BMJ

BMJ 2013;346:f457 doi: 10.1136/bmj.f457 (Published 29 January 2013)

RESEARCH

Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study


OPEN ACCESS

Oriana Ciani *PhD candidate*[1], Marc Buyse *chairman*[2], Ruth Garside *senior lecturer*[1], Toby Pavey *research fellow*[3], Ken Stein *professor*[1], Jonathan A C Sterne *professor*[4], Rod S Taylor *professor*[1]

[1]PenTAG, Institute for Health Services Research, University of Exeter Medical School, University of Exeter, Exeter EX2 4SG, UK; [2]International Drug Development Institute, Louvain-la-Neuve, Belgium and Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium; [3]School of Human Movement Studies, University of Queensland, Brisbane, QLD, Australia; [4]School of Social and Community Medicine, University of Bristol, Bristol, UK

Abstract

Objective To quantify and compare the treatment effect and risk of bias of trials reporting biomarkers or intermediate outcomes (surrogate outcomes) versus trials using final patient relevant primary outcomes.

Design Meta-epidemiological study.

Data sources All randomised clinical trials published in 2005 and 2006 in six high impact medical journals: *Annals of Internal Medicine*, *BMJ*, *Journal of the American Medical Association*, *Lancet*, *New England Journal of Medicine*, and *PLoS Medicine*.

Study selection Two independent reviewers selected trials.

Data extraction Trial characteristics, risk of bias, and outcomes were recorded according to a predefined form. Two reviewers independently checked data extraction. The ratio of odds ratios was used to quantify the degree of difference in treatment effects between the trials using surrogate outcomes and those using patient relevant outcomes, also adjusted for trial characteristics. A ratio of odds ratios >1.0 implies that trials with surrogate outcomes report larger intervention effects than trials with patient relevant outcomes.

Results 84 trials using surrogate outcomes and 101 using patient relevant outcomes were considered for analyses. Study characteristics of trials using surrogate outcomes and those using patient relevant outcomes were well balanced, except for median sample size (371 *v* 741) and single centre status (23% *v* 9%). Their risk of bias did not differ. Primary analysis showed trials reporting surrogate endpoints to have larger treatment effects (odds ratio 0.51, 95% confidence interval 0.42 to 0.60) than trials reporting patient relevant outcomes (0.76, 0.70 to 0.82), with an unadjusted ratio of odds ratios of 1.47 (1.07 to 2.01) and adjusted ratio of odds ratios of 1.46 (1.05 to 2.04). This result was consistent across sensitivity and secondary analyses.

Conclusions Trials reporting surrogate primary outcomes are more likely to report larger treatment effects than trials reporting final patient relevant primary outcomes. This finding was not explained by differences in the risk of bias or characteristics of the two groups of trials.

Introduction

Evidence for the effectiveness of treatments should ideally come from randomised clinical trials or systematic reviews of trials that assess final endpoints relevant to patients, such as survival or health related quality of life.[1 2] However, aspects of the design and conduct of randomised clinical trials have been shown to lead to overestimation of treatment effect size. These include inappropriate random sequence generation,[3] inadequate allocation concealment,[4 5] lack of blinding,[6] single centre status,[7 8] and the use of composite outcomes.[9]

Surrogate outcomes are often used in clinical trials as substitutes for final patient relevant outcomes. Advantages of surrogate outcomes over final outcomes are that they may occur faster or may be easier to assess, thereby shortening the duration, size, and cost of trials.[10 11] A key rationale for the use of surrogate outcomes in trials is not only substitution[12] but the prediction of treatment benefit in the absence of data on patient relevant outcomes.[13-15] Several drugs have been licensed on this

Correspondence to: O Ciani oriana.ciani@pcmd.ac.uk

Extra material supplied by the author (see http://www.bmj.com/content/346/bmj.f457?tab=related#webextra)

Details of search strategy
Characteristics of trials included in pair matched analysis
Funnel plot of included studies

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

basis—for example, statins (based on low density lipoprotein levels), AIDS drugs (based on HIV RNA or CD4 count levels), and cancer drugs (based on time to progression or disease-free survival).[16]

Despite the potential appeal of using surrogate outcomes, the use of such trials in policy making remains controversial.[11 17 18] Gefitinib, an orally administered epidermal growth factor receptor tyrosine kinase inhibitor, was approved by the Food and Drugs Administration in the United States for marketing in May 2003 for patients with non-small cell lung cancer based on the surrogate outcome of tumour response rate. The initial approved indication was for the treatment of patients who were refractory to established cancer treatments (both a platinum drug and docetaxel).[19] In 2005, however, data from two clinical studies became available showing no significant survival benefit; the FDA released a new labelling for gefitinib that prevented its use in new patients with non-small cell lung cancer, limiting its usage to only continuation in those patients with cancer who had already taken the medicine and whose doctor believed it was helping them.[20] Although often cited by sceptics, this and other such potentially complete failures of surrogate outcomes[21-23] remain relatively rare.

Given the growing pressure for faster access to innovative treatments for patients, reimbursement decisions for new treatments are now often made at or around the time of licensing, increasing pressure to rely on treatment effects from trials reporting surrogate primary outcomes.[24 25] Such reimbursement decisions often depend on the use of economic modelling to extrapolate treatment effects based on surrogate outcomes into an estimate of cost effectiveness, such as the incremental cost per quality adjusted life year (QALY), the metric currently recommended by the National Institute for Health and Clinical Excellence.[13 26]

We quantified and compared the treatment effects in a sample of randomised clinical trials reporting either a surrogate or final patient relevant primary outcome. We also compared the risk of bias in these two groups of trials.

