The grey zone of truth

Maarten L. Simoons*

Department of Cardiology – Thoraxcenter, Erasmus Medical Center Rotterdam, ’s Gravenskipswal 230, PO Box 2040, Rotterdam 3000 CA, The Netherlands

Online publish-ahead-of-print 6 February 2014

Registration of new drugs and devices, and our guidelines for management of cardiovascular disease, are mostly based on the results of clinical trials. Trial results are reported in great detail, with primary and secondary endpoints, event rates in the relevant treatment groups, relative risks or odds ratios with confidence intervals, and the corresponding statistical analyses. However, the busy clinician may remember only whether the trial was ‘significant’ ($P < 0.05$) or ‘non-significant’, white or black.

Some trial endpoints are clear and straightforward, such as total mortality. Other endpoints, however, require interpretation of clinical data in order to decide whether death was from cardiovascular disease, whether a myocardial infarction or ischaemic stroke occurred, whether hospital admission was for heart failure, or whether bleeding was ‘minor’ or ‘major’. This judgement can be made by the physician who treated the patient in a trial (investigator), or it may be made by an independent central ‘Clinical Event Committee’ (CEC). At first glance, it might be expected that such judgement should be consistent, since both the investigator and the CEC interpret the same data. However, the investigator may not be aware of the precise definitions of the events specified in the trial protocol. For example, in recent years the definition of myocardial infarction has evolved, with three reports by the European Society of Cardiology (ESC) together with several other organizations.1–3 In particular, different arbitrary choices have been made for the definition of procedure-related myocardial infarction. In the first international report, any elevation of markers of myocardial necrosis above the upper reference limit after a percutaneous coronary intervention (PCI) procedure was labelled ‘infarction’, while in the second and third reports an elevation exceeding three times and five times the upper reference limit, respectively, was arbitrarily chosen as a threshold to define procedure-related infarctions (Figure 1). This reflects differences in opinion and new evidence about the clinical implications of procedure-related myocardial necrosis.4–6 Similarly, it requires an expert to distinguish the different definitions for ‘major bleeding’ according to the International Society for Thrombosis and Hemostasis (ISTH),7 TIMI,8 GUSTO,9 and those in specific study protocols10 (Figure 2). The members of a CEC are trained to classify possible events such as a myocardial infarction, bleeding, stroke, or hospital admission according to the specific criteria specified in the study protocol, while the local investigator is likely to apply ‘clinical’ judgement.

Because the interpretation of clinical data may differ between the investigator and the CEC, the outcome of a trial may vary when different criteria (investigator call or CEC evaluation) are applied (Table 1). For example, the EPIC study, the first study to document the value of glycoprotein (GP) IIb/IIIa receptor blockers in patients undergoing PCI, was stopped upon recommendation of the Data Safety Monitoring Committee because of a major reduction in the primary endpoint, which was driven by a reduction in PCI procedure-related infarcts after CEC adjudication. However, if the investigator reports had been used, the treatment effect would have been smaller, and not statistically significant. In contrast, the results of IMPACT-II, with another GP IIb/IIIa receptor blocker, were statistically significant based on the investigator-reported infarcts, but not according to the CEC. Also the events in the CHARM-preserved study were not statistically significant according to the CEC-reported primary endpoint, while the investigator-reported events were significantly reduced.11 In PURSUIT, the difference between the new treatment and placebo was similarly reduced by the CEC evaluation of events, although both the CEC-reported

Figure 1 Summary of definitions for myocardial infarction.

Adapted from Thygesen et al.2 CABG, coronary artery bypass graft; MI, myocardial infarction; PCI percutaneous coronary intervention.

Table 1

<table>
<thead>
<tr>
<th>Marker</th>
<th>Spontaneous</th>
<th>Secondary</th>
<th>Death</th>
<th>PCI</th>
<th>STENT THROMB.</th>
<th>RESTENOSIS</th>
<th>CABG</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>3-5</td>
<td>5-10</td>
<td>&gt;10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The opinions expressed in this article are not necessarily those of the Editors of the European Heart Journal or of the European Society of Cardiology.

* Corresponding author. Tel: +31 10 704 0704, Fax: +31 10 463 5258, Email: msimoons@xs4all.nl

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author 2014. For permissions please email: journals.permissions@oup.com
endpoints and the investigator-reported events were significantly reduced by the GP IIb/IIIa receptor blocker12 (Table 1). More recently, the TRITON and PLATO trials10,13 reported a significant reduction in thrombotic events with two new antiplatelet drugs, which again was driven by a reduction in myocardial infarction evaluated by the CEC, while the findings would not have been significant following the investigator-reported events.14 It should be appreciated that, from the viewpoint of the sponsor, CEC adjudication in some trials ‘improved’ the results (EPIC, TRITON, and PLATO) and favoured registration of the new drug, while in other trials the results became less favourable (IMPACT-II, PURSUIT, and CHARM-preserved). In most trials presented in Table 1, the CEC event rates were higher than the investigator-reported event rates. Thus the CEC identified new events, not recognized as such by the investigators. In some trials, however, the CEC review had little impact (OASIS-5 and CURE) or reduced the event rates (CHARM-preserved and OASIS-5 control groups).

