Sharing some interim data in trial monitoring can mislead or unmask trial investigators: A scenario-based survey of trial experts

Victoria Borg Debono\textsuperscript{a,b}, Lawrence Mbuagbaw\textsuperscript{a,c}, James Paul\textsuperscript{b}, Norm Buckley\textsuperscript{b}, Lehana Thabane\textsuperscript{a,b,c,*}

\textsuperscript{a} Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street West, Health Science Centre-2C, Hamilton, Ontario, L8S 4K1, Canada
\textsuperscript{b} Department of Anesthesia, McMaster University, 1280 Main St. West, Health Science Centre-2V9, Hamilton, Ontario, L8S 4K1, Canada
\textsuperscript{c} Biostatistics Unit, St Joseph's Healthcare, St. Joseph's Healthcare Hamilton, Biostatistics Unit, 3rd. Floor, Martha Wing, Room H-325, 50 Charlton Avenue East, Hamilton, Ontario, L8N 4A6, Canada

\textbf{A B S T R A C T}

\textbf{Background:} Sharing masked interim results by the Data Safety Monitoring Board (DSMB) with non-DSMB members is an important issue that can affect trial integrity. Our survey’s objective is to collect evidence to understand how seemingly masked interim results or result extrapolations are interpreted and discuss whether these results should be shared at interim.

\textbf{Methods:} Conducted a 6 scenario-question survey asking trial experts how they interpreted three kinds of seemingly masked interim results or result extrapolation measures (interim combined event rate, adaptive conditional power and “unconditional” conditional power).

\textbf{Results:} Thirty-one current Consolidated Standards of Reporting Trials group affiliates were invited for survey participation (February 2015). Response rate: 71.0% (22/31). About half, 52.6% (95% CI: 28.9%–74.0%), (10/19), correctly indicated that the interim combined event rate can be interpreted in three ways (drug X doing better than placebo, worse than placebo or the same) if shared at interim. The majority, 72.2% (95% CI: 46.5%–99.7%), (13/18), correctly indicated that the adaptive conditional power suggests relative treatment group effects. The majority, 53.3% (95% CI: 26.6%–77.0%), (8/15), incorrectly indicated that the “unconditional” conditional power suggests relative treatment group effects.

\textbf{Discussion/Conclusion:} Knowledge of these three results or result extrapolation measures should not be shared outside of the DSMB at interim as they may mislead or unmask interim results, potentially introducing trial bias. For example, the interim combined event rate can be interpreted in one of three ways potentially leading to mistaken guesswork about interim results. Knowledge of the adaptive conditional power by non-DSMB members is telling of relative treatment effects thus unmasking of interim results.

\textbf{1. Introduction}

The Data Safety Monitoring Board (DSMB) is responsible for trial stewardship \cite{1,2}, typically charged with protecting participant safety and potential trial biases \cite{1,2}. An issue that can negatively affect trials is the introduction of bias if the DSMB were to share interim trial results or result extrapolations with non-DSMB members, especially those responsible for the trial’s conduct \cite{1,3,4}. Those individuals could potentially act upon that information, consciously or subconsciously, modifying the objectivity of the trial’s design to the point that the observed treatment difference is altered away from the truth. Conscious or subconscious alterations that introduce bias, by those non-DSMB members in the know of interim results, could be changes to treatment group adherence, endpoints, endpoint evaluation, accrual rates and enrollment, trial design, and the timing of trial termination \cite{1}. This is an especially serious issue for phase III trials because they are usually used to provide definitive evidence on efficacy and safety endpoints to inform practice or regulatory approvals \cite{5,6}.

A case described \cite{7} prompted us to investigate further the issue of sharing seemingly masked interim results or result extrapolations. The interim combined event rate (an interim result), and the adaptive conditional power and “unconditional” conditional power (both result extrapolations) provided at interim can be considered seemingly masked because they do not directly reveal the trial’s interim event
of the assumed control event rate prior to the start of the trial as it is in the protocol. Thus, to make the scenarios as realistic as possible, we included this information.

