

Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study

Jeppe Lerche la Cour, physician,^{1,3} Jesper Brok, paediatrician,^{1,2} Peter C Gøtzsche, professor³

¹Copenhagen Trial Unit, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark

²Pediatric Department, Hvidovre Hospital, Copenhagen

³Nordic Cochrane Centre, Rigshospitalet and University of Copenhagen

Correspondence to: J L la Cour
jeppelerche@hotmail.com

Cite this as: *BMJ* 2010;341:c3653
doi:10.1136/bmj.c3653

ABSTRACT

Objective To assess if authors of randomised clinical trials convey the fact that they have used surrogate outcomes and discussed their validity.

Design Cohort study.

Setting Six major general medical journals.

Participants Randomised clinical trials published in 2005 and 2006 that used a surrogate as a primary outcome.

Results Of 626 published randomised clinical trials, 109 (17%) used a surrogate as a primary outcome. Of these trials, 62 (57%, 95% confidence interval 47% to 67%) clearly reported that the primary outcome was a surrogate. Only 38 (35%, 26% to 45%) also discussed the validity of the surrogate.

Conclusion Only about one third of authors of randomised clinical trials that used a surrogate as a primary outcome reported adequately on the surrogate. Better reporting is needed.

INTRODUCTION

Surrogate outcomes may seem an attractive alternative to clinically relevant outcomes in clinical trials.¹ The reasons for this are that there may be more events and that the surrogate may occur faster or be easier to assess, thereby shortening a trial and sometimes making it more ethically acceptable.^{2,3} Surrogate outcomes are, however, associated with many problems. The main one arises when the intervention causes a “positive” response on the surrogate but has no effect, or a harmful effect, on the clinical outcome, or vice versa. Such discrepancies can lead to implementation of harmful treatments or to exclusion of beneficial interventions.

A surrogate outcome has been defined as “a laboratory measurement or a physical sign used as a substitute for a clinically meaningful outcome that measures directly how a patient feels, functions or survives.”¹ Examples are high cholesterol concentrations for mortality,³ glycated haemoglobin for complications from diabetes,⁴ or CD4 cell counts for AIDS related morbidity.³ These measures are all considered to be in the causal pathway for the clinical outcome. Sometimes, however, the surrogate may not be causally or strongly related to the clinical outcome, but only a concomitant factor, and thus may not predict the effect on the clinical outcome.³

Some criteria influence the validity of a surrogate outcome: if the surrogate has been shown to be in the causal pathway, it is considered biologically plausible^{1,3}; if a correlation has been established between the surrogate and the clinical outcome (for example, in observational studies)^{1,3,5}; and if one or, even better, many randomised trials on similar drug classes have established that the intervention’s effect on the surrogate captures the intervention’s effect on the clinical outcome.¹ Several statistical methods are used to assess these criteria and if they are fulfilled the validity of the surrogate is increased.⁵ However, even trials with thoroughly validated surrogates cannot match trials with clinical outcomes, as some uncertainty will always remain. Therefore results from trials based on surrogates can be difficult for doctors to apply in clinical practice.

Clinicians, and particularly pharmaceutical companies, often do not distinguish clearly between surrogate and clinical outcomes. It is therefore important that clinical trialists report whether they have used a surrogate and convey this information to the readers. Furthermore, readers will want to know if the surrogate has been validated and is sufficiently reliable to warrant changes in clinical practice, provided the harms of the treatment are acceptable.

We evaluated if authors of randomised clinical trials convey the fact that they have used a surrogate as a primary outcome and have discussed the surrogate’s validity. Manuscripts that reported on both were classified as adequate.

METHODS

We searched for all randomised clinical trials published in 2005 and 2006 in *JAMA*, *New England Journal of Medicine*, *Lancet*, *BMJ*, *Annals of Internal Medicine*, and *PLoS Medicine*. We chose these publications as they are broad journals with high clinical impact and their articles are easy to retrieve. We searched for the journal names through PubMed, limiting the search to “randomised controlled trials” and publication in 2005 and 2006.

We included only randomised clinical trials that used a surrogate as a primary outcome. An outcome was classified as a surrogate if it did not directly measure how a patient feels, functions, or survives—for

example, bone mineral content, level of liver enzymes, or results of a urine dipstick test.

Some subjectivity is involved in judging whether an outcome is truly a surrogate. All borderline cases were discussed among the authors, and only if all agreed was the outcome classified as a surrogate. We excluded trials that had used a primary clinical outcome (for example, mortality, nausea, daily function skills), those with a composite outcome if one or more of the individual outcomes were clinical, and those that had cost as a primary outcome.

