Journal of Clinical Epidemiology 160 (2023) 83−99

Journal of Clinical Epidemiology

REVIEW

Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate end points in trials: a scoping review

Anthony Muchai Manyara[a,*], Philippa Davies[b], Derek Stewart[c], Christopher J. Weir[d], Amber E. Young[b], Valerie Wells[a], Jane Blazeby[b,e,f], Nancy J. Butcher[g,h], Sylwia Bujkiewicz[i], An-Wen Chan[j,k], Gary S. Collins[l], Dalia Dawoud[m], Martin Offringa[g,n], Mario Ouwens[o], Joseph S. Ross[p,q], Rod S. Taylor[a,r], Oriana Ciani[s]

[a]*MRC/CSO Social and Public Health Sciences Unit, School of Health and Wellbeing, University of Glasgow, Glasgow, UK*
[b]*Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK*
[c]*Patient and Public Involvement Lead, Nottingham, UK*
[d]*Edinburgh Clinical Trials Unit, Usher Institute, University of Edinburgh, Edinburgh, UK*
[e]*Bristol NIHR Biomedical Research Centre, Bristol, UK*
[f]*University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK*
[g]*Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Canada*
[h]*Department of Psychiatry, University of Toronto, Toronto, Canada*
[i]*Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK*
[j]*Women's College Research Institute, Toronto, Canada*
[k]*Department of Medicine, University of Toronto, Toronto, Canada*
[l]*Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, Centre for Statistics in Medicine, University of Oxford, Oxford, UK*
[m]*Science, Evidence and Analytics Directorate, Science Policy and Research Programme, National Institute for Health and Care Excellence, London, UK*
[n]*Department of Paediatrics, University of Toronto, Toronto, Canada*
[o]*AstraZeneca, Mölndal, Sweden*
[p]*Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA*
[q]*Section of General Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA*
[r]*Robertson Centre for Biostatistics, School of Health and Well Being, University of Glasgow, Glasgow, UK*
[s]*Centre for Research on Health and Social Care Management, SDA Bocconi School of Management, Milan, Italy*

Accepted 20 June 2023; Published online 26 June 2023

Abstract

Objective: To synthesize the current literature on the use of surrogate end points, including definitions, acceptability, and limitations of surrogate end points and guidance for their design/reporting, into trial reporting items.

Study Design and Setting: Literature was identified through searching bibliographic databases (until March 1, 2022) and gray literature sources (until May 27, 2022). Data were thematically analyzed into four categories: (1) definitions, (2) acceptability, (3) limitations and challenges, and (4) guidance, and synthesized into reporting guidance items.

Results: After screening, 90 documents were included: 79% (n = 71) had data on definitions, 77% (n = 69) on acceptability, 72% (n = 65) on limitations and challenges, and 61% (n = 55) on guidance. Data were synthesized into 17 potential trial reporting items: explicit statements on the use of surrogate end point(s) and justification for their use (items 1−6); methodological considerations, including whether sample size calculations were informed by surrogate validity (items 7−9); reporting of results for composite outcomes containing a surrogate end point (item 10); discussion and interpretation of findings (items 11−14); plans for confirmatory studies, collecting data on the

Author contributions: P.D., C.J.W., A.Y., R.S.T., and O.C. contributed to funding acquisition. A.M.M., V.W., P.D., C.J.W., A.Y., R.S.T., and O.C. contributed to methodology. A.M.M., V.W., P.D., R.T., and O.C. contributed to data curation. A.M.M., R.T., and O.C. contributed to formal analysis and writing−original draft. J.B., N.J.B., S.B., A.W.C., G.S.C., D.D., M.O., M.O., and J.S.R. contributed to supervision and writing−review and editing. The views expressed in this article are those of the authors and not their employers or funders.

Funding: The development of SPIRIT and CONSORT extensions has been funded by the UK Medical Research Council (grant number MR/V038400/1). G.S.C. was supported by Cancer Research UK (program grant: C49297/A27294). J.M.B. was supported by the NIHR Bristol Biomedical Research center. S.B. was supported by Leicester NIHR Biomedical Research Centre. The funders had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the review for publication.

* Corresponding author. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Berkeley Square, 99 Berkeley Street, Glasgow G3 7HR, UK. Tel.:+07783121419.

E-mail address: Anthony.manyara@glasgow.ac.uk (A.M. Manyara).

https://doi.org/10.1016/j.jclinepi.2023.06.013

surrogate end point and target outcome, and data sharing (items 15−16); and informing trial participants about using surrogate end points (item 17).

Conclusion: The review identified and synthesized items on the use of surrogate end points in trials; these will inform the development of the Standard Protocol Items: Recommendations for Interventional Trials−SURROGATE and Consolidated Standards of Reporting Trials−SURROGATE extensions.

Keywords: Surrogate end points; Randomized controlled trials; Protocols; Reporting guidance; Scoping review; Design; SPIRIT-Surrogate; CONSORT-Surrogate; Validation

1. Introduction

According to the United States Food and Drug Administration and National Institutes of Health Biomarker Working Group, a surrogate end point is an end point used in randomized controlled trials (hereafter referred to as 'trials') to substitute for a patient-relevant target outcome (i.e., a direct measure of how an individual feels, functions, or how long they survive) that is expected to predict health benefit or harm based on pathophysiologic, epidemiological, therapeutic, or other scientific evidence [1]. Surrogate end points are used as primary outcomes in trials: (1) to improve efficiency, as they can be measured more often, more easily, and sometimes less invasively than patient-relevant target outcomes (hereafter referred to as 'target outcomes'); (2) to provide evidence of a treatment signal in early phase trials; and (3) to fast track approval of interventions and therapies in situations of high unmet medical need (life-threatening or irreversibly debilitating conditions or diseases) via what are commonly known as Accelerated Approval regulatory pathways [1].

Nevertheless, despite their appeal and benefit, there are concerns associated with surrogate end points. First, comparison of treatment effect sizes between surrogate end points and the linked target outcomes in trials found the former to be larger by approximately 46% which could lead to overestimation of clinical and cost-effectiveness of interventions approved on the basis of using surrogate end points [2]. Second, some surrogate end points have led to approval of interventions that have resulted in more harm than health benefit, due to unintended and unmeasured effects of the intervention, sometimes through mechanisms outside the known disease causal pathways [3−5]. Therefore, trials that use surrogate end points, especially as primary outcomes, should pay special attention to their design and reporting, including justification for use of a surrogate end point and providing evidence of its validation. An audit of trials published in 2006 and 2007 found that 57% ($n = 62/109$) stated that the primary outcome was a surrogate end point and only 35% ($n = 38/109$) discussed the validity of the surrogate [6]. Consequently, the authors recommended developing evidence-based reporting guidelines for trials using surrogate end points [6]. In general, there are two common reporting guidelines for trials: Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT), a checklist to report trial protocols [7], and Consolidated Standards of Reporting Trials (CONSORT), a checklist used to report completed trials [8]. However, these checklists and their extensions, including CONSORT-PRO [9] and the most recent SPIRIT/CONSORT Outcomes [10] provide no specific guidance for the reporting of trials using surrogate end points as primary outcomes.

An international multidisciplinary team of researchers and stakeholders is currently developing SPIRIT-SURROGATE and CONSORT-SURROGATE extensions. This development follows the Enhancing the QUAlity and Transparency Of health Research Network's guidance for developing a health research reporting guideline [11]; and the project protocol has been published [12]. A key initial step in developing a reporting guideline is reviewing the literature, including identifying previous relevant guidance and sources of bias in relevant studies [11]. Therefore, we sought to undertake a scoping review of current literature on definitions, limitations and acceptability in use, and guidance in reporting of surrogate end points in trials. The review's overarching purpose is to assess the current understanding, advice, and guidance on the use of surrogate end points in trials. Our specific research questions were: (1) How are surrogate end points defined?; (2) What are the limitations of using surrogate end points in trials?; (3) When is the use of surrogate end points acceptable?; and (4) What published advice and guidance exists on designing and reporting trial protocols and reports using surrogate end points?

2. Methods

The protocol for this scoping review has been published [13] and the review stages are briefly summarized below. The study is reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analysis for Scoping Reviews [14] (Table A.1). Protocol deviations with reasons are captured in Table A.2.

2.1. Literature searching

Relevant literature was identified through electronic bibliographic databases (Excerpta Medica Database [EMBASE], Medical Literature Analysis and Retrieval System Online [MEDLINE], Cochrane Methodology Register);

What is new?

