

Primary research

Experience collecting interim data on mortality: an example from the RALES study

Janet Wittes*, Jolie Palensky*, Debra Asner†, Desmond Julian‡, Jean-Pierre Boissel§, Curt D Furberg#, Henri Kulbertus¶, Stuart Pocock** and Barbara Roniker†

*Statistics Collaborative, Inc, Washington, DC, USA

†Pharmacia, Skokie, IL, USA

‡London, UK

§Hospital Neuro-Cardiologique, Lyon, France

#Wake Forest University School of Medicine, Winston-Salem, NC, USA

¶Hospital Universitaire de Sart Tilman, Liege, Belgium

**London School of Hygiene & Tropical Medicine, London, UK

Correspondence: Janet Wittes, PhD, Statistics Collaborative, Inc, 1710 Rhode Island Avenue, Suite 200, Washington, DC 20036, USA. Tel: +1 (202) 429 9267; fax: +1 (202) 429 9268; e-mail janet@statcollab.com

Received: 15 December 2000

Revisions requested: 25 January 2001

Revisions received: 26 January 2001

Accepted: 26 January 2001

Published: 7 February 2001

Curr Control Trials Cardiovasc Med 2001, **2**:59–62

© 2001 Wittes *et al*, licensee BioMed Central Ltd
(Print ISSN 1468-6708; Online 1468-6694)

Abstract

Introduction: The Randomized Aldactone Evaluation Study (RALES) randomized 822 patients to receive 25 mg spironolactone daily and 841 to receive placebo. The primary endpoint was death from all causes. Randomization began on March 24, 1995; recruitment was completed on December 31, 1996; follow-up was scheduled to continue through December 31, 1999. Evidence of a sizeable benefit on mortality emerged early in the RALES. The RALES data safety monitoring board (DSMB), which met semiannually throughout the trial, used a prespecified statistical guideline to recommend stopping for efficacy. At the DSMB's request, its meetings were preceded by an 'endpoint sweep', that is, a census of all participants to confirm their vital status.

Methods: We used computer simulation to evaluate the effect of the sweeps.

Results: The sweeps led to an estimated 5 to 8% increase in the number of reported deaths at the fourth and fifth interim analyses. The data crossed the statistical boundary at the fifth interim analysis. If investigators had reported all deaths within the protocol-required 24-h window, the DSMB might have recommended stopping after the fourth interim analysis.

Discussion: Although endpoint sweeps can cause practical problems at the clinical centers, sweeps are very useful if the intervals between patient visits or contact are long or if endpoints require adjudication by committee, reading center, or central laboratory.

Conclusion: We recommend that trials with interim analyses institute active reporting of the primary endpoints and endpoint sweeps.

Keywords: data safety monitoring boards, data sweeps, interim analysis, randomized clinical trials

Introduction

In randomized multicenter clinical trials monitored by an independent Data and Safety Monitoring Board (DSMB) charged with periodic evaluation of safety and efficacy, the data available for review typically include fewer events than have actually occurred when the database is closed. The discrepancies arise from either a failure of the investigators to report events soon after they occur or a failure of the adjudication committees to classify events quickly. This paper deals with an example of delays in reporting.

Many DSMBs prefer to review timely and relatively complete, but unaudited, data rather than fully audited, incomplete data, because delays in reporting endpoints may lead to unacceptable tardiness in identifying effects of therapy [1]. Some authors [2] contend that in well-run trials, late reporting occurs only infrequently. However, our experience indicates that even well-conducted long-term double-blind multicenter trials, particularly ones conducted internationally, often report many endpoints weeks or even months after they occur. Late reporting is common even when death is the primary endpoint. This problem leads to serious uncertainty in interim analysis if the follow-up is not equally intense in each of the various treatment arms. Because late reporting of the primary endpoint can distort the estimated effect of therapy, some clinical trialists perform an 'endpoint sweep', that is, a census of all participants to confirm the status of the primary endpoint, prior to each DSMB meeting [3]. Such sweeps are especially useful if the intervals between patient visits or contact are long or if endpoints require adjudication by committee, reading center, or central laboratory.

