Costing and Cost Analysis in Randomised Trials: Caveat Emptor

Daniel Polsky, Ph.D. and
University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Henry Glick, Ph.D.
University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Abstract

The objective of this paper is to provide an overview of the central issues with respect to cost valuation and analysis for a decision maker's evaluation of costing performed within randomized clinical trials. Costing involves specific choices for valuation and analysis that involve tradeoffs. Understanding these choices and their implications are necessary for proper evaluation of how costs are valued and analyzed within a randomized clinical trial that can not be assessed through a checklist of adherence to general principals..

The most common method of costing, resource costing, involves measuring medical service use in study case report forms and translating this use into a cost by multiplying the number of units of each medical service by price weights for those services. A choice must be made as to how detailed the measurement of resources will be. Micro-costing improves the specificity of the cost estimate, but it is often impractical to precisely measure resources at this level and the price weights for these micro units may not be available. Gross-costing may be more practical and price weights are often easier to find and are more reliable, but important resource differences between treatment groups may be lost in the bundling of resources. Price weights can be either nationally determined or they can be center-specific, but the appropriate price weight will depend on perspective, convenience, completeness, and accuracy. Identifying the resource types and the appropriate price weights for these resources are the essential elements to an accurate valuation of costs.

Once medical services are valued, the resulting individual patient cost estimates must be analyzed. The difference in the average cost between treatment groups is the important summary statistic for cost-effectiveness analysis both from the budgetary and social perspectives. The statistical challenges with cost data typically stem from its skewed distribution and the resulting possibility that the sample mean may be inefficient and possibly inappropriate for statistical inference. Multivariable analysis of cost is useful even if the data come from a randomized trial, but the same distributional problems that affect univariate tests of cost also affect use of cost as a dependent variable in a multivariable regression analysis. The Generalized Linear Model (GLM) overcomes many of the problems of more common cost models, but one must be cautious when applying this model because it is prone to misspecification and precision losses in data with a heavy-tailed log error term.

Attention to the appropriate cost valuation and analysis techniques reviewed here will help bring the same level of rigor and attention to the methodological issues in cost valuation as currently applied to clinical evidence within randomized trials.

In this cost conscious environment, less expensive therapies will displace more expensive ones unless there is evidence that higher cost therapies are justified by better outcomes. This need for justification has, in large part, driven the continued trend within randomized clinical trials

Corresponding author: Daniel Polsky, Ph.D., University of Pennsylvania, Division of General Internal Medicine, Blockley Hall, Rm. 1212, 423 Guardian Drive, Philadelphia, PA 19104, w 215-573-5752, f 215-898-0611, polsky@mail.med.upenn.edu.

to collect data on the costs of treatment and then to perform economic evaluations assessing the potential tradeoffs between costs and clinical effectiveness. The conclusions from analyses of costs and cost-effectiveness are sensitive to the methods of cost valuation and analysis. Confidence in the results produced by these analyses will depend on understanding whether appropriate analytical decisions were made and whether valid alternative methods might have changed the findings. Proper valuation and analysis of costs is not just a matter of following general costing principals; results are often sensitive to alternatives within the framework of acceptable principals The purpose of this review is to improve the understanding of the tradeoffs between the choices involved in costing and their potential implications.

I. VALUATION

I. A. Principals

The general principals for the valuation of costs include opportunity costs, consistency in perspective, and discounting. The economically appropriate cost of an intervention is best represented by its "opportunity cost." This cost is the value of opportunities forgone as a result of the resources consumed. It is often the case that the medical costs listed within accounting systems or medical charges assessed for a particular set of services delivered do not accurately represent these "opportunity costs." In this case, adjustments should be made so that the value of services represent their "opportunity cost" or amount that would be paid in exchange for goods or services deliveredd in a well functioning competitive market.