Key findings

- From 90 documents (peer-reviewed and gray literature), we categorized findings into four themes about the use of surrogate end points in trials: definitions, acceptability, limitations and challenges to use, and advice and guidance in design and reporting of trials.
- Data from these themes were synthesized into 17 items for designing and reporting of trials using surrogate end points as primary outcomes.

What this adds to what is known?

- Current reviews on surrogate end points in trials have been narrative in nature and restricted to specific clinical/health areas.
- There is no published comprehensive guidance for reporting trials using surrogate end points as primary outcomes.

What is the implication and what should change now?

- The items synthesized in this scoping review will inform the development of the Standard Protocol Items: Recommendations for Interventional Trials –SURROGATE and Consolidated Standards of Reporting Trials–SURROGATE extensions.
- Implementation of these extensions by trialists, journal editors, and peer reviewers can lead to improved reporting of trials and their protocols that use surrogate end points as primary outcomes.

Google and targeted website searches (e.g., Food and Drug Administration); hand-searching of reference lists of included records; and solicitation from experts. Databases search, without time limits, was done by an information specialist (V.W.) using a search strategy that combined ''surrogate end points'', ''guidelines'', and ''trials''–related search terms (Tables A.3–A.5). Website search involved various strategies: combination of search terms (e.g., ''surrogate end points'' AND ''guidance'') if advanced search was possible; broad searches (e.g., ''surrogate end points'') using the website search function; and browsing for websites without a search function. Last searches were done on March 1, 2022 (for databases) and May 27, 2022 (for websites) (Tables A.3–A.5). Additionally, we solicited documents related to surrogate end points from authors (N.J.B. and M.O.) who had completed a relevant scoping review on ''outcome reporting recommendations for trial protocols and reports'' [15].

2.2. Study selection

Search outputs from databases were exported to Covidence [16] and the titles, abstracts, and full texts screened independently by two reviewers (A.M.M., R.S.T., P.D., or O.C.). One reviewer (A.M.M.) screened outputs from relevant websites, reference lists, and solicited references. To balance between feasibility of screening website search outputs and relevancy, the first 100 hits in each search were screened for eligibility. We included documents that had data on limitations, acceptability, and guidance in reporting of surrogates in trials (any design/phase). Documents with data on limitations and acceptability of surrogates in non-trial areas such as health policy were deemed irrelevant to this review. Furthermore, to make the review feasible within project timelines, documents with only data on definitions of ''surrogate end points'' were excluded. Finally, documents in languages other than English and conference/meeting abstracts/slide presentations were excluded.

2.3. Data charting

One author (A.M.M.) extracted and charted data into Microsoft Excel. Data extracted included document characteristics [e.g., author, publication year, country, document type (e.g., review article, commentary, regulatory guidance), research area if specified, funding if stated] and verbatim information relevant to research questions (i.e., definition, limitations, acceptability, guidance on surrogate end points use). At the start of extraction, a subset of extracted data (in 10 documents) was checked for accuracy by a second reviewer (R.S.T. or O.C.) after which A.M.M. proceeded with extraction in all other documents.

2.4. Synthesis and reporting

A descriptive and thematic analysis of included studies was performed in Microsoft Excel [17]. Descriptive data (e.g., publication year, document type) were analyzed by one author (A.M.M.) using counts and percentages. Data on definitions, limitations, acceptability, and guidance were thematically analyzed into guidance items either by explicit statement (item is explicitly stated in data on guidance) or implication (item is implied from definitional, limitations, or acceptability findings). Three authors (A.M.M., R.S.T., and O.C.) independently read extracted data from 30 documents and summarized them into guidance items. They then met to discuss and agree on generated items after which one author (A.M.M.) summarized items from remaining documents. Definitions were analyzed as to whether they were cited from other source or authors' own definition; term used to refer to a ''surrogate'' (e.g., end point, outcome); what was considered a surrogate from the definition; and implied function of a surrogate from a definition, among other characteristics.

3. Results

3.1. Search results

Figure 1 shows the process of identification, screening, and inclusion of records. A total of 1,320 documents were identified via databases search and another 3,929 from other sources. At the title and abstract screening stage, 899 records from databases were considered not to be relevant to the review, whereas at the full-text screening stage, a total of 73 records were excluded due to not being relevant to the research questions (n = 62), conference abstracts (n = 6), and not being in English (n = 5). In total, 53 documents were included from databases search. For documents from other sources, 1,097 were assessed (only 100 hits for website outputs) and 1,060 were excluded mainly due to being irrelevant to research questions (n = 1,008) leaving 37 documents for inclusion. Table A.6 shows the 90 included records.

3.2. Descriptive characteristics of included records

Table 1 summarizes the characteristics of the 90 documents included in our review. Most were published in the last decade (n = 43, 48%), were journal publications (n = 83, 92%) and narrative reviews (n = 48, 53%), and were not specific to a particular clinical or research areas (n = 33, 37%).

3.3. Definitions of surrogate end points

Seventy one documents (of 90; 79%) provided a definition for a ‘‘surrogate end point’’. Thirty nine (of 71; 55%) provided their own definition, whereas 45% (n = 32) cited definitions from other literature. Table 2 shows definitions cited by at least two of the included records with the most cited being the National Institutes of Health Biomarkers Working Group definition [18] by 38% (n = 12/32).

Table 3 summarizes the characteristics of definitions. The term ‘surrogate end point’ was used in 83% of the records (75/90), although 50 of the records (56%) did not use one term exclusively. Furthermore, we only found two justifications for using one term over another: surrogate outcome rather than end point as latter was thought to imply an irreversible event, for example, death [21]; and surrogate observations rather than end points as ophthalmology trials have few true end points, involve ‘‘complex nonfatal events’’, and follow-up continues after these events occur [22].

Forty four definitions (of 71; 62%) considered surrogates to be biomarkers such as laboratory measures and radiographic images. Other measures regarded as surrogate end points were physical signs or physician-assessed measures in 30% of the records (n = 21); and intermediate outcomes in 12 records (17%) such as patient-reported outcomes (e.g., dyspnea relief) [23], function status (e.g., exercise tolerance in heart failure) [23,24], and symptoms, for example, angina pain in coronary heart disease [24].

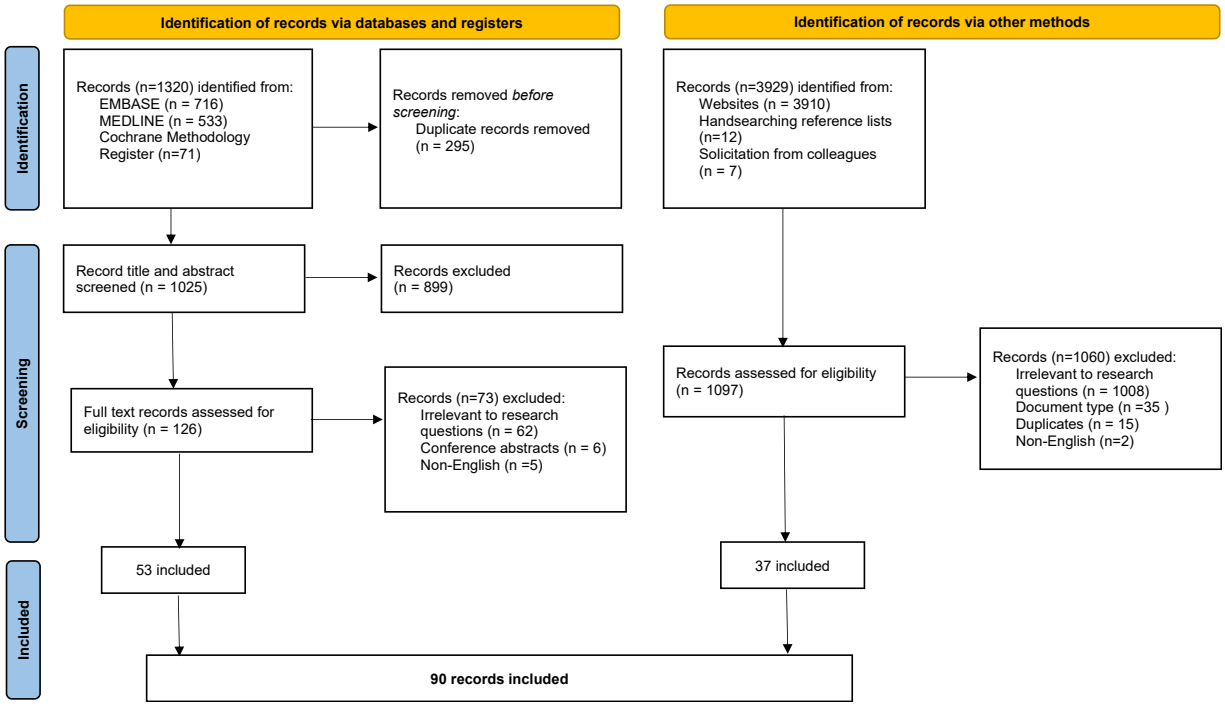

Fig. 1. Flow diagram showing record identification, screening, and inclusion. Exclusion on document types related to slide/PowerPoint presentations, blog, and news articles. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1. Characteristics of included records