Data extraction

We extracted data on characteristics of the trial and authors' comments. Trial characteristics included type of surrogate, journal source, main clinical area, sample size, length of follow-up, type of sponsor (for profit, mixed funding, or non-profit),⁶ and whether the experimental intervention was recommended for clinical use.⁶

Authors' comments included whether the authors: (1) reported that they had used a surrogate outcome—for example, the outcome was labelled as a “surrogate outcome,” “intermediate outcome,” or “non-clinical outcome,” or it was clearly understood in the context of the article that the outcome was a surrogate; and (2) reported on the surrogate's validity—for example, it was mentioned whether the surrogate outcome was considered validated or not validated, or the level of association, relation, or connection between the surrogate and clinical outcomes were discussed.

We noted only what the authors reported in the publication on validity of the surrogates. We did not assess the methods used for validation of the outcomes.

Statistical analyses

Our two main outcomes (reporting on use of a surrogate outcome and of its validity) are presented as proportions with 95% confidence intervals.

We explored if there was an association between authors reporting adequately and type of sponsor (for profit versus mixed funding and non-profit pooled together), and whether the treatment was recommended for clinical use. Exploratory analyses were carried out using Fisher's exact test, with $P < 0.05$ considered significant.

Procedure

One researcher (JLC) screened all abstracts and excluded those clearly reporting the use of a primary clinical outcome. For the remaining abstracts the full article was retrieved and assessed for eligibility. If no primary outcome was specified, we regarded the outcome used to estimate sample size as primary. We included trials with a surrogate used as a primary outcome. Data were extracted by one researcher (JLC) and entered into Excel sheets based on experience from a pilot study. JB or PG reassessed all extracted data. Disagreements about eligibility or data were resolved by discussion between all three authors. In case of no consensus on eligibility, we excluded the trial.

RESULTS

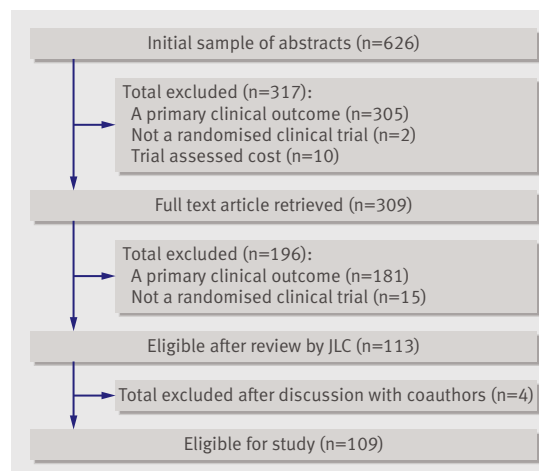
Overall, 626 citations were identified. Of these, 513 were excluded: 486 did not use a surrogate as a primary outcome, 17 were not randomised clinical trials, and 10 assessed cost (figure). Four of the remaining 113 trials were excluded after discussions among the authors, leaving 109 eligible trials (17% of all randomised clinical trials from the included journals). The table shows the characteristics of the included trials.

In 62 of the 109 trials (57%, 95% confidence interval 47% to 67%)^{w1-w109} the authors clearly conveyed that they had used a surrogate as a primary outcome: for example, “We selected a surrogate outcome, commonly used in studies of contrast induced nephropathy”⁷ and “Trial outcomes were cardiovascular disease risk factors, not clinical events.”⁸

Of these 62 trials, 38 (61%, 53% to 69%) discussed if their surrogate was validated or not: for example, “The measurement of carotid intima-media thickness is among the best validated of these surrogate endpoints”⁹ and “The relationship between efficacy measures assessed with the use of intravascular ultrasonography and clinical outcomes has not been fully explored.”¹⁰ Accordingly, a total of 38 trials (35%, 26% to 45%) reported adequately on surrogate outcomes.

Thirty three trials (30% of total) were sponsored by for profit organisations. Fifteen of these (45%) reported adequately on surrogates—that is, reported on both the use of and the validity of the surrogate. Seventy six trials (70% of total) were sponsored, 45 (59%) by mixed sponsors and 31 (41%) by non-profit organisations. Twenty three (30%) trials sponsored by non-profit organisations or mixed sponsors reported adequately on surrogates. Trials funded by for profit organisations were not significantly more likely to report inadequately on surrogates than trials funded by non-profit organisations (relative risk 0.67, 95% confidence interval 0.40 to 1.10).

