Pre-Treatment Health Related Quality of Life and Stratified Cost-Effectiveness Analysis

Joseph Schaafsma
Department of Economics
University of Victoria
Victoria, British Columbia, V8W 2Y2
Tel: (250)-721-8531
Fax: (250)-721-6214
Email: schaafsm@uvic.ca

December 2003

ABSTRACT

Introduction: This paper demonstrates that standard cost-effectiveness screening of pharmaceutical drugs may give misleading results if systematic data heterogeneity by base health related quality of life (BHRQoL), the HRQoL immediately prior to treatment, is ignored.

Study Design: It is argued that for some chronic diseases, such as rheumatoid arthritis, the magnitude of the HRQoL response to a specific pharmacotherapy may be negatively correlated with BHRQoL, and that the net cost of the pharmacotherapy may be positively correlated with BHRQoL. It is also argued that the value placed on a given increment in HRQoL declines as HRQoL rises. It is shown that if these factors are ignored cost-effectiveness analysis (CEA) based on sample averages will give misleading results, and that CEA applied to the data stratified by BHRQoL is indicated.

Methods: The case for sample stratification by BHRQoL, and for conducting cost-effectiveness analysis (CEA) at the stratum level, is illustrated with numerical examples using hypothetical data and the incremental net monetary benefit (INMB) criterion. Sensitivity analysis is used to explore the response of the INMB at the stratum level to different degrees of data heterogeneity across the strata.

Results: Due to cumulative effects, even relatively minor data heterogeneity by BHRQoL results in substantial differences in the cost-effectiveness of treatment across strata. CEA stratified by BHRQoL enables a more efficient and equitable allocation of public healthcare funding.

Keywords: Cost-effectiveness analysis, data heterogeneity, stratification, efficiency, equity

Running Title: Stratified cost-effectiveness analysis.
I. Introduction

The cost-effectiveness (CE) of a pharmacotherapy is commonly assessed by conducting a randomized clinical trial (RCT)\textsuperscript{[1]}. In this methodology patients are randomly assigned to either the drug or its comparator, and economic viability is determined by computing the incremental cost-effectiveness ratio (ICER) or, more recently, the incremental net monetary benefit (INMB)\textsuperscript{[2]}, using sample averages for the two arms of the trial. This standard methodology acknowledges that costs and benefits may vary across patients in the treatment group but assumes that such differences are randomly distributed so that sample averages produce an unbiased estimate of the treatment’s expected cost-effectiveness for a randomly selected patient. Confidence intervals can be constructed to determine the statistical significance of the presumed unbiased estimate\textsuperscript{[3-5]}.

This paper develops the case for stratifying cost-effectiveness analysis (CEA) by base health-related-quality-of-life (BHRQoL), the patient’s HRQoL just prior to treatment, when the objective of the pharmacotherapy is improved HRQoL. Stratified CEA is simply CEA applied to subsets of patients stratified (partitioned) by BHRQoL. The case for stratification rests on the argument that if the goal of a pharmacotherapy is improved HRQoL there are compelling reasons to expect that the drug’s cost-effectiveness will vary inversely with BHRQoL. CEA based on sample averages is thus not an unbiased estimate of the drug’s cost-effectiveness for a randomly selected patient, but understates/overstates cost-effectiveness if the patient’s BHRQoL is low/high.

CEA based on sample averages when the cost-effectiveness of a drug is in fact inversely related to BHRQoL will lead to incorrect policy decisions. For example, CEA
may indicate that a drug is not cost-effective and it may thus not be included in the provincial formulary for insured benefits. However, stratified CEA might identify a subset of patients for whom the treatment is cost-effective and for whom the drug should be an insured benefit. Conversely, CEA based on sample averages may indicate that a drug is cost-effective, and thus eligible for inclusion in the provincial formulary. However, had stratified CEA been conducted it is conceivable that treatment is found to be not cost-effective for an identifiable subset of patients, who should not have been funded. In short, stratified CEA assists in achieving equity (equal treatment of equals) and efficiency (funds are used to generate the greatest net benefit).

There are a number of compelling reasons for expecting that the cost-effectiveness of a pharmacotherapy aimed primarily at improved HRQoL varies inversely with the patient’s BHRQoL. First, the net cost of treatment likely is higher for patients with a higher BHRQoL. Second, the average HRQoL response to treatment likely declines as BHRQoL rises. Third, even if the first two reasons do not apply, it is conceivable that society’s willingness-to-pay (WTP) for an improvement in a patient’s HRQoL varies inversely with that patient’s BHRQoL. This, by itself, will generate an inverse relationship between a treatment’s cost-effectiveness, as measured by the INMB, and BHRQoL. The case for stratified CEA is developed in detail in the next section of this paper. The third section of the paper consists of a numerical simulation exercise that explores the sensitivity of stratified CEA results to the degree of heterogeneity across patients stratified by BHRQoL \(^6\). The paper ends with a brief conclusion and recommendation.
II. The Case for CEA Stratified by BHRQoL

A. BHRQoL and a Treatment’s Cost-Effectiveness

As noted, there are two reasons why a treatment’s cost-effectiveness likely declines as the patient’s BHRQoL rises. First, it is likely that, on average, the net treatment cost increases as BHRQoL rises. Second, the average HRQoL response to treatment likely declines as the patient’s BHRQoL rises. A rising net cost and a falling HRQoL response, as BHRQoL rises, implies a negative association between the treatment’s cost-effectiveness and the patient’s BHRQoL.

