ_C the standard deviation of the log costs. Whereas $\exp(\nu_C)$ is the geometric mean cost, which is not relevant to cost-effectiveness analyses [10], *exp*

($\nu_C + \tau_C^2/2$) is the mean cost on the original scale [7,12]. This explains the definition of μ_C in the last line of (3) above. Otherwise the model is similar to (1) and (2), and again has 10 parameters for a two-arm trial.

Implementation

Fitting these models by conventional maximum likelihood is straightforward for the (bivariate) normal case [16], but is more difficult for the gamma and lognormal distributions. Hence we chose to use Bayesian methods, implemented using Markov chain Monte Carlo (MCMC) methods in the software BUGS [17]. This also has the advantage of giving the natural interpretation of the CEAC [6] as providing the posterior probability, given the data, that the intervention is cost-effective. Bayesian methods require prior distributions for the parameters of the distributions. Here we use 'vague' priors, intended to be approximately non-informative, so that the resulting inferences essentially depend only on the data. Specifically, we use wide uniform priors for all the μ's, for the β's, for the $\log(\sigma)$'s in the normal distributions, for ρ_C in the gamma distribution, and for $\log(\tau_C)$ in the lognormal distribution [18].

Posterior distributions of quantities of interest for inferences about cost-effectiveness were derived from 20,000 MCMC iterations, after an initial 5000 iterations were discarded to ensure convergence [17]. These posterior distributions are summarised as means and 95% credible intervals. The quantities include the average cost difference (Δ_C), and the average effect difference (Δ_E), between the two arms of the trial. The distribution of (Δ_C , Δ_E) on the CE-plane was represented by a conventional contour plot, achieved by smoothing their joint posterior distribution [19]. The incremental net benefit, $INB(K) = K\Delta_E - \Delta_C$, is a function of the willingness-to-pay value K. The CEAC is a plot of the probability, given the data, that the intervention is cost-effective against K, that is of the probability that INB(K) is positive.

Results

Normal distributions fitted the effects data well in both arms of the trial (Figure 1). The cost data were considerably skewed (Figure 2); most costs were below $1000 but a few were above $10,000. A gamma or lognormal distribution clearly fitted the data much better than a normal distribution. However it is hard to judge by visual inspection which of the gamma or lognormal is better.

The estimated mean effects and costs, all fitted independently of each other, are shown in Table 1. A lower deviance (minus twice log likelihood) indicates a better fit to the data, so the results confirm that the gamma and lognormal distributions fitted the cost data much better than the normal distribution. In the X-ray group the fit of the gamma and lognormal distributions are almost the same, whereas in the rMRI group the lognormal distribution is a better fit. Looking back at Figure 2, one can see this is because the lognormal distribution, with its heavier right-hand tail, more easily encompasses the extreme costs observed in this group.

The inference about the mean costs is very substantially affected by choice of distribution. For example, if a lognormal rather than a gamma distribution is chosen for the costs in the X-ray group, the mean costs are estimated as about $450 greater (Table 1); this is despite the fact that both distributions fitted the data about equally well. In contrast, the choice of distribution makes little difference to the estimated mean cost in the rMRI group, despite the differences in how well the distributions fitted the data.

The combined models for costs and effects are shown as contour plots on the CE-plane in Figure 3. Contours enclosing 5%, 50% and 95% of the joint distribution of the mean cost and mean effect differences are shown, for each choice of assumed

distribution for the original cost data. The vertical scale gives the cost difference, showing that there is evidence of a higher mean cost in the rMRI group when assuming either a normal or gamma distribution for costs, but not when assuming a log normal distribution. The estimated effect difference (horizontal scale) is similar whatever distribution is chosen for costs, and is small. When using a lognormal distribution for costs, which causes an increase in the estimated mean cost in the X-ray group (Table 1), the contours on the CE-plane then become centred much more round the origin (Figure 3). Because the costs and effects were not substantially correlated in the trial data ($r = -0.14$ in the X-ray group and -0.08 in the rMRI group), the major axes of the ellipse-shaped contours are approximately horizontal and vertical. The deviances for these combined models of costs and effects (Figure 3) indicate that the lognormal distributions for costs fitted somewhat better than the gamma distributions, and that both of these fitted substantially better than the normal distributions.