Methods

We searched Medline through PubMed for randomised clinical trials published in 2005 and 2006 in six high impact (impact factor >14 in 2011 according to ISI Web of Knowledge) medical journals—that is, *Annals of Internal Medicine*, *BMJ*, *Journal of the American Medical Association*, *New England Journal of Medicine*, *Lancet*, and *PLoS Medicine* (see supplementary table 1 for details of search strategy). We purposively chose general (rather than specialist) medical journals as we sought to compare surrogate and final patient relevant outcomes across a broad range of medical conditions. Otherwise our two year sampling frame was based on a recent study that examined the reporting of surrogate outcomes in trials.[27] The study authors provided us with their listing of trials and we checked that this captured all randomised clinical trials from our database search.

Study selection

Two authors (OC, RST) independently undertook the inclusion and exclusion of trials (see box).

We classified studies into two groups according to whether the primary outcome was a surrogate one or a final patient relevant one. A final patient relevant outcome was defined as any outcome that captures "how a patient feels, functions or survives."[30] An outcome was consequently classified as a surrogate if it was a biomarker[12] (for example, low density lipoprotein cholesterol level) or an intermediate outcome (for example, progression-free survival)[13 31] judged to be a substitute for a final outcome. Where a trial did not state outcome primacy, we chose the one used for sample size calculation or the first outcome reported in the results section to be the primary outcome. As some subjectivity is involved in classification of outcomes as surrogate or final, two reviewers (OC, RST) resolved borderline cases by review of the full paper and discussion. To obtain comparable groups of trials using surrogate and final patient relevant outcomes, we used a hierarchical matching process to match each included surrogate outcome trial with a corresponding final patient relevant outcome trial, based on four criteria: the intervention clinical area, clinical population, journal, and publication year (see supplementary file for details of matched studies). After further detailed review of full papers, additional exclusions were necessary to omit trials with mixed primary outcome assessment and trials that were terminated early. As we sought to examine differences in treatment effects between trials using surrogate outcomes and those using final patient relevant outcomes, we excluded equivalence and non-inferiority designs.

Data extraction and risk of bias assessment

Using a predefined data extraction form we extracted data from the included trials on journal, sample size, patient population, type of intervention (drug, medical device, surgical procedure, health promotion activity, other therapeutic intervention[32]), duration of follow-up, centre status (single or multicentre), and sponsor (for profit, not for profit, or mixed[33]). For trials using surrogate outcomes we sought additional information on type of surrogate (imaging, histochemical/biochemical, instrumental, other), whether the authors explicitly 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 what authors reported in the publication on validity of the surrogate outcome.[13 34] We assessed risk of bias in terms of the adequacy of random sequence generation and concealment, statement of double blind placebo controlled trial, and use of intention to treat analyses. One reviewer (OC) initially undertook data extraction and risk of bias assessment, and this was then checked by a second reviewer (TP or RST).

Data analyses

We compared the treatment effects between the two trial types using several analytical approaches. In accord with previous studies, for our primary analysis we sought binary outcomes in each trial recorded as the number of patients and events in each arm.[3-7 35] Outcome events were recoded where necessary so that an odds ratio below 1.0 indicated beneficial effect of the intervention. Metaregression is a technique used to explore the relation between study characteristics (for example, sample size and journal of publication) and effect size.[36] We used random effects logistic metaregression models[35] to estimate ratios of odds ratios and 95% confidence intervals comparing treatment effects in trials using surrogate outcomes and final patient relevant outcomes. Ratio of odds ratios greater than 1.0 implied greater (more beneficial) treatment effects in the trials using surrogate outcomes than in the trials using final patient relevant outcomes. To take account of potential confounding, in our primary analysis we also included an adjusted analysis that incorporated predefined trial level covariates in the metaregression model—that is, clinical area of the treatment

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Inclusion criteria

- Randomised clinical trial
- Publication years 2005-06
- Journals (*Annals of Internal Medicine*, *BMJ*, *Journal of the American Medical Association*, *New England Journal of Medicine*, *Lancet*, *PLoS Medicine*)
- Interventional studies

Exclusion criteria

- Non-interventional studies (e.g. evaluations of screening or diagnostic tests)
- Economic evaluations
- Mixed primary outcomes (i.e. a primary outcome that comprised both a surrogate and final patient relevant outcome*)
- Multi-arm trials
- Secondary analyses
- Early terminated studies
- Equivalence or non-inferiority design
- No analysable data

*Two examples of mixed primary outcomes seen in this study were a composite endpoint of death or vein graft restenosis[28] and a composite of serum creatinine level, end stage renal disease, or death[29]

and patient population, intervention type, sponsor, journal, sample size, and mean follow-up time.

To assess the robustness of the primary analysis, we undertook several sensitivity analyses. Firstly, to maximise the number of studies in our analysis we first included trials that failed to report the number of patients and events in each arm but reported their results as a risk ratio (that is, relative risk, odds ratio, or hazard ratio). A pooled relative risk ratio comparing trials using surrogate outcomes with those using final patient relevant outcomes was estimated with inclusion of these trials. Secondly, for studies reporting continuous outcomes we first calculated a Cohen's standardised mean difference and associated standard errors. We then transformed the standardised mean differences to log odds ratios and combined them with studies reporting binary outcome or risk ratios using random effects meta-analysis.[36 37] Thirdly, we estimated log odds ratios and associated standard errors for each matched pair of trials, using the method of Bucher et al,[38] and combined across trials using random effects meta-analysis. Sensitivity analyses are reported as unadjusted ratio of odds ratios (or relative risk ratios) and adjusted for trial covariates.

