



**HAL**  
open science

## **A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials**

Oriana Ciani, Anthony M Manyara, Philippa Davies, Derek Stewart, Christopher J Weir, Amber E Young, Jane Blazeby, Nancy J Butcher, Sylwia Bujkiewicz, An-Wen Chan, et al.

### ► To cite this version:

Oriana Ciani, Anthony M Manyara, Philippa Davies, Derek Stewart, Christopher J Weir, et al.. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine*, 2023, 65, pp.102283. 10.1016/j.eclinm.2023.102283 . hal-04385165

**HAL Id: hal-04385165**

**<https://hal.science/hal-04385165>**

Submitted on 10 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials**

**Citation for published version:**

Ciani, O, Manyara, AM, Davies, P, Stewart, D, Weir, CJ, Young, AE, Blazeby, J, Butcher, NJ, Bujkiewicz, S, Chan, AW, Dawoud, D, Offringa, M, Ouwens, M, Hróbjartsson, A, Amstutz, A, Bertolaccini, L, Bruno, VD, Devane, D, Faria, CDCM, Gilbert, PB, Harris, R, Lassere, M, Marinelli, L, Markham, S, Powers, JH, Rezaei, Y, Richert, L, Schwendicke, F, Tereshchenko, LG, Thoma, A, Turan, A, Worrall, A, Christensen, R, Collins, GS, Ross, JS & Taylor, RS 2023, 'A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials', *EClinicalMedicine*, vol. 65, 102283. <https://doi.org/10.1016/j.eclinm.2023.102283>

**Digital Object Identifier (DOI):**

[10.1016/j.eclinm.2023.102283](https://doi.org/10.1016/j.eclinm.2023.102283)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

EClinicalMedicine

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials



Oriana Ciani,<sup>a,ap</sup> Anthony M. Manyara,<sup>b,ap</sup> Philippa Davies,<sup>c</sup> Derek Stewart,<sup>d</sup> Christopher J. Weir,<sup>e</sup> Amber E. Young,<sup>f</sup> Jane Blazeby,<sup>c,f,g</sup> Nancy J. Butcher,<sup>h,i</sup> Sylwia Bujkiewicz,<sup>j</sup> An-Wen Chan,<sup>k,l</sup> Dalia Dawoud,<sup>m</sup> Martin Offringa,<sup>h,n</sup> Mario Ouwens,<sup>o</sup> Asbjørn Hróbjartsson,<sup>p,q</sup> Alain Amstutz,<sup>r</sup> Luca Bertolaccini,<sup>s</sup> Vito Domenico Bruno,<sup>t</sup> Declan Devane,<sup>u,v</sup> Christina D. C. M. Faria,<sup>w</sup> Peter B. Gilbert,<sup>x</sup> Ray Harris,<sup>d</sup> Marissa Lassere,<sup>y</sup> Lucio Marinelli,<sup>z,aa</sup> Sarah Markham,<sup>ab</sup> John H. Powers,<sup>ac</sup> Yousef Rezaei,<sup>ad,ae,af</sup> Laura Richert,<sup>ag</sup> Falk Schwendicke,<sup>ah</sup> Larisa G. Tereshchenko,<sup>ai</sup> Achilles Thoma,<sup>aj</sup> Alparslan Turan,<sup>ak</sup> Andrew Worrall,<sup>d</sup> Robin Christensen,<sup>al</sup> Gary S. Collins,<sup>am</sup> Joseph S. Ross,<sup>an,ao</sup> and Rod S. Taylor,<sup>b,\*,ap</sup>

<sup>a</sup>Centre for Research on Health and Social Care Management, SDA Bocconi School of Management, Milan, Italy

<sup>b</sup>MRC/CSO Social and Public Health Sciences Unit, School of Health and Wellbeing, University of Glasgow, Glasgow, UK

<sup>c</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

<sup>d</sup>Patient and Public Involvement Partner, UK

<sup>e</sup>Edinburgh Clinical Trials Unit, Usher Institute, University of Edinburgh, Edinburgh, UK

