Published in final edited form as:

J Health Econ. 2008 July ; 27(4): 819–821. doi:10.1016/j.jhealeco.2008.05.002.

Future costs and the future of cost-effectiveness analysis

Alan M. Garber* and

VA Palo Alto Health Care System and Stanford University

Charles E. Phelps

University of Rochester

In the past decade, the use of cost-effectiveness (CE) analysis to guide resource allocation in health care has grown in the United Kingdom and in continental Europe. Its use in the United States has been far more limited. The political resistance to explicit consideration of costs in the Medicare's coverage decision making has all but buried efforts to make coverage decisions based on cost-effectiveness criteria, and explicit use of cost-effectiveness analysis is unusual among private U.S. health insurers. Nevertheless, this technique is attracting greater interest among policymakers and private health plans even in the United States, and its adoption elsewhere confirms that it is now seen as an essential tool to increase the efficiency of health expenditures. If it is really to be used, we should make sure that we get it right, and that its guidance will improve well-being.

Our previous article (Garber and Phelps, 1997) was motivated by two simple questions: does a cost-effectiveness criterion lead to potential Pareto-improvement, as is known to be the case when a cost-benefit criterion (positive net benefit) is applied, and if so, under what conditions? By this means, we hoped both to establish whether cost-effectiveness analysis provided a reasonable alternative to cost-benefit analysis as a guide to health expenditures, and to resolve methodological controversies – particularly whether unrelated future costs of care should be excluded from CE analysis or included. We proposed that the correct criterion for the use of CE analysis was to determine whether the allocations that resulted from application of the technique were optimal, in the sense that they represented a potential Pareto improvement, just as cost-benefit criteria could do. We argued that if the future costs were truly unrelated, it did not matter whether such costs were included or excluded, as long as the cutoff CE ratio, above which the costs of the health expenditure would exceed the value of the resulting benefits, were properly adjusted. In an article published in the same issue of JHE, Meltzer (1997) argued that the future unrelated costs – not only of health care, but of consumption more generally – must be included in the cost-effectiveness calculation in order to rank alternatives for optimal resource allocation.

Reasons for Disagreement and Consensus

In this issue of JHE, Lee offers an explanation for the differences between Meltzer's results and ours, highlighting the importance of the different assumptions that the two papers made about the nature of the intertemporal budget constraint. He concludes that future unrelated costs

*Corresponding author: Alan M. Garber, M.D., Ph.D., Henry J. Kaiser, Jr. Professor and Professor of Medicine, Director, Center for Primary Care and Outcomes Research, Director, Center for Health Policy, Stanford University, 117 Encina Commons, Stanford, CA 94305, Tele: 650-723-0920, Fax: 650-724-5182, garber@stanford.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

must be ignored, and further points out that empirical implementation of a recommendation to include future unrelated costs can be impracticable. Substituting earnings for income, for example, can lead to large errors whenever income and earnings deviate significantly, as they often do for the elderly. This is a welcome result for the practitioners of cost-effectiveness analysis.

Though they differ on the details of how future costs should be included, all three papers strongly support the contention that cost-effectiveness analysis can be used to make optimal decisions regarding the allocation of medical resources, under specific circumstances. Lee's paper goes a long way toward reconciling the different conclusions about future costs, and offers practical guidance toward the application of cost-effectiveness analysis. But since the assumptions used to reach these conclusions are strict, what should the would-be practitioner of cost-effectiveness analysis do, and how confident can we be that in the "real world" the answers given by a CE analysis will indeed be valid?

What is Truly Unrelated?

One of the most obvious questions is whether future costs are truly "unrelated". We defined future expenditures as "unrelated" if their optimal values were independent of previous expenditures, conditional on survival to the period in question. For a large class of interventions – preventive care comes immediately to mind – this is an untenable assumption, since their very purpose is to prevent disease and attendant expenditures in the future. For another large class of interventions, such as surgery and related care for the treatment of victims of accidents and violence, fully successful treatment is unlikely to affect future expenditures, apart from those that simply result from living longer. And then there is a very large class of interventions for which it is plausible that future expenditures might be independent, but it is also possible that they are not, and we frequently lack the data to be sure whether the future expenditures are unrelated. Does cancer treatment in early adulthood, for example, affect health expenditures at future ages? We now know that the risk of second malignancies is elevated among people treated for cancers earlier in life, but much of the time we cannot predict whether future patterns of medical expenditures will be affected by current expenditures, apart from their effect on longevity. An assumption that future costs will be unrelated is often plausible. In such analyses, we must be careful to distinguish between consequences of the *treatment per se* and (for example) underlying genetic risks or environmental risks such as smoking that may increase the risk of future cancers.

In many situations, the question of conditional independence may turn out to be unimportant. A careful description of outcomes and probabilities, as in a decision analysis, can be used to determine whether the cost-effectiveness estimate is sensitive to inclusion of unrelated future costs. One might speculate that since the elderly with chronic diseases account for such a disproportionate share of health expenditures, ignoring future unrelated costs is likely to have little effect. Presumably as more and more such studies are undertaken, some regularities will be discovered that will help practitioners gauge the importance or irrelevance of the costs of treating future illnesses.

