E C O N O M I C S B U L L E T I N

A new outcome measure for cost−utility analyses of screening programs

Carmen Herrero
Universidad de Alicante IVIE

Juan D. Moreno−Ternero
CORE, Université catholique de Louvain

Abstract

We show that, under some plausible assumptions, the gain in QALYs a screening program offers is a positive linear transformation of the program's sensitivity level. This result simplifies considerably the cost−utility analysis of mutually exclusive screening programs.

We thank Han Bleichrodt, Mike Drummond and Lars Østerdal for helpful comments and suggestions. Financial support from the Generalitat Valenciana, Ministerio de Educación y Ciencia and Fundación BBVA is gratefully acknowledged.
Citation: Herrero, Carmen and Juan D. Moreno−Ternero, (2005) "A new outcome measure for cost−utility analyses of screening programs." *Economics Bulletin*, Vol. 9, No. 7 pp. 1−8
Submitted: September 16, 2005. **Accepted:** September 19, 2005.
URL: http://www.economicsbulletin.com/2005/volume9/EB−05I10005A.pdf

1 Introduction

Cost–utility analyses (CUAs) that measure cost-effectiveness in costs per quality-adjusted life-year (QALY) have increasingly become the standard in the economic evaluation of health care programs (e.g., Neumann et al., 1996). The US Public Health Service's Panel on Cost-Effectiveness in Health and Medicine has recommended that, for analyses intended to inform resource allocation, a reference case should be included that measures cost-effectiveness as incremental costs/QALY, also known as incremental CU ratios (e.g., Gold et al., 1996).

However, the collection of utility data to quality-adjust years of life can be expensive and resource-intensive. Furthermore, there is no consensus about the method to be used to elicit health state utility weights (e.g., Bleichrodt and Johannesson, 1997; Bleichrodt, 2002) or about the discount rate to be used to obtain present values of future health benefits (e.g., Bleichrodt and Gafni, 1996; Cairns and van der Pol, 2000).

We show here that for an interesting class of health care programs, the so-called *screening* programs, incremental CU ratios can be obtained without the computation of QALYs, provided programs satisfy some mild assumptions. More precisely, we show that under three assumptions, the gain in QALYs a screening program offers is a positive linear transformation of an intrinsic property of the program, namely, the ability of the program to identify correctly those who *have* the disease. This allows circumventing the problems described above, besides of simplifying the analysis considerably.

2 The model

Screening is traditionally defined as testing a population of asymptomatic individuals to identify precursors of a disease. The subjects who test positive are sent on for further evaluation in a subsequent diagnostic test to determine whether they do, in fact, have the disease. Individuals can be partitioned into four groups (true positives, false positives, true negatives and false negatives), according to whether they do or do not have the disease and whether their screening tests are positive or negative. We can compute how likely an individual would belong to each of the four groups by using characteristics of the population (prevalence of the disease) and of the detection ability of the screening test (sensitivity and specificity). *Prevalence* (ρ) is the probability of an individual in the population being impaired. The *sensitivity* of the screening test (π^1) is the conditional probability that an individual with the disease is positively detected by the test. The *specificity* of the

test (π^2) is the conditional probability of an individual without the disease being correctly detected as negative in the test. Using these definitions, the probability of an individual being a true negative ($(1-\rho)\pi^2$), a true positive ($\rho\pi^1$), a false positive ($(1-\rho)(1-\pi^2)$) and a false negative ($\rho(1-\pi^1)$) can be easily expressed.

Let $S = \{s_1, ..., s_m\}$ be the set of mutually exclusive screening programs that are available for the early detection of a given disease. For all $s_j \in S$, let c_j denotes its cost.[1] For $s_i, s_j \in S$, we define their incremental *cost-sensitivity* (CS) ratio as $\frac{c_i - c_j}{\pi_i^1 - \pi_j^1}$.

In a CUA, the cost-effectiveness is measured in *costs-per-QALY-gained* (CU) ratios. Let Q be the QALY index that gives for each individual health profile its number of QALYs.[2] Let us denote by G_j^{TP} the group of true positive individuals after implementing s_j. Similarly, G_j^{FP} (G_j^{FN}) [G_j^{TN}] denotes the group of false positive (false negative) [true negative] individuals after implementing s_j. For each $k \in \{TP, FP, FN, TN\}$, let Q_j^k denote the expected number of QALYs for an individual in group G_j^k and ρ_j^k the probability of being in G_j^k. Then, the expected number of QALYs for an individual, after implementing $s_j \in S$, is $Q_j = \sum_k \rho_j^k Q_j^k$. Thus, for $s_i, s_j \in S$, their incremental CU ratio is given by $\frac{c_i - c_j}{Q_i - Q_j}$.

3 The result

We show in this section that, for some screening programs, CU ratios and CS ratios are essentially equivalent . More precisely, we show that for screening programs satisfying three assumptions, a CU ratio is a (positive) linear transformation of the corresponding CS ratio.