<sup>f</sup>Bristol NIHR Biomedical Research Centre, Bristol, UK

<sup>g</sup>University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, UK

<sup>h</sup>Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Canada

<sup>i</sup>Department of Psychiatry, University of Toronto, Toronto, Canada

<sup>j</sup>Biostatistics Research Group, Department of Population Health Sciences, University of Leicester, Leicester, UK

<sup>k</sup>Women's College Research Institute, Toronto, Canada

<sup>l</sup>Department of Medicine, University of Toronto, Toronto, Canada

<sup>m</sup>Science, Evidence and Analytics Directorate, Science Policy and Research Programme, National Institute for Health and Care Excellence, London, UK

<sup>n</sup>Department of Paediatrics, University of Toronto, Toronto, Canada

<sup>o</sup>AstraZeneca, Mölndal, Sweden

<sup>p</sup>Centre for Evidence-Based Medicine Odense (CEBMO) and Cochrane Denmark, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

<sup>q</sup>Open Patient Data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark

<sup>r</sup>CLEAR Methods Center, Division of Clinical Epidemiology, Department of Clinical Research, University Hospital Basel and University of Basel, Basel, Switzerland

<sup>s</sup>Department of Thoracic Surgery, IEO, European Institute of Oncology IRCCS, Milan, Italy

<sup>t</sup>Department of Minimally Invasive Cardiac Surgery, IRCCS Galeazzi – Sant'Ambrogio Hospital, Milan, Italy

<sup>u</sup>University of Galway, Galway, Ireland

<sup>v</sup>Health Research Board-Trials Methodology Research Network, University of Galway, Galway, Ireland

<sup>w</sup>Department of Physical Therapy, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

<sup>x</sup>Fred Hutchinson Cancer Centre, Seattle, USA

<sup>y</sup>St George Hospital and School of Population Health, The University of New South Wales, Sydney, Australia

<sup>z</sup>Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genova, Genoa, Italy

<sup>aa</sup>IRCCS Ospedale Policlinico San Martino, Genoa, Italy

<sup>ab</sup>Department of Biostatistics, King's College London, London, UK

<sup>ac</sup>George Washington University School of Medicine, Washington, USA

<sup>ad</sup>Heart Valve Disease Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran

<sup>ae</sup>Ardabil University of Medical Sciences, Ardabil, Iran

<sup>af</sup>Behyan Clinic, Pardis New Town, Tehran, Iran

<sup>ag</sup>University Bordeaux, INSERM, Institut Bergonié, CHU Bordeaux, BPH U1219, CIC-EC 1401, RECaP and Euclid/F-CRIN, Bordeaux, France

<sup>ah</sup>Charité Universitätsmedizin Berlin, Berlin, Germany

<sup>ai</sup>Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

<sup>aj</sup>McMaster University, Hamilton, Ontario, Canada

<sup>ak</sup>Department of Outcomes Research, Anesthesiology Institute, Cleveland Clinic, OH, USA

\*Corresponding author. Professor of Population Health Research, MRC/CSO Social and Public Health Sciences Unit & Robertson Centre for Biostatistics, School of Health and Well Being, Clarice Pears Building, Byres Road, University of Glasgow, Glasgow, G12 8TB, Scotland, United Kingdom.

E-mail address: [rod.taylor@glasgow.ac.uk](mailto:rod.taylor@glasgow.ac.uk) (R.S. Taylor).

<sup>ap</sup>Joint first authors.

<sup>al</sup>Section for Biostatistics and Evidence-Based Research, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen & Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark

<sup>am</sup>Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK

<sup>an</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA

<sup>ao</sup>Section of General Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

eClinicalMedicine  
2023;65: 102283

Published Online 17  
October 2023  
<https://doi.org/10.1016/j.eclinm.2023.102283>

## Summary

**Background** Interventional trials that evaluate treatment effects using surrogate endpoints have become increasingly common. This paper describes four linked empirical studies and the development of a framework for defining, interpreting and reporting surrogate endpoints in trials.

**Methods** As part of developing the CONSORT (Consolidated Standards of Reporting Trials) and SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) extensions for randomised trials reporting surrogate endpoints, we undertook a scoping review, e-Delphi study, consensus meeting, and a web survey to examine current definitions and stakeholder (including clinicians, trial investigators, patients and public partners, journal editors, and health technology experts) interpretations of surrogate endpoints as primary outcome measures in trials.