Realizing that event rates in clinical trials often differ if different methods for reporting are applied, the question arises of ‘which is the truth’, or ‘which is the best approach’ Poque et al.15 analysed the effects of event adjudication in 10 trials co-ordinated by McMaster University, and reported no significant differences in outcome of these ‘large, simple’ trials by either the investigator reports or the CEC. They conclude that the processes involved in clinical trials have become increasingly complex and expensive, and that empirical research is required to assess whether event adjudication, monitoring, and source data verification in large clinical trials are really needed. Also Granger et al.11 question the need for a complex adjudication process in large simple trials. In double-blind trials, both the investigators and the CEC will be unbiased and thus their interpretations will not be influenced by the specific treatment (active or placebo) of a given patient. In open trials, however, blinded evaluation by a CEC may be required to avoid biased assessment by investigators who know the actual treatment received by each patient. A closer look at Table 1 reveals that, although the event rates reported by investigators and CECs differ, and in some trials differ substantially (IMPACT-II, PURSUIT, and HOPE-2), the differences between the treatment groups remain directionally the same. In none of these trials was the treatment effect reversed by the CEC assessment. The relative risks or odds ratios were not much affected, although the nominal P-values differed substantially in some trials.

The manner in which events are collected and reported may have major implications for the interpretation of the trial results by the scientific and clinical communities, and in particular for the interpretation by the regulating authorities such as the European Medicines Agency (EMA) and the Food and Drug Administration (FDA). The

### Table 1: Comparison of investigator-reported and Clinical Event Committee-adjudicated events (primary endpoints) in 10 selected clinical trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Investigator Treated</th>
<th>Control</th>
<th>HR</th>
<th>P-value</th>
<th>Adjudicated Treated</th>
<th>Control</th>
<th>HR</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPIC</td>
<td>9.0</td>
<td>12.4</td>
<td>0.73</td>
<td>0.12</td>
<td>8.3</td>
<td>12.8</td>
<td>0.65</td>
<td>0.008</td>
</tr>
<tr>
<td>IMPACT-II</td>
<td>5.5</td>
<td>7.8</td>
<td>0.71</td>
<td>0.018</td>
<td>9.2</td>
<td>11.4</td>
<td>0.79</td>
<td>0.063</td>
</tr>
<tr>
<td>GUSTO-IIIB</td>
<td>8.4</td>
<td>9.6</td>
<td>0.88</td>
<td>0.016</td>
<td>8.9</td>
<td>9.8</td>
<td>0.89</td>
<td>0.058</td>
</tr>
<tr>
<td>PURSUIT</td>
<td>8.1</td>
<td>10.0</td>
<td>0.81</td>
<td>0.001</td>
<td>14.2</td>
<td>15.7</td>
<td>0.91</td>
<td>0.04</td>
</tr>
<tr>
<td>CHARM-pres.</td>
<td>21.4</td>
<td>24.7</td>
<td>0.85</td>
<td>0.028</td>
<td>22.0</td>
<td>24.3</td>
<td>0.89</td>
<td>0.12</td>
</tr>
<tr>
<td>OASIS-5</td>
<td>5.8</td>
<td>5.8</td>
<td>1.01</td>
<td>0.005*</td>
<td>5.8</td>
<td>5.7</td>
<td>1.01</td>
<td>0.007*</td>
</tr>
<tr>
<td>HOPE</td>
<td>14.9</td>
<td>18.3</td>
<td>0.79b</td>
<td>&lt;0.001</td>
<td>14.0</td>
<td>17.8</td>
<td>0.75b</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CURE</td>
<td>9.1</td>
<td>11.4</td>
<td>0.78b</td>
<td>&lt;0.001</td>
<td>9.3</td>
<td>11.5</td>
<td>0.79b</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TRITON</td>
<td>3.3</td>
<td>4.4</td>
<td>0.76</td>
<td>0.08</td>
<td>7.4</td>
<td>9.7</td>
<td>0.76</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PLATO</td>
<td>5.4</td>
<td>5.9</td>
<td>0.92</td>
<td>0.095</td>
<td>5.4</td>
<td>6.4</td>
<td>0.84</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

HR, hazard ratio; pres, preserved.
Treated, new, experimental therapy, control, placebo.
For references, see 11, 14, and 15 from which this table was adapted.

*P-value for non-inferiority.

Relative risk.
methods for event reporting in a specific trial, including the level of monitoring, source data verification, and adjudication, must be agreed upon when the trial is designed and the results must be reported accordingly. The EMA and FDA should verify that the procedures are followed as agreed. Nevertheless the authorities may, and should, evaluate the outcome of a trial taking into account the ‘grey zone’, the statistical significance as well as the clinical relevance of the events in the trial.

Current procedures for data collection, monitoring, source document verification, and event adjudication can make a trial cumbersome and expensive, and may prohibit execution of a specific trial. Therefore, let me repeat the conclusions from others mentioned above that it is time to revisit the procedures to be followed in large phase III trials. Indeed, in some trials, investigators, regulators, and industry may agree to avoid or reduce monitoring, source data verification, and CEC review, to reduce the complexity and costs of the trial. This will reduce the development costs of a new drug, facilitate phase III evaluation of a new drug, and it may reduce the price of the drug once it comes on the market. Also such simplification would facilitate the conduct of necessary trials for which sponsorship or other funding is difficult to find.16

Clinical scientists as well as clinicians should appreciate that there is no absolute truth in a clinical trial, that there remains a ‘grey zone’ of events which may, or arbitrarily may not be considered relevant for the outcome of a trial. For this reason, in many trials of antithrombotic therapy, bleeding rates are reported according to different classifications (ISTR, TIMI, and other major, minor, and clinically relevant bleeding). The experts developing the universal definitions of myocardial infarction2–3 recommend that myocardial infarctions should similarly be reported in different classes, related to the cause and size of myocardial necrosis. The reader can then better appreciate the grey zone and interpret the results more appropriately.

**Conflict of interest:** none declared.

**References**