The first assumption says, roughly, that utility does not decrease 'per se' by being referred to a screening program. In other words, the expected QALYs of a true (false) negative do not depend on the particular screening program being implemented. The second assumption says that early detection of the disease is advantageous at an individual level, and that this individual improvement is independent of the screening method chosen. The third assumption says that there are no utility differences between healthy individuals with different test results, i.e., between a false positive and a true negative individual. Formally,

[1] By costs we mean all sort of health care expenses associated with the program, such as the costs of the screening technique, and the costs of the final diagnostic test to which every individual that has been identified as positive by the screening is referred. Thus, the sensitivity and the specificity of the program influence its costs, i.e., $c_j = c_j(\pi_j^1, \pi_j^2)$.

[2] Note that Q depends on the particular elicitation method and discount rate chosen.

Assumption 1: *For each QALY index Q, and for all $s_j \in S$, we have*

$$Q_j^{TN} = \overline{Q}, \text{ and } Q_j^{FN} = \widehat{Q}.$$

Assumption 2: *For each QALY index Q, and for all $s_j \in S$, there exists $\gamma_Q > 0$ such that*

$$Q_j^{TP} - Q_j^{FN} = \gamma_Q.$$

Assumption 3: *For each QALY index Q, and for all $s_j \in S$, we have*

$$Q_j^{FP} = Q_j^{TN}.$$

The plausibility of these assumptions depends on the screening programs we are considering. They are fulfilled whenever the alternative programs are non-invasive and such that false positives are correctly identified in a short period of time. For instance, in the case of newborn and non-invasive screening programs, the three assumptions are sound.[3]

We now have the following result:

Theorem 1 *Under Assumptions 1 to 3, the incremental CU ratios are a (positive) linear transformation of the corresponding incremental CS ratios.*

Proof.

Fix $s_i, s_j \in S$ and a QALY index Q. For $k = i, j$, the expected number of QALYs for an individual after implementing s_k is given by:

$$Q_k = \rho\pi_k^1 Q_k^{TP} + (1-\rho)(1-\pi_k^2)Q_k^{FP} + \rho(1-\pi_k^1)Q_k^{FN} + (1-\rho)\pi_k^2 Q_k^{TN}.$$

Hence, the incremental utility from s_i to s_j is

$$\begin{aligned} Q_j - Q_i &= \rho\left(\pi_j^1 Q_j^{TP} - \pi_i^1 Q_i^{TP}\right) + (1-\rho)\left((1-\pi_j^2)Q_j^{FP} - (1-\pi_i^2)Q_i^{FP}\right) + \\ &\quad \rho\left((1-\pi_j^1)Q_j^{FN} - (1-\pi_i^1)Q_i^{FN}\right) + (1-\rho)\left(\pi_j^2 Q_j^{TN} - \pi_i^2 Q_i^{TN}\right). \end{aligned}$$

By Assumption 1,

$$(1-\pi_j^1)Q_j^{FN} - (1-\pi_i^1)Q_i^{FN} = (\pi_i^1 - \pi_j^1)\widehat{Q},$$

and

$$\pi_j^2 Q_j^{TN} - \pi_i^2 Q_i^{TN} = \left(\pi_j^2 - \pi_i^2\right)\overline{Q}.$$

By Assumptions 1 and 2,

$$\pi_j^1 Q_j^{TP} - \pi_i^1 Q_i^{TP} = (\widehat{Q} + \gamma_Q)(\pi_j^1 - \pi_i^1).$$

[3] See Herrero and Moreno-Ternero (2005) for further details in the particular case of newborn hearing screening programs.

By Assumptions 1 and 3,

$$(1-\pi_j^2)Q_j^{FP} - (1-\pi_i^2)Q_i^{FP} = \left(\pi_i^2 - \pi_j^2\right)\overline{Q}.$$

Thus,

$$\begin{aligned} Q_j - Q_i &= \rho(\widehat{Q} + \gamma_Q)(\pi_j^1 - \pi_i^1) + (1-\rho)\overline{Q}\left(\pi_i^2 - \pi_j^2\right) + \rho\widehat{Q}(\pi_i^1 - \pi_j^1) + (1-\rho)\overline{Q}\left(\pi_j^2 - \pi_i^2\right) \\ &= \rho(\pi_j^1 - \pi_i^1)\gamma_Q. \end{aligned}$$

Consequently, the incremental CU ratio is given by:

$$\frac{c_j - c_i}{Q_j - Q_i} = k \cdot \frac{c_i - c_j}{\pi_i^1 - \pi_j^1},$$

where $k = k(\rho, Q) = \frac{1}{\rho \cdot \gamma_Q} > 0$. Note that k is not screening method-specific. This says, in particular, that the order of incremental CU ratios coincides with the order of incremental CS ratios.[4] ■

4 Discussion

The first step in using CUA for mutually exclusive programs is to exclude those programs that are strictly dominated, i.e., those for which there exists another available program more effective (i.e., providing more QALYs) and less expensive (e.g., Weinstein, 1990). By Theorem 1, this simply means excluding programs that are more expensive and less sensitive than another available program (see footnote 4).

