International Journal of Technology Assessment in Health Care, **18:1** (2002), 112–119.

UNCERTAINTY IN COST-EFFECTIVENESS ANALYSIS

Probabilistic Uncertainty Analysis and Stochastic League Tables

Rob M. P. M. Baltussen
Raymond C. W. Hutubessy
David B. Evans
Christopher J. M. Murray
World Health Organization

Abstract

Interest is growing in the application of standard statistical inferential techniques to the calculation of cost-effectiveness ratios (CER), but individual level data will not be available in many cases because it is very difficult to undertake prospective controlled trials of many public health interventions. We propose the application of probabilistic uncertainty analysis using Monte Carlo simulations, in combination with nonparametric bootstrapping techniques where appropriate. This paper also discusses how decision makers should interpret the CER of interventions where uncertainty intervals overlap. We show how the incorporation of uncertainty around costs and effects of interventions into a stochastic league table provides additional information to decision makers for priority setting. Stochastic league tables inform decision makers about the probability that a specific intervention would be included in the optimal mix of interventions for different resource levels, given the uncertainty surrounding the interventions.

Keywords: Cost-effectiveness, Uncertainty analysis, Priority setting

As more prospective cost-effectiveness analysis (CEA) studies are undertaken, providing stochastic data on costs and effects, interest has grown in the application of statistical techniques to the calculation of cost-effectiveness ratios (CER). Several methods have been developed, including confidence planes (20), mathematical techniques (22), and the net health benefit approach (4).

However, it is important to recognize that many public health interventions do not lend themselves to the collection of sampled individual level data (by patient, health facility, region, etc.), especially in a developing-country context. For example, it is difficult to develop a feasible experimental design to identify the costs and effects of a national radio health education program or a policy to subsidize the use of essential pharmaceutical products. Many economic evaluations require nonstochastic parameter estimates and modeling assumptions.

The views expressed are those of the authors and not necessarily those of the organization they represent.

Typically, uncertainty stemming from the use of such nonsampled secondary data sources in CEA has been dealt with by sensitivity analysis (2;3).These deterministic analyses draw inferences from point estimates of variables, but interpretation is conditional upon a range of uncertainty that is assumed for critical variables. There are three major limitations to this approach: a) the analyst has discretion as to which variables and what alternative values are included; b) interpretation is essentially arbitrary because there are no comprehensive guidelines or standards as to what degree of variation in results is acceptable evidence that the analysis is robust; and c) variation of uncertain parameters one at a time carries a risk that interactions between parameters may not be captured (19).

This paper examines the application of probabilistic uncertainty analysis with Monte Carlo simulations in this context.[1] This builds on work already described in the literature (3;9;10;15) and requires that analysts assume some distributional form for costs and effects from which repeated samples are drawn to determine a distribution for the CER. The definition of an uncertainty range for CER is hampered by the instability of sample estimates of CERs, causing its mean to vary (6). This paper applies the simple percentile method—usually employed to estimate uncertainty ranges for CER in nonparametric bootstrapping—to estimate uncertainty intervals for CERs involving probabilistic uncertainty analysis. The approach is illustrated by constructing uncertainty intervals for seven hypothetical interventions in tuberculosis control.

In addition, this paper discusses how information on uncertainty should be communicated to policy makers. The abovementioned techniques present study results in terms of some type of uncertainty interval. However, little or no attention is paid to the question of how decision makers should interpret the results where uncertainty intervals overlap. The recently developed stochastic league tables inform decision makers about the probability that a specific intervention would be included in the optimal mix of interventions for various levels or resource availability, taking into account the uncertainty surrounding costs and effectiveness (14). This paper derives a stochastic league table for hypothetical interventions in tuberculosis control. We show that the incorporation of uncertainty ranges for CERs in a stochastic league table provides additional information to decision makers for priority setting.