These cost-effectiveness results on the CE-plane are re-expressed as CEACs in Figure 4. Again there is little apparent difference between results from the normal and gamma distribution assumptions for costs, despite the obvious difference in fits. Both indicate that the probability that rMRI is cost-effective is low (below 30%) at a willingness-to-pay of less that $1000, and never rises above 50% for any value of willingness-to-pay. Thus the conclusion, from these analyses, is that rMRI should not be introduced. However, using a lognormal distribution, the probability that rMRI is cost-effective is above 50% at a willingness-to-pay of $100 or greater. Formal decision-theoretic rules [20] would then suggest that rMRI should be introduced as a cost-effective strategy, provided the willingness-to-pay is at least $100.

To explore the sensitivity of inferences about cost-effectiveness to the assumed shape of the upper tail of the X-ray group cost distributions, we investigated the use of a truncated lognormal distribution. The lognormal distribution was truncated at various

high cutoff points, well beyond the range of the observed data, thereby assuming that costs higher than these cutoffs could not occur. Such truncation hardly affects at all the fit of the distribution to the observed data. Recall that, compared to using a normal distribution, using a full lognormal distribution estimated the mean cost in the X-ray group as about $450 greater (Table 1). When the lognormal distribution was truncated at twice the maximum observed cost, the estimated mean cost was only about $200 greater. Even truncating at five times the maximum observed cost, the mean was only about $300 greater, still less than the $450 greater figure obtained when using the full lognormal distribution. In contrast, truncating the normal or gamma distribution at twice, or five times, the maximum observed cost had almost no effect on the estimated mean cost compared to the use of full normal or gamma distributions. These results thus show substantial sensitivity of inferences to what is assumed about shape of the upper tail of the cost distribution, way beyond the observed data.

Discussion

We have shown that the conclusions from cost-effectiveness analyses are sensitive to the assumptions made about the distributions of cost. It might be argued that one should simply choose distributions that fit the data well, and then rely on conclusions derived from the consequent analyses. However, this is inadequate for two reasons. First, distributions that fit the cost data very differently can give similar inferences, as exemplified by the normal and gamma distributions for our back pain trial example. Second, distributions that fit the data similarly can give rise to very different inferences, as shown for the gamma and lognormal distributions in our example. The preferred choice of distribution was not very clear for the low back pain trial; although the lognormal distribution fitted the cost data somewhat better overall, this was mainly because of the presence of a few high cost outliers in the rMRI group.

Because inferences derived from cost-effectiveness analyses are conditional on the model being used, it is necessary to undertake a sensitivity analysis of how choice of distributions, amongst those that apparently fit the data well, affects the results. This however is easier said than done. Over what set of distributions should the sensitivity analysis be done? We have chosen three commonly used distributions, but there would be many others that could be used; there is no underlying theory that determines what might be a suitable choice of distribution for cost data. The advocacy of one particular distribution, for example the lognormal, as the distribution of choice for cost data, either in particular datasets [12] or in general [21], is not tenable. Thus existing cost-effectiveness analyses, which have not explored the issue of sensitivity to choice of distribution for modelling costs, should be regarded with caution.

Inferences about costs will often be sensitive to the shape and extent of the right-hand tail of the fitted distribution beyond the range of the data, amongst distributions that fit the observed data equally well. For example, the lognormal distribution generally gives higher estimated mean costs than the gamma or the normal, while the log-logistic (which has an even heavier right-hand tail than the lognormal) can give even more extreme results [18]. Our results obtained using a truncated lognormal distribution emphasise this sensitivity of inferences to what is assumed about the upper tail of cost distributions. Truncating a cost distribution at, say, twice the maximum observed cost might in fact be more realistic than allowing an arbitrarily long tail, since costs for individual patients must have some finite limit in practice. However, only with an unrealistically large amount of cost data could one ever be confident about choosing an appropriate model for the upper tail of their distribution.

Using non-parametric bootstrapping does not circumvent the issue of model choice. First a decision has to be made as to which statistic is to be monitored in the repeated samples. If this is the sample mean cost, the results will be similar to those based on a

parametric assumption of normality [7]. If the mean of a lognormal distribution is chosen, using the last line of model (3), then the results will be similar to those based on a parametric assumption of lognormality. Although non-parametric bootstrapping does not assume any particular form of distribution, the choice of statistic used implicitly does. Moreover the validity of the non-parametric bootstrapping depends on the observed distribution of the data being an adequate representation of the true underlying distribution which, in the case of extremely skew cost data, will only be true for large sample sizes [7].