Perspective is a second general principal. The base case perspective is the societal perspective [1], but costs can be valued from the perspective of any relevant economic stakeholder such as the payer, the provider, or the patient. Because the appropriate opportunity cost may be different depending on the perspective one adopts in the analysis, it is important to maintain consistency in the perspective from which costs are estimated. For example, while payment may not be an accurate representation of opportunity cost from the social perspective, it may be from the payer perspective. When costs are estimated the perspective (or perspectives) of the cost analysis should be stated and all costing should be consistent with the stated perspective (s).

Because costs may be incurred over several years and the timing of costs may differ between treatments, a general principal of costing is to equate the timing by adjusting for inflation and discounting future costs to account for time preference. These methods are described elsewhere.[2,3] When evaluating whether a cost analysis was done appropriately, it is important that these principals be followed. However, while it is necessary to follow these principals for best practices, it is equally important that the strategy by which medical resources are identified and valued be appropriately executed.

I. B. Strategies for valuation

The typical strategies for the valuation of cost are through determining actual costs, using billing records, or employing the resource costing method. All three strategies have advantages and disadvantages, but we discuss the issues of resource costing in depth because it is the most common strategy found in the literature.

While costs could be valued by summing the actual costs of health care services within a clinical trial -- e.g., by recording costs incurred by participants throughout the trial on their case report forms -- doing so is difficult. One difficulty has to do with availability of pricing data; a second has to do with whether any available pricing data represents opportunity costs. For example, those recording the data may, on occasion, be aware of health care charges, but charges generally do not equal cost[4] and their use may lead to an overstatement of the costs of treatment.

Adjusted billing records can be used for cost valuation. If the records obtained consistently cover the medical services used throughout the duration of the trial it may be possible to derive consistent measures of both medical service use and cost (or expenditure) from billing records.. For example if the subjects within the trial received their health care within a closed health system consistent and comprehensive billing records may be available from that health system. Alternatively, if the subjects within the trial are insured by companies able to provide claims records that contain bills for the key medical services provided, this may be another source for consistent and comprehensive billing records. This strategy has the advantage of being independent of fallible participant recall and may be more comprehensive in tracking service use across all relevant sites of care. Because some administrative databases aggregate service use, one should assess the methods for parsing costs that occurred prior to randomization, from randomization until the end of follow-up, and after the end of follow-up. In addition, proper conversion of payment or charges to costs may be necessary.[4] Most importantly, the administrative data should capture all relevant medical service use and costs incurred by the participant. If patients cross systems of care or payment sources the data source may be incomplete or the data sources may be difficult to aggregate.

The resource costing method involves measuring medical service use in study case report forms and translating this use into a cost by multiplying medical service use by price weights for those services. When costs are incurred across multiple health systems or payers, this method allows a consistent capture of resource use. In addition, the analyst can apply economic principals in deriving price weights. The issues confronted when applying this method are collecting data for appropriate resource units and then selecting the appropriate set of price weights for these units.

I. C. Resource costing and price weights

I. C. 1. The micro-costing and gross-costing continuum—There is no standard resource costing strategy. An extreme approach would be to identify, count, and price out every single health care service item consumed by each patient. Yet this strategy is rarely attempted in practice because the benefits from the specificity of the result are believed to be exceeded by the effort required to obtain the result. A more typical approach would be to identify, count, and price out health care encounters or other health care units that represent some aggregate of a bundle of service items (e.g. the average cost per hospital day or average cost per hospitalization). Greater bundling reduces investigator effort, but generally sacrifices specificity of the resulting cost estimate. The two ends of this continuum can be referred to as micro-costing method and the gross-costing method.[1]

Hospital days and hospital stays represent bundled resource units that typify the gross-costing method.[5–7] Under this method the same unit price is used for the bundled unit, even though some hospital days are more intensive than others and hospital stays differ in their duration and intensity. The unit price that is used is typically based on average costs in the population. As one moves along the continuum towards less bundling and the use of micro-costing, stays are disaggregated into days, and days are disaggregated into locations in the hospital and specific services that provided (e.g., counts of various procedures, drug prescriptions, and laboratory tests). The price information for these resource types can be derived from average manufacturing (AMP) or sales prices (ASP) for pharmaceuticals applied to specific drugs, Medicare's Physician Fee Schedules applied to specific procedures, and adjusted charges for bed days. Using Medicare's inpatient prospective payment system for inpatients stays based on diagnostic-related groups (DRGs) applied to hospitalizations could be considered another form of resource costing.[5,8] On the micro-costing extreme, one could count labs, provider contact time, and the specific dosing and frequency of use for each drug given along with other

relevant resource items. This specific level of resource costing will also require obtaining a unit price for each resource item.