The Randomized Aldactone Evaluation Study (RALES) [4], a 195-center double-blind randomized clinical trial, randomized 822 patients to receive 25 mg spironolactone daily and 841 to receive placebo. The primary endpoint was death from all causes. Scheduled length of follow-up ranged between 3 and 4.5 years. The DSMB for the RALES met twice yearly to review safety and efficacy. Evidence of a sizeable benefit on mortality emerged early in the trial (see Table 1, Look no.1). The protocol specified that the investigator should report a death within 24 h after its occurrence. Because the interval between protocol-specified visits was every 3 months through 1 year of study follow-up and every 6 months thereafter, the DSMB suspected that there was a delay in reporting deaths. It therefore asked the sponsor to confirm the vital status of participants 2 years after the first patient was randomized. Consequently, the sponsor asked all investigators to confirm the vital status of each of their participants on 31 December 1997 and again on 30 June 1998.

This paper describes the interim monitoring plan used in the RALES and the conduct of the endpoint sweeps. We show that these sweeps led to a 5 to 8% increase in deaths reported during the trial.

Methods

The RALES investigators randomized 1663 people with class III or IV heart failure on the scale of the New York Heart Association (NYHA) to receive spironolactone or placebo. Randomization began on 24 March 1995. Each participant was followed from randomization until the study's common closing date, which was planned to be December 31, 1999, but actually occurred on August 24, 1998. The study was designed with 80% power to detect a 17% reduction in mortality, assuming some noncompliance with study medication and complete ascertainment of vital status. The secondary endpoints were cardiac mortality, hospitalization for cardiac causes, the composite of cardiac death or cardiac hospitalization, and change in class of heart failure on the NYHA scale. Pitt *et al* [4] describe the protocol, the dosing scheme, and the results; here we focus on the planning and execution of the interim analyses.

Protocol-defined visits occurred at weeks 0, 4, 8, and 12; at months 3, 6, 9, and 12; and then every 6 months until the common close of the study. Investigators were to report deaths within 24 h of discovery. Originally, the trial was to continue until 540 deaths had occurred in the placebo group. The sponsor amended the protocol in March 1996 to continue randomizing until reaching a sample size of 2500 patients or until 31 December 1996, whichever occurred sooner. This change implied that the last patient randomized would be followed for 36 months unless the trial stopped early.

The DSMB, which was composed of three cardiologists (DJ, chair; J-P B, HK), a cardiovascular epidemiologist, (CF), and a statistician (SP), met semiannually throughout the trial to review data relating to safety and efficacy. Spironolactone, which has been in use since 1960, has a well-known adverse event profile. The most common adverse events are gynecomastia and other feminizing symptoms in males. The most serious expected adverse event associated with the use of spironolactone is hyperkalemia, because the drug acts as a potassium-sparing diuretic.

The board monitored mortality with a statistical guideline for stopping based on a Lan-DeMets [5] use function and an O'Brien-Fleming boundary [6] for efficacy at a two-sided α level of 0.05. It did not specify a formal boundary for safety, relying instead on its collective judgment to recommend early termination if spironolactone showed a net adverse effect. The board unblinded itself regarding the treatment code on 24 August 1996, its first planned interim analysis.

At the board's third interim analysis meeting, on 25 August 1997, the prespecified critical z-value for declaring efficacy of the drug was 3.67 (two-sided $P=0.00024$); the observed relative risk was 0.80 with a z-value of 2.55 (two-sided $P=0.011$), insufficient evidence to stop the trial. The board members were concerned that some

Table 1

Randomized Aldactone Evaluation Study (RALES) comparing spironolactone to placebo: observed and projected number of deaths and summary statistics at the interim analyses

