CLINICAL OUTCOMES ASSESSMENT

Adjusting for Patient Crossover in Clinical Trials Using External Data: A Case Study of Lenalidomide for Advanced Multiple Myeloma

K. Jack Ishak, PhD1,*, J. Jaime Caro, MD2, Mark T. Drayson, MBChB, PhD3, Meletios Dimopoulos, MD4, Donna Weber, MD5, Bradley Augustson, FRACP6, J. Anthony Child, MD7, Robert Knight, MD8, Gulnaz Iqbal, BSc9, Janet Dunn, PhD9, Arran Shearer, MSc8, Gareth Morgan, PhD, FRCP10

1United BioSource Corporation, Montreal, QC, Canada; 2United BioSource Corporation, Lexington, MA, USA; 3University of Birmingham, Birmingham, UK; 4Department of Clinical Therapeutics, University of Athens School of Medicine, Athens, Greece; 5Department of Lymphoma and Myeloma, M. D. Anderson Cancer Center, Austin, TX, USA; 6Sir Charles Gairdner Hospital, Nedlands, Australia; 7University of Leeds, Leeds, UK; 8Celgene UK and Ireland, UK; 9Warwick Medical School, UK; 10The Institute of Cancer Research, UK

ABSTRACT

Objectives: In some trials, particularly in oncology, patients whose disease progressions under the comparator treatment are crossed over into the experimental arm. This unplanned crossover can introduce bias in analyses because patients who crossover likely have a different prognosis than those who do not cross over; for instance, sicker patients not responding to standard therapy or those expected to benefit the most may be selectively chosen to receive the experimental treatment. Standard statistical methods cannot adequately correct for this bias. We describe an approach designed to minimize the impact of crossover, and illustrate this by using data from two randomized trials in multiple myeloma (MM).

Methods: The MM-009/010 trials compared lenalidomide and high-dose dexamethasone (Len + Dex) with dexamethasone alone (Dex). Nearly half (47%) of the patients randomized to Dex crossed over to Len with or without Dex (Len +/−Dex) at disease progression or study unblinding. Data from these trials was used to predict survival in an economic model evaluating the cost-effectiveness of lenalidomide. To adjust for crossover, the prediction equations were calibrated to match survival with Dex or Dex-equivalent therapies in trials conducted by the Medical Research Council (MRC) in the United Kingdom. To adjust for differences between the MM and MRC trial populations, a prediction equation was developed from the MRC data and used to predict survival by setting predictors to mean values for patients in the MM-009/010 trials. The expected survival with Dex without crossover was then predicted from the calibrated MM-009/010 equation (i.e., adjusted to match survival predicted from the MRC equation).

Results: The adjusted median overall survival predicted by the MRC equation was 19.5 months (95%CI, 16.6–22.9) for patients with one prior therapy, and 11.6 months (95% CI, 9.5–14.2) for patients with >1 prior therapy. These estimates are considerably shorter than was observed in the clinical trials: 33.6 months (27.1–NE) and 27.3 months (95% CI, 23.3–33.3) as of December 2005.

Conclusion: The calibration method described here is simple to implement, provided that suitable data are available; it can be implemented with other types of endpoints in any therapeutic area.

Keywords: bias, crossover, lenalidomide, multiple myeloma, survival, unexplained.

Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Crossover from one clinical trial study arm to another can occur because one arm is perceived to be better than another or in therapeutic areas where patients’ condition can change suddenly and require use of alternate therapy. This was noted in particular in studies of surgical interventions, including coronary artery bypass grafting (CABG) [1], where 38% of patients randomized to medical therapy received CABG during the course of a study. Similarly, patients randomized to undergo CABG sometimes refused surgery and instead were treated with medications [2]. This contamination of study arms leads to mixing of the effects and obscures the impact of the intervention being studied; furthermore, it can introduce more complex selection biases in the analyses of the study data because crossover is inherently related to patients’ condition and prognosis with the original treatment received.

Several approaches to handle crossovers have been considered, including restricting analysis to patients who adhered to their assigned therapy, grouping patients based on treatment received, censoring follow-up at crossover, transitioning patients to the group to which they crossed over (by changing their treatment group indicator when crossover occurs), and an intention-to-treat (ITT) analysis that groups patients as randomized [3]. There is no perfect solution to dealing with crossovers, although an ITT analysis has been recommended as the preferred approach, at least in cases where the number of crossovers is not excessive [4]. An ITT analysis preserves the baseline comparability of groups given by randomization, albeit at the cost of altering the interpretation of the estimated effect to encompass the potential impact of crossovers [3].