2.1.2. Constructing and administering scenario-based survey

We designed our survey to have scenario-based questions enabling the respondents to answer a multiple choice question, indicating how they interpreted three different kinds of interim results or result extrapolations regarding the relative treatment effects between treatment groups; in our case Drug X versus placebo. We asked the respondent to provide their interpretation for one kind of interim result or result extrapolation per scenario-based question. The definitions of the three kinds of interim result or result extrapolations were on the relevant survey pages for the respondent. We asked six scenario-based questions within the survey (See Appendix A: Scenario-based survey questions). We also had a general comments section under each question to allow the respondent to provide comments about the scenario or any other comment they may have had. The online survey was constructed and administered using fluidsurvey.com. We sent the first version of the online survey to 10 trial experts at McMaster University, Hamilton, Ontario for pilot testing for content validity, clarity and for any other feedback. Nine out of 10 of trial experts responded to the survey for pilot testing and feedback. We modified the online survey based on this feedback and created the final version of the online survey.

2.2. Sampling

2.2.1. Target group and sampling

The target focus group for this survey was trial experts and we contacted the Consolidated Standards of Reporting Trials (CONSORT) group in November 2014 to ask for permission to contact and solicit recent CONSORT members for their participation in our scenario-based survey. We chose members of CONSORT group because they are trial experts and as a group, they develop guidelines about the proper reporting of trials in journal publications. Writing such guidelines would require a member to have some appreciable understanding of the intricacies and workings of trials including interim analyses and possible information generated at trial interim. The CONSORT group sent out an initial email on our behalf in December 2014 based on their own mailing list, letting potential respondents know about the online survey, its purpose and the coming survey's email invitation. We first sent out the invitation to the online survey in February 2015 via Fluidsurveys.com and following the Dillman’s principles [10] a reminder email 2 weeks later to encourage a good response.

2.3. Data collection and analysis

We used FluidSurveys.com to disseminate the survey, and collect responses. A link to the survey through Fluidsurveys.com was sent to potential respondents via email. Responses were collected anonymously. The software used to analyse the results was integrated software within Fluidsurvey.com and Microsoft Excel 2010. We report results anonymously and in aggregate by count and percentages, indicating how many respondents chose a particular multiple-choice option stemming from a particular scenario-based question along with the a proportion's associated Fisher's Exact 95% Confidence Interval (CI). All respondents solicited were current members of the CONSORT group. We did not collect information on demographics to minimize respondent burden, and therefore unable to perform a subgroup analysis.

3. Results

Out of 31 invitations sent, we received 22 responses (16 complete responses and 6 partial or incomplete responses) for a total response rate of 71.0% (22/31). Fig. 1 (Fig. 1: Results from Survey) provides the
Table 1
Definitions of interim result and results extrapolations.