Characteristic	*N* (%)
Year of publication	
1989–1999	21 (23%)
2000–2010	25 (28%)
2011–2021	43 (48%)
Undated	1 (1%)
Authors' affiliation institution[a]	
University/research institute	64 (71%)
Hospital/healthcare facility	21 (23%)
Government agency including regulatory agencies	20 (22%)
Pharmaceutical company or consultancy	15 (17%)
Nonprofit organization	2 (2%)
Not reported/No affiliation	2 (2%)
Authors' regions[a]	
North America	48 (53%)
Europe	26 (29%)
International	19 (21%)
Asia	3 (3%)
South America	1 (1%)
Not reported	1 (1%)
Source of document	
Journal publication	83 (92%)
Gray literature	7 (8%)
Type of document	
Narrative review	48 (53%)
Commentary/Opinion/Editorial/Viewpoint	11 (12%)
Analysis of RCTs	9 (10%)
Guideline (regulatory/trial reporting)	9 (10%)
Working group output/Workshop proceedings	7 (8%)
Other[b]	6 (7%)
Research area of focus[a]	
General[c]	32 (36%)
Cancer	22 (24%)
Cardiovascular disease	9 (10%)
Infectious diseases including human immunodeficiecy virus	4 (4%)
Drug development	3 (3%)
Other[d]	23 (26%)
Funding source/Conflict of interest declaration[a]	
Not reported	36 (40%)
Public research funding	21 (23%)
Pharmaceutical company	18 (20%)
No conflict declared	11 (12%)
University/research institute	10 (11%)
Nonprofit organization	7 (8%)
Regulatory agency	4 (4%)

[a] Overlapping proportions.
[b] Document type frequency ≤2: survey, statistical methods articles, systematic review, and glossary.
[c] Not specific to one or two clinical or research areas.
[d] Document frequency ≤2 s: renal medicine; metabolic and endocrine; ophthalmology; hematology; musculoskeletal; dental; chronic obstructive pulmonary disease; critical care; dementia; global health; respiratory; sepsis; rare diseases; pediatrics; surgery; dermatology; and health policy.

Substitution and prediction of another outcome were implied as the key roles of a surrogate end point in 31 (44%) definitions. However, what was predicted and/or substituted for differed according to definitions. In 37 (of 71; 52%), surrogate end points substituted and/or predicted for what was referred to as "true" or "hard" or "clinical" or "clinically meaningful" end points or outcomes. Nineteen definitions (27%) clarified that such end points were direct measures of what an individual feels, functions, or survives. Mortality or survival was the most common example of the target outcome substituted for or predicted by a surrogate end point and others included morbidity, severe disability, symptoms, and health-related quality of life. Eleven definitions (16%) included a statement of the evidence base used for prediction which was "epidemiologic, therapeutic, pathophysiologic, or other scientific evidence"[1,18].

3.4. Acceptability of surrogate end points

Acceptability of surrogate end points in trials was discussed in 77% (69/90) of included reports. Findings were categorized in two broad themes as described below: considerations on use of surrogate end points and circumstances when use of surrogate end points is perceived to be acceptable. Textbox 1 summarizes findings under these themes.

3.4.1. Considerations in use of surrogate end points

First, 49 documents (of 69; 71%) reported the need for validation before surrogate end points are used. Two elements of validation were mentioned: (1) evidence of mechanistic or biologic plausibility rationale linking the disease pathway, surrogate end point, intervention, and target outcome (in 20/69 of records, 29%) and (2) evidence of statistical or quantitative validation which provides the magnitude of effect on the target outcome predicted by the surrogate end point (n = 26, 38%): see reviews on statistical validation in Table A.6 (References 56, 57, 67). Furthermore, six records (of 69; 9%) noted that validation of a surrogate end point in one intervention could be extrapolated to other interventions if the mechanism of action follows a similar disease causal pathway. A second consideration, contained in 13% of the records (n = 9), was whether a surrogate end point predicts interventions benefit and captures intervention harms of an intervention to inform the intervention's benefit-risk balance.

3.4.2. Circumstances when surrogate end points use is perceived to be acceptable

Twenty eight percent of documents (n = 19/69) noted that nonvalidated surrogate primary end points (reasonably likely to predict health benefit [1]) may be acceptable in evaluating interventions of "rare" or "life threatening conditions" with a high "unmet medical need"—a process

Table 2. Definitions cited by at least two of included records

Source	Definition	*N* (%)
NIH (2001) [18]	A biomarker that is intended to substitute for a clinical end point. A surrogate end point is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.	12 (38%)
Temple [19] (1999)[a]	A laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy.	10 (31%)
Prentice (1989) [20]	A response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true end point.	6 (19%)
BEST (2021) [1]	An end point that is used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives. A surrogate end point does not measure the clinical benefit of primary interest in and of itself but rather is expected to predict that clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.	3 (9%)

Abbreviations: BEST, biomarkers, end points, and other tools; NIH, national institutes of health.
[a] Definition was derived from the preamble to the FDA's proposed accelerated approval rule.

Table 3. Characteristics of definitions

Characteristic	*N* (%)
Term used[a]	
Surrogate end point	75 (83%)
Surrogate outcome	33 (37%)
Surrogate marker	27 (30%)
Surrogate measure	17 (19%)
Surrogate variable	6 (7%)
Other	5 (6%)
Exclusive use of one term	
No	50 (56%)
Yes	40 (44%)
What is considered a surrogate end point?[a]	
Biomarkers	44 (62%)
Physical sign	21 (30%)
Intermediate outcome	12 (17%)
Nonspecific term (End point/variables/measures/outcomes)	22 (31%)
Function of a surrogate end point?	
Both prediction and substitution	31 (44%)
Prediction	22 (31%)
Substitution	14 (20%)
Substitution or supplementation	1 (1%)
What is substituted and/or predicted for (as stated by authors)?[a]	
'True/Hard/Clinically meaningful end points'	37 (52%)
'How individual feels, functions, or survives'	19 (27%)
'Clinical benefit'	16 (23%)
'Effect/Lack of benefit or harm'	13 (18%)
'Patient-relevant/centered outcome'	7 (10%)
Explicit statement of need for evidence base for prediction/substitution	
No	60 (85%)
Yes	11 (16%)

Other includes surrogate indicator, surrogate observation, and surrogate parameter.
[a] Overlapping proportions.

commonly referred to as accelerated approval or orphan drug designation [1]. Examples of such conditions included cancer, heart failure, preterm labor, human immunodeficiecy virus, neurological conditions (e.g., Alzheimer's, amyotrophic lateral sclerosis)—see more examples in Table A.7. One article noted that there was limited consensus on what was an "unmet medical need" and potential use of surrogates in accelerated approvals being more than actual unmet needs [27]. Two documents called for a need for stakeholders, including industry experts and patient representatives, to clearly define circumstances when accelerated approvals are warranted [25,26]. Almost half of the records that discussed accelerated approvals (42%, $n = 8/19$) noted that such approvals were conditional subject to findings from postapproval/confirmatory studies. Two documents proposed ways to improve design and conduct of conditional approval trials: (1) trial participants should be informed before enrollment that the health benefit of the intervention is not confirmed and harm could exceed benefit [26] and (2) participant recruitment in confirmatory/postapproval studies should be completed before use of surrogate end points in accelerated approval [26,27].

Apart from accelerated approvals, 30% of the included records ($n = 21/69$) discussed the acceptability and utility of surrogate end points in Phase 1 and 2 trials to screen for drug efficacy, safety, and mechanism of action. Other circumstances when use of surrogate end points was acceptable were ethical feasibility reasons (e.g., procedures to assess the target outcome are harmful, excessively invasive, or uncomfortable), practicality and convenience, surrogate end point being less biased than target outcomes, and utility in pilot trials—see Textbox 1 for examples. One of the articles reporting a survey among diabetes patients found that majority preferred use of patient-important outcomes (diabetes-related complications) rather than surrogate end points (blood glucose measure) as primary outcomes in trials [36].