* Address correspondence to: K. Jack Ishak, United BioSource Corporation, 185 Dorval Avenue, Suite 500, Dorval, QC H9S 5J9 Canada. E-mail: jack.ishak@unitedbiosource.com.
1098-3015/$36.00 – see front matter Copyright © 2011, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc. doi:10.1016/j.jval.2011.02.1182
Even when crossovers are well documented, methods based on excluding patients, changing treatment group status, or censoring data can induce serious bias. For instance, excluding patients would likely introduce selection bias because these patients may have a different prognosis than those who did not crossover. By the same token, censoring could no longer be assumed to occur at random, as it is linked to crossover. These biases not only affect efficacy analyses, they also complicate the use of the trial data to inform other aspects, such as meta-analysis [5] or health-economic assessments [6].

More complex statistical methods have been proposed to deal with crossovers. Rank-preserving structural failure time models [7,8] and marginal structural models (or inverse probability weighting) [9] have been described as possible approaches to dealing with non-random treatment assignment and non-compliance, of which crossover is a specific form. These methods use specialized estimation techniques to re-establish randomization between treatment groups to derive an unbiased treatment effect estimate. Whereas these methods rely entirely on the data from the trial to correct the bias, we describe an alternative method that uses external information to minimize the impact of crossover in economic evaluations using predictive equations for clinical endpoints. External data in this approach provides a reference value for endpoints in the absence of crossover.

We describe and illustrate the method with an application in a pooled analysis of two randomized, phase III clinical trials in patients who have received at least one prior therapy for multiple myeloma (MM).

Methods

Case study

In two randomized clinical trials (MM-009/010) lenalidomide plus high-dose dexamethasone was compared with dexamethasone alone (Dex) in patients with MM who had failed at least one prior therapy [10–12]. Data from these trials were used in an economic evaluation of lenalidomide using a discrete event simulation (which was part of a submission to National Institute for Health and Clinical Excellence [NICE] in the United Kingdom). These analyses were based on time to progression (TTP) and post-progression survival (PPS), which, when combined, yielded the overall survival (OS) for patients who progress.

The simulation used predictive equations for TTP and PPS derived from the MM-009/010 trial data. The equations were obtained via parametric survival analyses and included assigned treatment, baseline characteristics, and best response achieved as predictors.

In the trials, particularly at unblinding (which occurred at disease progression in most cases), 47% of patients receiving dexamethasone alone crossed over to the lenalidomide arm. The crossover only occurred from the Dex arm. This high rate of crossovers led to an overestimate of survival in the dexamethasone group because it mixes in the benefits of lenalidomide. Thus, even if the prediction equations for TTP and PPS derived from the trial fit the observed data very closely, they would not reflect the true effect of adding lenalidomide to dexamethasone compared with dexamethasone alone.

The aim was, therefore, to adjust for the impact of crossover on the overall survival distribution predicted from the MM-009/010 equations for patients receiving Dex to match an external unbiased reference value (e.g., median survival time) – that is, an estimated of the survival in a similar population where outcomes are not affected by crossover. This is done by adding a calibration term (i.e., a coefficient that modifies the intercept of the equation) to the MM-009/010 equations and estimating the value that produces predictions that match a reference time (e.g., median) reflecting survival in the absence of crossover.

This reference value was derived by analyzing data from four trials conducted by the UK Medical Research Council (MRC) between 1980 and 2002 [13–17]. The MRC trials dataset comprised 2,942 patients starting treatment for MM with a median follow-up of 5.7 years. First, an equation was developed for overall survival as a function of predictors selected from various patient characteristics. Second, this equation was used to estimate the survival of the patients randomized to dexamethasone in the MM-009/010 trials by setting predictors in the MRC equation to their corresponding values for Dex patients. The estimated median survival was used as a reference value to calibrate the MM-009/010 equations. A summary of the method is illustrated in Figure 1.

Analysis of the MRC trials

Patients from the MRC trials were selected to match the lenalidomide trials’ inclusion criteria:

- All patients who received only first-line treatment in the MRC trials were excluded because the lenalidomide trials required prior failure of at least one treatment.
- All patients having failed an initial treatment and beginning a second one were included as the “one-prior therapy” subgroup.
- All patients having failed more than one treatment and beginning a third- or fourth-line treatment were included as the “multiple prior therapies” subgroup.