The bivariate normal model adopted here is the same as that introduced earlier by others [6]. We have extended this basic bivariate normal model to allow for other, more realistic, distributions of costs. The model could be similarly extended for other distributions of effects. We have made effects a function of costs, as implied by the regressions in the third line of each of models (1), (2) and (3). For model (1), whether effects are regressed on costs or vice-versa makes no difference; both represent the same bivariate normal distribution model. For the gamma model and lognormal model, however, slightly different results would ensue. In our case, where the correlations between costs and effects were low, it made no material difference. It is also possible to model effects as linearly related to log costs, rather than untransformed costs; again this made little difference in our example.

One advantage of the parametric modelling framework adopted here is the ease with which it extends to the inclusion of covariates. Including baseline covariates in the model for either costs or effects in a randomised trial not only adjusts for chance imbalance in the randomisation, but also should improve precision [22]. In the context of an observational study, adjustment for case-mix variables is essential in attempting to remove sources of confounding [22]. Inclusion of covariates can also be used to estimate cost-effectiveness within subgroups of patients in trials. A more complex extension would be to the case of censored effects and costs data [16]. These

are important areas that require further research. Meanwhile, the conclusion remains that, since inferences are not robust to choice of parametric distribution, sensitivity analyses with alternative distributions need to be undertaken before drawing conclusions about cost-effectiveness.

Acknowledgments

We thank Dr Jerry Jarvik and Dr Will Hollingworth for permission to use the low back pain trial data in our methodological study. The full costs and effects results from the trial are published elsewhere [13]; our retrospective analysis of the data is for expository purposes only.

References

1. National Institute for Clinical Excellence. Guide to the methods of technology appraisal. http://www.nice.org.uk/ (accessed August 2003).
2. Black WC. The CE plane: a graphic representation of cost-effectiveness. Medical Decision Making 1990; 10: 212-214.
3. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of health care interventions. Health Technology Assessment 1999, 3: 1-134.
4. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical Decision Making 1998; 18 suppl: S68-S80.
5. Van Hout BA, Al MJ, Gordon GS, Rutten F. Costs, effects and C/E ratios alongside a clinical trial. Health Economics 1994; 3: 309-319.
6. O'Hagan A, Stevens JW, Montmartin J. Bayesian cost effectiveness analysis from clinical trial data. Statistics in Medicine 2001; 20: 733-753.

7. Barber JA, Thompson SG. Analysis of cost data in randomised trials: an application of the non-parametric bootstrap. Statistics in Medicine 2000; 19: 3219-3236.
8. Laska EM, Meisner M, Siegel C. Statistical inference for cost-effectiveness ratios. Health Economics 1997; 6: 229-242.
9. Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. Journal of Health Services Research and Policy 1998; 3: 233-245.
10. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? British Medical Journal 2000; 320: 1197-1200.
11. O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trial data. Health Economics 2001; 10: 303-315.
12. O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? Health Economics 2003; 12: 33-49.
13. Jarvik JG, Hollingworth W, Martin B *et al*. Rapid magnetic resonance imaging vs. radiographs for patient with low back pain. Journal of the American Medical Association 2003; 289: 2810-2818.
14. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine 1995; 20: 1899-1908.
15. Evans M, Hastings N, Peacock B. Statistical distributions. New York: Wiley, 1993.
16. Willan AR, Lin DY. Incremental net benefit in randomised clinical trials. Statistics in Medicine 2001; 20: 1563-1574.
17. Spiegelhalter DJ, Thomas A, Best N. WinBUGS version 1.2. MRC Biostatistics Unit: Cambridge, UK, 1999.
18. Nixon R, Thompson SG. Parametric modelling of cost data in medical studies. Statistics in Medicine 2003; in press.

19. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, Series B 1991; 53: 683-690.

20. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics 1999; 18: 341-364.

21. Zhou XH. Inferences about population means of health care costs. Statistical Methods in Medical Research 2002; 11: 327-339.

22. Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1991.

