Summary points

Regulation of drugs by the UK regulatory agency is funded entirely by user fees

Strict rules are in place to ensure staff and committee members have no personal conflicts of interest

Although it has to protect commercial confidentiality, the agency can now disclose assessments used in its decisions

Increased transparency should further reassure the public that the agency's decisions are impartial

selective serotonin reuptake inhibitors in 2003-5 and hormone replacement therapy in 2004.[3 4]

In January 2005, the agency began to publish on the internet the summary data from the yellow card system for reporting adverse drug reactions as drug analysis prints (www.yellowcard.gov.uk/daps.html). It will consider legitimate applications for data on individual yellow card reports for research purposes, subject to ethical and scientific approval.

In addition, from November 2005, the agency will start producing UK public assessment reports on each medicine that it licenses. These will provide details of the clinical trials submitted as part of the application. The agency has also established a communications division headed by a board level director. One of the division's main tasks is the wider dissemination of information about the work of the agency and the safety and efficacy of licensed medicines.

Relationships between government, the health professions, and the public have evolved considerably in recent years. The setting up of the agency has been accompanied by wide ranging reviews of all aspects of medicines regulation. The changes now being implemented are intended to ensure that it continues to meet public expectations of scientific rigour, independence, and transparency in the years to come.

Contributors and sources: AB and KW have both worked in academic clinical pharmacology and the NHS for many years and have considerable experience in medicines regulation.This article arose from discussions on changes in UK medicines regulation. Both authors contributed to the writing of the article. AB is guarantor for the content of the article.

Competing interests: None declared.

1 House of Commons Health Committee. *The influence of the pharmaceutical industry. Fourth report of session 2004-05*. London: Stationery Office, 2005.
2 National Audit Office. *Safety, quality, efficacy: regulating medicines in the UK*. London: Stationery Office, 2003.
3 Committee on Safety of Medicines. *Report of the CSM expert working group on the safety of selective serotonin reuptake inhibitor antidepressants*. London: Stationery Office, 2005.
4 Medicines and Healthcare Products Regulatory Agency. Review of the evidence on long term safety of hormone replacement therapy. *Current Problems in Pharmacovigilance* 2004;30:4-7.

(Accepted 11 August 2005)

How well does the evidence on pioglitazone back up researchers' claims for a reduction in macrovascular events?

Nick Freemantle

Recent claims that pioglitazone prevents macrovascular events are based on a secondary outcome measure. But ignoring the primary outcome is statistically unsound

University of Birmingham, Birmingham B15 2TT
Nick Freemantle *professor of clinical epidemiology and biostatistics*
N.Freemantle@bham.ac.uk

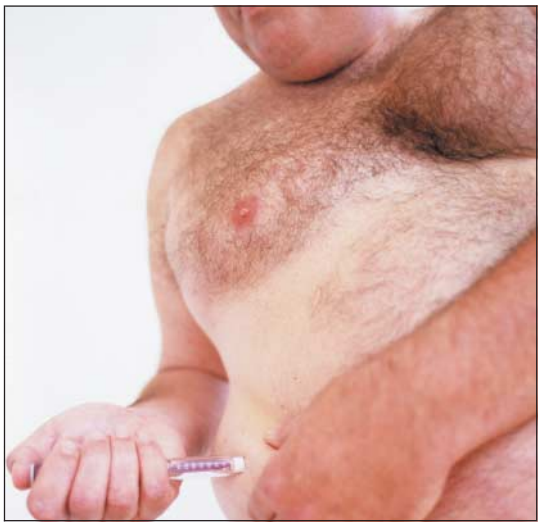
BMJ 2005;331:836–8

Last month, members of the steering committee of the prospective pioglitazone clinical trial in macrovascular events (Proactive) presented the results at the European Association for the Study of Diabetes meeting in Athens.[1] The audience, which overflowed from the meeting room, heard John Dormandy, chair of the steering committee, conclude that the trial had shown that pioglitazone, "Reduces the composite of all cause mortality, non-fatal myocardial infarction, and stroke." He commented: "We have now shown for the first time that oral glucose lowering medication can prevent macrovascular events." The audience seemed excited by these results and a consensus emerged that the results would change practice. The presentation was certainly positive and upbeat (as readers may judge for themselves from the webcast made available with the support of the study sponsors, Eli Lilly and Takeda[1]). Unfortunately, these conclusions are not based on robust standards for the interpretation of evidence from clinical trials.

