ELSEVIER

Journal of Clinical Epidemiology 63 (2010) 347–349

Journal of Clinical Epidemiology

TIMELY DETECTION OF ADVERSE ADVENTS IN THERAPEUTIC TRIALS

A probiotics trial on trial: the problem of timely detection of adverse advents in therapeutic trials

Raimond W.M. Giard[a,b,*]

[a]Department of Clinical Pathology, Maasstad Ziekenhuis, Rotterdam, The Netherlands
[b]Rotterdam Institute of Private Law, Department of Civil Law, Erasmus School of Law, Erasmus University Rotterdam, Rotterdam, The Netherlands

Accepted 15 December 2008

1. Introduction

The outcome of a therapeutic trial may be frustrating. In a recent Dutch randomized, double-blind, placebo-controlled multicenter trial, the effects of probiotics were studied to investigate their potential to diminish infectious complications in patients with predicted severe acute pancreatitis [1]. Desolately, the results of this study were quite contrary to expectations: there was no diminution of infectious complications, and furthermore, patients taking the probiotics had more than double the relative mortality risk. Finally, 33 of the 297 patients included in this trial died with an excess of 15 deaths occurring in the treatment group (Table 1).

These figures were made public in a press conference and drew much media attention in The Netherlands. Some statisticians and methodologists subsequently publicly expressed their strong condemnation, particularly stating that the data monitoring and safety committee (DMSC) had been negligent in its control function, missing an early trend of increased mortality in the treatment group at interim analysis using the wrong statistic [2]. If appropriate statistical tools had been used, they claim, this trial should have been stopped earlier resulting in fewer deaths. Was the DMSC really neglectful in its monitoring task? Could this not be a conclusion with the benefit of hindsight?

After this public criticism, several relatives of deceased trial participants announced that they would initiate a lawsuit for compensation because of this supposed negligence. To triumph in such a tort claim, the plaintiff must, among others, prove that the DMSC breached its duty by failing to adhere to generally accepted authoritative guidelines. As legal professionals will lack sufficient detailed knowledge on the planning, execution, and monitoring of a randomized clinical trial, any judicial ruling will invoke an expert medical opinion.

When scientific conduct is formally scrutinized, it invokes a particular—and perhaps novel—dimension of clinical epidemiology. But how? By its very nature, this basic medical science is a normative endeavor, a task which is essentially prospective: how to reach most favorable outcomes from medical interventions. A judgment of the conduct of a DMSC is different: it is retrospective, and the analysis and subsequent conclusions are now to be understood in a legal context. For this evaluation of the conduct of a DMSC, we first need a preset standard of its tasks and processes for this type of trial. This benchmark not only must be realistic and meet the essential medical requirements, but also be legally understandable and usable. Secondly, we need an established distinguishing method for objective retrospection, because it is always easy to be wise after the event [3]. In the course of this enquiry, we meet two interrelated forms of bias: outcome and hindsight bias [4]. In this article, I will comment how careful appraisal of the work of a DMSC can be done in relation to the controversies that relate to both statistical issues and decision making [5].

2. Detecting harm: a challenging task for the data monitoring and safety committee

As trial safety is of paramount importance, a DMSC should focus principally on issues of harm [6]. Are there convincing arguments, while carefully balancing benefits and harms, to stop this trial because of harmful side effects? An increased all-cause mortality in the intervention group is of course a principal reason for such a decision. But how and when? For evaluating differences in mortality in a comparative trial, we can opt between two essentially different routes: the use of traditional frequentist statistics or

* Corresponding author at: Department of Clinical Pathology, Maasstad Ziekenhuis, PO Box 9119, 3007 AC Rotterdam, The Netherlands. Tel.: +31-0-291-15-16; fax: +31-0-291-1035.
E-mail address: rwm.giard@xs4all.nl

0895-4356/10/$ – see front matter
doi: 10.1016/j.jclinepi.2008.12.012

Table 1
Tabulation of the interim and final results of the probiotics study [1]

Allocation	Evaluation point	Deceased (%)	Alive (%)	Total (%)
Probiotics	Interim	14 (15)	80 (85)	94 (100)
	Final	24 (16)	128 (84)	152 (100)
Placebo	Interim	6 (7)	84 (93)	90 (100)
	Final	9 (6)	135 (94)	144 (100)

The interim data were presented at a press conference.

of Bayesian methods [7,8]. In the frequentist statistics approach, there is a further choice between significance testing or the use of confidence intervals (CIs). In case of hypothesis testing, do we use a one-sided or a two-sided test?