This methodologic work on estimating uncertainty is part of the larger World Health Organization (WHO) concept of Generalized-CEA (18). The WHO proposes to provide policy makers with a simple set of results that are more generalizable across settings by evaluating the costs and effectiveness of new and existing interventions, compared to the starting point of doing none of the current interventions, called the "null." This removes the constraint that the current intervention mix must be continued and eliminates differences in starting points, which traditionally makes the results of incremental analyses difficult to transfer across settings.

PROBABILISTIC UNCERTAINTY ANALYSIS BY MONTE CARLO SIMULATIONS

Probabilistic uncertainty analysis using Monte Carlo simulations has been well described elsewhere (3;9;10;15). Most applications assume a distributional form (e.g., normal, uniform, binominal) for each estimated (but nonsampled) variable. Repeated samples are then drawn from these distributions to determine an empirical distribution for some construct of the variables, such as CERs.

To illustrate the procedure, consider a hypothetical example first presented in Murray et al. (18) related to four interventions for tuberculosis: a) passive case detection and treatment with directly observed short-course therapy (DOTS) (A); b) bacille Calmette-Guérin (BCG) vaccination at 50% coverage (B1); c) BCG vaccination at 75% coverage (B2); and d) BCG vaccination at 100% coverage (B3). In addition, three other mutually exclusive

Table 1. Costs and Health Effects of Interventions

Intervention	Costs	Effects
B1	180	200
B2	325	300
B3	600	400
A	550	500
AB1	631	600
AB2	726	650
AB3	952	700

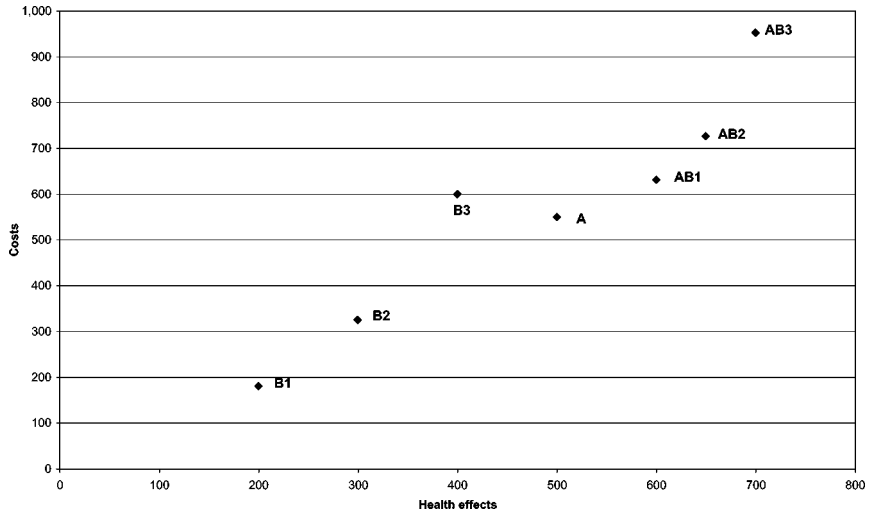

Figure 1. Costs and health effects of interventions.

options are presented: DOTS combined with the different levels of BCG coverage, i.e., AB1, AB2, and AB3, respectively. Costs and health effects interact: the variable costs component of DOTS decreases when the vaccination is given, and fewer cases of tuberculosis will occur. The health effects of BCG vaccination will be fewer in the presence of a treatment program, because many of the deaths from tuberculosis expected in the absence of treatment will be avoided. Total costs and health effects of the interventions at the population level are presented in Table 1 and Figure 1. To reflect uncertainty, costs are assumed to be normally distributed with a standard deviation of 100; health effects are assumed to be normally distributed with a standard deviation of 100. The covariance is assumed to be zero.

The procedure to generate a sample distribution for the incremental CER from expanding the intervention, for example, BCG coverage from 50% to 75% (B1 → B2), is as follows:

1. Take one sample of costs (C) and health effects (E) from the distribution of costs and effects from B1: C_{B1} and E_{B1}, and one sample of cost and effects from the distribution of costs and effects of B2: C_{B2} and E_{B2};
2. The sample estimate of the incremental CER is then given by C_{B2}-C_{B1} divided by E_{B2}-E_{B1}; and
3. Repeating this process a large number of times gives a vector of sample estimates that is the empirical sampling distribution of the incremental CER statistic.