1. Net treatment cost varies positively with BHRQoL

The net cost of a pharmacotherapy is the sum of the direct and indirect treatment costs minus the costs that are no longer incurred as a result of improved health status. For simplicity it is assumed that the direct and indirect costs of treatment, while variable across patients, are uncorrelated with BHRQoL. Thus, a positive correlation between net treatment cost and BHRQoL must be due to variations in the costs no longer incurred as a result of the positive HRQoL response to treatment.

There are a number of reasons why costs eliminated as a result of improved HRQoL are larger the lower BHRQoL. Taking chronically ill rheumatoid arthritis (RA) patients as a case in point, the lower the BHRQoL the higher the use of healthcare goods and services such as GP and specialist visits, pain control drugs, therapy sessions, splints/casts/canes, and corrective surgery\textsuperscript{[7-10]}. Disease related non-healthcare costs are also negatively correlated with the RA patient’s BHRQoL. As the severity of RA increases, the need for homemaker services and for modifications to the home environment increases, as does the expense to facilitate independent living inside and outside the home\textsuperscript{[11]}. Lost productivity
also increases with the severity of RA. The lower the BHRQoL the less likely that the patient will be a productive member of society, either in the labor force or in a volunteer capacity \[^{[12]}\]. Improved patient productivity reduces the net cost of treatment. Clearly, when improved HRQoL is the goal of treatment and BHRQoL varies across patients, there are compelling reasons for expecting the net cost of successful treatment to vary positively with BHRQoL if the direct and indirect treatment costs are uncorrelated with BHRQoL.

2. *Average HRQoL response to treatment declines as the patient’s BHRQoL rises*

When the primary purpose of pharmacotherapy is improved HRQoL, such as in the treatment of RA or Crohn’s disease, it is conceivable that the lower the patients’ BHRQoL the larger the average positive BHRQoL response to treatment. For example, consider two sets of patients with BHRQoL of 0.2 and 0.6, respectively, on a scale from 0 (death) to 1 (perfect health), who receive the same pharmacotherapy to improve HRQoL. For the former, the potential improvement in HRQoL ranges from 0.0 to 0.8 and potential deterioration, due to adverse side effects, ranges from 0.0 to -0.2. For the latter, the potential improvement ranges from 0.0 to 0.4, and potential deterioration ranges from 0.0 to -0.6. Clearly, the scope for a large positive average response to treatment is greater for the group of patients with a BHRQoL of 0.2 than for the group with a BHRQoL of 0.6. Whether the former actually has a larger average positive response to treatment than the latter will depend on the specifics of the pharmacotherapy and the condition being treated. In the case of some chronic conditions, such as RA, where the HRQoL of patients receiving best supportive care (BSC) ranges from very low to moderate \[^{[13]}\] the hypothesis that the average HRQoL response to an effective pharmacotherapy varies inversely with BHRQoL is sufficiently plausible that it deserves to be considered in the CEA.
B. The Dollar Value of the increment in HRQoL.

In cost-utility analysis it is standard practice to add the changes in quality-adjusted-life-years (QALYs), attributable to treatment vis-à-vis placebo, across patients to compute the total gain in QALYs. The average incremental net cost of treatment is then divided by the average QALYs gained to compute the dollar cost per QALY gained. Treatment is considered cost-effective (not cost-effective) if the dollar cost per QALY is below (above) some cut-off. This technique implicitly assumes that the dollar value placed on a year of life varies proportionately with HRQoL, and implies that the marginal valuation of HRQoL, the value placed on a unit increment in HRQoL, is constant and independent of BHRQoL.

A constant marginal valuation of HRQoL, regardless of BHRQoL, is inconsistent with economic theory. Standard economic theory suggests that the marginal value of a good declines as the consumption of that good rises. This implies that as a patient’s BHRQoL rises the value placed on yet another unit increment in HRQoL declines, i.e. the marginal valuation of HRQoL varies inversely with BHRQoL. Arguably, a larger value is placed on, for example, a 0.1 increment in HRQoL if BHRQoL is 0.3 than if it is 0.7. If this is the case, QALYs cannot be summed across patients for policy analysis. Rather, the individual patient QALY gains should be converted to dollar values using the marginal valuations based on the patients’ BHRQoL. If the marginal valuation of HRQoL varies inversely with BHRQoL, it is conceivable that a treatment that generates the same increase in HRQoL for all patients, and has the same net cost for all, could be cost-effective for patients with a low BHRQoL but not for patients with a high BHRQoL. Thus, regardless
of whether or not the net cost of treatment varies directly with BHRQoL, stratified CEA is needed if the marginal valuation of HRQoL varies with BHRQoL.