Often the time horizon of evaluation within a clinical trial is not limited to a hospital stay or may be in other settings entirely. In these cases, the valuation of costs should include non-negligible medical services delivered regardless of the setting. For example, for an episode of care that may include inpatient and outpatient care gross-costing may have three resource types: hospital stays, physician visits, and emergency room visits while micro costing may have a long list of resource types within each of these bundles.

I. C. 2. Choice of the level of bundling—During the design phase of a clinical trial, one must define the level of bundling that will be used to identify the health care resources that make up the cost of medical care. This level will fall somewhere on the micro-costing/gross-costing continuum. This decision should involve maximizing accuracy of the estimate of the cost difference between the randomized treatments subject to the resource constraints of the investigator. Accuracy of costing could be improved by use of more detailed micro resource units. On the other and it could be improved with more accurate unit prices or counts of resources used which may be more likely for bundled resource units.

This decision is most important for those resource items that may drive the cost difference between treatment groups. If within a bundled unit (eg., hospitalization), one treatment group incurs greater costs, this difference would be lost in a valuation strategy that provided a single price for that bundled unit. An extreme example is when the intervention being evaluated is a more resource intensive procedure conducted in the hospital. In this case the extra resources devoted to the treatment group would not be counted if the entire hospitalization is treated as a single resource type. On the other hand, if the intervention is in the outpatient setting and hospitalizations are rare, it may be appropriate to treat the hospitalization as a single bundled unit, particularly if the types of hospitalizations would be unlikely to differ between treatment groups.

Often these decisions depend on the availability of price weights. The unit needs to be bundled at a level that matches the price weight. If the only prices available are for hospitalizations for different admitting diagnoses, then the resources should be bundled at this same level. A mismatch presents grave problems for proper valuation. And even when the resource unit matches the price weight, one must evaluate whether the level of aggregation of the unit hides true cost differences.

There are times when maximizing the accuracy of the cost estimate needs to be weighed against the goal of precision. This tradeoff is most evident in the choice between including all medical costs or disease specific costs only. If costs that are not relevant for the disease under consideration are included, then the variance of the cost estimate may be increased (worse precision) without any gain in accuracy. However, the medical costs that are relevant for a specific treatment may not be known. By excluding these costs, the gain in precision may come at the expense of accuracy.

I. C. 3. Choice of price weights—Price weights can either be nationally determined, as in the use of national tarrifs or they can be center-specific. The decision as to which type of price weights to use will depend on perspective, convenience, completeness, and accuracy. Price weights can also determine the generalizability of the findings. Do they represent the costs of the environment relevant for the economic decision? For example, if the resource allocation decision is a national decision, an analysis that used price weights representative of average national costs would be preferable to price weights that may only be relevant for academic centers within that country.

A commonly used and well accepted source of price weights for costing out medical service use are national tariffs. Examples of national tariffs include diagnosis related group (DRG) payments in Australia or the United States or health resource group (HRG) payments in Great Britain. When available, national tariffs have all of the advantages that one looks for in a price weight. They are typically publicly available which makes them convenient; they usually provide price weights for most if not all of the services that are measured in the trial; within individual countries, they are usually developed by use of a common methodology. The fact that these tariffs represent what is spent by governments may also be considered an advantage, particularly when the perspective is that of governmental decision making bodies. Some authors have raised concerns that national tariffs do not equal cost because of differences in costs between centers or the level of aggregation provided in different data sources.[9–11] There remains a debate about whether these deviations from costs matter in practice.[12–15]

Center-specific price weights are another commonly used source of price weights for costing out medical service use. These price weights are often obtained from the institutions in which the trial was conducted. One of the primary advantages of center-specific price weights is that they provide a more accurate estimate of the cost that was actually incurred within the trial. As Goeree and colleagues have reported, the valuation of costs are sensitive to the number of centers from which price weights are derived.[16] Therefore, this strategy is best employed when price weights are derived systematically and comprehensively across centers. If the results of the trial are to be generalized to settings that have nothing in common with the centers participating in the trial, then the use of center-specific price weights collected from study sights may limit the generalizability of the findings.