| Look no. | Data cutoff (d-mo-y) | Meeting date (d-mo-y) | Observed | | | Estimated information time* | z-value | | Observed <i>P</i> (two-sided) |
|--|----------------------|-----------------------|------------------|------------------|---------------------|-----------------------------|----------|---------------------|-------------------------------|
| | | | Placebo deaths | Treatment deaths | Relative risk | | Critical | Observed | |
| Interim analysis cut-offs with the sweeps as they occurred | | | | | | | | | |
| 1 | 09-8-96 | 24-8-96 | 70 | 52 | 0.76 | 0.12 | 6.38 | 1.58 | 0.11 |
| 2 | 10-3-97 | 17-3-97 | 136 | 109 | 0.83 | 0.24 | 4.43 | 1.69 | 0.092 |
| 3 | 14-8-97 | 25-8-97 | 224 | 175 | 0.80 | 0.34 | 3.67 | 2.55 | 0.011 |
| 4 | 26-3-98 | 30-3-98 | 304 | 241 | 0.81 | 0.48 | 3.04 | 3.02 | 0.0026 |
| 5 | 14-8-98 | 24-8-98 | 351 | 269 | 0.78 | 0.57 | 2.79 | 3.75 | 0.00018 |
| Estimated interim analysis cutoffs without the sweeps [†] | | | | | | | | | |
| 4a | 26-3-98 | 30-3-98 | 281 (275–289) | 222 (214–228) | 0.81 (0.76–0.85) | 0.45 | 3.16 | 2.93 (2.33–3.69) | 0.0034 |
| 5a | 14-8-98 | 24-8-98 | 333 (325–341) | 256 (248–264) | 0.79 (0.74–0.83) | 0.55 | 2.81 | 3.59 (2.87–4.33) | 0.00034 |
| Interim analysis cutoffs that would have occurred if we had known the true numbers and times of deaths | | | | | | | | | |
| 1b | 09-8-96 | 24-8-96 | 81 | 59 | 0.75 | 0.14 | 5.88 | 1.80 | 0.072 |
| 2b | 10-3-97 | 17-3-97 | 189 | 140 | 0.76 | 0.26 | 4.24 | 2.75 | 0.0060 |
| 3b | 14-8-97 | 25-8-97 | 257 | 201 | 0.80 | 0.38 | 3.46 | 2.82 | 0.0048 |
| 4b | 26-3-98 | 30-3-98 | 330 | 254 | 0.79 | 0.51 | 2.95 | 3.56 | 0.00038 |
| 5b | 14-8-98 | 24-8-98 | 383 | 283 | 0.76 | 0.60 | 2.72 | 4.46 | 0.000008 |

This three-panel table shows (a) the data observed in the RALES trial at each of the five interim analyses; (b) simulated data for the fourth and fifth interim analysis had the sweeps not occurred; and (c) the data that would have been observed had all the deaths been reported within 24 h of their occurrence, as specified by the protocol. *No. of events that have occurred divided by the no. of events that will occur up to the planned end of the trial. Because the number of events that will occur is unknown at the time of a DSMB meeting, one must estimate it on the basis of data observed thus far and assumptions made about the future. [†]Observed values are median number of events (in 1000 simulated iterations of the trial), with ranges shown in parentheses. The 1000 iterations of the simulation assuming neither sweep showed a probability of 18% and 100% that the data would have first crossed the prespecified stopping boundary at the fourth and fifth interim analyses, respectively. These numbers refer to looks 4a and 5a (data not shown).

deaths might remain unreported, especially among the patients who were lost to follow-up and those who had not come to their 12-month visit. Given the strong trends in the data and the consistent patterns observed across subgroups of interest, the board suspected that the data would cross the prespecified stopping boundary before the planned end of the study. To avoid crossing the statistical boundary with uncertainty remaining about the number of unreported deaths, the board requested that each investigator provide a census of vital status as of 31 December 1997. In order not to alert the sponsor or investigators that the data were showing a strong trend, the board worded its request in terms of the need for a 'standard two-year' accounting of data. At the DSMB meeting in March 1998, after the 'sweep', the Board announced it found the process so valuable that it requested sweeps prior to each subsequent meeting. Because the board recommended stopping early, the study had only one additional sweep of deaths.

We used computer stimulation to evaluate the effect of the sweeps. We assumed that the patterns observed up to August 1998 would have persisted throughout the study and simulated 1000 replications of a RALES trial under these assumptions.

Results

The RALES showed fewer deaths on spironolactone than on placebo throughout the entire course of the study. The relative risk of mortality remained close to 0.80 (see Table 1).

At the fifth interim analysis, in August 1998, 620 deaths were reported. Assuming the protocol-projected mortality rates, 1080 deaths would be expected at the planned end of the study, yielding an estimated 'information time' of 620/1080 or 0.57. The actual number of deaths that would have occurred had the study continued to its planned end was of course unknowable; therefore, the statistical center (JP, JW) provided a range of projected number of deaths to examine the results' sensitivity to the assumptions. The range, from 518 to 648, corresponded to a range of critical z-values from 2.70 to 3.08. The observed z-value of 3.75 was greater than the z-value for any of the projected number of deaths. Therefore, even if the projection of 1080 were incorrect, the data would have crossed the formal statistical boundary for efficacy.

The DSMB recommended termination of the study on 24 August 1998, when a total of 620 deaths had been reported: 351 in the placebo group and 269 in the spironolactone group. The chair of the study and the

sponsor concurred with this recommendation. After RALES ended, the investigators and sponsor reported an additional 46 deaths that had occurred prior to midnight, 2 August 1998. If all deaths had been reported within the 24-h window required by the protocol, the DSMB might have recommended stopping after the fourth interim analysis, when the critical and observed z-values would have been 2.95 and 3.56, respectively (see Table 1, Look no. 4b).