Textbox 1 Summary of findings on acceptability of surrogate end points

Considerations in use of surrogate end points

1. Surrogate end points should be validated through either or both of the following:
   - ■ Evidence of mechanistic or biological plausibility rationale.
   - ■ Evidence of statistical validation—see reviews on this in Table A.6 (Reference 56, 57, 67).
     - * A validated surrogate end point in one intervention may be extrapolated to other interventions if mechanism of action has a similar causal pathway.
2. Consider whether using a surrogate end point predicts benefits and captures harms of an intervention to inform the overall intervention's benefit-risk balance.
   - * Interventions may have unintended consequences not on surrogate-target outcome causal pathways; and surrogate end points improve trial efficiency by reducing sample size and follow-up period which limits a trial's ability to capture harms of intervention under evaluation—see Limitations section and Figure 2.

Circumstances when surrogate end points use is perceived to be acceptable.

1. Accelerated approval of interventions of rare, life-threatening conditions with a high unmet medical need can use nonvalidated surrogate end points that are reasonably likely to predict benefit as primary outcomes.
   - ■ Examples of such conditions included cancer, heart failure, preterm labor, human immunodeficiecy virus, neurological conditions (e.g., Alzheimer's, amyotrophic lateral sclerosis)—see more examples on Table A.7.
   - ■ Need for consensus on the ''unmet medical need'' and circumstances for accelerated approval by all stakeholders including patient and public involvement representatives [25,26].
   - ■ Such approvals are or should be followed by postapproval trials to confirm benefit.
     - * Participants in accelerated approval trials need to be informed of use of surrogate end points (including potential harm or lack of benefit) before enrollment to trial [26].
     - * Postapproval trials should be fully recruited before approval of intervention [26,27].
2. Utility of surrogate primary end points in Phase 1 and 2 trials.
   - * Biomarkers used as surrogate end points enable study of interventions' mechanism of action and reduce sample size and trial period; hence, fewer participants exposed to harmful interventions and faster termination of ineffective ones.
3. It can be ethically feasible, practical, or convenient to use surrogate end points as primary outcomes when
   - ■ Target outcome requires long follow-ups such as in rare diseases.
     - ○ Included records did not define what was a ''long follow-up'' and an included record from European Medicines Agency defined rare diseases as disease affecting only a few thousand or fewer people in the European Union [28]. It seems that ''rare condition'' is a qualitative term without a straightforward interpretation.
   - ■ Procedures to assess the target outcome are harmful, excessively invasive, or uncomfortable.
   - ■ Ethically impossible to evaluate therapies for biological or chemical weapons exposure [29].
   - ■ Participants cannot self-report symptoms such as in pediatric trials [25].
   - ■ The objective of intervention is to intervene early and prevent/delay mortality such as cancer screening interventions [30].
4. Target outcomes may have limitations in some circumstances.
   - ■ Crossover design and rescue therapies may confound survival as the target outcome [31] for example, in cancer a clinician may introduce a rescue treatment due to nonresponse that might confound the intervention-control comparison in terms of overall mortality.
   - ■ Quality of life measures (if regarded as target outcomes) maybe subjective and complex to measure [32].
     - * Quality of life measures are direct measures of what an individual feels or functions, hence important target outcomes.
   - ■ Mortality may be confounded in evaluation of preventive interventions (e.g., due to background noise from other aging conditions in a disease-specific preventive intervention initiated in middle age) [33]; and death is not

always a target outcome for people who prefer low morbidity and better quality of life over few additional years of survival [33].

- Some interventions directly impact surrogate end points, such as innovations to improve disease testing and impacts on target outcomes need additional interventions [34], for example, improvement in disease testing as a surrogate end point will only result to mortality benefit if positive disease is well treated.

5. Utility in pilot trials to inform conduct of a larger trial.
   * Trials should not be labeled as pilot trials simply because they used surrogate end points [35].
   * A note related to relevant point.

3.5. Limitations and challenges in use of surrogate end points

Limitations and challenges in surrogate end point use were covered in 72% of the records (65/90). Figure 2 summarizes the limitations (and their reasons) and challenges in use of surrogate end points and how they may interact. First, 44 documents (of 65, 68%) noted that surrogate end points may fail to predict target outcome(s) or capture intervention effect(s). The reasons that could have resulted in this failure were drawn from a framework by Fleming and DeMets [4] (Figure 2). Second, 40% of the records (n = 26) reported that use of a surrogate end point could fail to provide information about the safety profile of interventions and/or result to approval of harmful interventions. This failure maybe due to an intervention's unintended consequences not in the known disease causal pathway (mentioned in eight documents, 12%) and reduction of sample size and follow-up (n = 10, 15%) which is common in trials using surrogate end points. Third, 12 of the included documents (18%) stated that surrogates may overestimate or underestimate intervention effect. This intervention effect uncertainty could be due to small study effects and various sources of bias [measurement error, evaluation bias, informative censoring, attrition bias, confounding, interpretation bias] (noted in 13 documents, 20%) which could blur the relationship between the surrogate and target outcome. A final limitation was synthesized based on the second and third limitation: surrogate end points make it difficult to assess the risk-benefit ratio of an intervention.

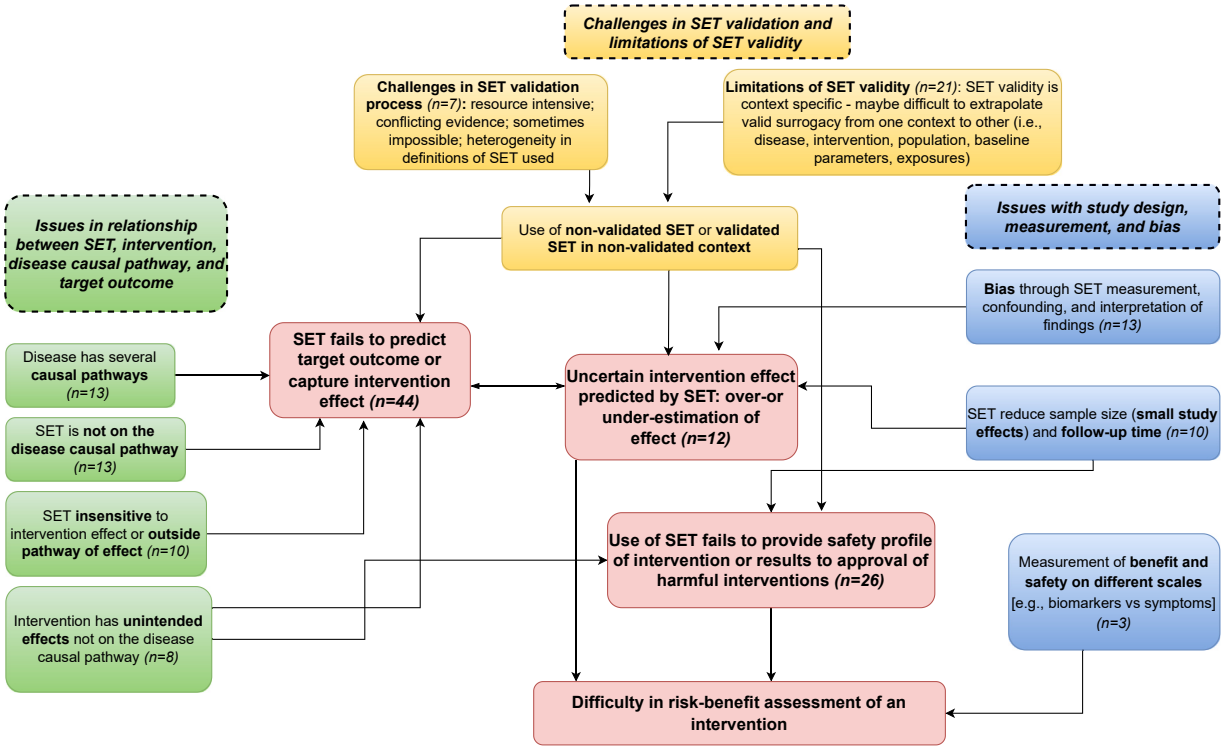

Fig. 2. Limitations (pink boxes) and challenges (yellow boxes) of use of surrogate end points (SET in the figure). Reasons for limitations include issues with relationship between surrogate end point, intervention, disease causal pathway, and target outcome (green boxes) and issues with study design, measurement, and bias (blue boxes). Figure drawn using draw.io. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Some records reported challenges and limitations of surrogate end point validity: seven records (of 65; 11%) noted challenges in validation process (i.e., resource intensive, sometimes impossible, heterogeneity in definition of surrogate end points used in trials, conflicting evidence); and 21 documents (32%) mentioned that surrogate validity is context-specific making it difficult to extrapolate valid surrogacy from one context to another, for example, diseases/disease stages, interventions, exposures, population, baseline characteristics, or previous treatment history. These challenges and limitations may result in use of nonvalidated surrogate end points or validated surrogates in nonvalidated contexts which could result in the surrogate failing to predict the target outcome, uncertainty in the intervention effect, or approval of interventions that are harmful [37] (Figure 2).