Variables reflecting patients’ characteristics were selected to match those available from the MM-009/010 trials as closely as possible. These included age, sex, Durie-Salmon disease stage, presence of lytic bone lesions (at start of the first line of treatment), performance status, maximum response, M-protein (g/L), and beta-2 microglobulin levels (dichotomized at 2.5 mg/L), both at start of first line, and at progression with each treatment. Best response achieved was also considered.

In the MRC trials received various medications, such as melphalan (M7), melphalan and prednisone (MP), ABCM (adriamycin, BCNU, cyclophosphamide, and melphalan), cyclophosphamide, VAMP (vincristine, Adriamycin, and methylprednisolone), VAD (vincristine, Adriamycin, and dexamethasone), or HD (dexamethasone and prednisone) among others. Ideally, only data from patients receiving dexamethasone monotherapy would be included in the analyses, but this would have yielded a relatively small sample. To fully leverage the data, all patients were considered and a log-rank test was performed to assess whether survival differed among regimens containing dexamethasone and those not using this drug. Calendar trends were
tested to ensure that survival from the older trials could be pooled with the more recent studies and applied to the MM trials. This was done by comparing overall survival by year of enrollment starting from the first line of treatment received in the MRC trials for all patients for whom survival data was available. Restricting this assessment to patients who fail first-line treatment (i.e., the subset used in the main analyses) would be subject to selection bias because starting second line treatment in later years may be associated with better response to first line treatment.

OS equations were derived from the MRC data separately for one-prior and multiple-prior subgroups. Distributions commonly used to fit survival data (exponential, Weibull, log-normal, log logistic) were tested and the one that best fit the observed data was selected. Best fit was assessed based on the log-likelihood statistic and visual inspection of the observed and predicted survival distribution with each of the functions. Potential predictors were first identified by univariate equation where each variable is included on its own. Significant predictors were then included together into a multiple regression equation. The final equation included only predictors that were statistically significant (P < 0.10) in the multiple regression analysis. For categorical variables, it is possible that the coefficients for some levels are statistically significant, even though others are not. In such cases, levels with non-significant coefficients were combined with adjacent levels. For example, in the equation for the multiple prior group, Eastern Cooperative Oncology Group (ECOG) levels 0 and 1 did not differ significantly, and were combined, so that the equation only has a coefficient for 2–3 versus 0–1. The resulting equations were used to estimate median OS times for the patients randomized to dexamethasone in the MM-009/010 trials by setting the predictors to the corresponding mean values (proportions for categorical variables) derived from the pooled MM-009/010. These median times were used as the reference values for calibration.

**Calibration of MM-009/010 equations**

Analyses of the MM-009/010 trials yielded three equations: two for TTP (for one- and for multiple-prior therapies) and a third for PPS. Separate PPS equations for patients with one and multiple prior therapies were not derived due to limited number of deaths. Instead, number of prior therapies (one vs. multiple) was included as a predictor in the PPS equation. The two TTP equations were based on a Weibull distribution and the PPS one was included as a predictor in the PPS equation. The two TTP distributions comprised of the MRC trials for all patients for whom survival data was available. However, restricting this assessment to patients who fail first-line treatment (i.e., the subset used in the main analyses) would be subject to selection bias because starting second line treatment in later years may be associated with better response to first line treatment.

### Table 1 – Baseline characteristics of patients in the MRC trials.

<table>
<thead>
<tr>
<th>Characteristic at start of treatment (n [%], unless otherwise stated)</th>
<th>One prior group</th>
<th>Multiple prior group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years)</td>
<td>64.7</td>
<td>65.1</td>
</tr>
<tr>
<td>Male</td>
<td>617 (56.6)</td>
<td>216 (57.6)</td>
</tr>
<tr>
<td>Performance status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(mapped to ECOG)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>231 (26.8)</td>
<td>47 (16.9)</td>
</tr>
<tr>
<td>1</td>
<td>329 (38.2)</td>
<td>91 (32.7)</td>
</tr>
<tr>
<td>2–3</td>
<td>302 (35.0)</td>
<td>140 (50.4)</td>
</tr>
<tr>
<td>Mean M-protein (g/L)</td>
<td>29.1</td>
<td>38.6</td>
</tr>
<tr>
<td>Beta-2 microglobulin &gt;2.5 mg/L</td>
<td>859 (95.6)</td>
<td>293 (95.1)</td>
</tr>
<tr>
<td>Lytic bone lesions (at first-line treatment)</td>
<td>709 (74.5)</td>
<td>246 (73.9)</td>
</tr>
<tr>
<td>Durie-Salmon stage (at first-line treatment)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>44 (4.0)</td>
<td>18 (4.8)</td>
</tr>
<tr>
<td>II</td>
<td>104 (9.5)</td>
<td>33 (8.8)</td>
</tr>
<tr>
<td>III</td>
<td>868 (79.6)</td>
<td>299 (79.7)</td>
</tr>
</tbody>
</table>