**Findings** Current surrogate endpoint definitional frameworks are inconsistent and unclear. Surrogate endpoints are used in trials as a substitute of the treatment effects of an intervention on the target outcome(s) of ultimate interest, events measuring how patients feel, function, or survive. Traditionally the consideration of surrogate endpoints in trials has focused on biomarkers (e.g., HDL cholesterol, blood pressure, tumour response), especially in the medical product regulatory setting. Nevertheless, the concept of surrogacy in trials is potentially broader. Intermediate outcomes that include a measure of function or symptoms (e.g., angina frequency, exercise tolerance) can also be used as substitute for target outcomes (e.g., all-cause mortality)—thereby acting as surrogate endpoints. However, we found a lack of consensus among stakeholders on accepting and interpreting intermediate outcomes in trials as surrogate endpoints or target outcomes. In our assessment, patients and health technology assessment experts appeared more likely to consider intermediate outcomes to be surrogate endpoints than clinicians and regulators.

**Interpretation** There is an urgent need for better understanding and reporting on the use of surrogate endpoints, especially in the setting of interventional trials. We provide a framework for the definition of surrogate endpoints (biomarkers and intermediate outcomes) and target outcomes in trials to improve future reporting and aid stakeholders' interpretation and use of trial surrogate endpoint evidence.

**Funding** SPIRIT-SURROGATE/CONSORT-SURROGATE project is Medical Research Council Better Research Better Health (MR/V038400/1) funded.

**Copyright** © 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Surrogate endpoints; Target outcomes; Intermediate outcomes

## Introduction

The recent accelerated approval by the US Food and Drug Administration (FDA) of two biologics for the treatment of early Alzheimer's disease (aducanumab and lecanemab), with a third under review (donanemab), has brought back into sharp focus the controversy over the use of surrogate endpoints<sup>aq</sup> in informing healthcare decision-making. Evidence from placebo-

controlled randomised controlled trials (RCTs) showing an effect on reduction in amyloid plaque protein in the brain was deemed as sufficient evidence for their marketing authorisation, despite mixed findings of their effects on cognitive impairment.<sup>2-5</sup>

In the case of this example, the trial endpoint of amyloid load (assessed by magnetic resonance imaging) is a biomarker, i.e., 'a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention'.<sup>6</sup> Over the last three decades, biomarkers have become widely accepted by global regulators as surrogate endpoints, substituting for treatment effects on the trial target outcomes of interest,

<sup>aq</sup>Other descriptive terms used with surrogate include 'outcome'; 'marker'; 'measure'; 'observation'; 'parameter'. Surrogate endpoints also referred to as 'early', 'replacement', 'proxy', 'substitute' endpoints/outcomes/measures/markers.<sup>1</sup>

### Research in context

#### Evidence before this study

Over the last three decades, an increasing proportion of medical products have been approved by global regulators based on interventional trials using validated biomarkers (e.g., LDL cholesterol, blood pressure, tumour response) as a surrogate endpoint in place of the treatment effect on target outcome of interest (e.g., all-cause mortality). There has been less consideration, however, of the wider use of ‘intermediate outcomes’ that can include measures of function or symptoms (e.g., angina frequency, exercise tolerance)—as potential surrogate endpoints in trials. As a result, current definitions, and interpretational frameworks for surrogate endpoints in trials may either be lacking or be too narrow.

#### Added value of this study

This study presents an analysis examining current definitions and an assessment of how stakeholders (including clinicians, trial investigators, patients and public partners, journal

editors, and health technology experts) interpret trial endpoints as surrogates or target outcomes. Based on these findings, we provide a framework that extends the definition and interpretation of surrogate endpoints in trials beyond biomarkers to include intermediate outcomes.