C. Skewed Distribution of Patients by BHRQoL.

If, as postulated above, a drug’s benefits and net costs are correlated with the patient’s BHRQoL there is a third reason for stratified CEA. In any prospective random sample the distribution of patients by BHRQoL is generally not uniform but skewed to the left. For example, again using RA patients as a case in point, Young et al. \cite{13} report that in a prospective random sample of 732 patients 9.4% had marked functional loss at presentation, compared with normal function in 33% of the sample. Thus, a CEA based on sample averages may understate a drug’s cost-effectiveness for patients with a low BHRQoL not only because for this group the dollar value of the benefits are understated and net costs are overstated, but also because the calculation reflects the preponderance of the less favorable benefit and net cost data for patients with a higher BHRQoL. This problem can be largely circumvented with stratified CEA.\cite{2}

III. A NUMERICAL ILLUSTRATION WITH HYPOTHETICAL DATA

A. The Hypothetical Data

The concepts discussed above are illustrated with hypothetical data. Data for a base case (no systematic data heterogeneity by BHRQoL) and controlled deviations from this case (progressively more pronounced data heterogeneity by BHRQoL) are constructed. These data are then used to examine the sensitivity of cost-effectiveness results to the degree of systematic heterogeneity.

The hypothetical data is constructed for 600 patients. The distribution of these patients, by BHRQoL, is shown in columns (1) and (2) of Table 1 and is skewed towards
lower values of BHRQoL on the grounds that in any prospective sample the more severe cases will be under-represented. It is assumed that within each stratum patients are uniformly distributed across the interval.

Table 1 also depicts the four alternative patterns assumed for the relationship between the HRQoL response to treatment and BHRQoL. The first, Case (1), assumes that there is no relationship between the response to treatment and BHRQoL. Regardless of BHRQoL the patient’s gain in HRQoL (QG1) is randomly assigned, without replacement, from the set of QG1 values that are uniformly distributed across the interval 0 – 0.2.\(^3\) Thus, for each stratum the average gain is 0.1. Cases (2) – (4) incorporate an inverse relationship between the average gain in HRQoL and BHRQoL by using a wider interval for the gain when BHRQoL is low and reducing the size of the interval as BHRQoL increases. This is achieved by transforming QG1 into QG2, QG3, and QG4 using equations (1) – (3), respectively, in footnote (2) of Table 1. Case (2) assumes the steepest gradient in the HRQoL response to treatment as BHRQoL rises, Case (3) assumes a moderate gradient, and Case (4) assumes a minimal gradient.

Note that in the “steep” gradient case, Case (2), the average gain in HRQoL for patients in the lowest stratum is 32% larger than when there is no gradient at all, and the average gain for patients in the highest stratum is 31% smaller. These are not large differences across the same strata for the no gradient and “steep” gradient cases. However, comparing the gain across the lowest and highest strata within the “steep” gradient case, the HRQoL gain for the former is almost double the gain for the latter. Such differences are ignored in CEA based on sample averages, but are picked up by stratified CEA.
Table (2) shows the four patterns for the mean cost avoided as a result of improved HRQoL. In each of the four cases the cost of BSC is $6480 when HRQoL = 0.1. However, the rate at which this cost declines to 0 as HRQoL rises to 1 varies across the cases. In Case (1) it is assumes that this cost declines by $720 for each 0.1 gain in HRQoL. Cases (2) – (4) assume that, for a given increase in HRQoL, the cost of BSC declines by a larger amount when BHRQoL is low than when it is high. The cost equations for the four cases are shown in footnote (1), Table (2). The slope of the negative gradient between the cost of BSC and HRQoL declines in absolute terms across Cases (2), (3), and (4).

At the patient level, the cost avoided, as a result of improved HRQoL, is obtained by simulating the relevant cost equation, once with BHRQoL, and then with the patient’s improved HRQoL, and computing the difference. To keep the number of cases manageable, mean cost patterns were computed for only 4 of the 16 possible cases.⁴ In each of the 4 cases, matching underlying assumptions were used, e.g., the assumption of a steep decline in the HRQoL response to treatment as BHRQoL rises was combined with the assumption of a steep decline in the cost of BSC as HRQoL improves. Thus, the differences across the four cases in the mean cost avoided, given BHRQoL, are due not only to differences in how costs decline as HRQoL improves, but also to differences across the cases in the mean HRQoL response to treatment.