The second step is to rank programs according to their effectiveness (i.e., number of QALYs they provide) and then compute the incremental CU ratio for each successively more effective program. If any of these incremental ratios turns out to be less than the previous one in the sequence of increasingly effective mutually exclusive programs, then the less effective one is ruled out by *extended dominance*, and it should never be implemented irrespective of the amount of resources available (e.g., Garber, 2000). This algorithm results in a sequence of programs with increasing incremental cost-effectiveness ratios. The optimal decision rule is to move up the list of incremental ratios and implement successively more effective (and expensive) programs until the resources are exhausted (e.g., Johannesson and Weinstein, 1993).[5]

[4] Note also that the ranking of programs according to the QALYs they provide coincides with the ranking of programs according to their sensitivity levels, as $Q_j > Q_i \iff \pi_j^1 > \pi_i^1$.

[5] This decision rule applies under the assumption of divisibility of programs with constant returns to scale (e.g., Birch and Gafni, 1992). See Elbasha and Messonnier (2004) for the resulting rule after relaxing this assumption.

By Theorem 1, we can simplify this algorithm as follows. First, rank programs according to their sensitivity levels. Then, calculate the incremental CS ratio for each successively more *sensitive* program. Ruling out programs by extended dominance, this alternative algorithm results in a sequence of programs with increasing incremental CS ratios. The optimal decision rule would be to move up the list of incremental ratios and implement successively more sensitive (and expensive) programs until resources are exhausted. Theorem 1 guarantees that both algorithms produce the same outcomes.

We acknowledge, however, that the alternative algorithm just presented is only valid for the case of mutually exclusive programs and it could not be used for most general decision contexts in which a number of clusters of mutually exclusive programs is considered.[6]

The alternative algorithm just described only uses statistical properties of each screening program. This resembles the main method in the literature on the measurement of the accuracy of diagnostic systems: the so-called *ROC analyses* (e.g., Sweets, 1988). A ROC analysis assesses the value of diagnostic tests by deciding where to put the threshold when using the test. This requires consideration on two factors: the total number of errors (type I errors, or false negative cases, and type II errors, or false positive cases) made, and the relative importance of errors. The threshold is chosen so that we minimize the total error rate, i.e., the weighted aggregation of type I errors and type II errors, where weights are chosen to reflect their relative importance. Our Assumption 3 implies that the importance of false positives is negligible. Thus, a ROC analysis of screening programs satisfying our assumptions would also rank them according to their sensitivity levels. This is precisely what we obtain from Theorem 1 in the case in which we rank programs according to the QALYs they offer.

References

[1] Bleichrodt, H., 2002. A New Explanation for the Difference Between SG and TTO Utilities. Health Economics 11, 447–456.

[2] Bleichrodt, H., Gafni, A., 1996. Time preference, the discounted utility model and health care. Journal of Health Economics 15, 49–67.

[6] The reason being that the scalar k that appears in the proof of Theorem 1 depends on the prevalence of the disease.

[3] Bleichrodt, H., Johannesson, M., 1997. Standard Gamble, Time Trade-off and Rating Scale: Experimental Results on the Ranking Properties of QALYs. Journal of Health Economics 16, 155–175.

[4] Birch, S. and A. Gafni, 1992, Cost-effectiveness/utility analyses: Do current decision rules lead us to where we want to be?, Journal of Health Economics 11, 279-296.

[5] Cairns, J., van der Pol, M., 2000. Negative and zero time preference for health. Health Economics 9, 171–175.

[6] Elbasha E, Messonnier, M, 2004. Cost-effectiveness analysis and health care resource allocation: decision rules under variable returns to scale. Health Economics 13, 21–35.

[7] Garber, A.M., 2000. Advances in cost-effectiveness analysis of health interventions, in Culyer A.J., Newhouse J.P. (Eds.), Handbook of Health Economics. North Holland. Amsterdam. Chapter 4.

[8] Gold, M.R., Siegel, J.E., Rusell, L.B., Weinstein, M.C., 1996. Cost-effectiveness in health and medicine. Oxford University Press. New York.

[9] Herrero C., Moreno-Ternero J.D., 2005, Hospital Costs and Social Costs: A Case Study of Newborn Hearing Screening. Investigaciones Económicas 29, 203–216.

[10] Johannesson, M., Weinstein, M.C., 1993, On the decision rules of cost-effectiveness analysis. Journal of Health Economics 12, 459-467.

[11] Neumann P.J., Stone P.W., Chapman R.H., Sandberg E.A., Bell C.M., 2000 The quality of reporting in published cost–utility analyses, 1976–1997. Annals of Internal Medicine; 132: 964–972.

[12] Sweets, J.A., 1988, Measuring the accuracy of diagnostic systems. Science 240(4857), 1285–1293.

[13] Weinstein, M.C., 1990, Principles of cost-effective resource allocation in health care organizations, International Journal of Technology Assessment in Health Care 6, 93-103.