ECOG, Eastern Cooperative Oncology Group; MRC, UK Medical Research Council.

* Performance status in the MM-009/010 trials was measured using the Eastern Cooperative Oncology Group (ECOG) scale. Because the MRC trials used their own performance status scale, the levels of the MRC performance status scale were mapped to the ECOG scale: asymptomatic, ECOG 0; minimal symptoms, ECOG 1; restricted activity, ECOG 2–3; bedridden, ECOG 4.
Results

MRC data were analyzed for the 1628 patients who achieved plateau/stable disease from the total 2942 patients starting first-line therapy. Of these, 1090 patients started second-line therapy and comprised the one-prior therapy group. Data were available for 375 patients with multiple prior therapies; of these, 269 were third-line treatment and 106 were fourth-line treatment. The characteristics of these two groups are described in Table 1. Average follow-up was 24.4 months from start of second-line treatment, during which 1015 of the 1090 second-line patients (93%) died. The average follow-up in the multiple prior therapy group was 14.2 months, and 354 of the 375 (94.4%) died. The survival distributions in the two groups are shown in Figure 2. The median overall survival (time to death) was 16.1 and 9.2 months in the one- and multiple-prior groups, respectively.

The best fitting distribution form for both one and multiple prior therapy groups was an exponential function. Gender, Durie-Salmon stage, and presence of lytic bone lesions were not retained for second-line equation (P < 0.10) (Table 2). The equation yields a median survival of 17.0 months for the MRC population, compared with the observed 16.1 months (Fig. 3). The prediction equation for the multiple prior therapy group is summarized in Table 2. Age, gender, Durie-Salmon stage, M-protein and presence of lytic bone lesions were not retained (P < 0.10). The equation also had very good fit to the data, as suggested by predicted median OS of 9.4 months compared with 9.2 observed (Fig. 4).

Applying the equations to patients randomized to dexamethasone in the lenalidomide trials (Table 3) yielded a median of 19.5 months (95% CI, 16.6–22.9) for patients with one prior therapy, and 11.6 months (95% CI, 9.5–14.2) for patients with multiple prior treatments. These estimates were substantially lower than observed in the MM-009/010 trials: 35.3 and 27.3 months, respectively [18] (based on the December 2005 cut of the MM-009/010 data).

Dexamethasone was part of the second-line treatment regimen for 103 patients (9.5%) in the MRC trials. The regimen of patients not on dexamethasone consisted of M7/MP (24%), ABCM (19%), cyclophosphamide (19%), VAMP or VAD (13%), or HDP (2%), with the remainder on other treatments. Dexamethasone was part of the treatment regimen for 59 (15.7%) patients in the multiple prior groups. Survival for patients on regimens involving dexamethasone was compared with that of those not on dexamethasone-containing regimens. Survival did not differ significantly (P = 0.88 for one-prior group, and P = 0.13 for multiple prior group).

Differences in survival by year in which treatment was initiated for patients entered into the MRC trials (Fig. 5) were not statistically significant (log-rank test, F = 0.40). In fact, the ordering of curves suggests poorer survival among patients starting treatment after 1995.