In a secondary analysis, we classified the trials using surrogate outcomes and final patient relevant outcomes according to whether the reported result for the primary outcome was positive (the treatment group was superior to control, $P \leq 0.05$), negative (the control group was superior to treatment, $P \leq 0.05$), or neutral (no significant difference between groups, $P > 0.05$). We then compared the outcomes using an unadjusted logistic regression model and a model adjusted for study level covariates.

For both primary and secondary analyses, when not explicitly stated, we considered the latest available follow-up. Trials with multi-arms were included if it was possible and clinically meaningful to pool the number of events across arms towards a unique comparator—for example, different dosages of the same drug versus placebo arm. When treatment effect estimates and their variability were only shown graphically we used the open source software WinDig version 2.5 to extract numbers from the graphically presented information.

A regression test for funnel plot asymmetry was performed to assess small study effects and potential publication bias.[39] All analyses were run in Stata SE version 12.

Results

A total of 639 titles and abstracts were identified, 511 of which were identified as eligible for the study. Of these, 27% (n=137) were judged to use a surrogate primary outcome. After matching and exclusions, 185 trials contributed to the quantitative analyses (fig 1⇓). See the supplementary file for a list of included trials. The fidelity of matching of trials using surrogate outcomes and those using final patient relevant outcomes seemed to be retained in these 185 trials (table 1⇓) and in the subgroup of trials reporting binary primary outcomes (data not shown). In both groups, drug interventions (surrogate: 58%; final: 61%, P=0.33) and not for profit sponsorship (surrogate: 58%; final: 56%, P=0.86) were most common. Trials using final patient relevant outcomes had a larger median sample size than trials using surrogate outcomes (P<0.001) and were also more likely to be multicentre (P=0.01). Follow-up between the two groups of trials did not differ significantly (P=0.73). Although the duration of the trials was similar overall, when chronic conditions, such as cardiovascular disease, cancer, and endocrine disorders, were considered, the trials using surrogate outcomes had a shorter median follow-up (255 *v* 730 days, P=0.03).

Trials with surrogate outcomes used laboratory biomarkers (52/137, 38%) such as prostate specific antigen; imaging (35/137, 25.5%) such as left ventricular ejection fraction; or instrumental endpoints (35/137, 25.5%) such as body weight. Fifteen trials were judged to use an intermediate outcome to substitute for a final outcome (15/137, 11%), such as disease-free survival or rate of ovulation. In fewer than half of the cases (36/84, 43%) authors explicitly stated they used a surrogate outcome and provided criteria or references for its validation in 28 out of 84 trials (35%).

Comparison of treatment effects

Primary analysis

Overall, 134 trials (51 surrogate outcome trials and 83 final outcome trials) reporting binary outcomes in the primary analysis were included. The pooled odds ratio for the primary outcome in the surrogate trials was 0.51 (95% confidence interval 0.42 to 0.60; I^2=91.2%, P<0.001) compared with 0.76 (0.70 to 0.82; I^2=89.8%, P<0.001) in trials using final patient relevant outcome. On average the treatment effect estimate was 47% higher in the trials using surrogate outcomes than in the trials using final patient relevant outcomes (ratio of odds ratios 1.47, 95% confidence interval 1.07 to 2.01, P=0.02). This

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

difference remained after adjustment for characteristics of the trials (1.46, 1.05 to 2.04, P=0.03, table 2⇓).

Sensitivity analyses

After incorporating trials with risk ratios as reported by the authors, 143 trials (57 surrogate outcome trials and 86 final outcome trials) were included. The treatment estimates in the trials using surrogate outcomes remained higher than in the ones using final patient relevant outcomes (unadjusted relative risk ratios 1.38, 95% confidence interval 1.12 to 1.71, P<0.01; adjusted relative risk ratios 1.36, 1.08 to 1.70, P<0.01). After combining continuous and binary outcomes in the overall sample, the estimated ratio of odds ratios showed a similar direction of effect (unadjusted 1.44, 95% confidence interval 0.83 to 2.49, P=0.20; adjusted 1.48, 0.83 to 2.62, P=0.18). A total of 43 pairs of matched trials were available for a paired analysis, with a pooled ratio of odds ratios of 1.38 (1.01 to 1.88, P=0.04).

Secondary analysis

Trials using surrogate outcomes were more likely to obtain a positive result (as stated by the authors) in favour of the treatment (52/84, 62%) than trials using final patient relevant outcomes (37/101, 37%). The odds ratio of reporting a positive result in the trials using surrogate outcomes compared with trials using final patient relevant outcomes was 2.17 (95% confidence interval 1.20 to 3.92, P=0.01) and 2.43 (1.29 to 4.57, P<0.01) after adjustment for the characteristics of the trials.

Influence of trial characteristics

No trial characteristics showed a statistically significant association with treatment effect estimates in bivariate metaregression models, in addition to type of primary endpoint, except for sample size and single centre status. Harbord's modified test showed similar levels of small study bias in both the trials using surrogate outcomes (t=3.63; P=0.001) and the trials using final patient relevant outcomes (t=2.79; P=0.007; see the funnel plot in the supplementary file). We undertook a retrospective analysis including centre status as trial level covariate in the primary analysis adjusted model, resulting in a ratio of odds ratios of 1.28 (95% confidence interval 0.96 to 1.72, P=0.09). There was no evidence of an interaction between the primary analysis ratio of odds ratios of trials using surrogate outcomes compared with trials using final patient relevant outcomes and the characteristics of the trials (fig 2⇓).