In addition to the choice of an evaluative method, we also have to settle on a decision threshold. Can we apply the ''customary'' P value of 0.05 for establishing or refuting increased mortality or use a 95% or 99% CI? Should we adopt a more sensitive strategy to detect increased mortality because we cannot afford to be wrong? But won't that be at the cost of more false alarms? In this whole process, we have to deal with various value judgments: about choices of methods, of measures, of balances, and how much uncertainty we want to tolerate [9].

In this trial, the primary endpoint was the total numbers of infectious complications, and mortality was a secondary endpoint. The statistical analysis was performed on the basis of an intention-to-treat principle [10].

Is there a superior method to spot increased mortality? If we try the interim trial results from Table 1 to check for a divergence in mortality, different statistical models produce inconsistent results. The simplest ''rough and ready'' test shows a P value of 0.07 [11]. With significance testing, the difference between the two mortality proportions is only statistically significant with one-sided testing ($P = 0.03$), but not with two-sided testing ($P = 0.07$). With the use of CIs, the 95% CI for the difference between the two proportions is −0.02 and 0.18. Of the four methods, only one-sided testing provided a warning signal, knowing the results of this trial. Lack of data in the paper circumvents a dynamic Bayesian analysis in this case. But how to decide at interim? It is this methodological multiplicity that produces fertile ground for theoretical disputes.

Here we are confronted with the divergence between what we do want to know and what we really can know. After all, we are doing experiments, not knowing what will happen but hoping that an intervention will be effective. As Feinstein clearly pointed out, all the outcomes from various comparative methods are basically stochastic, and quantitative boundaries cannot be rigidly fixed. What is the real significance—in its most literal sense—of the magnitude of an observed difference [12]? Hence, decision making by the DMSC to stop a trial because of detected harm is complex and inherently limited.

3. Tackling uncertainties and the play of chance

Trials are there to combat therapeutic ignorance [13]. This disputed trial was intended to confront the uncertainties about the effect of probiotics in a defined class of patients. However, for reliable definite answers on research questions, we must pass a phase of insecurity. Especially in the incipient stage of the trial with small numbers of patients, chance may well account for differences in outcomes between groups. With increasing numbers of participants, the limits of oscillations will accordingly narrow. Examples of early signs of apparent benefit or harm disappearing when the study was continued are known, described as ''regression to the truth'' [5,8,14]. Trials could therefore be stopped too early for the wrong motive [5,6,13].

There are several reasons for a DMSC to stop an ongoing trial: evident superiority of the experimental treatment, futility of the trial or emerging harm [6,8]. However, to stop a trial for any of these reasons, we must avoid jumping to conclusions, and proof beyond reasonable doubt is clearly needed [14]. Stopping because of harm is particularly difficult, emotionally and rationally [5]. A worrying negative trend brings about a psychological obstacle, because the ideas and expectations based on earlier studies were pointing in an opposite direction. This blockage has to be removed.

Then there is the logic of the process. Predefined statistical boundaries are a directive for decision making, but not the sole ground for it. A totality of evidence is needed to stop a trial prematurely, drawn from both the current trial and from previous studies and external information that may arise during the conduct of the study that raises serious safety issues [14]. Both numeracy and wisdom are required for members of a DMSC and the two keywords of this whole process are: uncertainty and complexity. Yet, before we can get to any conclusions, we will always need a sufficient number of patients.

4. Troubling torts

Human subject research is increasingly confronted with litigation [15]. A bad outcome always is the prime reason for such a liability action. But a well-thought-out legal opinion must take into account both the processes and outcomes of the work of the DMSC. However, while looking back, the horrific ending gives rise to outcome bias and some sort of blinded review of the process is to be preferred [3]. As the psychologist Edwards formulated it more than 20 years ago: ''A good decision cannot guarantee a good outcome. All real decisions are made under uncertainty. A decision is therefore a bet, and evaluating it as good or not must depend on the stakes and the odds, not on the outcome [16].'' Hence, the prime thing to scrutinize is the process, for which there are legal and procedural regulations and scientific schemes. In a trial, both the expert and the judge have to apply a standard of reasonably prudent care, and the DMSC cannot be blamed for negligence for using this benchmark without hindsight. As discussed earlier,

we try hard to model these stakes and odds mathematically. But is there any such method that will reasonably tell us in time when things go wrong? Are nasty surprises, such as this one, of the probiotic trial preventable at all times? DMSCs are accountable for safety monitoring, but they are not in control of it [17].