The analog in multinational trials is that price weights are often collected from a subset of countries that participated in the trial. Within each country, they are often derived from a single source, which in the same trial can vary between center-specific price weights in some countries to national tariffs in others. The countries from which price weights are collected might be ones that: enroll a large number of participants in the trial; represent the spectrum of economic development among countries that participated in the trial; have readily available price weights; or have regulators that require a submission for reimbursement. In industry-sponsored studies, they also may include countries in which the study sponsor wishes to make economic claims. Resource costing requires a price weights for every service we intend to cost out in our analysis. However, it may be costly to obtain well documented and accurate price weights for all services in all centers and countries that participated in the trial. Typically, price weights are obtained for the most common or costly services in the largest centers or countries and the remaining price weights are either imputed using mean price weight for similar services or using a prediction model.[17,18] Imputation error is minimized when we obtain estimates for fewer types of services from as many countries as is feasible, rather than by obtaining estimates for more types of services in fewer countries.[18]

II. ANALYSIS

Once medical services are valued and summed across resource types, the resulting cost estimates must be analyzed. The difference in mean cost is the important parameter for cost-effectiveness analysis both from the budgetary and social perspectives.[19,20] It is the difference in the mean between the treatment groups that provides the most accurate assessment of budgetary impact because it allows decision makers to calculate the total cost of adopting a therapy. Minimization of the difference in mean cost (and maximization of the difference in mean effect) yields Kaldor-Hicks social efficiency.[21] Other parameters such as median cost -- the cost above and below which the cost of half the patients lie -- may be useful in describing the data, but do not provide information about the total cost that will be incurred by treating all patients nor the cost saved by treating patients with one therapy versus another. They thus

are not associated with social efficiency. The question is how to estimate the difference in means (and its precision) within the sample of a population that comes from the patients enrolled in a randomized clinical trial. While the sample mean is always unbiased, other statistics have been proposed as alternatives to the sample mean because it is common for its distribution to be highly skewed. What needs to be stressed here is that no matter what method we use for drawing inferences about cost differences, the between-group difference in mean cost remains the parameter of interest.

While most introductory discussions of the analysis of continuously scaled data simplify the analysis by assuming that the sampling distribution of the parameter is normally distributed. The analysis of health care cost is complicated by the fact that the data are generally right skewed with a long, heavy, right tail.[22] There are a number of reasons why cost is routinely characterized by a skewed distribution. First, cost cannot be negative, which places a bound on the lower tail of the distribution. Second, there is often a nontrivial fraction of study participants that requires substantially more medical services than the norm which make up the long right tail. Third, there usually are a small number of participants -- often referred to as outliers --among whom a catastrophic event occurs that leads to a cost that is several standard deviations above the mean.

II. A. Univariate analysis of cost

The most common univariate tests of the sample mean are the two sample t-test or --when we are comparing more than two treatments -- one-way analysis of variance (ANOVA). (When we refer to sample mean in general we refer to the arithmetic mean.) One of the assumptions that underlies these tests is that the sampling distribution of the parameter is normally distributed. However, in moderately large samples, in samples of similar size and with similar skewness, or samples where skewness is not too extreme, these tests have been shown to be robust to violations of this assumption.[19,23–25]