### Table 2 - Prediction equation for OS in patients with one or multiple prior therapies in the MRC trials.

<table>
<thead>
<tr>
<th></th>
<th>One prior group</th>
<th>Multiple prior group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.81 (0.35)</td>
<td>2.68 (0.35)</td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>-0.023 (0.005)</td>
<td>-0.0001</td>
</tr>
<tr>
<td>Performance status (mapped ECOG)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 versus 0</td>
<td>-0.26 (0.10)</td>
<td>NA</td>
</tr>
<tr>
<td>2–3 versus 0</td>
<td>-0.56 (0.10)</td>
<td>NA</td>
</tr>
<tr>
<td>2–3 versus 0–1</td>
<td>-0.26 (0.13)</td>
<td>0.044</td>
</tr>
<tr>
<td>Mean M-protein (g/L)</td>
<td>0.0049 (0.002)</td>
<td>0.008</td>
</tr>
<tr>
<td>Beta-2 microglobulin &gt;2.5 mg/L (yes versus no)</td>
<td>0.53 (0.21)</td>
<td>0.011</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>0.19 (0.02)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Variable not included in the equation.
ECOG, Eastern Cooperative Oncology Group; MRC, UK Medical Research Council; NA, not applicable; OS, overall survival.
Discussion

Data from an external source (MRC multiple myeloma trials from 1980 to 2002) were used to adjust equations derived from two lenalidomide trials and to estimate OS with dexamethasone treatment in the absence of crossover. The adjusted median survival was 19.5 months for patients in the lenalidomide trials who had failed one prior therapy, and 11.6 months for those with multiple prior treatments, 42% and 59% lower (respectively) than what was observed with crossover in the trials. This suggests crossover can have a substantial impact on overall survival. This is consistent with the large benefit observed for lenalidomide for TTP: 13.4 versus 4.6 months in MM-009/010 [18].

Beyond its use for calibration, the overall survival equation derived from the MRC data may be used to predict survival for myeloma patients starting treatment with standard therapies in current day in other settings (e.g., for studies with short follow-up). Our analyses did not reveal a difference in the survival of patients receiving treatment regimens that included dexamethasone [20]. Analysis of patient-level data allows proper handling of differences in populations and other aspects of study design (e.g., calendar period). Reference values can be derived with respect to design, methodology, and populations. Attempts to correct for differences in populations based on aggregate information (e.g., mean values of baseline characteristics) would be crude at best. The desired reference values may not be reported exactly. Furthermore, if crossover is common in the therapeutic area, other studies may be prone to the same biases. For instance, crossovers were observed in another recent myeloma study, the assessment of proteasome inhibition for extending remissions (APEX) trial, which compared bortezomib with dexamethasone [20]. Analysis of patient-level data allows proper handling of differences in populations and other aspects of study design (e.g., calendar period). Reference values can be derived from appropriate subsets that match the profiles of the original trial equations.

Both the MRC and Mayo Clinic data support the use of historical data as a robust indicator of the survival likely to be achieved today with traditional therapies for multiple myeloma.

The method described in this article adds to the body of approaches to deal with crossover in trials. Other methods like rank-preserving structural failure time models and marginal structural models use complex estimation procedures to derive unbiased treatment effect estimates. Our approach uses more standard statistical methods but relies on external data expected to be free of bias to calibrate equations derived from the trial data. Although a reference value to calibrate may sometimes be found in the literature, this is prone to problems. Published studies may differ substantially from the trials from which the original trial equations are derived with respect to design, methodology, and populations. Attempts to correct for differences in populations based on aggregate information (e.g., mean values of baseline characteristics) would be crude at best. The desired reference values may not be reported exactly. Furthermore, if crossover is common in the therapeutic area, other studies may be prone to the same biases. For instance, crossovers were observed in another recent myeloma study, the assessment of proteasome inhibition for extending remissions (APEX) trial, which compared bortezomib with dexamethasone [20]. Analysis of patient-level data allows proper handling of differences in populations and other aspects of study design (e.g., calendar period). Reference values can be derived from appropriate subsets that match the profiles of the original trial equations.
The predictive equations for OS derived from the MRC data source provided the reference values for calibration. It may be asked why not use the MRC equation directly instead of going through the calibration process to get unbiased survival estimates. The primary reason to calibrate the original MM-009/010 equations is to preserve the shape of the distributions observed in these studies. Although the parametric survival equations capture both the shape and the scale of the distributions of the outcomes, only the scale parameter is related to predictors of risk. At the same time, the shape is assumed to be unique to the entire population and may differ between the MM-009/010 trials and the MRC studies. Furthermore, the economic analyses where the MM-009/010 equations were used were structured to predict OS as the sum of time to progression and post-progression survival (Fig. 1). Using OS predictions directly from the MRC equation would imply ignoring the observed time to progression with dexamethasone in the MM-009/010 trials, which were not affected by crossover.