We used the statistical program @RISK 4.0^{TM} (Palisade Decision Tools) to run the analyses.

https://doi.org/10.1017/S0266462302108105 Published online by Cambridge University Press

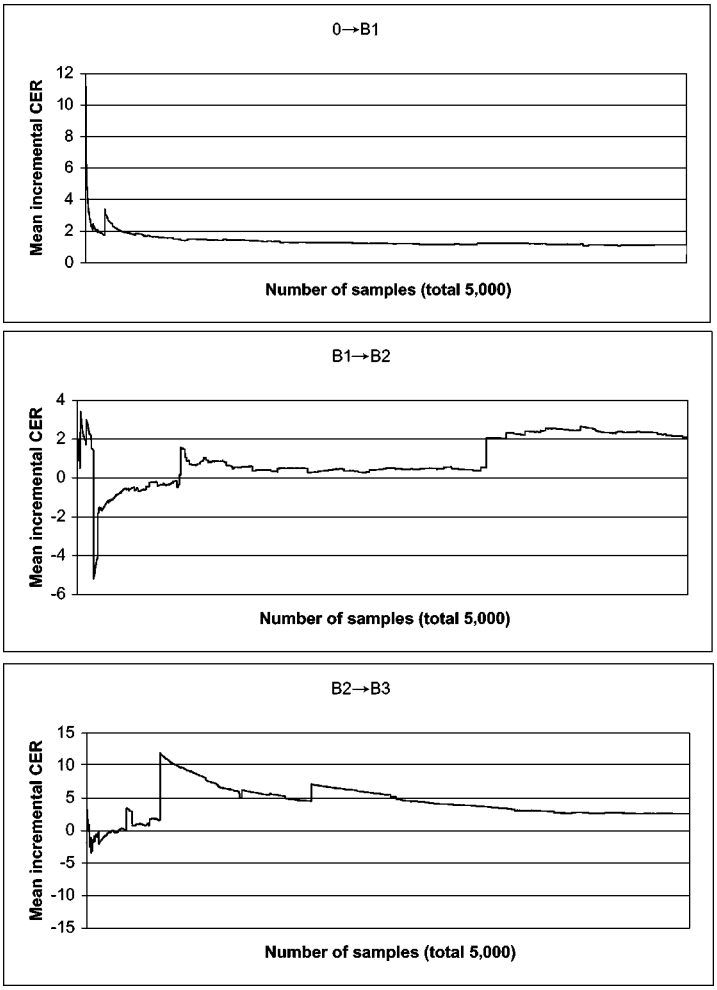

Figure 2. Mean incremental CERs of three interventions as a function of the number of samples.

There is little stability in these CER estimates where the distributions of costs or health effects overlap—some simulations will produce negative net health effects and some positive net health effects, for example. This can lead to positive or negative incremental CERs. Figure 2 shows the mean value of the sampled estimates of the incremental CERs for $0 \rightarrow$ B1, B1 $\rightarrow$ B2, and B2 $\rightarrow$ B3. For the assumed ranges of costs and health effects, even after a large number of samples, some means do not stabilize. The mean CER of $0 \rightarrow$ B1 is relatively stable because the origin is fixed, and costs and health effects of intervention B1 constitute its only source of uncertainty (i.e., it is an average ratio).

The simple percentile method allows us to estimate confidence intervals in the presence of these unstable means. This approach takes the $100(\alpha/2)$ and the $100(1-(\alpha/2))$ percentile values of the bootstrap distribution as the upper and lower confidence limits for the CER (6). Table 2 shows the 90% confidence intervals for the seven mutually exclusive alternatives. The occasional very high values for the incremental CER of the expansion from AB1 $\rightarrow$ AB2 is the reason why its confidence interval does not include its mean value.