<table>
<thead>
<tr>
<th>Interim combined event rate</th>
<th>Adaptive conditional power</th>
</tr>
</thead>
<tbody>
<tr>
<td>The total number of events observed at some planned interim point into the trial divided by the total number of participants admitted at that same planned interim point (e.g. a planned interim point can be six months from the start of the trial or after enrolling a certain number of participants).</td>
<td>The probability of rejecting the null hypothesis of no effect by the end of the trial (i.e. finding a statistically significant effect in favour of the intervention), at some predetermined interim point in the trial when the adaptive conditional power is scheduled to be calculated.</td>
</tr>
<tr>
<td>Example:</td>
<td>Therefore the Interim Combined Event Rate at the trial’s interim analysis, six months from the start of the trial, is 11.4%</td>
</tr>
<tr>
<td>• Total # of Deaths in both the placebo group and new intervention group, six months from the start of the trial = 80</td>
<td>The assumption made is that the observed interim effect (i.e. relative risk reduction) in the trial will remain the same till the end of the trial.</td>
</tr>
<tr>
<td>• Total # of Participants in both the placebo group and the new intervention group, six months from the start of the trial = 700</td>
<td>Example statement:</td>
</tr>
<tr>
<td>• Calculation: 80/700 = 0.114 or 11.4%</td>
<td>Given the interim data (data collected 2 years into the trial that is planned to last for 3 years), and assuming the observed interim effect (i.e. relative risk reduction) at the two year point to be the true effect for the remainder of the trial, the probability of rejecting the null hypothesis of no effect (i.e. finding a statistically significant effect in favour of the intervention) at the end of the trial is 60%.</td>
</tr>
<tr>
<td>Therefore: Adaptive conditional power:</td>
<td>The following pieces of information are used to calculate Adaptive Conditional Power at trial interim:</td>
</tr>
<tr>
<td>• Control event rate and experimental event rate</td>
<td>• Control event rate and experimental event rate</td>
</tr>
<tr>
<td>• Information Fraction; a ratio of the planned sample size and the number of patients recruited in trial at the interim analysis</td>
<td>• Information Fraction; a ratio of the planned sample size and the number of patients recruited in trial at the interim analysis</td>
</tr>
<tr>
<td>• Z score and B value at interim</td>
<td>• Z score and B value at interim</td>
</tr>
<tr>
<td>• Drift parameter</td>
<td>• Drift parameter</td>
</tr>
<tr>
<td>“Unconditional” conditional power</td>
<td>The following pieces of information are used to calculate Unconditional Conditional Power at interim:</td>
</tr>
<tr>
<td>The probability of correctly rejecting the null hypothesis of no effect at the end of the trial (i.e. finding a statistically significant effect in favour of the intervention) and accepting the alternative hypothesis when indeed the alternative hypothesis is true, at some interim point in the trial.</td>
<td>1. The hypothesized treatment effect at the design stage (i.e. relative risk reduction) of the trial, assuming the hypothesized treatment effect at the design stage to be true and correct for the remainder of the trial;</td>
</tr>
<tr>
<td>The following pieces of information are used to calculate Unconditional Conditional Power at interim:</td>
<td>2. The sample size calculated at the design stage for the trial AND;</td>
</tr>
<tr>
<td>1. The hypothesized treatment effect at the design stage (i.e. relative risk reduction) of the trial, assuming the hypothesized treatment effect at the design stage to be true and correct for the remainder of the trial;</td>
<td>3. The combined event rate calculated at the trial’s interim, assuming this rate to be true for the remainder of the trial.</td>
</tr>
<tr>
<td>2. The sample size calculated at the design stage for the trial AND;</td>
<td>Example statement:</td>
</tr>
<tr>
<td>3. The combined event rate calculated at the trial’s interim, assuming this rate to be true for the remainder of the trial.</td>
<td>Given the interim data (data collected 2 years into the trial that is planned to last for 3 years), and assuming the observed interim effect (i.e. relative risk reduction) at the two year point to be the true effect for the remainder of the trial, the probability of rejecting the null hypothesis of no effect (i.e. finding a statistically significant effect in favour of the intervention) at the end of the trial is 60%.</td>
</tr>
</tbody>
</table>

Fig. 1. Results from Survey.

Survey Questions

Please see Appendix A for more details on full questions.
results from our survey for each of the six scenario-based questions asking respondents what they interpreted as the most compatible answer, in regards to the relative treatment effects between two treatment groups, based on the interim results or result extrapolations provided in the scenario.

Fig. 1 summarizes the results for Question 1 to Question 6. Question 1 is in regards to how trial experts interpreted the effect of Drug X relative to the placebo after seeing the interim combined event rate of 0.34. Just over half of the respondents, 52.6% (95% CI: 28.9%–74.0%), (10/19), correctly assumed “D. All of the above”, that one of the three options (A. drug X doing better than placebo, B. worse than placebo or C. the same) could be true. These 3 possible options were also demonstrated to be so empirically from the responses to Questions 4, 5 and 6, as summarized in Fig. 1 (See Appendix A for details about the questions), as additional information in these questions was given about event rates in both the Drug X group and the placebo group. This is information typically not given as it is unmasking of event rates per group. It should be noted for Question 1 that 42.1% (95% CI: 20.3%–64.3%), (8/19), of respondents answered answer “E. None of the Above”. Though the more correct answer is “D. All of the Above”, in that any one of the three options (A. drug X doing better than placebo, B. worse than placebo or C. Drug X and the placebo group are performing the same) could be true of the relative interim results, one could have also interpreted that none of the options, A, B or C are correct as a sole answer by themselves. This hypothesis seems to be supported by the fact that most of the respondents correctly answered questions 4, 5 and 6.

For Question 4 (see Fig. 1 and Appendix A), the majority of respondents, 87.5% (95% CI: 61.7%–98.3%), (14/16), correctly assumed Option A that Drug X is performing better than placebo when given two additional pieces of information where the Drug X event rate given was 0.291 and the placebo event rate given was 0.389. This is a 25% relative risk reduction in 28-day mortality between Drug X and the placebo, hypothesized and specified in the trial protocol, as indicated in the scenario presented. For Question 5, the majority of respondents, 86.7% (95% CI: 59.5%–98.2%), (13/15), correctly assumed Option B that Drug X is performing worse than placebo when given two additional pieces of information where the Drug X event rate given was 0.389 and the placebo event rate given was 0.291. This is a 25% relative risk increase in 28-day mortality, contrary to what was hypothesized and specified in the trial protocol, as indicated in the scenario presented. And for Question 6, the majority of respondents, 75.0% (95% CI: 47.6%–92.2%), (12/16), correctly assumed Option C that Drug X and the placebo group are performing the same when given two additional pieces of information where the Drug X event rate given was 0.337, and the placebo event rate given was 0.343. This is a 2% relative risk reduction in 28-day mortality, different from what was hypothesized and specified in the trial protocol, as indicated in the scenario presented.