3.6. Advice and guidance in design and reporting of trials using surrogate end points

Advice or guidance in design and reporting of trials using surrogate end points was mentioned in 61% of the records (n = 55/90). Additionally, we synthesized the guidance and advice by implication based on findings from the definitions, limitations, and acceptability themes described above. Table 4 lists the guidance/advice either explicitly mentioned in records or implied from definitions, limitations, and acceptability findings along with examples of excepts from records and references in Table A.6 that covered each theme. In total, 17 guidance/advice items were synthesized from findings (Figure 3).

Seventeen documents (of 55; 31%) had an explicit call for providing justification of the surrogate end point selected using validation evidence. Another common recommendation was to make available the trial data on both surrogate and target outcomes to aid in surrogate validation. This could be achieved through (1) designing trials to collect both outcomes (suggested in 16% of the records [n = 9/55]) and (2) efforts to make these data available were recommended in 10 documents (18%). The need to estimate the predicted effect on the target outcome based on the effect on the surrogate was suggested in 16% (n = 9/55) of the records. Other items included stating that the primary outcome was a surrogate (stated in 5/55, 9%), interpretation of findings of the trial in the context of using a surrogate end point as a primary outcome (stated in 6/55, 11% and implied in 54/65, 83%), and informing participants that the trial was designed to evaluate the intervention effect using a surrogate end point (stated in 3/55, 6%)—see Table 4 for all the 17 items.

4. Discussion

We performed a scoping review of the literature on the use of surrogate end points in trials to explore how surrogate end points are defined, their reported limitations and acceptability, and current advice and guidance on the design and reporting. Based on 90 documents identified, we synthesized findings into 17 broad guidance/advice items that could be considered in the design and reporting of trials using surrogate end points as a primary outcome measure. These guidance items recommend inclusion of a statement on surrogacy in trial reports where the primary outcome is a surrogate end point and the target outcome that is being substituted for is also defined; justification for the selected surrogate end point, including evidence of validation in the setting in which the surrogate was being used; clarifying if sample size calculation was informed by surrogate validity; interpretation of findings of the trial in the context of using a surrogate as the primary outcome, including consideration of the benefit-risk ratio of the trial intervention; and informing participants that the trial was designed to evaluate an intervention's effect on a surrogate end point.

We found heterogeneity in how surrogate end points were defined, including how they were referred to, what was considered a surrogate, and what a surrogate substituted and/or predicted for. This finding is consistent with a review by Shi and Sargent (2009), who based on a collection of eight definitions, found differences in surrogate end point references, usage, measures regarded as surrogates, and what made a surrogate valid [47]. In our review, although biomarkers and mortality were the most common type of measures considered as surrogates and target outcomes, respectively, there was inconsistency in how intermediate outcomes, including symptoms and functional outcomes, were regarded; some definitions considered them surrogate end points, whereas others viewed them as target outcomes. Such inconsistences may lead to misinterpretation of trial findings and misalignment with health technology assessment efforts. For instance, EUnetHTA guidelines recommend reimbursement decisions to be based on 'long-term' and 'final end points'—mortality and severe morbidity, for example, fractures, stroke, myocardial infarction—rather than 'short-term outcomes' [48,49]. Therefore, there is need for a more complete, inclusive, and consensus-driven surrogate end point definition to inform better interpretation and use of trial findings.

One of the key recommendations identified by this review was the need to justify use of a surrogate end point using evidence of validation, that is, that the surrogate end point is a reliable substitute and predictor of the target outcome. Nevertheless, the current review also identified challenges in validation, that is, resource intensive as surrogate and target outcome data are needed, and validity of a surrogate is context-specific. Two solutions to aid in surrogate validation were identified from included documents: design of trials to collect data on both surrogate and target outcomes, even in a subset of the sample, and when these data are collected making it freely accessible for future

Table 4. Guidance or advice on design and reporting of trials using surrogate end points based on explicit statements or implications of findings on definitions, limitations, and acceptability

Guidance/Advice	Source of guidance/advice	*N* (%)	Illustrative excerpts [see Table A.6 for references]
1. State that primary outcome is considered a surrogate end point	Explicit recommendation for trialists to state whether outcome is a surrogate or target outcome	5/55 (9%)	*Clinical trialists and systematic reviewers need to be clearer in their reporting as to whether outcomes are surrogate or final patient* [2] *[Reference 3, 22,40, 51, 84]*
2. State the target outcome for which the surrogate end point is a substitute/predictor for	Explicit recommendation to specify the target outcome for which the surrogate end point is a substitute	4/55 (7%)	*The use of biomarkers as surrogate end points in a clinical trial requires the specification of the clinical end points that are being substituted* [18] *[Reference 13, 16, 40, 85]*
	Implication from the functions of a surrogate end point based on different definitions: substitution and/or prediction	67/71 (94%)	*Surrogate outcomes are often used in clinical trials as substitutes for final patient relevant outcomes. A key rationale for the use of surrogate outcomes in trials is not only substitution but the prediction of treatment benefit in the absence of data on patient relevant outcomes* [2] *[Reference 1, 3-5, 7-18, 20, 22, 23, 25, 27, 29-31, 35, 36, 39, 40, 42, 44, 45, 47-50, 52, 53, 55-69, 71-75, 79-89]*
3. Clear definition of the surrogate end point used	Explicit recommendation to clearly define surrogate end point used	2/55 (4%)	*Clearly define any clinical surrogate outcome(s) used* (e.g., *number of transfusions/numbers of thrombocytopenic days, corrected count increments)* [38] *[Reference 9, 14]*
	Implication of a challenge of surrogate end point validation process: heterogeneity in definitions of surrogate end points used	3/65 (5%)	*Disadvantages of [progression-free survival] PFS as clinical trials end points: Definitions vary among studies* [39] *[Reference 47, 52, 76]*
4. State the practical reason(s) for using a surrogate end point as a primary outcome	Explicit need to justify practical reason for using a surrogate end point	5/55 (9%)	*In the study report, (as well as in the study protocol), a justification for any deviations from the principles laid down in regulatory guidance is expected as well as justification for any alternative study design chosen* (e.g., *choice of surrogate end point, lack of randomization, lack of control group) and a justification for the statistical considerations should be given.* [28] *[Reference 13, 21, 24, 37, 80]*
	Implication from circumstances in which use of surrogate end points is acceptable such as in accelerated approval; Phase I/II trials; ethical or feasibility reasons; target outcome is biased; pilot trials	48/69 (70%)	*In addition, OS [overall survival] is potentially confounded or diminished by effective post progression therapies (including crossover). These limitations of the OS end point are the primary motivations for the use of surrogate end points in oncology clinical trials* [31] *[Reference 1-3, 5-7, 9-14, 17, 25, 29, 34, 35, 38-41, 44, 46-48, 50, 51, 55, 59-64, 70, 72-78, 80, 81, 83-85, 89]*
5. Justification for selected surrogate			
a) Evidence of validation	Explicit recommendation to justify surrogate end point use by providing evidence of validation	17/55 (31%)	*Reports of clinical trials should therefore state whether their collected outcomes are surrogate end points and provide a clear rationale for the selection of these surrogates, including reference to biological plausibility and evidence of validation* [40] *[Reference 3, 10, 13, 19, 21, 24, 33, 36, 38, 47, 51, 53, 84-88]*

(*Continued*)