Risk of bias

The risk of bias between trials using surrogate outcomes and those using final patient relevant outcomes did not differ significantly (table 3⇓). Further adjustment of the metaregression estimates for these four factors did not change the inference of the primary analysis comparing trials using surrogate and final patient relevant outcomes—that is, risk of bias adjusted ratio of odds ratios 1.45 (95% confidence interval 1.06 to 1.99, P=0.02). No risk of bias characteristic was found to be significantly associated with treatment effect and there was no interaction with the primary analysis ratio of odds ratios of trials using surrogate outcomes compared with those using final patient relevant outcomes (fig 2).

Discussion

We provide empirical evidence that trials using surrogate primary outcomes report larger treatment effects than a matched sample of trials using final patient relevant primary outcomes. We analysed a cohort of randomised clinical trials categorised according to whether their primary outcome was surrogate or a final patient relevant endpoint and matched on the basis of key characteristics of the trials. On average, trials using surrogate outcomes reported treatment effects that were 28% to 48% higher than those of trials using final patient relevant outcomes. Furthermore, we found that surrogate trials were twice as likely to report positive treatment effects as the final outcome trials. These findings were not explained by differences in risk of bias or other trial characteristics and are comparable with the level of exaggeration of treatment effect attributed to inadequate allocation concealment.[5] Although, as anticipated, we found that trials with a patient relevant primary outcome were more likely to have larger sample sizes, two groups of trials had similar average follow-up times. However, when limited to trials of chronic conditions, follow-up was longer for trials using final patient relevant outcomes. Given the range of interventions and outcomes included, substantial statistical heterogeneity was evident in treatment effects in both types of trials.

Comparison of our findings with previous studies

Few studies have empirically compared trials reporting surrogate and final primary outcomes. One study[33] reviewed 324 consecutive cardiovascular trials published in major general medical journals between 2000 and 2005. In accord with the findings of the present study, trials reporting surrogate primary outcomes were more likely to report a positive treatment effect (77 out of 115 trials, 67%) than trials reporting final patient related primary outcomes (113 out of 209 trials, 54%, P=0.02). A recent systematic review of anti-tumour necrosis factor agents for rheumatoid arthritis[40] compared the methodological quality of trials reporting surrogate primary outcomes with those with final primary outcomes. In contrast with the present study, a difference in study quality between the two groups of trials was seen; the mean percentage of items met in the consolidated standards of reporting trials (CONSORT) statement was lower for studies with surrogate outcomes than with final patient relevant outcomes (62.5 *v* 70.7, P=0.03). However, as this systematic review included both randomised and non-randomised trials, and fewer studies with surrogate outcomes were randomised (63% *v* 74%), this finding is likely to be confounded.

Several reasons may explain why trials assessing surrogate endpoints showed larger treatment effects than trials assessing final endpoints. The first relate to small study effects—the tendency for smaller studies to show a large treatment effect.[41] As expected, we found a smaller sample size for trials using surrogate outcomes than for trials using final patient relevant outcomes. However, our results remained consistent after adjustment for the sample size of the trials. A second reason may relate to publication bias. Although we observed substantive small study bias and therefore potential publication bias, the extent of this bias seemed similar across the two sets of trials. Furthermore, it may be argued that published results based on surrogate outcomes are more likely to be positive as the requirements for publication of such results are generally more stringent, whereas results based on (definitive) final endpoints are more likely to be published regardless of the trial's findings. Thirdly, trials using surrogate outcomes may be of lower methodological quality than trials using final patient relevant outcomes and therefore more prone to exaggeration of effect size.[40] However, a comparison of risk of bias between the two

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

groups of randomised clinical trials showed no differences in random sequence generation, allocation concealment, blinding, and intention to treat analysis. Furthermore, our results were consistent after adjustment for these risks of bias dimensions. Fourthly, two recent meta-epidemiological studies have shown that single centre trials are more likely than multicentre trials to lead to larger intervention effects.[7 8] In our sample a higher proportion of the trials using surrogate outcomes were single centre trials compared with the trials using final patient relevant outcomes. However, an additional retrospective sensitivity analysis including adjustment for centre status showed consistent results, with use of surrogate outcomes still associated with larger treatment effects.

Finally, the treatment effect may be truly larger in trials with surrogate outcomes than with final patient relevant outcomes. In the continuum of health outcomes measures, biomarkers and intermediate outcomes can be identified as disease centered measures, reflecting the biology of the disease process and the underlying mechanism of disease.[42] Assuming the surrogate outcomes lie in the causal pathway between the onset of the disease and the final patient relevant outcome, they are generally more proximal (closer) to the disease and therefore more sensitive to the effect of interventions with therapeutic purposes.

Limitations of the study

As our sample of randomised clinical trials was drawn from six high impact general medical journals over two specific consecutive calendar years, the findings may lack generalisability. We purposively chose general medical journals so as to compare surrogate and final patient related outcomes across a range of medical conditions. Although the choice of publication year would not be expected to influence treatment effects, trials published in high impact journals, although contributing a relatively small proportion of all published trials,[43] are more likely to report newsworthy results.[44] However, it is unclear how this would have influenced the generalisability of our findings. High impact journals might be expected to publish trials of lower risk of bias. We observed higher methodological quality of trials in our sample compared with a representative sample of trials indexed in PubMed,[44] therefore it could be argued our findings are less likely to be susceptible to confounding by other aspects of trial methodological quality.