Table 1: Effects (change in backpain score) and costs ($) from the low back pain trial, fitted separately to various distributions.

Group		Distribution	Mean	95% credible interval	Deviance
X-ray	Effects	Normal	3.93	(2.97,4.89)	1239
(n=185)	Costs	Normal	1652	(1332,1963)	3461
		Gamma	1675	(1413,1975)	3185
		Log-normal	2106	(1576,2846)	3181
rMRI	Effects	Normal	3.89	(2.87,4.90)	1258
(n=180)	Costs	Normal	2121	(1644,2604)	3616
		Gamma	2152	(1817,2549)	3283
		Log-normal	2100	(1660,2677)	3230

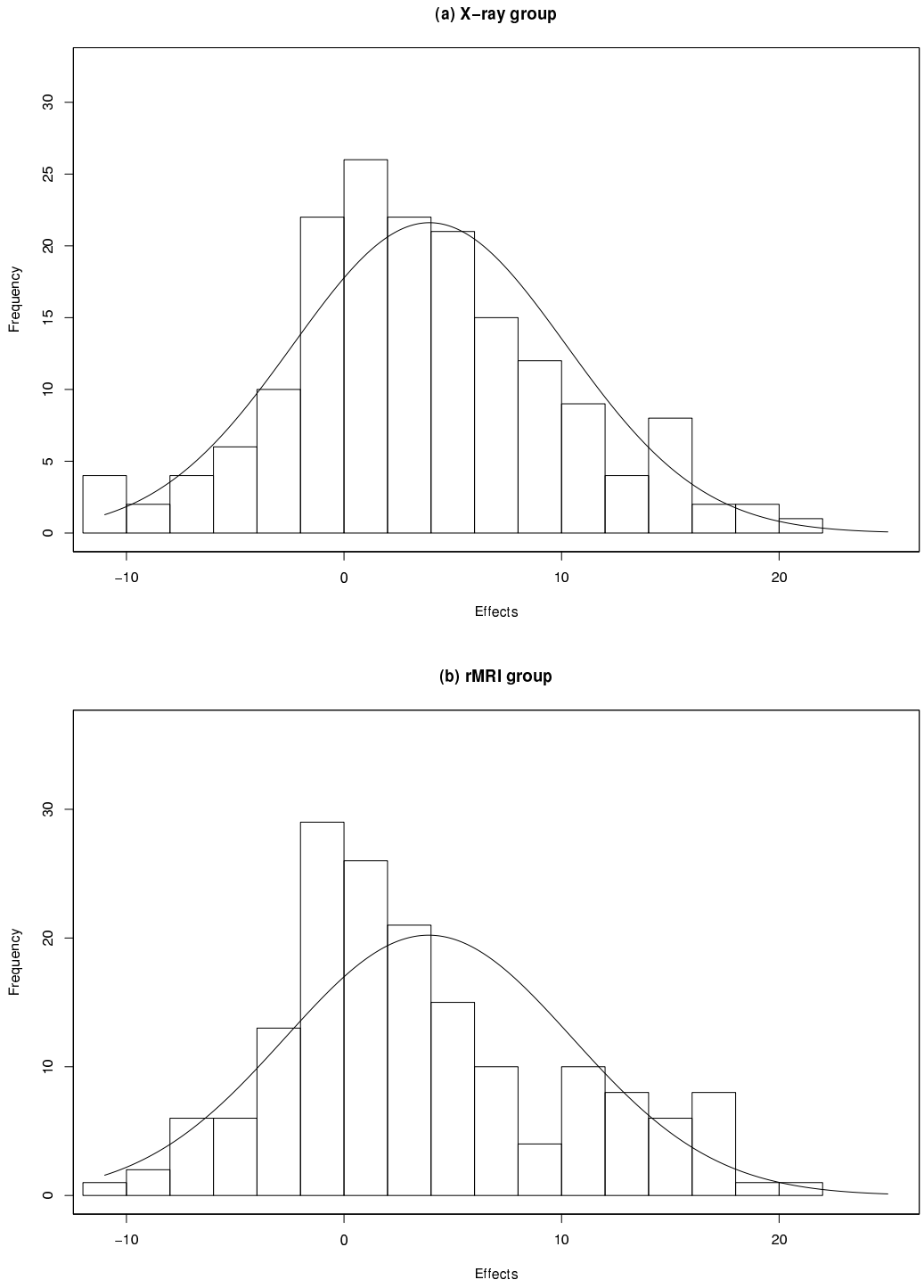

Figure 1: Effects data from the low back pain trial, along with fitted densities from a normal distribution.