Second Best Issues in a Subsidized Environment

One of the assumptions embedded in all three papers is that future medical expenditures are chosen optimally. Indeed, any method for determining optimal medical expenditures, as in every other setting in which such techniques are used, is necessarily subject to qualification in a second-best world. A common use of cost-effectiveness analysis is to mimic the demand that a perfectly informed individual would have for a treatment, given no price distortions and with optimal indemnity insurance. The indemnity insurance component is crucial because it enables individuals to reallocate funds to states of health in which the marginal impact of expenditures,

whether for health or other forms of consumption, would be very high. But what if there are subsidies or other distortions in some of those states of health? For example, some Americans might have more insurance than is optimal when they qualify for Medicare, since participation in Part A is mandatory and participation in the other components is heavily subsidized. Then optimality might require, for example, using a relatively low CE threshold during years of Medicare eligibility, and it suggests that an individual's demand at younger ages for life-saving interventions will be higher than otherwise in order to be able to capture the subsidized benefits of Medicare. This would be a fruitful area for future research.

It is no easier to develop second-best solutions in the context of medical care than it is in other arenas. Yet it is also possible that some of the distortions – such as apparently excessive health insurance – are not really distortions when externalities are appropriately considered. After all, much of the impetus supporting health coverage for the uninsured is a belief that there is a collective benefit from extending access to medical care to all members of society. If this were merely a matter of income redistribution, there would be equally strong or stronger support for an income subsidy, such as a negative income tax. If we think that individuals would under-consume medical care because they ignore the externalities, the tax subsidy to health insurance and mandatory participation in Medicare may correct a distortion, rather than causing one. In many situations, the distortions caused by subsidies and other distortions in pricing and demand will cause greater challenges to the application of CE analysis than do issues about unrelated future costs.

Practical Implementation

In the 1990s, the practical obstacles to the application of CE analysis were great. Although they remain significant, better and more extensive individual data are becoming available, and are poised to become very widely available in the coming decade. Britain's National Health Service, Kaiser-Permanente and the Department of Veterans Affairs in the United States, and many other providers and payers throughout the world, have made massive investments in electronic health records and information infrastructure. These investments may produce the information needed to overcome perhaps the largest practical barrier to the application of CE analysis in health care settings: the tremendous heterogeneity in the magnitude of benefits from most interventions when applied in typical patient populations.

In our earlier paper, we noted that one of the most important reasons for the failure of a CE-based allocation rule to lead to optimal expenditures is the use of a single CE ratio to determine whether to use an intervention in a heterogeneous population. Then the intervention might be given to some people who derive tremendous benefit, but also to many whose benefits will be small and whose value would be less than the marginal cost of the intervention. Our limited sample of available studies suggested that the largest swings in CE ratios occurred with shifts along the extensive margin – extending the use of diagnostic tests or treating people with lower prior probability of disease, for example. With better information about individual health conditions, age, and other factors that determine incremental health benefit, it should be possible to more precisely define the cost-effectiveness of an intervention at the individual level, rather than at the population level.

Cost-Effectiveness and Performance Incentives

Another new development makes cost-effectiveness analysis more relevant, and more applicable, than ever before. Pay for performance (P4P) is designed to align provider incentives with patient well-being. Its rationale is largely that of CE – ideally, it rewards providers for producing outcomes, rather than paying them for units of services, whether those services are appropriate or of high quality. Ideally, the performance criteria and the reimbursement rates for meeting them should be based on a utility-enhancing criterion, and as our earlier work and

that of Meltzer and Lee affirm, this leads to the use of a cost-effectiveness test. Insofar as performance-based payments match payment to benefit, the need to deny a treatment to some people who benefit but not others can often be avoided.

Recently payment by results became a serious option for expensive biological compounds. Facing the likelihood that the National Health Service in Britain would not pay for its drug Velcade for the treatment of multiple myeloma, Johnson & Johnson proposed to accept payment only for those patients who had a meaningful clinical response. This mode of payment means that patients with high and low expected benefits can be treated, since the cost is restricted to those with (ex post) benefit (Garber and McClellan, 2007). Conventional unit pricing with a cost-effectiveness threshold would have meant that fewer patients who might benefit would be treated, only those whose expected benefit would meet a standard cost-effectiveness threshold. Although P4P is relatively new and in most settings the performance incentives are small, the widespread interest in this approach suggests that cost-effectiveness analysis can and should have a large role to play in provider compensation and patient reimbursement design, not only coverage decisions, in the future. Resolving methodological difficulties in CE analysis remains important for its future, but so does its successful implementation, which is likely to get easier as time goes on.

Acknowledgements

Alan Garber's work is supported in part by the Department of Veterans Affairs, by an Investigator Award in Health Policy Research from the Robert Wood Johnson Foundation, and by grants AG 17253 and AG24957 from the National Institute on Aging.

References

- Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* 1997;16:1–31. [PubMed: 10167341]
- Garber AM, McClellan MB. Satisfaction guaranteed - "payment by results" for biologic agents. *New England Journal of Medicine* 2007;357:1575–77. [PubMed: 17942870]
- Meltzer DM. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* 1997;16:33–64. [PubMed: 10167344]