In addition, we compared the treatment effects of a matched sample of randomised clinical trials reporting surrogate primary outcomes and final primary outcomes. Alternatively, we could have compared the treatment effects between surrogate and final outcomes within the same trials or meta-analyses of homogeneous trials (a meta-epidemiological analysis).[45] Although such a "within trial" comparison would minimise confounding by study population, intervention type, and risk of bias, this approach has problems. Firstly, trials are generally powered to detect statistically significant differences in their primary endpoint. Trials with surrogate primary outcomes may be underpowered for final patient relevant outcomes and thus lead to imprecision in the estimation of the comparative treatment effect of surrogate and final outcomes. Secondly, where within trial meta-analysis comparisons of surrogate and final outcomes have been performed, they have been limited to a single treatment (or treatment class) in one specific disease area.[46-49] In this study we sought to address a different question—that is, in the absence of final patient relevant outcomes, what is the potential effect of relying on the treatment effects based on surrogate outcomes across a range of medical conditions and interventions and surrogate outcome types? To maximise their comparability we matched the cohorts of surrogate and final trials on the basis of key study characteristics, such as disease and intervention area.

Finally, the classification of primary outcomes as surrogate or final patient relevant involves an element of subjective judgment. For example, change in body mass index[50] and carbon monoxide confirmed smoking abstinence rate[51] were both classified as surrogate outcomes, having assumed the patient relevant outcomes in these cases to be long term decline in lung function and lung cancer, respectively. To minimise assessment bias, two reviewers independently applied a standard outcome definition of surrogate and final patient relevant outcomes across all trials, with discussion and consensus on any initial disagreements. To our knowledge this is the first empirical study designed to deal with this subject and therefore our results should be verified in another sample.

Implications of the study

The potential for surrogate outcomes to impact on healthcare policy making and the consequent diffusion of treatments into practice is shown by the fact that 27% of the randomised clinical trials identified during the two year study period reported surrogate primary outcomes. In health technology assessment reports, both in the United Kingdom[13] and internationally,[52] some 1 in 20 base their clinical and economic conclusions on evidence from surrogate outcomes alone. That trial based surrogate outcomes can lead to substantive overestimation of treatment effects that would have been seen if evidence on patient relevant outcomes was available is a salutary message for policy makers when weighing up the evidence from use of surrogate outcomes in their licensing and coverage decisions. Several recent drug appraisals by the National Institute for Health and Clinical Excellence have relied on evidence of clinical effect derived solely from surrogate outcomes.[53-55] Our findings reinforce the importance of formally evaluating the acceptability of biomarkers and intermediate outcomes as valid surrogate outcomes and quantifying the association of treatment effect between the surrogate and patient relevant final outcomes and its uncertainty. The statistical validation of surrogates and the quantification of the relation between surrogate and final patient relevant outcomes are key problems tackled in NICE's update of its methodological guidance for technology assessment.[56] The updated version of methods guidance makes several requirements for health technology assessment producers when faced with clinical trials with evidence based on surrogate outcomes (that is, a systematic review of the evidence to support the validity of the surrogate outcome, clear statement of how the relation between surrogate and final outcome is modelled in determining cost effectiveness) and exploration of the additional uncertainty associated with this prediction on cost effectiveness estimates.[26]

Clinical trialists and systematic reviewers need to be clearer in their reporting as to whether outcomes are surrogate or final patient relevant, and appropriately frame any conclusions of superiority of interventions when based on surrogate outcomes alone. Others have recently suggested that guidance on surrogates should be incorporated into the CONSORT statement.[27] Novel and adaptive approaches to trial design are needed that allow surrogate endpoints to continue to be used as primary outcomes, while also providing evidence on their validation against patient relevant outcomes.[57]

Conclusions

In the absence of data on final outcome, policy makers need to interpret intervention effects based on surrogate outcomes with

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

caution. Although our results have highlighted the risks, they support the application of methods for the validation and quantification of the relations between surrogate and final patient relevant outcomes in licensing and reimbursement decisions on new and existing treatments.

Contributors: OC and RST conceived and designed the study. OC screened the titles and abstracts, data extracted papers, ran the analyses, and drafted the manuscript. RST screened the titles and abstracts and checked data extraction and analyses. TP checked data extraction. MB and JAS advised on methods of data analysis. All authors commented on drafts of the manuscript. RST is guarantor.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work. OC is currently receiving a Peninsula College of Medicine and Dentistry Doctoral studentship.

Ethical approval: Not required.

Data sharing: The dataset is available from the corresponding author at oriana.ciani@pcmd.ac.uk.