Of special interest are interventions that are weakly dominated on the basis of the point estimate of their CER but have a wide confidence interval. In such cases, some

https://doi.org/10.1017/S0266462302108105 Published online by Cambridge University Press

Table 2. Sample Incremental CERs for Seven Interventions

Intervention	Mean	Minimum	Maximum	90% confidence interval	
0 → B1	1	−379	340	0	4
B1 → B2	2	−912	5,131	−8	8
B2 → B3	3	−961	6,358	−11	13
B3 → A	−1	−1,046	635	−5	5
A → AB1	1	−4,469	1,727	−5	6
AB1 → AB2	21	−20,545	109,129	−7	7
AB2 → AB3	−4	−30,827	6,662	−11	12

simulations might show them to be no longer dominated. In Figure 1, consider intervention B3. Because its mean is located northwest of intervention A, it appears to be strongly dominated. However, the uncertainty range of the incremental CER of B3 → A ranges from −$5 to $5 per unit of health effect and thus includes positive values. Therefore, we cannot be sure that B3 should be excluded from the set of alternatives under consideration.

COMBINING PROBABILISTIC UNCERTAINTY ANALYSIS WITH NONPARAMETRIC BOOTSTRAP PROCEDURES

In the situation in which individual level data are available for some component of costs or effects, one feasible approach is to combine probabilistic uncertainty analysis with nonparametric bootstrapping to estimate a total "uncertainty range"[2] for CERs (15). The use of nonparametric bootstrapping has been advocated by many authors (5;6;8;16) and has been extensively applied to empirical data (1;5;6;7;8;9;13;17;21;22;23;24) Unlike probabilistic uncertainty analyses, the bootstrap approach is a nonparametric method that makes no distributional assumptions concerning the statistic in question. Instead it employs the original data in a resampling exercise in order to give an empirical estimate of the sampling distribution of that estimate.

The basic concept behind nonparametric bootstrapping is to treat the study sample as if it were the population, the premise being that it is better to draw inferences from the sample in hand rather than make potentially unrealistic assumptions about the underlying population. Using the nonparametric bootstrap approach, successive random draws are taken *with replacement* from the study sample data. As such, the fact that an observation has been selected does not preclude it from being selected again for the same resample, which leads to the construction of different bootstrap resamples. The statistic of interest and its distribution is calculated from these resamples. The number of bootstrap resamples, B, should at least be 1,000 to construct confidence intervals, in order to ensure that the tails of the empirical distributions are filled (6). An important advantage of the nonparametric bootstrap approach is that it is of no consequence whether the original sample is a well-behaved distribution because it forms its own probability density function.

To illustrate the combination of probabilistic uncertainty analysis and nonparametric bootstrapping, consider a CEA with costs being the product of vectors of unit prices and resource utilization. By defining a probability distribution of unit prices, and with resource utilization and effectiveness data stemming from sampled data, a total uncertainty range can be estimated by combining probabilistic uncertainty analysis with nonparametric bootstrapping. To start with, a large number (B) of samples of size n_p of sets of unit prices are obtained by random sampling from the prior distributions, and the mean price is calculated for each of the B samples. Similarly, B bootstrap samples of size n_q are taken from the resource utilization and effectiveness data, and the mean resource utilization and effectiveness is estimated for each of the B bootstrap samples. Then B replicates of the CER can be obtained by

combining B bootstraps of both resource utilization and effectiveness data with the B sets of prices sampled from the prior distributions. These are then used to calculate a percentile interval.

UNCERTAINTY AT THE ALLOCATION LEVEL

Traditionally, the above results, as reported in Table 2, are placed in a single league table to inform decision makers about the relative value of a set of (mutual exclusive) interventions. Rank ordering in the league table approach is usually made on the basis of point estimates of CE alone (11). However, in our example of tuberculosis control, uncertainty intervals overlap and it is not clear how decision makers should interpret such information. This problem was also faced in many practical situations, including the study by Goodman and Mills (12), which incorporated the estimated uncertainty interval for their estimates of the cost-effectiveness of interventions against malaria; however, because the intervals overlapped, the authors were unwilling to suggest which ones should be given preference in the event of a shortage of resources (12). We believe that additional information is obtained in the data used to produce the uncertainty intervals, which could be used to guide policy makers more than by simply saying that no decision could be made because confidence intervals overlap.