The differences in the marginal cost patterns implied by equations (1) – (4) in footnote (1), Table 2, are not large. For example, the decline in the cost of BSC, when HRQoL improves from 0.2 to 0.3, is $720, $990, $855 and $774, respectively. These are not large differences. However, combining the cost gradients with the gradients for the gain in HRQoL magnifies the dispersion in mean cost avoided across the four cases, for
any given stratum. Moreover, within a given case, the combination of the two gradients can have major implications at the margin. For example, in Case (2), mean costs avoided are $1203 and $345, respectively, for patients in the lowest and highest stratum. Again, CEA based on sample averages does not pick up such differences; stratified CEA analysis does.

Table (3) shows the four patterns for the mean WTP (the mean value placed on the improvement in the patients’ HRQoL as a result of treatment) by stratum. At the patient level, the WTP was computed with the relevant equation shown in footnote (1), Table 3. Two simulations were again needed; the first with the patient’s BHRQoL and the second with the post-treatment HRQoL. The value of the gain in HRQoL is the difference between the two simulated values. Each of the four cases assumes that a life in full health is valued at $70,000. The cases differ in how rapidly this value declines as HRQoL declines. In the constant marginal WTP case, Case (1), it declines by $7,000 for each 0.1 reduction in HRQoL. In Cases (2), (3) and (4), the decline in the value placed on a life year as HRQoL declines, starts off below $7,000 but ultimately rises above $7,000, from $3,500, $5,250 and $6300, respectively, when HRQoL drops from 1.0 to 0.9, to $10,500, $8,750 and $7,700, respectively, when HRQoL declines from 0.2 to 0.1.

The mean WTP values for a given case in Table 3 combine the WTP pattern specified by the relevant equation in footnote (1) of the table with the pattern for the HRQoL gain shown for the corresponding case in Table 1. The interaction of these two gradients for a given case again has major implications for what happens at the margin. For example, for Case (2), the marginal WTP for a 0.1 increment in HRQoL is $9,625 at BHRQoL = 0.2 compared to $5,250 at BHRQoL = 0.7, a difference of $4,375. The
difference between the mean WTP for the realized gain in HRQoL between the lowest and highest strata is $8,344. This much larger difference captures not only the difference in the marginal valuation of a given increase in HRQoL, but also the larger HRQoL response to treatment when BHRQoL is low.

It is clear from Table 3 that even if a minimal gradient is assumed for the inverse relationship between the gain in HRQoL and BHRQoL, and between marginal WTP and BHRQoL, there are major cost-effectiveness implications. For example, the mean dollar value placed on the response to treatment is $1,739 larger for patients in the lowest stratum than for patients in the highest stratum. CEA based on sample averages ignores such differences; stratified CEA does not.

Finally, it is assumed that the sum of direct and indirect costs of treatment is $7,200.

**B. The Incremental Net Monetary Benefit and Sensitivity Analysis**

In this section the hypothetical data developed above are used to illustrate how stratified CEA gives much richer results than CEA based on sample averages. The illustration uses the INMB criterion. The INMB is the dollar value placed on the gain in HRQoL plus the costs eliminated as a result of improved quality of life minus the direct and indirect costs of treatment. The INMB, rather than the ICER, is used since it allows one to convert individual gains in HRQoL to dollar values before aggregating and averaging. Additional advantages of INMB over ICER are outlined by Hoch, Briggs and Willan \[^2\]. It is assumed that costs and benefits repeat annually, but for simplicity the illustrations use one year’s data only. Discounting future flows can be incorporated in the analysis, but would unnecessarily complicate the presentation.
The INMB results, using sample averages and stratified data, are shown in Table 4. The computations were conducted for the four cases in Tables (1) – (3), and for a fifth case where a different gradient is used for each of the underlying relationships. Standard errors are shown in parentheses under the INMB values. Two sources of variation give rise to these standard errors. First, as discussed above, within each stratum patients have different BHRQoL and vary in their response to treatment. This creates differences across patients in the value of the gain in HRQoL and in costs avoided. These differences are accentuated in Cases (2) – (4) by the non-linearities built into the equations for these cases. Second, when the INMB is calculated for a patient a random error term, drawn from a normal distribution with a mean of 0 and standard deviation of 4000, is included to capture the fact that costs avoided at the patient level are not an exact function of the gain in HRQoL.

Case 1: Identical mean gains in HRQoL, constant MWTP, constant MC avoided.

When the mean gain in HRQoL is identical across strata, and the MWTP and MC avoided are constant, i.e., Case (1) Table 4, stratification is irrelevant and the INMB is $520 regardless of whether it is based on sample averages or on averages within strata. The standard errors indicate that the INMB estimate is statistically significant at the 1% level when it is based on sample averages, but not when it is based on strata averages. The latter is simply the result of the smaller sample sizes at the strata level. By construction, the population distributions are the same across the strata and efficient estimation should use the pooled (full sample) data.