1 Pocock SJ. *Clinical trials: a practical approach* . Wiley, 1996.
2 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.
3 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
4 Herbison P, Hay-Smith J, Gillespie WJ. Different methods of allocation to groups in randomized trials are associated with different levels of bias. A meta-epidemiological study. *J Clin Epidemiol* 2011;64:1070-5.
5 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
6 Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.
7 Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Ann Intern Med* 2011;155:39-51.
8 Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ* 2012;344:e813.
9 Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007;334:786.
10 Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res* 2008;17:303-40.
11 Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996;125:605-13.
12 Biomarkers Definition Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89-95.
13 Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports. *Int J Technol Assess Health Care* 2009;25:6-13.
14 Burzykowski T, Buyse M. An alternative measure for meta-analytic surrogate endpoint validation. In: Burzykowski T, Molenberghs G, Buyse M, eds. *The evaluation of surrogate endpoints* . Springer, 2005:323-40.
15 Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol* 2006;7:741-6.
16 Chakravarty A. Regulatory aspects in using surrogate markers in clinical trials. In: Burzykowsi T, Molenberghs G, Buyse M, eds. *The evaluation of surrogate endpoints* . Springer, 2005.
17 Moynihan R. Surrogates under scrutiny: fallible correlations, fatal consequences. *BMJ* 2011;343:d5160.
18 Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ* 2011;343:d7995.
19 US Food and Drug Administration. Label and approval history Iressa. 2003. www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Search.DrugDetails.
20 US Food and Drug Administration. Gefitinib (marketed as Iressa) information. 2005. www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm110473.htm.
21 Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. *N Engl JMed* 1989;321:406-12.
22 Barter PJ, Caulfield M, Eriksson M, Grundy SM, Kastelein JJ, Komajda M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl JMed* 2007;357:2109-22.
23 Giaccone G, Herbst RS, Manegold C, Scagliotti G, Rosell R, Miller V, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial—INTACT 1. *J Clin Oncol* 2004;22:777-84.
24 Czoski-Murray C, Warren E, Chilcott J, Beverley C, Psyllaki MA, Cowan J. Clinical effectiveness and cost-effectiveness of pioglitazone and rosiglitazone in the treatment of type 2 diabetes: a systematic review and economic evaluation. *Health Technol Assess* 2004;8:iii, ix-x, 1-91.
25 Pavey T, Hoyle M, Ciani O, Crathorne L, Jones-Hughes T, Cooper C, et al. Dasatinib, nilotinib and standard-dose imatinib for the first-line treatment of chronic myeloid leukaemia: systematic reviews and economic analyses. *Health Technol Assess* , 2012;16:iii-iv, 1-277, doi:10.3310/hta16420.
26 National Institute for Health and Clinical Excellence. 2011/12 review of the guide to the methods of technology appraisal. TA methods guide review: supporting documents. 2012. www.nice.org.uk/aboutnice/howwework/devnicetech/TAMethodsGuideReview.jsp?domedia=1&mid=C673A1A2-19B9-E0B5-D4C41B8F1C064D54.
27 La Cour JL, Brok J, Gotzsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. *BMJ* 2010;341:c3653.
28 Alexander JH, Hafley G, Harrington RA, Peterson ED, Ferguson TB Jr, Lorenz TJ, et al. Efficacy and safety of edifoligide, an E2F transcription factor decoy, for prevention of vein graft failure following coronary artery bypass graft surgery: PREVENT IV: a randomized controlled trial. *JAMA* 2005;294:2446-54.
29 Hou FF, Zhang X, Zhang GH, Xie D, Chen PY, Zhang WR, et al. Efficacy and safety of benazepril for advanced chronic renal insufficiency. *N Engl JMed* 2006;354:131-40.
30 Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, eds. *Clinical measurement in drug evaluation* . Wiley, 1995.
31 Ciani O, Taylor RS. Surrogate, friend or foe? the need for case studies of the use of surrogate outcomes in cost-effectiveness analyses. *Health Econ* 2013;22:251-2. Published online 24 May 2012.
32 National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal* . NICE, 2008.
33 Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000-2005. *JAMA* 2006;295:2270-4.
34 Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. 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. Evidence-Based Medicine Working Group. *JAMA* 1999;282:771-8.
35 Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513-24.
36 Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions* 5.1.0 [updated March 2011]. Available from www.cochrane-handbook.org. Cochrane Collaboration, 2011.
37 Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;19:3127-31.
38 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683-91.
39 Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25:3443-57.
40 Nobre MR, da Costa FM. Surrogate outcomes are associated with low methodological quality of studies of rheumatoid arthritis treated with antitumour necrosis factor agents: a systematic review. *Evid Based Med* 2011;17:3-7.
41 Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;53:1119-29.
42 Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol* 2007;34:607-15.
43 Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet* 2005;365:1159-62.
44 Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA* 2002;287:2847-50.
45 Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997;315:617-9.
46 Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 2005;23:8664-70.
47 Johnson KR, Freemantle N, Anthony DM, Lassere MN. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE). *J Clin Epidemiol* 2009;62:328-36.
48 Hughes MD, Daniels MJ, Fischl MA, Kim S, Schooley RT. CD4 cell count as a surrogate endpoint in HIV clinical trials: a meta-analysis of studies of the AIDS Clinical Trials Group. *AIDS* 1998;12:1823-32.
49 Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;25:5218-24.
50 Poustie VJ, Russell JE, Watling RM, Ashby D, Smyth RL. Oral protein energy supplements for children with cystic fibrosis: CALICO multicentre randomised controlled trial. *BMJ* 2006;332:632-6.
51 Tonstad S, Tonnesen P, Hajek P, Williams KE, Billing CB, Reeves KR. Effect of maintenance therapy with varenicline on smoking cessation: a randomized controlled trial. *JAMA* 2006;296:64-71.
52 Mangiapane S, Velasco Garrido M. Use of surrogate end points in HTA. *GMS Health Technol Assess* . 2009;5:Doc12.
53 National Institute for Health and Clinical Excellence. Leukaemia (chronic myeloid, first line)—dasatinib, nilotinib and standard-dose imatinib: final appraisal determination document. 2012. http://guidance.nice.org.uk/TA/Wave24/15/FAD/FinalAppraisalDetermination/pdf/English.
54 National Institute for Health and Clinical Excellence. Renal transplantation—immuno-suppressive regimens (adults). 2004. www.nice.org.uk/guidance/TA85.
55 National Institute for Health and Clinical Excellence. Hepatitis B (chronic)—adefovir dipivoxil and pegylated interferon alpha-2a (TA96). 2006. http://guidance.nice.org.uk/TA96.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

What is already known on this topic

Surrogate outcomes are used to substitute and predict for a final patient relevant outcome in clinical trials

Failures of specific surrogate outcomes have been reported in the literature

Licensing and coverage decisions of health technologies often rely on evidence based on surrogate outcomes

What this study adds

Trials reporting surrogate primary outcomes are more likely to report larger treatment effects than trials reporting final patient relevant primary outcomes

In the absence of patient relevant outcomes, policy makers should rely on validated surrogate outcomes and take into account the potential uncertainty in their prediction of treatment benefit and harm

56 National Institute for Health and Clinical Excellence. Review of the guide to the methods of technology appraisal. 2012. www.nice.org.uk/aboutnice/howwework/devnicetech/TAMethodsGuideReview.jsp?domedia=1&mid=CB13DD0D-19B9-E0B5-D4AB0011AEE2B0E7.