Use of these tests is common in the economic trial literature. Doshi and colleagues[26] found that in 2003 50% of the published economic assessments based on the analysis of patient level data collected within clinical trials used t-tests or ANOVA. Nevertheless, because cost data often have a skewness or kurtosis that violate the normality assumption, and because there are no specific thresholds or ranges for sample size or skewness for determining when parametric tests will be reliable, many analysts have rejected these parametric tests. They often have turned to comparing cost by use of nonparametric tests of other characteristics of the distribution that are not as affected by the nonnormality of the distribution. Examples of such tests include the Mann-Whitney U (or Wilcoxon rank-sum) test, the Kolmogorov-Smirnov test, and the Kruskall-Wallis test. The problem with the use of these tests is that while they tell us that some measure of the cost distribution differs between the treatment groups, such as its shape or location, they do not necessarily tell us that the sample means differ. Some authors who have adopted nonparametric tests of other characteristics of the distribution have ignored the fact that the resulting p-values need not be applicable to the sample mean. For example, Doshi and colleagues[26] found that a number of authors reported cost estimates in terms of sample means and the difference in sample means but derived a p-value for this difference by use of the Mann-Whitney U test.

When the cost distribution is not normal, some analysts have transformed cost in an attempt to make the distribution of the resulting transformed variable more normal. Examples of such transformations include the logarithmic, square root, or reciprocal transformation. When we make such transformations, we estimate and draw inferences about the difference on the transformed scale, but our fundamental goal is to apply these estimates and inferences to the difference in sample mean of untransformed cost. Thus we need to be concerned about issues that may undermine this applicability. The logarithmic transformation of cost is widely used

in health care cost analysis. The most common reason for its use is that if the resulting distribution of log cost is normal, t-tests of log cost may be more efficient than t-tests of nonnormally distributed untransformed cost. Some analysts also use a log transformation because the difference in log cost -- or the coefficient estimates for explanatory variables in multivariable analysis of log cost -- can be interpreted as percentage differences.

Unfortunately, log transformation raises several issues for the analyses of cost. First, when observations have 0 cost, their log cost is undefined. Many researchers address this problem by adding an arbitrary constant such as 1 to all cost observations before taking logs. However, one should always determine whether the magnitude of the arbitrary constant affects one's conclusions, and the presence of a substantial proportion of zeroes can make this approach problematic. We could instead analyze such data by use of a two-part model. [27] Second, transformation of costs does not typically yield a normal distribution. Third, and possibly most important, estimates and inferences about log cost apply more directly to the geometric mean of cost than they do to the arithmetic mean. A proper transformation can retransform the mean of the logs to obtain an estimate of arithmetic mean cost, but this is rarely done correctly[19, 29] and even these common retransformations may provide unbiased estimates when the distribution of log cost is not normal.

These problems for estimation translate into problems for inference about differences in the sample mean of cost. When one uses a t-test to evaluate log cost, the resulting p-value has direct applicability to the difference in the sample mean of log cost. This p-value generally also applies to the difference in the geometric mean of cost (i.e., one sees similar p-values for the difference in the sample mean of logs and the difference in the geometric mean). However, the p-value for the log is not directly applicable to the difference in arithmetic mean of untransformed cost. As with estimation, the exception to this rule is the unlikely case where the distributions of the log costs to be differenced are normal and have equal variance.

Thus, in the presence of inequality of variance on the log scale (or on any scale based on a power transformation of costs), comparisons of log cost are not appropriate for drawing statistical inferences about differences in arithmetic mean cost.[19,28] If the sample mean is the statistic of interest for decision making, and if the underlying assumptions of parametric tests of this mean are violated, one alternative is to adopt a nonparametric test of sample means. An increasingly popular test of this kind is the non-parametric bootstrap.[29,30] But, in the face of small sample sizes and highly skewed distributions, O'Hagan and Stevens[31] have raised concerns about whether the sample mean is "necessarily a good basis for inference about the population mean." As they indicate, if the true underlying population distribution is known, alternatives, such as the use of Bayesian analysis may be superior. The practical issue with cost data, however, is that the true distribution is typically not known. Simulations suggest that in these circumstances, the sample mean generally remains the estimator of choice. [32,33]