Case 2: Pronounced Negative Gradient for each of the Three Underlying Relationships

Case (2) uses the most pronounced negative gradient in our analysis for each of the three relationships underlying the data: the inverse relationship to BHRQoL of the mean
response to treatment, marginal WTP, and marginal cost avoided. As was shown in Tables (2) and (3), cumulatively these gradients imply substantial variation across strata in costs avoided and the dollar value of the gain in HRQoL. The INMB values for Case (2) reflect these differences and range from $5698 for the lowest stratum to – $3504 for the highest stratum, a difference of $9202. At the full sample level, the INMB is – $339 and not statistically significant. Thus, CEA based on sample averages would lead to the conclusion that treatment is not cost-effective. However, stratified CEA, being sensitive to differences at the margin, indicates that treatment is cost-effective for the first three strata but not for the last three.

The results for Case (2) also illustrate how the skewed distribution of patients by BHRQoL in the full sample affects the sample results. Had patients been uniformly distributed across the six strata, the INMB for the full sample would have been a simple average of the six INMB values at the strata level, i.e., $733. However, because a large proportion of the patients are in the upper strata, the negative INMB values for these strata are weighted more heavily in the CEA and the full sample INMB is – $339. Thus, if stratified CEA is indicated by systematic data heterogeneity by BHRQoL, a reinforcing factor for conducting stratified CEA is the skewed distribution of patients by BHRQoL.

**Case (3) and (4): Moderate and Minimal Negative Gradients**

Cases (3) and (4), Table (4), show that as the gradient for the three relationships underlying the data becomes flatter, the gradient of the INMB values across the strata also flattens, and the INMB value based on the full sample averages rises. However, even with the minimal gradient (for which the average marginal effects for patients in the lowest and highest strata differ from the constant marginal effects by about 6% and -6%, respectively)
the INMB values for the lowest and highest strata differ by $1918. Thus, even very minor heterogeneity in the data by BHRQoL can result in substantial variation in INMB values across strata.

Reducing the slope of the gradients underlying the data reduces the statistical significance of the INMB values for the three lowest strata: the standard error declines due to less variability in the data as the slope of the gradient flattens, but mean INMB declines even more. When the minimal gradient is used, Case (4), none of the INMB values is statistically significant. However, if the results for the lowest three strata are pooled the mean INMB value, $1010, is statistically significant (SE = $429).

Case (5): Mixed Negative Gradients for the Three Underlying Relationships

Case (5) uses a different gradient for each of the three data generating relationships. The minimal gradient is specified for the relationship between the HRQoL response to treatment and BHRQoL (equation (3) in footnote (1), Table 1) on the grounds that this relationship is probably the most difficult to substantiate on “a priori” reasoning. Recall that for the minimal gradient case (Case (4), Table 1), the mean gain in HRQoL is 6.3% larger for the lowest stratum, and 6.2% smaller for the highest stratum relative to the no gradient case. Thus, the minimal gradient represents a very minor deviation from the case where the response to treatment is uncorrelated with BHRQoL.

The moderate gradient is used for the relationship between the marginal WTP and BHRQoL (Equation (3), footnote (1), Table 3). This implies that for HRQoL equal to 0.2 and 0.7, the marginal WTP for a 0.1 increase in HRQoL is $8,313 and $6,125, respectively, compared to $7,000 when the marginal WTP is uncorrelated with HRQoL. If the principle
of declining marginal WTP is accepted the moderate gradient is likely a very weak expression of it.

The inverse relationship between the cost of BSC and HRQoL is the easiest to argue on “a priori” grounds, and thus the most pronounced gradient is used for it (equation (2), footnote (1), Table 2). However, even here the gradient is not all that steep. Compared to the constant marginal-cost-avoided case ($720 for a 0.1 unit increase in HRQoL, regardless of BHRQoL), the cost avoided for a 0.1 unit increase in HRQoL is $990 and $540, for BHRQoL 0.2 and 0.7, respectively. In dollar terms, these are relatively small deviations from the constant marginal cost case.

The last row in Table (4) shows the mean INMB values for the mixed gradient case. For the full sample the INMB is positive but small and not statistically significant. If the CEA were to stop at this level the conclusion would be that the drug is not cost-effective. However, at the strata level, the INMB for the two lowest strata is positive and statistically significant. The drug is thus clearly cost-effective for these patients. For the two upper strata the INMB is negative, the drug is thus clearly not cost-effective for BHRQoL > 0.6. For the two middle strata the INMB is positive but not statistically significant. Pooling the results for the two middle strata does not change this conclusion (INMB = $459, S.E. = $429). However, splitting the third stratum down the middle, and computing the INMB for each half indicates that treatment is cost-effective for $0.4 \leq \text{BHRQoL} < 0.45$ (INMB = $1463, S.E. = $880, p = 0.052), and not cost-effective for $0.45 \leq \text{BHRQoL} < 0.5$ (INMB = $167, S.E. = $958). Thus, for what could be considered a reasonable approximation of real data, it has been shown that CEA based on sample averages can indicate that a drug is not
cost-effective, while stratified CEA indicates that the drug is cost-effective for the subset of patients with a BHRQoL < 0.45.