57 Renfro LA, Carlin BP, Sargent DJ. Bayesian adaptive trial design for a newly validated surrogate endpoint. *Biometrics* 2012;68:258-67.

Accepted: 29 October 2012

Cite this as: *BMJ* 2013;346:f457

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: http://creativecommons.org/licenses/by-nc/2.0/ and http://creativecommons.org/licenses/by-nc/2.0/legalcode.

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Tables

Table 1| Characteristics of trials using surrogate primary outcomes and final patient relevant primary outcomes. Values are numbers (percentages) unless stated otherwise

Characteristics	Surrogate outcomes (n=84)	Patient relevant outcomes (n=101)
Intervention clinical area:		
Cardiovascular	20 (24)	25 (25)
Endocrinology	4 (5)	3 (3)
Gastrology and hepatology	9 (11)	10 (10)
Infectious disease	18 (21)	21 (21)
Nephrology and urology	1 (1)	4 (4)
Neurology	0 (0)	2 (2)
Obstetrics	5 (6)	4 (4)
Oncology	5 (6)	2 (2)
Other	17 (20)	24 (23)
Pulmonology	5 (6)	6 (6)
Population clinical area:		
Cardiovascular	25 (30)	25 (25)
Endocrinology	4 (5)	7 (7)
Gastrology and hepatology	8 (10)	9 (9)
Infectious disease	13 (15)	15 (14)
Nephrology and urology	0 (0)	5 (5)
Neurology	0 (0)	1 (1)
Obstetrics	7 (8)	7 (7)
Oncology	5 (6)	2 (2)
Other	16 (19)	25 (25)
Pulmonology	6 (7)	5 (5)
Journal:		
Annals of Internal Medicine	8 (10)	6 (6)
BMJ	7 (8)	11 (11)
Journal of the American Medical Association	20 (24)	22 (22)
Lancet	21 (25)	19 (19)
New England Journal of Medicine	27 (32)	42 (41)
PLoS Medicine	1 (1)	1 (1)
Publication year:		
2005	40 (48)	49 (49)
2006	44 (52)	52 (51)
Centre status:		
Single centre	19 (23)	9 (9)*
Multicentre	65 (77)	92 (91)*
Intervention:		
Drugs	49 (59)	61 (60)
Medical devices	7 (8)	7 (7)
Surgical procedures	4 (5)	8 (8)
Health promotion activities	7 (8)	2 (2)
Other therapeutic technologies	17 (20)	23 (23)
Sponsor:		
Profit	24 (29)	29 (29)
Not for profit	49 (59)	56 (55)
Mixed	11 (12)	16 (16)

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 (continued)

Characteristics	Surrogate outcomes (n=84)	Patient relevant outcomes (n=101)
Median (interquartile range) sample size	371 (162-787)	741 (300-4731)†
Median (interquartile range) follow-up (days)	255 (133-540)	180 (35-730)

*χ^2 test, P=0.01.

†Mann-Whitney U test, P<0.001.

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Table 2| Comparison of treatment effects of trials using surrogate outcomes with trials using final patient relevant outcomes: primary and sensitivity analyses

	Risk ratio* (95% CI)		Ratio of odds ratios or relative risk ratio (95% CI)	
Method of analysis	**Surrogate outcomes**	**Patient relevant outcomes**	**Unadjusted**	**Adjusted†**
Primary analysis:				
Binary outcomes (51 surrogate *v* 83 patient relevant)	0.51 (0.42 to 0.60)	0.76 (0.70 to 0.82)	1.47 (1.07 to 2.01)	1.46 (1.05 to 2.04)
Sensitivity analyses:				
Inclusion of risk ratios as reported by authors (57 *v* 86)	0.56 (0.48 to 0.65)	0.80 (0.75 to 0.86)	1.38 (1.12 to 1.71)	1.36 (1.08 to 1.70)
Inclusion of continuous outcomes (84 *v* 101)	0.46 (0.39 to 0.54)	0.68 (0.62 to 0.74)	1.44 (0.83 to 2.49)	1.48 (0.83 to 2.62)
Binary outcomes, matched pairs (43 *v* 43)	0.48 (0.39 to 0.59)	0.68 (0.61 to 0.77)	1.38 (1.01 to 1.88)	—

*Pooled using DerSimonian and Laird random effects meta-analyses.

†Adjusted for trial level characteristics of clinical area of intervention, patient population, type of intervention, sponsor, journal, mean sample size, and mean follow-up time.

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Table 3| **Summary of risk of bias assessment for trials reporting biomarkers or intermediate outcomes (surrogate outcomes) versus final patient relevant primary outcomes**

	No (%) of trials		
Quality assessment	**Surrogate outcomes (n=84)**	**Patient relevant outcomes (n=101)**	**P value***
Intention to treat analysis	62 (74)	83 (82)	0.17
Adequate randomisation sequence generation	54 (64)	65 (64)	0.99
Adequate randomisation allocation concealment	61 (73)	74 (73)	0.92
Double blinding/placebo control	42 (50)	43 (43)	0.31

*χ^2 test.