II. B. Multivariable analysis of cost

Multivariable models are most important when the subjects in the treatment groups differ with respect to baseline characteristics [34]. Even when the randomization achieves balance in baseline characteristics, there remain several reasons why a multivariable model is justifiable. In fact, multivariable analysis of cost may be superior to univariate analysis because by explaining variation due to other causes it improves the power for tests of differences between groups [35–37]. It also facilitates subgroup analyses such as for participants with more and less severe disease; different countries/centers; etc.[38]. Finally, it accounts for potentially large and influential variations in economic conditions and practice patterns by provider, center, or country that may not be entirely balanced by randomization in the clinical trial [39, 34]. As with univariate tests of costs, identifying the most appropriate multivariate model for

the analysis of costs will depend on addressing the issues that arise from the skewed cost distribution. [20]

Ordinary least squares (OLS) regression on untransformed cost (which in the absence of other covariates is equivalent to the univariate t-test) has been the most frequently used model for multivariable analysis of cost in randomized clinical trial settings.[26] The advantages of OLS models are that they are easy to implement in any computer software and the incremental cost estimates are easy to calculate, because the coefficient for the treatment indicator provides a direct estimate of the adjusted difference in sample mean cost between the treatment groups. Yet OLS regression models are difficult to defend when analyzing cost data because these models results can be highly sensitive to extreme cases; they can be prone to overfitting; and the results may not be robust both in small to medium sized datasets and in large datasets with extreme observations.

One alternative that appears frequently in the literature is OLS regression predicting the log of cost. While Manning and Mullahy have identified a limited set of situations in which log OLS may be a preferred model,[340] in many cases the estimates and inferences derived for log cost will not be applicable to sample mean cost (e.g., when the distribution of log cost is not normal and when the variance in log cost differs between treatment groups). The problems related to differences in the variance arise primarily because the mean of the logs does not equal the log of the means.

Generalized Linear Models (GLMs) [341–42] are a class of models that can be used to analyze costs on the log scale while maintaining applicability to the sample mean cost[43,44]. The GLM can predict the log of mean costs directly which overcomes the problem faced by the OLS of log costs because with the GLM, a simple retransformation of the predicted results represents the sample mean while the transformation of the predicted results from an OLS of log costs does not. The GLM maintains applicability to the sample mean cost regardless of the scale from which costs are estimated. This is because the GLM covariates are always combined in a linear fashion. They are subsequently scaled with a link function which determines the scale in which the covariates are assumed to act on the outcome. For example, a log link assumes covariates act multiplicatively on the mean. In addition to the link function, a distribution function (or family) is needed to characterize a particular GLM model. The distribution function describes the shape of the distribution of the outcome and the relationship between its variance and mean. A GLM model analogous to the OLS has a identity link function and a normal distribution while a GLM model analogous to the log OLS has a log link and a normal distribution. The caution here, however, is that the GLM model is prone to misspecification and precision losses in datasets with cost distributions having a heavy-tailed log error term. [45]

III. CONCLUSION

The methods of cost valuation and analysis are not consistently applied. For cost valuation this is largely because many decisions in the design of the cost valuation methodology will depend on the specific context of the study. To strictly adhere to any single method would not yield more accurate or precise estimates of cost. What is important is to understand the tradeoffs involved when making design decisions and when evaluating cost valued and analyzed from a clinical trial. For cost analysis the inconsistency in the application of methods is largely due to the difficult problems that result from the highly skewed cost distribution. When the goals of economic analysis are budgetary or making socially optimal decisions, cost comparisons should involve sample means and the statistical tests applied to those means should be tests appropriate for means. Attention to the appropriate cost valuation and analysis techniques reviewed here will help bring the same level of rigor and attention to the methodological issues

in cost valuation as currently applied to clinical evidence within randomized trials. Yet cost analysis is typically insufficient when analyzing a clinical trial because costs of an intervention should be assessed in relation to its effectiveness through cost-effectiveness analysis. We note that there are unique complexities in the statistical assessment of cost-effectiveness in trials that have not been covered here.