IV. Discussion

Patient heterogeneity and its relevance for CEA are beginning to receive attention in the literature [2, 14-16]. However, this paper is the first to make the case for stratification by BHRQoL. Furthermore, this paper adds to the standard case for stratification (based on systematic differences in either net treatment costs, or the response to treatment, or both) the two additional factors of declining marginal WTP and a skewed distribution by BHRQoL of the patients in the sample. Additionally, this paper strengthens the case for stratification by demonstrating that even relatively minor systematic heterogeneity can cumulatively result in substantial differences in the cost-effectiveness of treatment across strata. Finally, the proposed stratification criterion, BHRQoL, is likely to be more acceptable to policy makers than stratification criteria such as the patient’s age [2, 15-16], race [2], or sex, on the grounds that it is consistent with directing healthcare funding to those who need it most.

Stratified CEA with real RCT data will require that the data for both arms of the trial are stratified by BHRQoL. The HRQoL response to treatment, and the net cost of treatment, can then be computed at the stratum level as the difference between means across the same stratum in the control and treatment arms. This raises the important issue of whether stratification requires larger samples to get a sufficient number of observations within each sub-sample to support the desired confidence in the statistical results. The cost of a clinical trial rises with sample size, and this higher cost is an offset to efficiency and equity gains from stratification. Whether or not stratification requires a larger sample is an
empirical issue. Stratification removes one source of variation in the data and can result in higher levels of statistical significance in tests at the stratum level than at the full sample level, despite the smaller number of observations at the stratum level. In fact, it is conceivable that under certain circumstances stratification allows the researcher to reduce the overall sample size. This would occur if there is very little variability in the data within strata but substantial variation across strata.

Whether or not stratification by BHRQoL is indicated can be checked by separately regressing individual net treatment costs, and individual gains in HRQoL, on BHRQoL. The larger the absolute value of the coefficients on the BHRQoL variable in these equations, and the tighter the fit, the more compelling the case for stratified CEA. Minimally, when CEA is based on sample averages, these regressions should routinely be conducted and reported to provide valuable context for assessing the likelihood that there is a systematic bias in the cost-effectiveness results based on sample averages.

The need to check for structure in data heterogeneity prior to CEA increases with the direct cost of the drug and with the variability in the patients’ BHRQoL. The higher the cost of a drug the less likely that it will be cost-effective for a diverse sample of patients and the greater the need to evaluate its cost-effectiveness by sample subsets. The greater the variability in BHRQoL the less homogeneous the patient sample, and thus the more misleading the full sample INMB value may be. These are important issues when considering some of the new, expensive pharmacotherapies for chronic diseases such as RA and Crohn’s disease. These drugs can cost in excess of $16,000 CDN annually\textsuperscript{[17]}, and base HRQoL can vary substantially across patients\textsuperscript{[13]}. However, even if net treatment costs and the HRQoL response to treatment show no systematic relationship to BHRQoL,
Stratified cost-effectiveness analysis

stratified CEA is nevertheless indicated if the marginal WTP for improved HRQoL declines with BHRQoL.

In conclusion, it should be noted that CEA based on sample averages essentially attempts to test two hypotheses simultaneously with one set of results. The explicit null hypothesis is that the drug is not cost-effective. The second, implicit, hypothesis is that the criteria for including a patient in the trial have been set correctly. If the RCT fails to reject the explicit null hypothesis, it is not clear whether this is because the pharmacotherapy is not cost-effective, or whether the criteria for inclusion in the trial were set incorrectly. At best, failure to reject the null hypothesis is a conditional decision, i.e. the treatment is not cost-effective given the selected mix of patient characteristics. CEA analysis stratified by BHRQoL tests not only whether a drug is cost-effective, but also for whom (by BHRQoL) it is cost-effective.