The authors recommended the experimental intervention in 29 trials (27% of total). Of these, six trials (21%) reported adequately on surrogates. Eighty of the trials (73% of total) did not recommend the



Selection of trials

Characteristics of included randomised clinical trials that used a surrogate as primary outcome

Characteristics	No (%) of trials
Journal source:	
<i>New England Journal of Medicine</i>	45 (41)
<i>Lancet</i>	21 (19)
<i>JAMA</i>	18 (17)
<i>Annals of Internal Medicine</i>	13 (12)
<i>BMJ</i>	10 (9)
<i>PLoS Medicine</i>	2 (2)
Clinical area:	
Cardiovascular	27 (25)
Infectious disease	20 (18)
Endocrinology	13 (12)
Gastrology and hepatology	11 (10)
Pulmonology	8 (7)
Obstetrics	7 (6)
Nephrology and urology	5 (5)
Oncology	4 (4)
Neurology	3 (3)
Other	11 (10)
Sample size:	
0-100	17 (16)
>100-300	30 (28)
>300-1000	41 (38)
>1000-5000	17 (16)
>5000	4 (4)
Type of surrogate:	
Blood sample	40 (37)
Imaging	27 (25)
Lung function	6 (5)
Other	36 (32)
Length of follow-up:	
0-30 days	13 (12)
>30 days-6 months	33 (30)
>6 months-1 year	29 (27)
>1 year	34 (31)

intervention and 33 of these trials (40%) reported adequately on surrogates. Trials reporting inadequately on surrogates were not significantly more likely to recommend the intervention than trials reporting adequately on surrogates (1.93, 0.90 to 4.14).

DISCUSSION

One in five randomised clinical trials published in six general major medical journals during 2005 and 2006 used a surrogate as a primary outcome. Of these, only one third reported adequately that they had used a surrogate and discussed the surrogate's validity. In contrast with our hypothesis, we did not find that trials sponsored by for profit organisations or trials that recommended the experimental intervention were more likely to report inadequately on surrogates. However, our study had limited power for this comparison.

Strengths and weaknesses of the study

We assessed all published randomised clinical trials (n=626) during two years in six major general medical

journals. We chose these journals because they have a high impact on clinical decisions and because they do not focus on a single discipline of medicine such as journals oriented to a particular specialty. Hence the readers cannot be expected to have prior knowledge about surrogates beyond their clinical area, which strengthens the need to mention when a surrogate has been used and to discuss its validity.

Only one researcher screened all trial abstracts for eligibility. This may have led to a few relevant trials that used borderline surrogates being overlooked, but clear cut cases were not likely to have been missed. Two authors independently assessed the trials for final inclusion. Four trials were excluded because no consensus was reached. This raises the possibility that the prevalence of trials using surrogate outcomes is marginally higher than our finding.

Two authors independently extracted data to minimise the risk of errors. For seven trials, disagreements occurred about whether the authors had reported adequately on the use of a surrogate or its validation, but consensus was obtained through discussion.

It is not always clear whether trialists report adequately on the use and the validity of surrogates. Sometimes it is conveyed in the context of the publication—for example, “though whether such a small difference is clinically relevant is doubtful.”¹¹ In this example, the validity of the surrogate was not reported directly, but the authors did clearly state that the benefit shown on the surrogate did not necessarily translate into clinical benefit. In such cases, we erred on the conservative side and classified the study as showing adequate reporting.

Comments on results

To our knowledge, this is the first study to assess the extent to which surrogate outcomes are used. However, the number of included trials was moderate. Given the shortcomings of surrogates it is surprising that they are used as primary outcomes in about one fifth of published randomised clinical trials. One reason may be the involvement of for profit organisations in many trials. These organisations have an interest in using surrogate outcomes, as it shortens the trial, makes it less costly, and speeds up the implementation of new interventions. Also, trialists may not consider if the outcome supports what the trial aims to clarify and may therefore, inappropriately, choose a surrogate as a result of lack of guidance on trial design. Other reasons for the frequent use of surrogates are that the trial is small owing to lack of adequate resources, the trial is only hypothesis generating, the surrogate is truly validated (although this can almost always be disputed), or the clinical outcomes are impossible to assess owing to ethical or practical reasons. We did not investigate whether or not the use of surrogates was justified in the included trials.

That about only one third of all trials report adequately on both the use of a surrogate and its validity is notable, especially considering how little it takes to be phrased: “compared with glimepiride, pioglitazone reduced carotid artery intima-media thickness progression, a validated surrogate end point for coronary

WHAT IS ALREADY KNOWN ON THIS TOPIC

A surrogate outcome can sometimes be a relevant substitute for a clinically meaningful outcome in randomised trials

Uncritical use of surrogate outcomes can be misleading and has resulted in implementation of harmful interventions

New drugs can be approved for commercial use in Europe and the United States on the basis of surrogate outcomes

WHAT THIS STUDY ADDS

Surrogate outcomes are often used as main outcomes in randomised trials

Only one third of authors of randomised clinical trials that use surrogate outcomes report adequately on the surrogate.