References

1. Gold, MR.; Siegel, JE.; Russel, LB.; Weinstein, MC., editors. Cost-Effectiveness in Health and Medicine. Oxford: Oxford University Press; 1996.

2. Drummond, MF.; Sculpher, MJ.; Torrance, GW.; O'Brien, BJ.; Stoddart, GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2005.

3. Glick, HA.; Doshi, JA.; Sonnad, SS.; Polsky, D. Economic Evaluation in Clinical Trials. Oxford: Oxford University Press; 2007.

4. Finkler S. The Distinction Between Costs and Charges. Annals of Internal Medicine 1982;96:102–109. [PubMed: 7053682]

5. Medical Research Council Laparoscopic Groin Hernia Trial Group. Cost-utility analysis of open versus laparoscopic groin hernia repair: results from a multicentre randomized clinical trial.[comment]. British Journal of Surgery 2001;88(5):653–61. [PubMed: 11350435]

6. Fleming MF, Mundt MP, et al. Benefit-cost analysis of brief physician advice with problem drinkers in primary care settings.[comment]. Medical Care 2000;38(1):7–18. [PubMed: 10630716]

7. Halpern MT, Palmer CS, et al. Cost-effectiveness of gemifloxacin: results from the GLOBE study. American Journal of Health-System Pharmacy 2002;59(14):1357–65. [PubMed: 12132563]

8. Simon GE, Manning WG, et al. Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. Archives of General Psychiatry 2001;58(2):181–7. [PubMed: 11177120]

9. Beck EJ, Beecham J, Mandalia S, Griffith R, Walters MDS, Boulton M, Miller DL. What is the cost of getting the price wrong? J Public Health Medicine 1999;21:311–7.

10. Heerey A, McGowan B, Ryan M, Barry M. Microcosting versus DRGs in the provision of cost estimates for use in pharmacoeconomic evaluation. Expert Rev Pharmacoeconomics Outcomes Res 2002;2:29–33.

11. Shwartz M, Young DW, Siegrist R. The ratio of costs to charges: how good a basis for estimating costs? Inquiry 1995;32:476–81. [PubMed: 8567084]

12. Rittenhouse BE, Dulissec B, Stinnett AA. At what price significance? The effect of price estimates on statistical inference in economic evaluation. Health Econ 1999;8:213–9. [PubMed: 10348416]

13. Reed SD, Friedman JY, Gnanasakthy A, Schulman KA. Comparison of hospital costing methods in an economic evaluation of a multinational clinical trial. Int J Technology Asessment in Health Care 2003;19:395–406.

14. Chumney ECG, Biddle AK, Simpson KN, Weinberger M, Magruder KM, Zelman WN. The Effect of cost construction based on either DRG or ICD-9 codes or risk group stratification on the resulting cost-effectiveness ratios. Pharmacoeconomics 2004;22:1209–16. [PubMed: 15606227]

15. Taira DA, Seto TB, Siegrist R, Cosgrove R, Berezin R, Cohen DJ. Comparison of analytic approaches for the economic evaluation of new technologies alongside multicenter clinical trials. Am Heart J 2003;145:452–8. [PubMed: 12660668]

16. Goeree R, Gafni A, Hannah M, Myhr T, Blackhouse G. Hospital selection for unit cost estimates in multicentre economic evaluations. Does the choice of hospitals make a difference? Pharmacoeconomics 1999;15:561–572. [PubMed: 10538329]

17. Schulman K, Burke J, Drummond M, Davies L, Carlsson P, Gruger J, Harris A, Lucioni C, Gisbert R, Llana T, Tom E, Bloom B, Willke R, Glick H. Resource costing for multinational neurologic clinical trials: methods and results. Health Econ 1998;7:629–38. [PubMed: 9845256]

18. Glick HA, Orzol SM, Tooley JF, Polsky D, Mauskopf JO. Design and analysis of unit cost estimation studies: How many hospital diagnoses? How many countries? Health Economics 2003;12:517–27. [PubMed: 12825205]

19. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ 2000;320:1197–200. [PubMed: 10784550]

20. Ramsey S, Willke R, Briggs A, Brown R, Buxton M, Chawla A, Cook J, Glick H, Liljas B, Petitti D, Reed S. Best Practices for economic analysis alongside clinical trials: an ISPOR RCT-CEA Task Force report. Value in Health 2005;8:521–33. [PubMed: 16176491]

21. Hurley, J. An overview of normative economics of the health sector. In: Culyer, AJ.; Newhouse, JP., editors. Handbook of Health Economics. Vol. Chapter 2. Vol. 1A. Amsterdam: Elsevier Science B.V; 2000. p. 61

22. Duan N. Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc 1983;78:605–10.

23. Gayan AK. Significance of difference between the means of two non-normal samples. Biometrika 1950;37:399–408. [PubMed: 14801064]

24. Pearson ES, Please NW. Relation between the shape of population distribution and the robustness of four simple test statistics. Biometrika 1975;62:223–41.

25. Geary RC. Testing for normality. Biometrika 1947;36:353–369.

26. Doshi JA, Glick HA, Polsky DP. Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. Value in Health. In press.

27. Mullahy J. Much Ado Abut Two: Reconsidering Retransformation And The Two-Part Model In Health Econometrics. J Health Econ 1998;17(3):247–281. [PubMed: 10180918]

28. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. J Health Econ 1998;17(3):283–95. [PubMed: 10180919]

29. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med 2000;19:3219–36. [PubMed: 11113956]

30. Davison, AC.; Hinkley, DV. Bootstrap Methods and their Application. Cambridge: Cambridge University Press; 1997.

31. O'Hagan A, Stevens JW. Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? Health Econ 2003;12:33–49. [PubMed: 12483759]

32. Conigliani C, Tancredi A. Semi-parametric modelling for costs of health care technologies. Statistics in Medicine 2005;24(20):3171–84. [PubMed: 15568210]

33. Briggs A, Nixon R, Dixon S, Thompson S. Parametric modelling of cost data: some simulation evidence. Health Economics 2005;11:421–428. [PubMed: 15685641]

34. Gail MH, Wieand S, Pianados S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika 1984;71:431–444.

35. Senn SJ. Covariate imbalance and random allocation in clinical trials. Stat Med 1989;8:467–475. [PubMed: 2727470]

36. Vazquez-Polo FJ, Negrin Hernanez MA. Gonzalez Lopez-Valcarcel B. Using covariates to reduce uncertainty in the economic evaluation of Clinical Trail Data. Health Econ 2005;14:545–557. [PubMed: 15497202]

37. Newhous JM. Estimation efficiency with omitted covariates in generalized linear models. J Am Statist Assoc 1998;93(443):1124–1129.

38. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. Health Econ 2005;14:1217–1229. [PubMed: 15945043]

39. Newhous JM, Jewell NP. A geometric approach to assess bias due to omitted covariates in generalized linear models. Biometrika 1993;80:807–816.

40. Manning WG, Mullahy J. Estimating log models: To transform or not to transform? J Health Econ 2001;20:461–94. [PubMed: 11469231]

41. Nelder JA, Wedderburn RWM. Generalized linear models. J Roy Statist Soc A 1972;135:370–84.

42. McCullagh, P.; Nelder, JA. Generalized Linear Models. 2. London: Chapman & Hall; 1989.

43. Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. J Health Econ 1999;18:153–71. [PubMed: 10346351]

44. Barber J, Thompson S. Multiple regression of cost data: use of generalized linear models. J Health Serv Res Policy 2004;9:197–204. [PubMed: 15509405]

45. Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. Biostatistics 2005;6:93–109. [PubMed: 15618530]
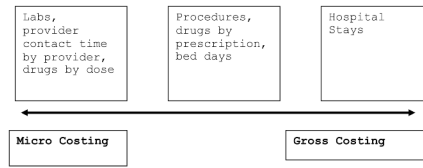
Figure 1.
Resource Costing Continuum for hospital stays