We propose the use of stochastic league tables. The approach provides the probability of inclusion of a specific intervention in the optimal mix of interventions given the uncertainty surrounding the intervention. The construction of stochastic league tables requires four steps and is described in more detail elsewhere (14). In a first step, CERs are calculated for the respective programs by drawing single samples from distributions of both costs and health effects, using Monte Carlo simulations. Second, based on these samples, the optimal mix of interventions is defined, applying resource allocation decision rules as described in Murray et al. (18). Third, this exercise is repeated 10,000 times to obtain a distribution of rank orders of interventions, given a certain resource level. This provides information on the probability of the cost-effectiveness of interventions. Fourth, this procedure is repeated for various resource levels. This provides a so-called "budget expansion path," which shows that different interventions will be chosen at different resource levels.

Table 3 summarizes the results as probabilities (in percentages) that a particular intervention in tuberculosis would be included in the optimal set at different resource availability levels. Probabilities of inclusion of an intervention depend on the resource availability, its costs, and its relative cost-effectiveness. In our example, at a resource level of 100, intervention B1 is chosen in 19% of all cases (i.e., costs of B1 are less than 100 in 19% of the cases, and other interventions are too costly to be chosen). If resource availability increases, the probability of inclusion of other interventions also increases. Note that intervention B3,

Table 3. Stochastic League Table, with Probabilities of Inclusion of Interventions (%) for Different Resource Availability

Intervention	Resource level											
	100	200	300	400	500	600	700	800	900	1,000	1,100	1,200
B1	19	64	57	36	15	3	0	0	0	0	0	0
B2	0	6	36	58	42	10	1	0	0	0	0	0
B3	0	0	0	1	8	12	5	2	1	1	0	0
A	0	0	0	5	27	38	22	11	6	4	3	3
AB1	0	0	0	0	7	31	43	37	27	20	16	16
AB2	0	0	0	0	0	7	29	47	48	37	32	29
AB3	0	0	0	0	0	0	0	3	17	38	48	53

https://doi.org/10.1017/S0266462302108105 Published online by Cambridge University Press

which would not be considered in a deterministic approach because of strong dominance, is now chosen in a low number of cases at certain resource levels. At the highest resource level (1,200), AB3 is chosen in 53% of all cases. Decision makers can use this information to prioritize interventions should more resources become available for health care.

Stochastic league tables present decision makers the probability that an intervention will be included in the optimal choice and are therefore more informative than traditional league tables, which simply present uncertainty ranges (and may leave decision makers indecisive when they overlap). They also allow decision makers to better evaluate the impact of trading off the efficiency goal against other objectives such as reducing health inequalities in their selection of interventions. For example, the stochastic league table informs decision makers that they are not likely to lose much in terms of efficiency if, at a resource level of 1,000, they decide to select AB2 rather than AB3 for equity reasons.

DISCUSSION

This paper presents an extension and generalization of previously described methods examining uncertainty in cost-effectiveness studies. Whereas previous studies applied the concept of bootstrapping and Monte Carlo simulations in the contexts of clinical trials, here we apply them to decision models, analyzing cost-effectiveness based on any combination of primary and secondary data. Given the prevailing scarcity of sampled data on costs and effects of many public health interventions in developing countries, we propose the use of probabilistic uncertainty analysis using Monte Carlo simulations, in combination with nonparametric bootstrapping techniques where appropriate.

The reporting of some type of uncertainty range of CER in individual studies ignores the question of how policy makers should interpret the results where uncertainty results overlap. The paper has shown that cost-effectiveness uncertainty ranges of interventions in tuberculosis control overlap and that decision making is difficult in such a situation. The stochastic league table is a new way of presenting uncertainty around costs and effects to decision makers. This paper shows that it provides additional information beyond that offered by the traditional treatment of uncertainty in CEA, presenting the probability that each intervention is included in the optimal mix for given levels of resource availability.