In regards to the two questions about the interim result extrapolations, Fig. 1 also summarizes the results about how trial experts interpreted the efficacy of Drug X relative to the placebo after seeing the adaptive conditional power (Question 2) and the “unconditional” conditional power (Question 3). For Question 2, the majority of respondents, 72.2% (95% CI: 46.5%–89.7%), (13/18), correctly assumed that an adaptive conditional power of 99% meant that Drug X is performing better than placebo (Option A) according to the assumptions used to calculate the adaptive conditional power (See Appendix A and Table 1 for definitions of interim measures). However, for the “Unconditional: Conditional Power” (see Fig. 1 and Appendix A), the results indicate that there is more confusion about the meaning of this measure. A minority of respondents for Question 3 answered correctly “E. None of the above”, 40.0% (95% CI: 16.3%–64.9%), (6/15). The majority of respondents incorrectly answered, “A. Drug X is performing better than placebo” 53.3% (95% CI: 26.6%–77.0%), (8/15) and one respondent, 6.7% (95% CI: 2.0%–23.2%), (1/15), answered, “C. Drug X and the placebo group are performing the same”.

4. Discussion

4.1. Key findings

Our results empirically show that sharing the interim combined event rate is a well-understood measure but is one that can be interpreted in one of three ways, when presented by itself, without additional knowledge about interim control event rate and the interim new intervention event rate; information which is not shared during a trial with non-DSMB members because it is unmasking of group effects. There are three assumptions that can be made about relative treatment effects (A. drug X doing better than placebo, B. worse than placebo or C. the same) when just given the interim combined event rate. These three possible assumptions were demonstrated to be empirically plausible from the responses to Questions 4, 5 and 6 when additional knowledge about interim control event rate and the interim new intervention event rate (Drug X) was provided. Knowledge of the interim combined event rate can lead non-DSMB members to guess about how a trial is progressing. Sharing this interim result with a non-DSMB member, such as a trial investigator, who could very well make one of three different assumptions or guesses about the interim relative treatment effect, may in turn influence a change in their behaviour towards the operation of the trial and hence introduce bias. Thus, we believe the interim combined event rate should not be shared with non-DSMB members by the DSMB because of the potential assumptions that could be made by non-DSMB members about relative treatment effects, which may lead to introducing trial bias.

As for the adaptive conditional power, it is clear that trialists understand this measure and that the higher the adaptive conditional power, based on the assumptions given in our scenario, the more likely Drug X was performing better than placebo. With this evidence, the adaptive conditional power is a dangerous measure to share because it is unmasking as it indirectly gives a non-DSMB member an idea how a treatment group is doing relative to another. This knowledge could influence a change in a non-DSMB member’s behaviour regarding the operation or conduct of the trial thus introducing trial bias.

The majority of respondents were unclear about how to interpret the “unconditional” conditional power. It is believed that the majority of respondents being unclear on this measure have to do with it not being very familiar to many, as it does not appear to be used often. The reason option “E. None of the above” is correct for question 3 is because the “unconditional” conditional power is not giving you any information about the relative efficacy between treatment groups. Simply put, it tells you the power your trial will have to answer your primary research question if you were to complete the trial given the sample size, the hypothesized effect size determined at the trial’s design stage and the combined event rate at interim; much like the power calculation done before the trial commences. It only gives the non-DSMB member some reassurance that your trial will have the desired amount of power needed to sufficiently answer your primary question once it has reached it planned sample size and is complete. We found one case of it being used [7] to reassure the sponsor that the trial underway will be properly powered to answer the primary question given that the predetermined sample size is reached. Most people also associate an interim power calculation with the adaptive conditional power and may think the “unconditional” conditional power is a variation of the adaptive conditional power and that it may convey the same information. We know this is not the case, however comments made for this question seem to suggest that trialists think it is similar to the adaptive conditional power. Knowledge of the “unconditional” conditional power can lead to non-DSMB members misinterpreting how a trial is progressing. This too may in turn influence a change in their behaviour towards the operation of the trial and hence introduce bias. Because of the confusion around how to interpret the “unconditional” conditional power, it may
not be a good measure to share as one could confuse the measure as suggestive of the relative treatment effect between treatment groups.