With patient-level data from the MRC, it was possible to select a population similar to the MM-009/010 trials, and to derive a prediction equation to adjust for differences in patient characteristics. Only few predictors were available for consideration, and only those that were available in the MM-009/010 trials could be used in the MRC equations. Some variables had to be adapted to coincide with the definitions in the MM-009/010 trials (e.g., performance status). It needs to be acknowledged that there may be some unmeasured baseline characteristics associated with mortality that differed between the MRC and MM-009/010 and, thus, could account for part of the observed differences in survival. Thus, it is possible that part of the difference in the observed median survival and the one predicted from MRC is not completely attributable to the crossover effect, but rather may be due to residual differences in the populations not captured in the equations. This may also occur if a predictor was available in the MRC trials but not in the MM-009/010, so that possible differences between the two populations cannot be adjusted completely. No such variables were identified, however, in this study; all variables with an a priori clinical basis for being a predictor were available in both sources. This may not always be the case, however. When the predictors available in the two sources differ, it is important to explore the potential influence of the non-common variables.

Another limitation is that only about 10% of patients used from the MRC trials received regimens that included dexamethasone. Mortality in these patients was not distinguishable, however, from that of patients who received regimens that did not include dexamethasone. Thus, using the combined population for these analyses was justified.

The applications of the method are broader than the case study described in this article. For instance, the same technique can be used to calibrate equations to predict outcomes for other medications and make external comparisons [21]. The calibration term can be thought of as a measure of comparison, like a hazard ratio. Furthermore, our case study was complicated by the use of TTP and PPS equations to predict OS. The same approach could have been used if OS had been modeled directly, in which case an algebraic solution would have been possible to derive the value of the calibration term. Finally, the technique is not only applicable to survival equations. The same strategy can be applied to adjust other types of equations, like those derived from logistic or linear regressions. A key question is whether calibration should be based on the mean rather than the median (or other percentiles). The mean is certainly more appropriate for outcomes analyzed with linear or logistic regression models, as these are inherently based on modeling the means of the underlying distributions. For time to event outcomes, however, we believe the median, or percentiles of the distributions are more appropriate measures. Using the mean for these outcomes poses two possible problems. Event times tend to have a skewed distribution with a long tail that can greatly influence the mean. Furthermore, the tail of the distribution is usually not observed (follow-up is limited in studies); thus, the tails of predicted distributions are not always directly supported by data but rather projected based on the pattern of the earlier portions of the distribution where the observed data lie. As a result, the predicted means are more susceptible to be affected by errors in prediction. Predicted percentiles such as the median are less prone to these types of problems because they are not affected by the tail of the distribution and can be chosen to lie within the observed range of the data or at least very close to this. Therefore, we recommend using centiles for calibration with time-to-event equations, although we realize that there may be diverging views. For instance, in NICE’s appraisal of the lenalidomide submission [22], the review committee noted that the choice of calibration at the mean versus median was a matter of scientific judgment, but ultimately favored use of the mean. This was justified by the fact that incremental cost-effectiveness ratios are based on means. Furthermore, they argued that because the overall survival distribution was observed nearly completely in the MRC data, the mean could be estimated accurately. We chose to calibrate OS at the median as this represents the middle of the survival distribution. Other factors should be taken into account in deciding a point for calibration. For instance, if a comparison of the reference and original trial survival curves reveal an early deviation, a point in the earlier part of the curve may be more appropriate. Furthermore, although the distributions in our example were exponential, only the scale parameter needed to be manipulated. More complex forms, like Weibull or Gompertz, involve both a scale and shape parameter. Although usually only the scale parameter is related to predictors (via regression) one may consider calibrating by changing the shape parameter depending on what aspect of the distribution is thought to be most affected by crossover.

Conclusions
The calibration method described in this article adds to the existing set of methods to deal with crossover in trials. The approach is relatively simple to implement and readily extendable to any type of statistical models (other than survival regressions). Though ideally implemented with patient-level data, it may also be used with published information. Its application to the MM-009/010 trials suggests that crossover had a substantial impact on survival estimates.

Acknowledgments
The authors wish to thank the patients in the MM-009/010 and MRC trials as well as the investigators participating in these studies.

Source of financial support: Celgene UK and Ireland provided funding for portions of the study.

References