To assess how the sweep affected the interim analyses, we assumed that the patterns observed up to August 1998 would have persisted throughout the study; that is, 80% of those identified in the sweep would have been identified at the subsequent interim analysis, another 10% two meetings later, and the final 10% at the end of the study. We simulated 1000 replications of a RALES trial under these assumptions. Table 1 (Look nos 4a and 4b) shows the data that would have been reported. Even without the sweeps, the data were highly likely to have crossed the monitoring boundary at the fifth interim analysis. In terms of decision-making, therefore, the sweeps probably did not change the behavior of the board. The sweeps did, however, lead to a much more complete set of data to form the basis of the decision and much more security, on the part of the DSMB, that its decision was correct. In the RALES, the data showed no evidence of differential delay according to treatment or of a differential rate of events in the deaths reported with a delay and those reported on time. Thus, the final results showed essentially the same reduction in mortality as the DSMB observed, but because of the larger number of deaths, the final *P*-value was lower than that observed at the interim analysis.

In the RALES, the sweeps led to sizeable increases in the number of reported deaths, and hence in the statistical power, at both the fourth and fifth interim analyses. At the fourth interim analysis, the sweep led to an 8% increase in the number of reported deaths (from an estimated 503 to an observed 545). If the sweep had identified all deaths, the increase would have been 16% (503 to 584). At the fifth interim analysis, the increase in number of deaths observed was about 5% (589 to 620), with a potential increase of 13% (589 to 666) if the sweeps had identified all deaths.

Discussion

Our experience in the RALES emphasizes three aspects of data collection for interim analysis – the need to project the total number of events in a trial of fixed duration, the importance of timely collection of data regarding the primary endpoint, and the practical difficulties of reporting endpoints in trials with long-term follow-up. To project the total number of deaths that would have occurred in the RALES if the trial had continued to its planned end, we augmented the assumptions in the protocol with ranges of reasonable other assumptions. If crossing the statistical boundary is sensitive to the estimated number of deaths

that would occur, then a DSMB should be uncomfortable recommending stopping.

The DSMB for the RALES, recognizing the importance of timely data, requested endpoint sweeps. The board was initially concerned that its request for endpoint sweeps would raise suspicion that it was positioning itself for early stopping. After the study ended, the DSMB learned that the process had in fact alerted neither the sponsor nor the investigators to such a possibility. Indeed, those carrying out the study felt burdened by the request for sweeps. The sponsor and the sites would have found them easier to perform if the operations manual had originally incorporated them. In the RALES, a monitor reviewed charts at each clinical site. This procedure, though consistent with the Good Clinical Practice guidelines of the International Conference on Harmonisation (see <http://www.ifpma.org/ich5.html>), is not consistent with rapid reporting of endpoints. We recommend that trials with interim analyses institute active reporting of the primary endpoints and endpoint sweeps, along with DSMB meetings.

Continuous identification of endpoints that require independent adjudication poses more problems for interim analyses than does mortality. If the primary endpoint for the RALES had included, say, hospitalization due to worsening heart failure, endpoint sweeps would have been much more difficult. The investigators reported many of the hospitalizations many months after the study ended. Trials with an adjudication committee charged with assessing the primary endpoint and a DSMB charged with interim analysis must incorporate methods for timely adjudication of endpoints in order for the DSMB to discharge its responsibilities. The institution of formal interim analysis must be accompanied by resources that allow rapid collection of relevant data. While we strongly recommend sweeps, we also recommend that the procedures of a trial specifically describe the sweeps to help alleviate the practical problems of identifying the endpoint status of people at times other than at their planned clinic visits.

References

1. Wittes J. **Data safety monitoring boards: a brief introduction.** *Biopharm Rep* 2000, **8**:1–7.
2. Meinert C. **Clinical trials and treatment effects monitoring.** *Controlled Clin Trials* 1998, **19**:515–522.
3. Hallstrom A, McBride R, Moore R: **Toward vital status sweeps: a case history in sequential monitoring.** *Stat Med* 1995, **14**: 1927–1931.
4. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, Palensky J, Wittes J: **The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators.** *N Engl J Med* 1999, **341**:709–717.
5. Lan K, DeMets D: **Discrete sequential boundaries for clinical trials.** *Biometrika* 1983, **70**:659–663.
6. O'Brien P, Fleming T: **A multiple testing procedure for clinical trials.** *Biometrics* 1979, **35**:549–556.