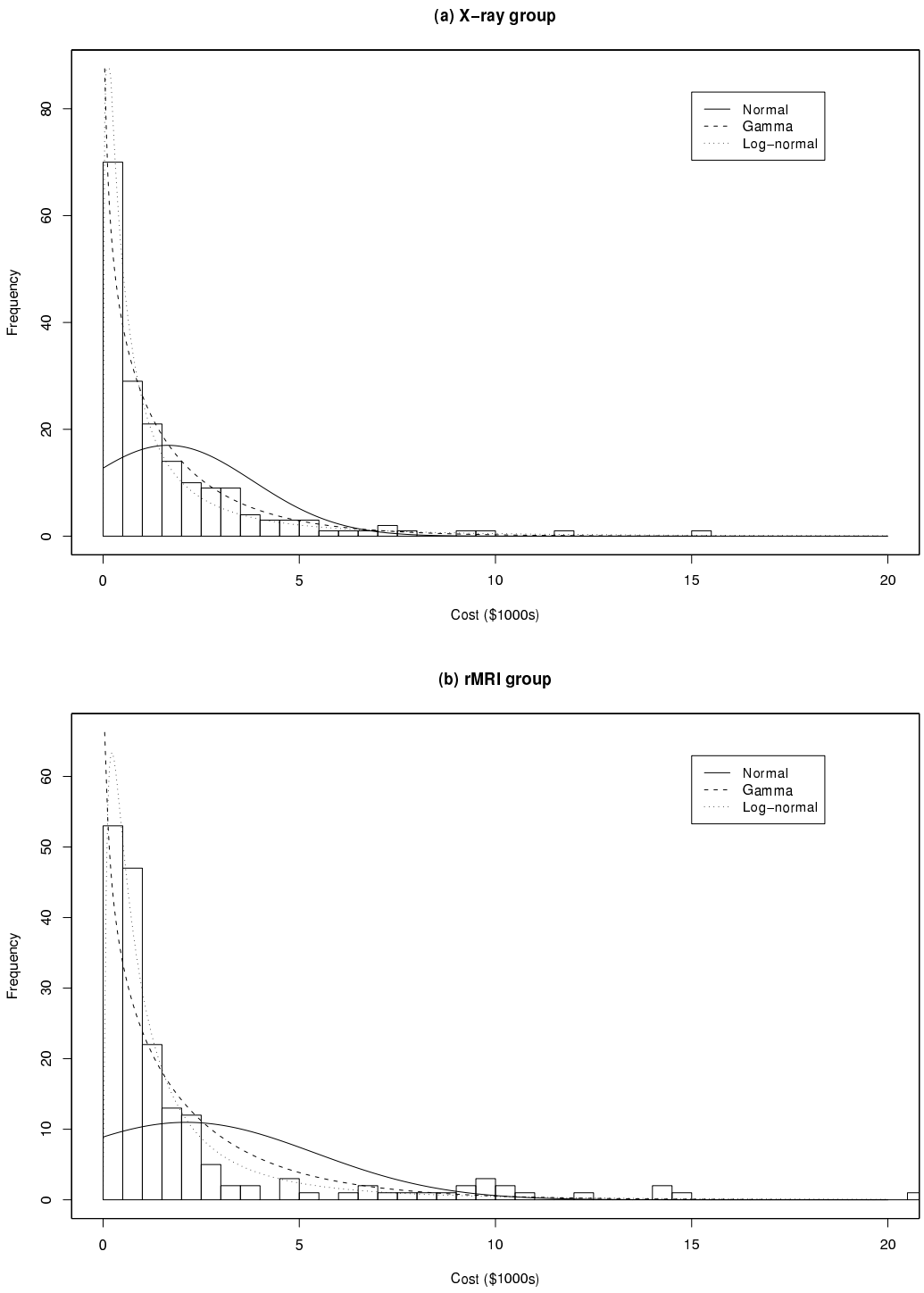

Figure 2: Cost data from the low back pain trial, along with fitted densities from normal, gamma and lognormal distributions.