V. CONCLUSION

When there is systematic heterogeneity in the sample data by BHRQoL, CEA based on sample averages systematically understates/overstates the cost-effectiveness of treatment for identifiable subsets of patients in the sample, and results in an inefficient and inequitable allocation of the healthcare budget. This problem can be mitigated if the data are stratified by BHRQoL and CEA is conducted at the stratum level. Stratification is indicated if the marginal WTP for improved HRQoL declines as HRQoL rises, or if the regression of individual gains in HRQoL, and/or of individual net treatment costs, on BHRQoL yields statistically significant results. Due to cumulative effects, even relatively minor systematic heterogeneity across patients can result in substantial differences in the cost-effectiveness of treatment across the strata. Furthermore, unlike CEA based on
sample averages, stratification by BHRQoL allows one to test not only whether the drug is cost-effective, but also for whom it is cost-effective.
Footnotes

1 If one year of life in full health is valued at $70,000, one year of life at HRQoL = 0.3 is valued at $21,000, or one year of life at HRQoL = 0.6 is valued at $42,000.

2 The problem will still exist within each stratum.

3 Within an interval the QG1 values are equidistant from each other and the number of values equals the number of patients in that interval.

4 The four patterns for the response to treatment in Table (1), and the four cost equations underlying Table (2), generate 16 mean cost avoided patterns.
References


Table 1. The Distribution of Patients by BHRQoL and the Four Patterns for Mean Improvement in HRQoL

<table>
<thead>
<tr>
<th>BHRQoL</th>
<th>Number of patients</th>
<th>Mean BHRQoL</th>
<th>Mean gain in HRQoL attributable to treatment when the Gradient between the Gain in HRQoL and BHRQoL is:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (1) Flat$^1$</td>
</tr>
<tr>
<td>0.2 – 0.299</td>
<td>50</td>
<td>0.25</td>
<td>0.1</td>
</tr>
<tr>
<td>0.3 – 0.399</td>
<td>70</td>
<td>0.35</td>
<td>0.1</td>
</tr>
<tr>
<td>0.4 – 0.499</td>
<td>90</td>
<td>0.45</td>
<td>0.1</td>
</tr>
<tr>
<td>0.5 – 0.599</td>
<td>110</td>
<td>0.55</td>
<td>0.1</td>
</tr>
<tr>
<td>0.6 – 0.699</td>
<td>130</td>
<td>0.65</td>
<td>0.1</td>
</tr>
<tr>
<td>0.7 – 0.799</td>
<td>150</td>
<td>0.75</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>0.2 – 0.799</strong></td>
<td><strong>600</strong></td>
<td><strong>0.56</strong></td>
<td><strong>0.1</strong></td>
</tr>
</tbody>
</table>

1 Individual patient HRQoL gains (QG1) were randomly assigned from a pre-determined set that were equidistant from each other over the interval 0 – 0.2 (see text).

2 Individual patient HRQoL gains were computed as follows:

\[
\text{Case (2)} \quad QG2 = (1.6250 - 1.250 \times \text{BHRQoL}) \times QG1
\]

\[
\text{Case (3)} \quad QG3 = (1.3125 - 0.625 \times \text{BHRQoL}) \times QG1
\]

\[
\text{Case (4)} \quad QG4 = (1.1250 - 0.250 \times \text{BHRQoL}) \times QG1
\]

where QG1 is the patient’s HRQoL response to treatment from Case (1) (see text).
Table 2. The Distribution of Patients by BHRQoL and the Four Patterns for Mean Costs Avoided as a Result of Treatment.

<table>
<thead>
<tr>
<th>BHRQoL</th>
<th>Number of patients</th>
<th>Mean BHRQoL</th>
<th>Case (1) Constant&lt;sup&gt;1,2&lt;/sup&gt;</th>
<th>Case (2) “rapidly” declining&lt;sup&gt;1,2&lt;/sup&gt;</th>
<th>Case (3) “moderately” declining&lt;sup&gt;1,2&lt;/sup&gt;</th>
<th>Case (4) “minimally” declining&lt;sup&gt;1,2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 – 0.299</td>
<td>50</td>
<td>0.25</td>
<td>720</td>
<td>1203</td>
<td>951</td>
<td>810</td>
</tr>
<tr>
<td>0.3 – 0.399</td>
<td>70</td>
<td>0.35</td>
<td>720</td>
<td>989</td>
<td>852</td>
<td>772</td>
</tr>
<tr>
<td>0.4 – 0.499</td>
<td>90</td>
<td>0.45</td>
<td>720</td>
<td>799</td>
<td>760</td>
<td>736</td>
</tr>
<tr>
<td>0.5 – 0.599</td>
<td>110</td>
<td>0.55</td>
<td>720</td>
<td>628</td>
<td>672</td>
<td>700</td>
</tr>
<tr>
<td>0.6 – 0.699</td>
<td>130</td>
<td>0.65</td>
<td>720</td>
<td>476</td>
<td>589</td>
<td>665</td>
</tr>
<tr>
<td>0.7 – 0.799</td>
<td>150</td>
<td>0.75</td>
<td>720</td>
<td>345</td>
<td>511</td>
<td>631</td>
</tr>
<tr>
<td>0.2 – 0.799</td>
<td>600</td>
<td>0.56</td>
<td>720</td>
<td>640</td>
<td>671</td>
<td>698</td>
</tr>
</tbody>
</table>