Table 4. Continued

Guidance/Advice	Source of guidance/advice	*N* (%)	Illustrative excerpts [see Table A.6 for references]
	Implication of considerations in use of surrogate end points: they should be validated	49/69 (71%)	*Before a surrogate end point can replace a clinical end point for evaluating an experimental treatment in a phase III clinical trial, it must be formally validated to show that the treatment effect on the surrogate end point reliably predicts the treatment effect on the clinical end point* [31] *[Reference 1, 3, 5, 7, 9-16, 20, 25, 26, 32, 37-39, 43, 45-50, 52, 57-62, 64-69, 71-74, 81, 83, 85, 87-89]*
	Implication of a common limitation of surrogates: failure to adequately predict the target outcome, hence should be validated before use	44/65 (68%)	*Surrogates may fall short in their ability to predict outcomes in hard end points* [27] *[Reference 3, 5, 7, 9, 10-13, 15, 18, 26, 29, 31-33, 35, 37, 39, 42, 47, 48, 52, 53, 55, 57, 58, 60, 61, 68, 69, 71, 73-75, 79-81, 83, 85-90]*
b) Evidence of being specific to setting used, for example, intervention, disease, population	Explicit recommendation to clarify context (intervention/population/ disease state) where the surrogate validity holds	6/55 (11%)	*The use of biomarkers as surrogate end points in a clinical trial requires the specification of the clinical end points that are being substituted, class of therapeutic intervention being applied, and characteristics of population and disease state in which the substitution is being made* [18] *[Reference 16, 18, 24, 38, 47, 48]*
	Implication from a challenge/ limitation of surrogates: surrogate use and its validity is context specific	21/65 (32%)	*This illustrates that although a surrogate has been validated for one drug and one clinical outcome, it is not automatically valid for other drugs or clinical outcomes* [6] *[Reference 5, 13, 17, 18, 20, 30, 40, 47, 48, 49, 53, 58-60, 63, 67, 72, 76, 79, 83, 89]*
	Implication of acceptability of use of validated surrogate in one intervention if mechanism of action follows a similar disease casual pathway	7/69 (10%)	*Once validated for a particular intervention, the relationship may be considered valid for other interventions thought to affect the disease through the same pathway* [21] *[Reference 7, 11, 49, 52, 65, 87, 88]*
6. Outline surrogate end points measurement (including when, how, and by whom) and justification	Explicit recommendation for valid and reliable measurement of surrogate end points	5/55 (9%)	*For surrogate outcomes, specific details related to measurement and timing should be clearly stated* [38] *[Reference 4, 24, 25, 54, 81]*
	Implication from reasons for uncertainty in surrogates predicting the intervention effect: sources of bias	13/65 (20%)	*Mis-measured surrogate outcomes may produce mis-measured survival estimates* [41] *[Reference 8, 12, 17, 28, 31, 52, 62, 66, 69, 72, 74, 76, 81]*
7. State if the sample size calculation was explicitly informed by statistical metrics of surrogate validity (such as the surrogate threshold effect [STE] or its equivalent)	Explicit recommendation on calculation of sample size or caution on small sample sizes	3/55 (6%)	*In trials using surrogate end points the sample size has to be based not only on the expected treatment efficacy but also on the validity of the surrogate marker used* [33] *[Reference 37, 44, 66]*
	Implication from a limitation of surrogates: over or underestimation in predicting target outcome	12/65 (18%)	*Several reasons may explain why trials assessing surrogate end points showed larger treatment effects than trials assessing final end points. The first relate to small study effects: As expected, we found a smaller sample size for trials using surrogate outcomes than for trials using final patient relevant outcomes* [2] *[Reference 2, 13,15, 22, 23, 29, 42, 48, 63, 73, 81, 84]*
8. Outline baseline characteristics of the population and the surrogate end point	Explicit call to identify skewed baseline characteristics of the surrogate	1/55 (2%)	*Also, the statistician must be aware that the baseline characteristics of patients who enter a trial may not be constant over the course of*

(*Continued*)

Table 4. Continued

Guidance/Advice	Source of guidance/advice	*N* (%)	Illustrative excerpts [see Table A.6 for references]
			the study. In particular, a shift tends to occur in trials in which the very ill prevalent cases enter earliest, whereas the less ill incident cases enter later. Such a subtle temporal shift in baseline distributions affects the distribution of outcome, for it often induces greater variability than expected [42] *[Reference 66]*
	Implication from limitations in use of some surrogates whose treatment effect may be influenced by baseline characteristics	1/65 (2%)	*Unlike other surrogate markers of cardiovascular disease, the ability of carotid IMT to reflect treatment effects may be influenced by the choice of the patient population, baseline parameters, and the previous treatment history of trial participants* [43]*[Reference 53]*
9. State analysis methods used to improve robustness of findings	Explicit call for caution of limitations introduced by bias and use of specific analytical methods to improve robustness of findings such as sensitivity analyses	4/55 (7%)	*In considering PFS as a clinical trial end point in large randomized phase II and phase III clinical trials, we must be mindful of the limitations with regard to the potential for bias and informative and interval censoring, the necessity of sensitivity analyses in evaluating the robustness of conclusions based on PFS* [39] *[Reference 54, 66, 76, 78]*
	Implication from reasons for uncertainty in surrogates predicting the intervention effect: sources of bias	13/65 (20%)	*Factors such as measurement error, evaluation bias, attrition bias, or informative censoring may weaken the association between the surrogate and hard end point such that its predictive value is lost* [27] *[8, 12, 17, 28, 31, 52, 62, 66, 69, 72, 74, 76, 81]*
10. If using surrogate as part of a composite end point, report results in absolute numbers	Explicit call for reporting of results in absolute numbers given the complexities of surrogate and composite end points and relative reporting	2/55 (4%)	*Overall, our study shows that the use of surrogate and composite end points and end points involving disease-specific mortality is common. In addition, articles frequently report results in relative numbers. These findings highlight the need for educational efforts to ensure that readers understand the complexities of these end points and of relative risk reporting. Finally, medical journals may consider instituting editorial policies mandating the reporting of results in absolute numbers* [44] *[Reference 29, 42]*
11. Provide an estimate of the predicted effect on the target outcome based on the observed effect on the surrogate end point	Explicit guidance on stating the magnitude of treatment effect associated with use of a surrogate end point	9/55 (16%)	Table 1. *Users' Guide for a Surrogate End Point Trial. Are the Results Valid? What Were the Results? How Large, Precise, and Lasting Was the Treatment Effect?* [45] *[Reference 8, 21, 24, 28, 47, 63, 71, 86, 87]*
	Implication from the main limitation of surrogates: failure to predict the target outcome or uncertainty in intervention effect	49/65 (75%)	*See excerpts and reference numbers from implications of limitations under item 5a and 7 above*
12. Consider whether the trial sample size and follow up period is sufficient to adequately capture potential harms of the intervention being tested	Explicit call to consider the relationship between surrogate end point and intervention including off-target effects of intervention and a commitment to monitor for unintended effects	3/55 (6%)	*The control group should reflect the current treatment standard when the trial was designed. But we need to do this with a willingness and commitment to monitor and report survival results, recognizing that what happened in the BELLINI trial is a sobering reminder of the limitations of surrogate end points, especially as we engage with increasingly complex strategies such as combining chemotherapy with immunotherapy to treat cancer* [5] *[Reference 24, 66, 70]*

(*Continued*)

Table 4. Continued

Guidance/Advice	Source of guidance/advice	*N* (%)	Illustrative excerpts [see Table A.6 for references]
	Implication from a limitation of use of surrogate end points: could fail to provide safety profile of intervention and/or result to approval of harmful interventions	26/65 (40%)	*These unintended mechanisms can readily cause the effect on the true clinical outcome to be inconsistent with what would have been expected solely on the basis of evaluation of surrogate end points. These mechanisms are insidious because they are often unanticipated and unrecognized* [4] *[Reference 7, 10, 11, 13, 15, 17, 35, 37, 40, 55, 58, 61, 66, 68, 70-72, 75, 79, 81, 83-85, 87-89]*
13. Interpretation of findings of the trial in the context of using a surrogate primary end point including its known validity and the potential benefit-risk ratio of the tested intervention for participants	Explicit recommendation to consider the benefit-risk trade-off of using a surrogate and factor limitations of surrogate while making conclusions	6/55 (11%)	*Thus, the ENHANCE trial further highlights the importance of understanding the objective strengths and limitations of surrogate markers in the context of decision-making in cardiovascular drug development and patient care* [43] *[Reference 2, 22, 53, 65, 68, 80]*
	Implication from limitations of surrogates: they may not accurately predict target outcomes and fail capture the unintended effects of an intervention	54/65 (83%)	*Surrogate End Points Do Not Always Provide an Adequate Indication of the Safety Profile of a Substance. Although there may be a correlation between the surrogate end point and the clinical end point, it is not possible to make a conclusive estimate of the benefit/risk ratio for a drug on the basis of a surrogate end point. Surrogate end points do not completely indicate the safety profile of a drug* [24] *[Reference 2, 3, 5, 7, 9, 10-13, 15, 17, 18, 22, 23, 26, 29, 31-33, 35, 37, 39, 40, 42, 47, 48, 52, 53, 55, 57, 58, 60, 61, 63, 66, 68-72, 73-75, 79-81, 83-90]*
14. Generalisability of findings to other populations, settings, intervention types and if applicable, specify population that would most benefit from trial findings	Explicit call for caution in generalizing evidence from surrogate end points	1/55 (2%)	*We need to explicitly lower unreasonable expectations of the impact of innovations, when surrogate end points are used, and when findings (including of RCTs) may not be transferable beyond specific and similar context* [34] *[Reference 68]*
	Implication from a challenge in use of validated surrogate end points: they are context-specific hence difficult to extrapolate from one context to another	21/65 (32%)	*See relevant excerpt and reference numbers on item 5b above*
15. State if there are explicit to plans to extend follow up or conduct subsequent analyses/ studies to verify benefit of current findings on the patient relevant target outcome	Explicit guidance to conduct subsequent studies/analyses with patient relevant outcomes to verify health benefit and link future work to current publication	4/55 (7%)	*Furthermore, analyses based on even the best surrogate cannot fully replace a long-term analysis based directly on mortality, because the natural history assumptions need to be verified. Subsequent analyses, based on the actual cause specific mortality, are essential* [30] *[Reference 2, 23, 62, 88]*
	Implication from acceptability of surrogates: confirmatory studies would be required after an intervention is approved through the accelerated approval pathway	9/69 (13%)	*Hold the FDA responsible for ensuring that clinically meaningful end points are used in confirmatory trials (the FDA claims this is already the case)* [26] *[Reference 10, 17, 35, 37, 38, 50, 55, 69, 74]*
16. Collection and availability of both surrogate and target outcome data			
a) Design of trials that collect both surrogate and target outcome data	Advise for trials to collect both surrogate and target outcome data even if it is for a subset of participants to aid in evaluation of the surrogate	9/55 (16%)	*Due to the difficulty in obtaining true outcomes on many subjects, the methods we have proposed have useful applications in clinical trials. Designing studies such that surrogate outcomes are collected on all patients and true*