#### Implications of all the available evidence

The focus of surrogate endpoint use in trials has traditionally been directed by regulatory bodies such as the US Food and Drug administration (FDA) and European Medicines Agency (EMA). However, given the importance of trial findings to clinical practice and policy, the paradigm and importance of surrogate endpoints extends beyond the regulatory setting to patients, clinicians, payers, and other stakeholders. In the future, the design and interpretation of surrogate endpoints in trials need more attention on their broader definition, reporting, interpretation, and relation to target outcomes.

measuring how patients feel, function, or survive.<sup>7–9</sup> In contrast, the RCTs of aducanumab and lecanemab primary endpoint of cognitive impairment (assessed using the Clinical Dementia Rating–Sum of Boxes (CDR–SB)<sup>10</sup>), can be directly perceived by and has potential value for patients, clinicians, and other stakeholders—a trial target outcome. This primary endpoint can also be considered an intermediate outcome, substituting for longer-term target outcomes, such as Alzheimer’s disease, irreversible disability requiring institutionalised care and/or premature death. Measuring effects on longer-term target outcomes would require trials of much longer follow up, larger sample size, and higher costs. Thus, the trials’ use of cognitive function as an intermediate outcome can be interpreted as a surrogate endpoint, too.<sup>11–13</sup> Dependent on the definition used, it is estimated that between 20 and 80 percent of published trials use surrogate endpoints as primary outcomes, the most common areas of application being cardiovascular disease, cancer, and infectious disease.<sup>1,14,15</sup>

Reporting of healthcare interventional trials, and appropriate application of their findings to inform practice and policy, centrally depend on clarification as to how biomarkers and intermediate outcomes are defined and interpreted as a surrogate endpoint or target outcome. As part of a UK Medical Research Council-funded project to develop the CONSORT (Consolidated Standards of Reporting Trials) and SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) extensions for RCT protocols and final reports using surrogate endpoints, we undertook four linked empirical studies to examine the definition and interpretation of surrogate endpoints in trials.<sup>16,17</sup> This included a scoping review of current surrogate endpoint

definitions, rated for their acceptability and clarity as part of an e-Delphi study, and assessment in a hybrid consensus meeting with an extension of the e-Delphi study to gauge how a sample of international stakeholders (including clinicians, trial investigators, patients and public partners, journal editors, and health technology experts) approached the judgment of intermediate outcomes in RCTs as a surrogate endpoints or target outcomes. We present a summary of these empirical studies and, using their findings, propose a definitional framework to inform the better reporting of trials using surrogate endpoints and to aid stakeholders’ (including patients, clinicians, regulators, and payers) in their interpretation of surrogate endpoint evidence.

### Methods

The methods for developing SPIRIT and CONSORT extensions for RCTs that use a surrogate endpoint as a primary endpoint were guided by the EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network’s recommended steps for developing health research reporting guidelines<sup>26</sup> and are previously reported elsewhere.<sup>16,17</sup> The project received ethical approval from the University of Glasgow College of Medical, Veterinary, and Life Sciences Ethics Committee (Project No: 200210151). Methodology of our four linked empirical studies focusing on surrogate endpoint definitions and their interpretation in trials are summarised in the box below.

#### Scoping and targeted reviews

Details of scoping review methods are reported in detail elsewhere.<sup>1</sup> In summary, literature was identified using

Prompt	Methodological decisions and rationale
At start of the project, the SPIRIT/CONSORT-SURROGATE project team identifies the need to have a comprehensive definition of surrogate endpoints	Inclusion of a research question to explore how surrogate endpoints are defined in the scoping review
Scoping review identifies different definitions in the literature	Inclusion of rating of definitions in two rounds of e-Delphi survey
Consensus in two definitions and lack of consensus in four other definitions rated in e-Delphi survey	Inclusion of a session to discuss definitions in the consensus meeting
Difficulty in deciding whether intermediate outcomes of recent trials, identified in the targeted review, are surrogate endpoints or target outcomes	Discussion and categorisation of six intermediate outcomes as surrogate endpoints or target outcomes in the consensus meeting
Lack of consensus on the best definition of surrogate endpoints and categorisation of intermediate outcomes as surrogate endpoints/target outcomes in the consensus meeting	Extension of exercise to categorise the six intermediates as either surrogate endpoint/target outcomes to participants of a e-Delphi survey.
Lack of consensus in categorisation of intermediate outcomes in the web survey	Based on all evidence gathered, conceptualisation of a framework for surrogate endpoint definition (see Fig. 3) and criteria in interpreting intermediate outcomes as surrogate endpoints or target outcomes (see Table 3)

electronic bibliographic databases (Excerpta Medica Database [EMBASE], Medical Literature Analysis and Retrieval System Online [MEDLINE], Cochrane Methodology Register) up to March 1st, 2022; Google and targeted website searches (e.g., US FDA) up to May 27th, 2022; hand searching of reference lists of included records; and solicitation from experts including the core project team (OC, PD, AMM, RST, AEY or CJW). Full texts were screened independently by two reviewers (OC, PD, AMM, or RST) on their consideration of limitations and acceptability of surrogate endpoints and reference to surrogate definitions recorded.