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Figures

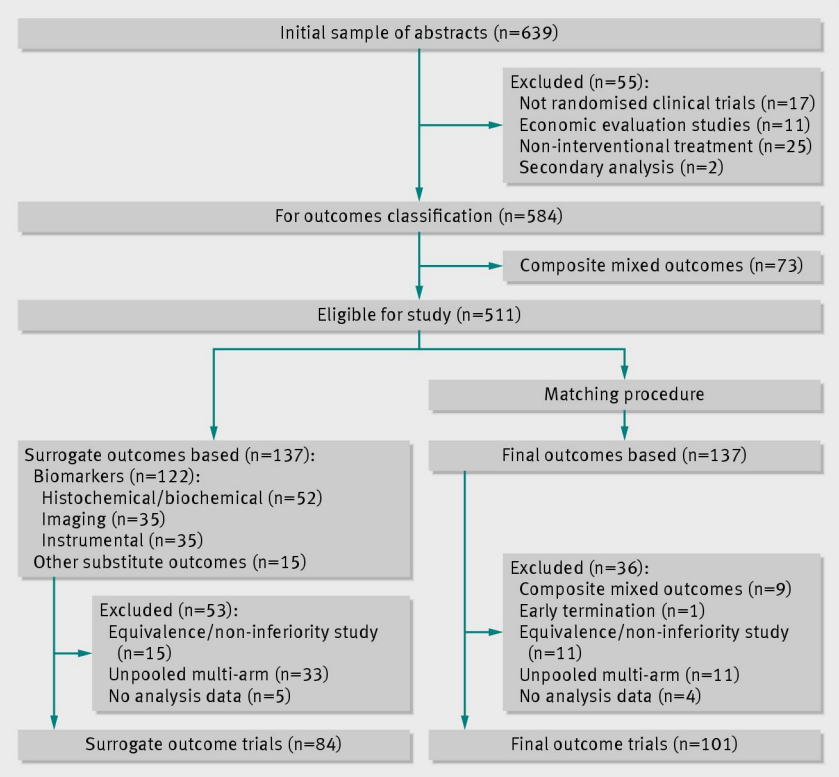
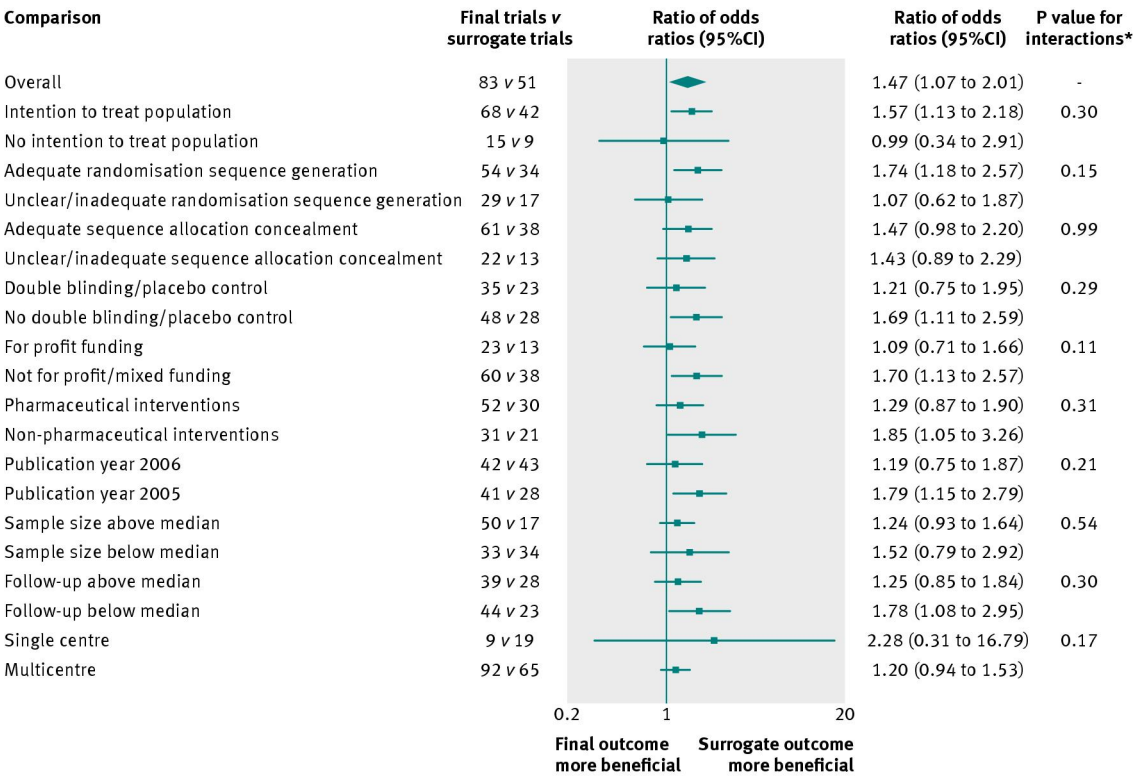


Fig 1 Flow of studies through inclusion process



* P values from test of interaction between type of primary outcome and trial characteristics

Fig 2 Ratio of odds ratios comparing treatment effect estimates in trials using surrogate outcomes versus trials using final primary end points stratified by key trial characteristics. *P values from tests of interaction between type of primary outcome and trial characteristics

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ: first published as 10.1136/bmj.f457 on 29 January 2013. Downloaded from https://www.bmj.com/ on 13 May 2025 at Department GEZ-LTA Erasmushogeschool. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.