(*Continued*)

Table 4. Continued

Guidance/Advice	Source of guidance/advice	*N* (%)	Illustrative excerpts [see Table A.6 for references]
			outcomes only collected on a subsample of patients can save on trial costs and ensure that an adequate number of patients are enrolled [41] *[Reference 8, 17, 30, 39, 42, 44, 49, 63, 90]*
b) If surrogate and patient relevant outcome data were collected, make data available for evaluation of the surrogate	Explicit identification of the need of both trial level surrogate and patient relevant outcome data for validation of surrogates	10/55 (18%)	*To better accomplish validation studies for specific intervention and specific indication, we need a large pool of clinical trials data in which both OS and the surrogate endpoints were captured. An easy access to the data from completed clinical trials from many pharmaceutical industry or government spooned trials for research use is the key to such success* [46] *[Reference 25, 30, 43, 49, 56, 67, 79, 81, 82, 84]*
17. Inform participants/patients before enrollment that trial was designed to evaluate an intervention's effect using a surrogate end point (rather than a target outcome)	Explicit recommendation to inform/ educate trial participants on limitations of using surrogates in participant information sheets and consent forms	3/55 (6%)	*Patients should, therefore, be educated throughout the informed consent process that higher response rate, better depth of response, or longer time to progression does not necessarily mean that a new treatment will prolong life. As investigators, it is our duty to patients that trials are designed to show clear clinical benefit, and that any reliance on surrogate end points is communicated clearly to patients with the attendant risks* [5] *[Reference 42, 55, 70]*

The items were synthesized for randomized controlled trials but could be relevant to nonrandomized trials, observational studies, and other studies using surrogate outcomes.

validation. Although it may be impractical to require trialists to collect additional specific outcomes, it would be appropriate for trialists to ensure open access availability of participant-level surrogate and target outcomes data (when both collected) to allow future surrogate validation. It is worth noting that although surrogate end point validation traditionally has been carried out using trial data [40] and available validation methods have been developed for

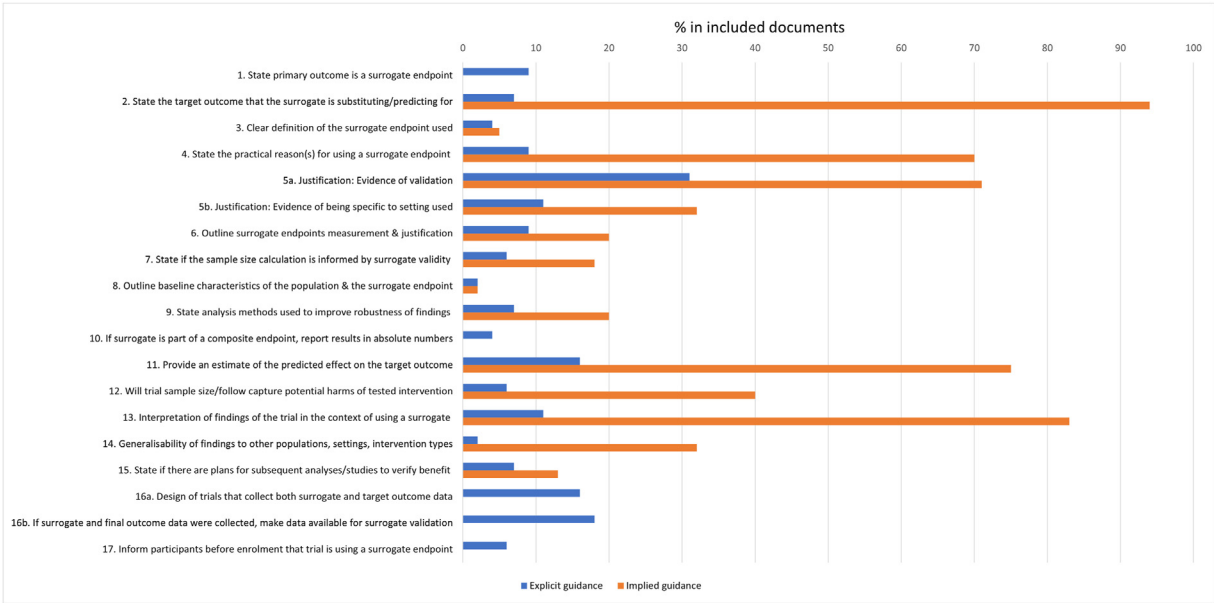

Fig. 3. The 17 synthesized trial design and reporting items and percentage proportion of explicit or implied guidance from the included documents. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

such data [50], there has been considerable interest in use of real-world data, from observational or other nonrandomized studies, when assessing new health technologies with research carried out into methods for using real-world data to enhance evidence base in surrogate end point validation [51].

Another key recommendation of this review was increased involvement of patients and public in trials that use surrogate end points. First, we found recommendations for involving patient representatives, among other stakeholders, on consensus building on what should constitute ''unmet medical need'' in regulatory accelerated approval pathways [25,26]. Additionally, we did not find a comprehensive definition from included records on what constituted a ''rare disease'' or ''long follow-up'' necessitating a consensus-driven definition of these issues. Second, the review synthesized a reporting guidance item that would require trialists to inform participants that the trial is designed to use a surrogate end point rather than a target outcome. This would involve educating participants on surrogate end points and their potential limitation of failing to predict clinical benefit or provide a safety profile of an intervention. Nevertheless, involvement could start earlier, at the point of selection of primary outcomes as recent research has reported that patients and health professionals agreed with the choice of primary outcome by trialists only 28% of the time in breast cancer and nephrology trials [52]. Although one of the articles included in this review found that patients preferred use of patient important outcomes rather than surrogate end points as primary outcomes in trials [36], the review also identified various practical and ethical reasons when use of surrogates as primary outcomes is justifiable. In such circumstances, selection of surrogate end points as primary outcomes can be informed by relevant core outcome sets [52], if available, or consensus among stakeholders.

A strength of our review is the collation of scientific and gray literature with no restriction to clinical area, regions, or time periods. However, our review had some limitations; for example, exclusion of literature published in languages other than English. Nevertheless, we included documents in English from regions where the primary language is not English: Europe, Asia, and South America. Furthermore, we may have missed some relevant literature based on our search strategy, such as failure to use ''early end points'' in searches. However, our review did not aim to be exhaustive but to identify important items for consideration when using surrogate end points in trials [13]; and we included literature from other sources such as reference list of included records, hence a low risk of missing important literature. Additionally, data extraction was not done in duplicate and literature search was limited to publicly available documents. Furthermore, tight project timelines did not allow for consultation and patient and public involvement in the review as had been planned. Nevertheless, we had input from the projects patient and public involvement lead who is one of the co-authors (D.S.). Also, we included an article relating to patients preferences on types of outcomes [36] and synthesized an item on patient and public engagement (item 17). Finally, our review would have benefited from wider solicitation of insights from other stakeholders including trial sponsors and regulatory assessors whose views may not be captured in publicly available literature. Nevertheless, our gray literature screening included six documents from regulatory bodies. Additionally, we had representation from industry in our project team and a co-author of this article (M.O.).