On the one hand, consumer groups—the potential patients—increasingly ask for innovative efficient and safe medical research. On the other hand, society is increasingly risk aversive. As a consequence of this, we observe a rise of litigation, and this may lead to more rules for and restrictions on human subject research that may ultimately impede scientific progress [15].

5. Living forward

There is much at stake here. Chalmers recently urged to increase our efforts to protect patients from unproven but used treatments. Despite the tremendous progress of medicine, there are still so many therapeutical questions to be answered. Changing attitudes and growing restrictions may curtail therapeutic research [13]. But how to respond to the problem of minimizing the risk for participants in trials? As discussed, the play of chance may well limit our abilities to avert harm. There has been much theorizing on different statistical models. But have we, while using data of already completed trials, compared the results of these methods at interim analyses? Proponents of the Bayesian approach claim a continuous learning process enabling early modification of trials that could help to decide when to stop a therapeutic trial [7,18]. Others however are skeptical of this methodology [19].

The escalation of litigation urges us to further reflect on our prospective and retrospective methods. The famous words from the Danish philosopher Kierkegaard are especially pertaining to this safety monitoring practice of clinical trials: ‘‘Life can only be understood backward, but it must be lived forward.’’

References

[1] Besselink MG, van Santvoort HC, Buskens E, Boermeester MA, van Goor H, Timmerman HM, et al. Dutch Acute Pancreatitis Study Group. Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. Lancet 2008;371:651–9.

[2] Gill RD. Statistics, ethics, and probiotica. Statistica Neerlandica 2009;63:1–12. Available at: http://arxiv.org/abs/0804.2522v1.

[3] Dekker S. The field guide to understanding human error. Aldershot: Ashgate; 2006.

[4] Henriksen K, Kaplan H. Hindsight bias, outcome knowledge and adaptive learning. Qual Saf Health Care 2003;12(Suppl 2): ii46–50.

[5] Pocock SJ. Current controversies in data monitoring for clinical trials. Clin Trials 2006;3:513–21.

[6] Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic. J Clin Epidemiol 2008;61:241–6.

[7] Berry DA. Bayesian clinical trials. Nat Rev Drug Discov 2006;5: 27–36.

[8] Todd S, Whitehead A, Stallard N, Whitehead J. Interim analyses and sequential designs in phase III studies. Br J Clin Pharmacol 2001;51: 394–9.

[9] Strech D, Tilburt J. Value judgments in the analysis and synthesis of evidence. J Clin Epidemiol 2008;61:521–4.

[10] Besselink MG, Timmerman HM, Buskens E, Nieuwenhuijs VB, Akkermans LM, Gooszen HG. Dutch Acute Pancreatitis Study Group. Probiotic prophylaxis in patients with predicted severe acute pancreatitis (PROPATRIA): design and rationale of a double-blind, placebo-controlled randomised multicenter trial [ISRCTN38327949]. BMC Surg 2004;4:12.

[11] Pocock SJ. The simplest statistical test: how to check for a difference between treatments. BMJ 2006;332:1256–8.

[12] Feinstein AR. P-Values and confidence intervals: two sides of the same unsatisfactory coin. J Clin Epidemiol 1998;51:355–60.

[13] Chalmers I. Confronting therapeutic ignorance. BMJ 2008;337: 246–7.

[14] McCartney. Leaping to conclusions. BMJ 2008;336:1213–4.

[15] Mello MM, Studdert DM, Brennan TA. The rise of litigation in human subjects research. Ann Intern Med 2003;139:40–5.

[16] Edwards W. How to make good decisions. Acta Psychol 1984;56: 5–27.

[17] DeMets DL, Fleming TR, Rockhold F, Massie B, Merchant T, Meisel A, et al. Liability issues for data monitoring committee members. Clin Trials 2004;1:525–31.

[18] Kpozèhouen A, Alioum A, Anglaret X, Van de Perre P, Chêne G, Salamon R. Use of a Bayesian approach to decide when to stop a therapeutic trial: the case of a chemoprophylaxis trial in human immunodeficiency virus infection. Am J Epidemiol 2005;161: 595–603.

[19] Howard G, Coffey CS, Cutter GR. Is Bayesian analysis ready for use in phase III randomized clinical trials? Beware the sound of the sirens. Stroke 2005;36:6122–3.