Better reporting on surrogates is needed to avoid misleading conclusions and uncritical acceptance of new treatments

artery disease and cardiovascular risk.”¹⁴ Reasons for inadequate reporting may be that authors do not consider it important, focus and guidance on this issue is minimal, or the statistical validation of surrogates is considered complicated.⁵ We would expect less adequate reporting on surrogates in small specialty journals, in part because the use of certain surrogates has been tacitly assumed to be adequate. Therefore, in the opinion of some editors and peer reviewers, authors are not required to mention that the outcome is a surrogate and discuss its validity in every manuscript.

To improve reporting on surrogates, journal editors could guide authors in their instructions to authors, and there could be guidance on surrogates in the CONSORT (Consolidated Standards of Reporting Trials) statement to which many journals refer.^{12 13}

Problems in defining an outcome as clinical or surrogate

Defining an outcome as clinical or surrogate can be difficult. For example, “length of stay in hospital”¹⁴ is intuitively a surrogate outcome for morbidity but might be perceived by some patients and clinicians as a clinical outcome. Furthermore, some patients may argue that they get the best treatment and support at the hospital whereas others prefer home care. Thus, this outcome may not measure directly how patients feel and it is difficult to classify solely as clinical or surrogate.

Another example is incidence of cancer,¹⁵ which depending on study design can be clinical or surrogate. The diagnosis cancer will affect how a patient feels and it is therefore a clinical outcome in most trials. Conversely, the incidence of cancer is a predictor of mortality and can be used as a substitute measure (surrogate) for this clinical outcome in some trials. Thus to classify an outcome as surrogate or clinical trialists need to be aware of what the trial aims to clarify.

Yet another example concerns the ambiguity in determining the cut off for a clinically relevant outcome when dealing with continuous scales. Postpartum haemorrhage is defined as a blood loss of 500 ml or more.¹⁶ It seems obvious that the higher the amount of blood loss, the higher the risk of morbidity, but is 500 ml clinically relevant? If the blood loss makes the

patient afraid, is it then a clinical measure? Or should we choose a cut-off point that reflects when blood loss becomes life threatening? Haemorrhage illustrates the problems with using apparently clinical outcomes measured on continuous scales, as it may be difficult to assess the clinical relevance and classify the outcome as clinical or surrogate.

Validation

In general it is not sufficient to ensure that the surrogate and the clinical outcome are correlated.¹⁷ Ideally, surrogates used in randomised clinical trials should capture the full effect of the intervention on the clinical outcome, and all effects of the intervention should be mediated through or captured by the surrogate.^{5 18} However, even trials based on validated surrogates may not be able to capture unexpected important harmful effects of the intervention. Thus, even in cases with apparently validated surrogates, challenges are faced. Glycated haemoglobin is considered a surrogate of cardiac events in type 2 diabetes. Accordingly, lowering the HbA_{1c} level reduces the risk of cardiac events. However, recently a randomised clinical trial on rosiglitazone (Avandia; GlaxoSmithKline) found that the drug lowered HbA_{1c} levels but increased the composite risk of myocardial infarction, angina, or sudden death.¹⁹ This illustrates that even though a surrogate has been validated for one drug and one clinical outcome, it is not automatically valid for other drugs or clinical outcomes. As with clinical interventions, surrogates that have undergone validation through a meta-analysis approach are more reliable than other surrogates.¹⁵ However, even trials with thoroughly validated surrogates can never be as reliable as trials using clinical outcomes.

Approval on the basis of surrogates

In Europe and the United States new drugs can be approved for commercial use without having shown effects on clinical outcomes. The European Medicines Agency has no specific regulations for this but has guidelines for new drugs, which state that a positive benefit-risk balance is a key requirement to obtain a marketing authorisation in the European Union. This process generally lacks transparency. New drugs can be approved by the US Food and Drug Administration without having shown effects on clinical outcomes, through the accelerated approval programme used for serious illness: “FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on a surrogate endpoint that is reasonably likely, based on epidemiologic, therapeutic, pathophysiologic . . .” or other evidence, to predict clinical benefit.²⁰ It was through this process that rosiglitazone was approved, as HbA_{1c} was considered reasonably likely to predict mortality. This highlights the uncertainty related to approval of drugs based on surrogate outcomes. In the case of serious illnesses where no good treatments exist, these procedures may seem

acceptable, although they will always be risky. When clinically effective similar drugs are on the market, however, it is important to make sure that evidence on clinical outcomes (both benefits and harms) is sufficient before marketing a new drug.