4.2. Findings compared to similar studies

There were no other studies found that empirically evaluated how commonly generated pieces of interim results or extrapolations are interpreted by trial experts. This study is unique in its ability to evaluate how three interim results or extrapolations used are interpreted by trial experts with a survey asking hypothetical scenario-based questions, using real trial interim information.

4.3. Key limitations

In regards to limitations of our study, the comments we received by respondents mentioned that it would have been helpful to provide confidence intervals for our estimate of the interim combined event rate and the additional information regarding the individual group event rates we provided for our scenarios in Questions 4, 5 and 6 (see Appendix A). It was noted that there was too much uncertainty to judge the difference of the true effect in the absence of confidence intervals. This is true; however, we do not believe this would have changed the respondents’ answers because it is generally known at interim that the confidence intervals will be wide since the precision of the estimates will be low when only half the needed sample size is enrolled.

Though we had a good response rate for our survey, it is likely that we did not have a big enough sample size. However, this survey was designed to focus on a specific group of trial experts familiar with interim trial analyses. Due to the detailed nature of the questions and the feedback we received in the survey’s testing stage, we knew that such questions would be best answered by trial experts who have familiarity and knowledge of the workings of interim analyses and the kind of interim results generated. A larger and more general survey given to those interested or are involved in trials, asking their views on the usefulness and the need to share certain types of interim results or results extrapolations, can be done to further understand if there are cases that may warrant sharing of such information or not.

4.4. Implications for practice

Trials are susceptible to bias and it is important to have a protocol with safeguards in place to prevent the introduction of biases that could alter trial results generated, away from the most true effect size, especially in phase III trials used to generate definitive results on efficacy and safety endpoints that provide evidence for practice and regulatory approval. As previously noted, there are some circumstances in the literature where interim results may be shared by the DSMB with non-DSMB members and knowledge gained form this survey may have implications on what is shared in those circumstances [8]. In cases where there may be a request from non-DSMB members to have interim results or extrapolations shared with them by the trial’s DSMB, we do not recommend sharing the interim combined event rate, “unconditional” conditional power or the adaptive conditional power. The reasons why are as follows. The interim combined event rate is a well-understood measure and the majority of respondents correctly indicated that having it alone, without knowledge of the interim control event or the interim new intervention event rate, can be interpreted in one of three ways. This measure thus provides no useful information and only invites the mistaken opportunity for guesswork about how one group is doing compared to another. The “unconditional” conditional power is a measure that is mostly misinterpreted. Moreover, the adaptive conditional power seems to be well-understood and is unmasking of interim relative treatment effects, based on the empirical evidence we collected and evaluated. Knowledge of any of these three kinds of interim results or extrapolations may influence a change in behaviour in those responsible for the operation or conduct of the trial or those who participate in the trial in some way and hence, may introduce trial bias. If information had to be shared with a particular non-DSMB member, safeguards should be in place that prevents other non-DSMB members directly responsible for the operation or conduct of the trial or those participating in the trial in some way, from knowing such interim results or extrapolations.

5. Conclusion

From this survey, we have some empirical evidence to suggest that the interim combined event rate, the adaptive conditional power and the “unconditional” conditional power should not be shared. The interim combined event rate and the adaptive conditional power are well-understood measures. However, the interim combined event rate can suggest any one of three plausible relative group effects at interim if shared at interim making it a useless measure to share that invites guesswork about relative effects. The adaptive conditional power is unmasking of relative treatment effects at interim. The “unconditional” conditional power on the other hand is misinterpreted most likely because the measure is unfamiliar. There is a danger with the DSMB sharing any of these three measures with non-DSMB members as it may lead non-DSMB members to consciously or subconsciously alter their behaviour towards a trial, possible introducing trial bias.

Funding

This project received full funding support from the Canadian Institutes for Health Research (CIHR). They were not involved in this study’s design, the collection, analysis or interpretation of this survey data, writing this report, or in deciding to submit this report for publishing.

Competing interests

The authors declare that they have no competing interests.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.conctc.2017.05.005.

References