NOTES

[1] There is confusion in the literature as to the definition of sensitivity analysis, on the one hand, and uncertainty analysis on the other. We argue that sensitivity analysis refers to uncertainty about social choices, such as the discount rate or the inclusion of productivity costs. Uncertainty analysis refers to variation in the distribution of costs and effects (stemming from either nonsampled or sampled data). Following that definition, we prefer to use the term *probabilistic uncertainty analysis* rather than *probabilistic sensitivity analysis* to describe the process of drawing repeated samples from nonsampled data, i.e. from some *a priori* defined distributional form of costs and/or effects. We use the term *nonparametric bootstrapping* only in relation to drawing samples from sampled data.

[2] Instead of calling this a confidence interval, the term *uncertainty range* could be used since such an interval incorporates both uncertainty related to sampled and nonsampled data.

REFERENCES

1. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ*. 1998;7:723-740.
2. Briggs A, Sculpher M. Sensitivity analysis in economic evaluation: A review of published studies. *Health Econ*. 1995;4:355-371.
3. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Econ*. 1994;3:95-104.
4. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ*. 1999;8:257-261.

5. Briggs AH, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. *Health Technol Assess*. 1999;3:1-134.
6. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health Econ*. 1997;6:327-340.
7. Campbell MK, Torgerson DJ. Bootstrapping: Estimating confidence intervals for cost-effectiveness ratios. *Q J Med*. 1999;92:177-182.
8. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: An example from a randomized trial. *Stat Med*. 1996;15:1447-1458.
9. Critchfield GC, Willard KE, Connelly DP. Probabilistic sensitivity analysis methods for general decision models. *Comput Biomed Res*. 1986;19:254-265.
10. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation: A practical approach. *Med Decis Making*. 1985;5:157-177.
11. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: More harm than good? *Soc Sci Med*. 1993;37:33-40.
12. Goodman CA, Mills AJ. The evidence base on the cost-effectiveness of malaria control measures in Africa. *Health Policy Plan*. 1999;14:301-312.
13. Hunink MG, Bult JR, de Vries J, Weinstein MC. Uncertainty in decision models analyzing cost-effectiveness: The joint distribution of incremental costs and effectiveness evaluated with a nonparametric bootstrap method. *Med Decis Making*. 1998;18:337-346.
14. Hutubessy RC, Baltussen RM, Barendregt JJ, Evans DBB, Murray CJ. Stochastic league tables: Communicating cost-effectiveness results to decision makers. *Health Econ.* 2001;10:473-477.
15. Lord J, Asante MA. Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. *Health Econ*. 1999;8:323-333.
16. Manning WG Jr, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In Gold MR, Siegel JE, Russel LB, Weinstein MC, eds. *Cost-effectiveness in health and medicine*. 1st ed. New York: Oxford University Press; 1996;247-175.
17. Mennemeyer ST, Cyr LP. A bootstrap approach to medical decision analysis. *J Health Econ*. 1997;16:741-747.
18. Murray CJ, Evans DB, Acharya A, Baltussen RM. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Econ.* 2000;9:235-251.
19. O'Brien BJ, Drummond MF. Statistical versus quantitative significance in the socioeconomic evaluation of medicines. *PharmacoEconomics*. 1994;5:389-398.
20. O'Brien BJ, Drummond MF, Labelle RJ, Willan A. In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care*. 1994;32:150-163.
21. Obenchain RL, Melfi CA, Crompton DW, Bueshing DP. Bootstrap analyses of cost-effectiveness in antidepressant pharmacotherapy. *PharmacoEconomics*. 1997;11:464-472.
22. Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: A comparison of four methods. *Health Econ*. 1997;6:243-252.
23. Severens JL, De Boo TM, Konst EM. Uncertainty of incremental cost-effectiveness ratios: A comparison of Fieller and bootstrap confidence intervals. *Int J Technol Assess Health Care*. 1999;15:608-614.
24. Tambour M, Zethraeus N. Bootstrap confidence intervals for cost-effectiveness ratios: Some simulation results. *Health Econ*. 1998;7:143-147.