In conclusion, this review provides practical guidance on the design and reporting of trials using surrogate end points for trialists, methodologists, editors, and decision makers. The guidance items identified in this review have been rated in a two-round Delphi survey [12], which will in turn contribute to the final published SPIRIT-SURROGATE and CONSORT-SURROGATE extension checklists.

Declaration of competing interest

Sylwia Bujkiewicz is a member of the NICE Decision Support Unit and NICE Guidelines Technical Support Unit. She has served as a paid consultant, providing methodological advice, to NICE, Roche, IQVIA, and RTI Health Solutions, received payments for educational events from Roche, and has received research funding from European Federation of Pharmaceutical Industries and Associations and Johnson & Johnson. Mario Ouwens works for and has shares in AstraZeneca. Joseph Ross is an Associate Editor at BMJ and co-founder (unpaid) of medRxiv; research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology, from the Food and Drug Administration for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation program (U01FD005938); from the Agency for Healthcare Research and Quality (R01HS022882); from the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644); and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International; expert witness at the request of Relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. Nancy Butcher has received consulting fees from Nobias Therapeutics, Inc.

Acknowledgments

We are indebted to Professor Amber Young (University of Bristol) for her contribution to the planning and conduct

of the CONSORT/SPIRIT-SURROGATE project and who sadly passed away in September 2022.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2023.06.013.

References

[1] FDA-NIH Biomarker Working Group. BEST (biomarkers, endpoints, and other tools) resource; 2016. Available at https://www.ncbi.nlm.nih.gov/books/NBK326791. Accessed August 6, 2023.

[2] Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JAC, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. BMJ 2013;346:f457.

[3] Cohen D. Rosiglitazone: what went wrong? BMJ 2010;341:c4848.

[4] Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med 1996;125:605−13.

[5] Kumar S, Rajkumar SV. Surrogate endpoints in randomised controlled trials: a reality check. Lancet 2019;394:281−3.

[6] La Cour JL, Brok J, Gotzsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. BMJ 2010;341:380.

[7] Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158:200−7.

[8] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.

[9] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA 2013;309:814−22.

[10] Butcher NJ, Monsour A, Mew EJ, Chan AW, Moher D, Mayo-Wilson E, et al. Guidelines for reporting outcomes in trial reports: the CONSORT-outcomes 2022 extension. JAMA 2022;328:2252−64.

[11] Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010;7(2):e1000217.

[12] Manyara AM, Davies P, Stewart D, Weir CJ, Young A, Butcher NJ, et al. Protocol for the development of SPIRIT and CONSORT extensions for randomised controlled trials with surrogate primary endpoints: SPIRIT-SURROGATE and CONSORT-SURROGATE. BMJ Open 2022;12(10):e064304.

[13] Manyara AM, Davies P, Stewart D, Wells V, Weir C, Young A, et al. Scoping and targeted reviews to support development of SPIRIT and CONSORT extensions for randomised controlled trials with surrogate primary endpoints: protocol. BMJ Open 2022;12(10):e062798.

[14] McGowan J, Straus S, Moher D, Langlois EV, O'Brien KK, Horsley T, et al. Reporting scoping reviews-PRISMA ScR extension. J Clin Epidemiol 2020;123:177−9.

[15] Butcher NJ, Mew EJ, Monsour A, Chan AW, Moher D, Offringa M. Outcome reporting recommendations for clinical trial protocols and reports: a scoping review. Trials 2020;21(1):620.

[16] Covidence. Covidence systematic review software. Melbourne, Australia: Veritas Health Innovation; 2021.

[17] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19−32.

[18] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89−95.

[19] Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? JAMA 1999;282:790−5.

[20] Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med 1989;8:431−40.

[21] Gøtzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. Int J Technol Assess Health Care 1996;12(2):238−46.

[22] Hillis A, Seigel D. Surrogate endpoints in clinical trials: ophthalmologic disorders. Stat Med 1989;8:427−30.

[23] Greene SJ, Mentz RJ, Fiuzat M, Butler J, Solomon SD, Ambrosy AP, et al. Reassessing the role of surrogate end points in drug development for heart failure. Circulation 2018;138:1039−53.

[24] Kleist P. The impact of biomarkers and surrogate end points on regulatory approval of new drugs. Heart Drug 2002;2(5):235−45.

[25] Hey SP, Kesselheim AS, Patel P, Mehrotra P, Powers JH. US Food and Drug Administration recommendations on the use of surrogate measures as end points in new anti-infective drug approvals. JAMA Intern Med 2020;180(1):131−8.

[26] Lenzer J, Brownlee S. Should regulatory authorities approve drugs based on surrogate endpoints? BMJ 2021;374:n2059.

[27] Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? BMC Med 2017;15(1):134.

[28] EMEA. Guideline on clinical trials in small populations. London: EMEA; 2007.

[29] Svensson S, Menkes DB, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. JAMA Intern Med 2013;173(8):611−2.

[30] 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(4):178−85.

[31] Zhao F. Surrogate end points and their validation in oncology clinical trials. J Clin Oncol 2016;34:1436−7.

[32] Fleming TR. Surrogate markers in AIDS and cancer trials. Stat Med 1994;13:1423−35. discussion 37.

[33] Cleophas TJ, Zwinderman AH, Chaib AH. Novel procedures for validating surrogate endpoints in clinical trials. Curr Clin Pharmacol 2007;2(2):123−8.

[34] Pai M, Schumacher SG, Abimbola S. Surrogate endpoints in global health research: still searching for killer apps and silver bullets? BMJ Glob Health 2018;3:e000755.

[35] Campbell M, Lancaster G, Eldridge S. A randomised controlled trial is not a pilot trial simply because it uses a surrogate endpoint. Pilot Feasibility Stud 2018;4(1):1−4.

[36] Murad MH, Shah ND, Van Houten HK, Ziegenfuss JY, Deming JR, Beebe TJ, et al. Individuals with diabetes preferred that future trials use patient-important outcomes and provide pragmatic inferences. J Clin Epidemiol 2011;64:743−8.

[37] Furgerson JL, Hannah WN, Thompson JC. Challenge of surrogate endpoints. South Med J 2012;105:156−60.

[38] Meyer E, Delaney M, Lin Y, Morris A, Pavenski K, Tinmouth A, et al. A reporting guideline for clinical platelet transfusion studies from the BEST Collaborative. Transfusion 2013;53(6):1328−34.

[39] Villaruz LC, Socinski MA. The clinical viewpoint: definitions, limitations of RECIST, practical considerations of measurement. Clin Cancer Res 2013;19(10):2629−36.

[40] Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to review the role of surrogate end points in health policy: state of the art and the way forward. Value Health 2017;20(3):487−95.

[41] Zee J, Xie SX. Assessing treatment effects with surrogate survival outcomes using an internal validation subsample. Clin Trials 2015;12:333−41.

[42] Wittes J, Lakatos E, Probstfield J. Surrogate endpoints in clinical trials: cardiovascular diseases. Stat Med 1989;8:415−25.

[43] Colatsky TJ. Reassessing the validity of surrogate markers of drug efficacy in the treatment of coronary artery disease. Curr Opin Invest Drugs 2009;10(3):239−44.

[44] Hochman M, McCormick D. Endpoint selection and relative (versus absolute) risk reporting in published medication trials. J Gen Intern Med 2011;26:1246−52.

[45] 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.

[46] Zhang J, Pilar MR, Wang X, Liu J, Pang H, Brownson RC, et al. Endpoint surrogacy in oncology phase 3 randomised controlled trials. Br J Cancer 2020;123:333−4.

[47] Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol 2009;14(2):102−11.

[48] Assessment ENfHT. Endpoints used in relative effectiveness assessment of pharmaceuticals: clinical endpoints. Final version [internet]. Copenhagen: EUnetHTA; 2013.

[49] EUnetHTA 21 − individual practical guideline document D4.4 − OUTCOMES (ENDPOINTS). EUnetHTA; 2023.

[50] Bujkiewicz S, Achana F, Papanikos T, Riley R, Abrams K. Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. Technical Support Documents. London: National Institute for Health and Care Excellence (NICE); 2019.

[51] Wheaton L, Papanikos A, Thomas A, Bujkiewicz S. Using Bayesian evidence synthesis methods to incorporate real-world evidence in surrogate endpoint evaluation. Med Decis Making 2023;43:539−52.

[52] Treweek S, Miyakoda V, Burke D, Shiely F. Getting it wrong most of the time? Comparing trialists' choice of primary outcome with what patients and health professionals want. Trials 2022;23(1):537.