The trial

The trial studied over 5000 patients with inadequately controlled type 2 diabetes randomised to receive pioglitazone or matched placebo. Participants had raised cardiovascular risk, and most were receiving treatment for cardiovascular disease. The minimum planned exposure to study treatment was 2.5 years. The trial seems to have been carried out to a high standard, as we should expect from an industry sponsored trial that was conducted largely to answer safety concerns among regulatory agencies. So why are the conclusions unsafe?

The answer lies with the choice of composite outcome measure and the undue emphasis given to a secondary end point which provided contrasting results to that from the prespecified primary outcome. It is well known that multiple testing can lead to spurious results. Each statistical test in a neutral trial is the equivalent of rolling a 20 sided dice on which one side is denoted as a success. The more times the dice is



IAN HOOTON/SPL

Oral glucose lowering drugs may prevent patients needing insulin injections but the evidence on macrovascular events is still questionable

rolled, the greater the chance of success ending face-up at some point. This problem is avoided by predefining a primary outcome measure, which is the single test used to calculate type 1 error.[2] Of course, the primary outcome must be defined before the data are available and any analysis of results is performed, as was the case in the Proactive trial.

It is increasingly popular (and sensible) to identify a principal secondary outcome, defining where to look next. The primary outcome is not necessarily the most important outcome clinically in a trial, but it is the most important outcome statistically and is the one on which the main interpretation of the trial is based. During the presentation of the Proactive results in Athens we saw the primary outcome being outcast, like a crazy aunt,[2] because it didn't give the desired answer.

Assessing composite outcome measures

The primary outcome in the Proactive study was the composite of all cause mortality, non-fatal myocardial infarction (including silent myocardial infarction), stroke, major leg amputation (above ankle), acute coronary syndrome, coronary artery bypass graft or percutaneous coronary intervention, and leg revascularisation. The P value reported at the conference for this outcome was $P=0.10$, which is above the maximum conventional value for significance (0.05).

The principal secondary outcome was the composite of all cause mortality, non-fatal myocardial (excluding silent myocardial infarction), and stroke. The P value for this was described to be significant (0.03), and the conclusions were drawn from this finding. However, when the primary outcome is not significant, all the available α or type 1 error has been "spent," and none is left over for the principal secondary outcome. In other words, the secondary outcome is only nominally significant and should in all but exceptional circumstances be considered exploratory and hypothesis generating rather than hypothesis testing.

The correct interpretation of clinical trial results is not simply a case of following a rule book, although careful attention to the rules can help prevent inappropriate conclusions. Those with bayesian leanings seem to be unimpressed with the concept of α spending, and in any case exceptionally it is appropriate to reach different conclusions on the results of a trial from those described by the primary outcome. For example, the regulatory programme of trials conducted for carvedilol in heart failure used a six minute walk test as the primary outcome.[3] This proved a poor choice, as across five randomised trials considered by the US Food and Drug Administration the primary outcome was consistently neutral, although there were statistically overwhelming benefits in the secondary outcomes describing all cause mortality, left ventricular remodelling, New York Heart Association classification, patient and physician global improvement scales, hospital admission, and heart failure symptom score. Indeed, the six minute walk test was one of only two outcomes that were not highly significant across the trial programme. The FDA considered the case for the licensing of carvedilol in heart failure to be sufficient to set aside their standard requirement of two randomised clinical trials showing significant results on the primary outcome measure.

The conclusions drawn from the Proactive trial are based on a much weaker premise, as I will explain below. Composite outcomes have the advantage of increasing the statistical power of time to event analyses, but only when the included outcomes move in the same direction. In addition, they avoid the need to select a single outcome when several related outcomes may be expected to reflect the effects of a treatment. They also have disadvantages, principally in interpretation and when, unexpectedly, the selected components of the outcome do not all reflect treatment modifying effects.[4] Composite outcomes are most useful when they reflect a common biological process and when they can be referred to with an understandable single label—for example, macrovascular events.