<sup>1</sup> At the individual patient level, the total cost of BSC, given the patient’s HRQoL, was computed as follows:  
Case (1) \[ TC_1 = 7200 \times (1 - \text{HRQoL}) \]  
Case (2) \[ TC_2 = 7,650 - 12,150 \times \text{HRQoL} + 4,500 \times \text{HRQoL}^2 \]  
Case (3) \[ TC_3 = 7,425 - 9,675 \times \text{HRQoL} + 2,250 \times \text{HRQoL}^2 \]  
Case (4) \[ TC_4 = 7,290 - 8,190 \times \text{HRQoL} + 900 \times \text{HRQoL}^2 \]  

<sup>2</sup> At the individual patient level, the cost saved is obtained by twice simulating the cost equation, once with the patient’s BHRQoL and then with post-treatment HRQoL, and computing the difference.
Table 3. The Distribution of Patients by BHRQoL and the Four Patterns for Mean Willingness-to-Pay for the Gain in HRQoL.

<table>
<thead>
<tr>
<th>BHRQoL</th>
<th>Number of patients</th>
<th>Mean BHRQoL</th>
<th>Mean WTP (in $’s) for the Gain in HRQoL when the Marginal Gain in HRQoL and Marginal WTP are both:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (1) Constant^{1,2}</td>
</tr>
<tr>
<td>0.2 – 0.299</td>
<td>50</td>
<td>0.25</td>
<td>7,000</td>
</tr>
<tr>
<td>0.3 – 0.399</td>
<td>70</td>
<td>0.35</td>
<td>7,000</td>
</tr>
<tr>
<td>0.4 – 0.499</td>
<td>90</td>
<td>0.45</td>
<td>7,000</td>
</tr>
<tr>
<td>0.5 – 0.599</td>
<td>110</td>
<td>0.55</td>
<td>7,000</td>
</tr>
<tr>
<td>0.6 – 0.699</td>
<td>130</td>
<td>0.65</td>
<td>7,000</td>
</tr>
<tr>
<td>0.7 – 0.799</td>
<td>150</td>
<td>0.75</td>
<td>7,000</td>
</tr>
<tr>
<td>0.2 – 0.799</td>
<td>600</td>
<td>0.56</td>
<td>7,000</td>
</tr>
</tbody>
</table>

^{1} At the individual patient level, the value placed on the patient’s HRQoL was computed as follows: Case (1) \( V_1 = 70,000 \times \text{HRQoL} \) (1)  
Case (2) \( V_2 = -4,375 + 118,125 \times \text{HRQoL} - 43,750 \times \text{HRQoL}^2 \) (2)  
Case (3) \( V_3 = -2,187.5 + 94,063 \times \text{HRQoL} - 21,875 \times \text{HRQoL}^2 \) (3)  
Case (4) \( V_4 = -875 + 79,625 \times \text{HRQoL} - 8,750 \times \text{HRQoL}^2 \) (4)

^{2} At the individual patient level, the WTP for the gain in HRQoL is obtained by twice simulating the relevant WTP equation in footnote (1), once with the patient’s BHRQoL and then with post-treatment HRQoL, and computing the difference.
Table 4. INMB Simulation Results Based on Various Assumptions about Disease Costs Eliminated, WTP, and Patient Distribution by BHRQoL. Direct and Indirect Costs of Treatment = $7,200.

<table>
<thead>
<tr>
<th>Case</th>
<th>Assumed relationship with BHRQoL</th>
<th>Incremental Net Monetary Benefit (standard errors in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean HRQoL Gain</td>
<td>Marginal WTP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>no relationship</td>
<td>no relationship</td>
</tr>
<tr>
<td>2</td>
<td>pronounced negative gradient</td>
<td>pronounced negative gradient</td>
</tr>
<tr>
<td>3</td>
<td>moderate negative gradient</td>
<td>moderate negative gradient</td>
</tr>
<tr>
<td>4</td>
<td>minimal negative gradient</td>
<td>minimal negative gradient</td>
</tr>
<tr>
<td>5</td>
<td>minimal negative gradient</td>
<td>moderate negative gradient</td>
</tr>
</tbody>
</table>

* Statistically significant at the 5% level in a one-tailed test.
# Statistically significant at the 1% level in a one-tailed test.
Footnotes

1 If one year of life in full health is valued at $70,000, one year of life at HRQoL = 0.3 is valued at $21,000, or one year of life at HRQoL = 0.6 is valued at $42,000.
2 The problem will still exist within each stratum.
3 Within an interval the QG1 values are equidistant from each other and the number of values equals the number of patients in that interval.
4 The four patterns for the response to treatment in Table (1), and the four cost equations underlying Table (2), generate 16 mean cost avoided patterns.