Conclusion

Our study shows that about one in five published randomised clinical trials use surrogate outcomes and that only one third of these report that they have used a surrogate and discuss its validity. Thus readers should be aware of conclusions in randomised clinical trials that are based on hidden surrogates. Better reporting on surrogates might reduce unwarranted conclusions and uncritical acceptance of new treatments.

Contributors: JLC, JB, and PG conceived and designed the study. JLC found the relevant articles. JLC, JB, and PG extracted and analysed data. JLC and JB drafted the review. JB and PG revised the review. JLC is guarantor. All authors had full access to the data and take responsibility for the data analyses.

Funding: This study was supported by the Copenhagen Trial Unit, Centre for Clinical Intervention Research, and the Nordic Cochrane Centre, Denmark. All authors worked independently of any funders. Competing interest: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any company for the submitted work; no financial relationships with any companies that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

- 1 Bucher HC, Guyatt GH, Cook DJ, Holbrook A. Users' guides to the medical literature XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. *JAMA* 1999;282:771-8.
- 2 Gluud C, Krogsgaard K. Would you trust a surrogate respondent? *Lancet* 1997;349:665-6.
- 3 Gøtzsche PC, Torri ALV, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care* 1996;12:238-46.
- 4 Theodore Mazzone T, Meyer PM, Feinstein SB, Davidson MH, Kondos GT, D'Agostino RB, et al. Effect of pioglitazone compared with glimepiride on carotid intima-media thickness in type 2 diabetes: a randomized trial. *JAMA* 2006;296:2572-81.

- 5 Wier CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006;25:183-203.
- 6 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusion in randomised drug trials—a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-8.
- 7 Burns KEA, Chu MWA, Novick RJ, Fox SA, Gallo K, Martin CM, et al. Perioperative N-acetylcysteine to prevent renal dysfunction in high-risk patients undergoing CABG surgery: a randomized controlled trial. *JAMA* 2005;294:342-50.
- 8 Lawrence J, Appel LJ, Sacks FM, Carey VJ, Obarzanek E, Swain JF, et al. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids results of the OmniHeart randomized trial. *JAMA* 2005;294:2455-64.
- 9 Mazzone T, Unverdorben S, Degenhardt R, Bulitta M, Gouni-Berthold I. Effect of policosanol on lipid levels among patients with hypercholesterolemia or combined hyperlipidemia: a randomized controlled trial. *JAMA* 2006;295:2262-9.
- 10 Nissen SE, Tuzcu EM, Brewer HB, Sipahi I, Nicholls SJ, Ganz P, et al. Effect of ACAT inhibition on the progression of coronary atherosclerosis. *N Engl J Med* 2006;354:1253-63.
- 11 Moll E, Bossuyt PM, Korevaar JC, Lambalk CB, van der Veen F. Effect of clomifene citrate plus metformin and clomifene citrate plus placebo on induction of ovulation in women with newly diagnosed polycystic ovary syndrome: randomised double blind clinical trial. *BMJ* 2006;332:1485-90.
- 12 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.
- 13 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
- 14 Crotty M, Whitehead CH, Wundke R, Giles LC, Ben-Tovim D, Phillips PA. Transitional care facility for elderly people in hospital awaiting a long term care bed: randomised controlled trial. *BMJ* 2005;331:1110.
- 15 Cuzick J, Cafferty FH, Edwards R, Møller H, Duffy SW. Surrogate endpoints for cancer screening trials: general principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial. *J Med Screen* 2007;14:178-85.
- 16 Høj L, Cardoso P, Nielsen BB, Hvidman L, Nielsen J, Aaby P. Effect of sublingual misoprostol on severe postpartum haemorrhage in a primary health centre in Guinea-Bissau: randomised double blind clinical trial. *BMJ* 2005;331:723-8.
- 17 Baker SG, Kramer BS. A perfect correlate does not a surrogate make. *BMC Med Res Methodol* 2003;3:16.
- 18 Gluud C, Brok J, Gong Y, Koretz RL. Hepatology may have problems with putative surrogate outcome measures. *J Hepatol* 2007;46:734-42.
- 19 Rosen CJ. The rosiglitazone story—lessons from an FDA advisory committee meeting. *N Engl J Med* 2007;357:844-6.
- 20 Food and Drug Administration. Drug approvals and databases. 2008. www.fda.gov/cder/rdmt/accappr1.htm.

Accepted: 5 May 2010