The Proactive trial included two definitions of macrovascular events in the primary and principal secondary outcomes. When the first one did not work out, we were offered a second, along with strongly put arguments that the second definition was to be preferred. But these arguments were made after the data had been analysed. Had the effects of treatment been real and substantial we could have expected consistent results across all important cardiovascular outcomes. For example, if pioglitazone really reduces

Summary points

The Proactive trialists claim to have shown that pioglitazone reduces macrovascular events

The results for the primary composite outcome measure were insignificant

Conclusions based on the secondary outcome do not have sufficient statistical strength to prove an association

Judgment should be reserved until the results are published in an academic journal

macrovascular events, it is surprising that it had no effect on all cause mortality, especially given that over 350 deaths were observed in the trial.

Further review

The results of the trial will be published in the *Lancet*. Publication should enable a more informed and detailed debate on the safety and efficacy of pioglitazone in poorly controlled type 2 diabetes, and, hopefully, a shift from sound bite to science.

Contributors and sources: NF has worked as the study statistician on several drug and device trials in diabetes and other clinical areas. He has undertaken methodological work in the interpretation of clinical trials and the use of composite outcome measures. He was in the audience for the presentation of the Proactive study, and the article arose from this and subsequent discussions.

Competing interests: NF has received funding for research and consultancy from a number of pharmaceutical and device companies that manufacture products for diabetes, including Takeda Pharmaceuticals, which co-sponsored Proactive. He has received research funding from the UK Department of Health and various medical charities for relevant work. He is an editorial adviser to the *BMJ*.

1 Official PROactive results website www.proactive-results.com/index.htm (accessed 15 Sep 2005).
2 Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ* 2001;322:989-91.
3 Fisher LD. Carvedilol and the Food and Drug Administration (FDA) approval process: the FDA paradigm and reflections on hypothesis testing. *Controlled Clin Trials* 1999;20:16-39.
4 Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;289:2554-9.

(Accepted 21 September 2005)

At the frontier of biomedical publication: Chicago 2005

Kristina Fišter

Last month the fifth congress on peer review and biomedical publication was held in Chicago. The presentations highlighted that we still have plenty of room to improve the quality of published research

BMJ, London WC1H 9JR
Kristina Fišter
Roger Robinson editorial registrar
kfister@bmj.com

BMJ 2005;331:838–40

Evidence started to matter in biomedical publishing soon after it came to matter in medicine—relatively recently. The first international congress on peer review and biomedical publication was held in Chicago in 1989. At the time of the third congress, in 1997, only 146 original scientific articles had been published on peer review, of which 22 were prospective studies and 11 randomised controlled trials.[1] Since then, the body of evidence has been growing, with about 200 abstracts indexed in Medline a year.[2] We now have plenty of evidence to support the contention that peer review is "expensive, slow, subjective and biased, open to abuse, patchy at detecting important methodological defects, and almost useless at detecting fraud or misconduct."[3] The evidence on how to improve the process is scarce. What did the fifth congress add?

Industry funding

Some of the presented research looked into what happens when the pharmaceutical industry sponsors meta-analyses—the top of the hierarchy of evidence. Yank and colleagues analysed the agreement between results and conclusions in 71 meta-analyses of antihypertensive drugs published between 1966 and 2002.[4] In about a third, authors disclosed financial ties with the pharmaceutical industry. Meta-analyses sponsored by industry were five times more likely than those funded by other sources to report conclusions favouring the study drug when such conclusions were not supported by the results. Meta-analyses funded by academic institutions showed no disagreement between the results and conclusions. Richard Smith, former editor of the *BMJ*, said: "It's a marvellous study and very disturbing." This indicates an embarrassing editorial failure, commented Yank. But she refused to be drawn on the identity of the worst offending journals.

Another study compared quality and conclusions in pairs of meta-analyses of the same drugs for treating the same disease, one Cochrane systematic review and the other sponsored by the manufacturing drug company.[5] Despite the limitations—only eight pairs of meta-analyses met the inclusion criteria and the study wasn't blinded—the results were compelling. None of the Cochrane reviews and all of the industry sponsored meta-analyses concluded without reservation that the study drug was better than the comparison treatment. "Patients—to the barricades," said Peter Gøtzche towards the end of his presentation.