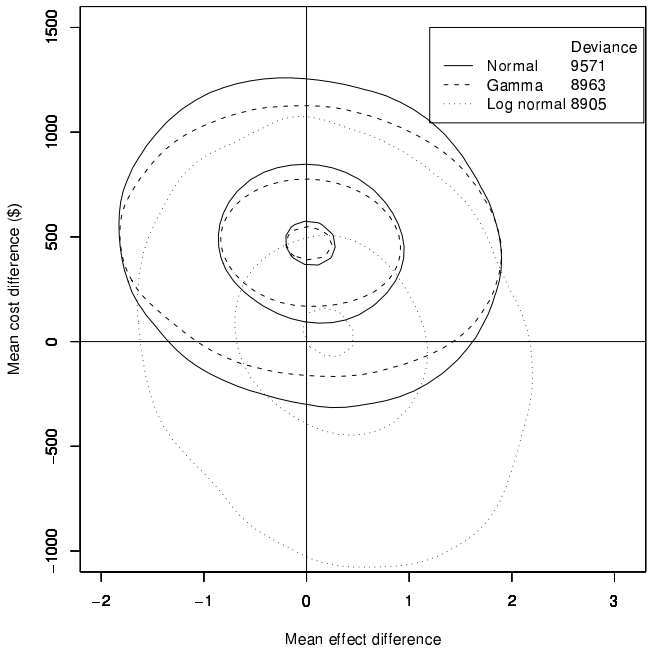

Figure 3: Contour plots of cost-effectiveness density: rMRI minus X-ray. Effects are fitted by a normal distribution; costs are fitted by normal, gamma and lognormal distributions. Contours are drawn to include 5%, 50% and 95% of the joint distribution in each case.

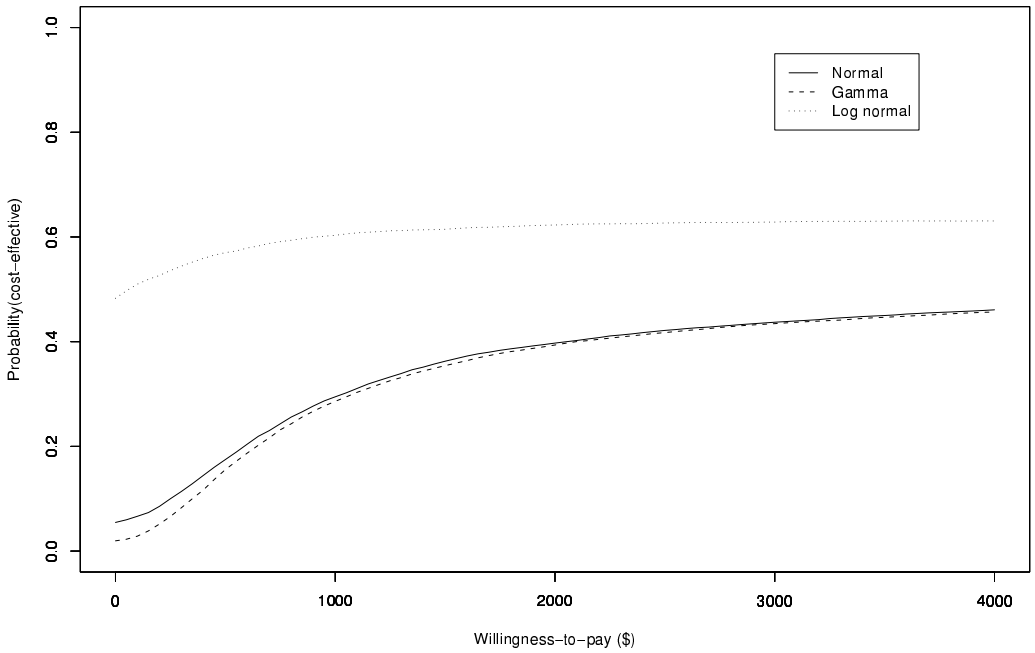

Figure 4: Cost-effectiveness acceptability curves of rMRI compared to X-ray. Effects are fitted by a normal distribution; costs are fitted by normal, gamma and lognormal distributions.