We also undertook a targeted review to identify RCTs that have used surrogate endpoints as primary outcomes. MEDLINE through PubMed was searched for RCT full reports and protocols published from January 2017 to June 2022 in six high-impact general medical journals (Annals of Internal Medicine, BMJ, Journal of the American Medical Association, New England Journal of Medicine, Lancet, and PLoS Medicine) and two journals that commonly publish protocols: BMJ Open and Trials. The reports and protocols were exported to Covidence and screened for using a surrogate endpoint as a primary outcome by two reviewers (OC, PD, AMM, or RST). Publications identified by the targeted review were used to inform our selection of RCTs with intermediate outcomes.

### e-Delphi study

The primary objective of our e-Delphi study was to obtain consensus on SPIRIT and CONSORT extension items informed by the scoping review. Given the lack of consensus on surrogate endpoint definitions identified by our scoping literature review, we elected to include a selection of definitions to be rated as part of our e-Delphi study. The study was conducted online using DelphiManager software (version 5.0), maintained by the COMET (Core Outcome Measures in Effectiveness Trials) initiative, [www.comet-initiative.org/delphimanager/](http://www.comet-initiative.org/delphimanager/). A range of stakeholders were

invited to participate including clinicians, trial investigators and methodologists, patient and public involvement (PPI) partners, health technology assessment experts, funding panel members, and journal editors. We included use purposive and snowball (nonprobability) sampling) to identify participants.<sup>24</sup> Identification strategies included: 1) professional contacts known to the research team; 2) relevant professional bodies and networks; 3) relevant conferences and meetings; 4) authors of records included in the scoping and targeted reviews; 5) a call for participants on the project website and social media pages; and 6) asking registered participants to share the link with other people, networks or organisations that would be interested in participating. Participant inclusion criteria were: 1) expertise in surrogate endpoints (through authored literature) or self-reported interest and basic understanding of the concept of surrogacy; 2) registered interest, in English (although international participation was sought), to participate during the allocated period. We had no exclusion criteria. Two rounds of e-Delphi process (round 1: 24th August 2022–10th October 2022; round 2: 31st October 2022–11th December 2022) were used to achieve consensus on our SPIRIT and CONSORT extension items.

Two definitions were added to the four definitions identified by the scoping review. First, given involvement in its development and publication by project team members (OC & RST), the surrogate endpoint definition by Ciani (2017) was added.<sup>11,12</sup> The Ciani definition was considered important as it comes from the health technology assessment community perspective in contrast to the other definitions that focus on the regulatory perspective. Second, during round 1, one of the respondents drew our attention to the recently completed Banff 2022 workshop.<sup>25</sup> The Banff definition was therefore also added to round 2.

e-Delphi survey participants rated each definitions comprehensiveness (completeness, inclusivity, and clarity) using a 9-point Likert scale. We used the

following consensus thresholds for definition acceptance: consensus on comprehensiveness:  $\geq 70\%$  participants scoring 7–9 and  $< 15\%$  participants scoring 1–3; consensus on lack of comprehensiveness:  $\geq 70\%$  participants scoring 1–3 and  $< 15\%$  of participants scoring 7–9; no consensus: failure to achieve both the above. Participants could provide free text comments alongside on their rating of each definition.

### Consensus meeting

The overarching aim of the consensus meeting was to ratify SPIRIT and CONSORT extension items that reached consensus during the e-Delphi study and allow discussion of items that did not reach consensus. The meeting closely followed the EQUATOR Network guidance on conducting consensus meetings and included the 13 project team members and 20 purposively selected stakeholders (see Acknowledgements).<sup>26</sup> The project team selected participants from those who completed e-Delphi round 2 and expressed their interest based on their availability to attend and the need to ensure international and multidisciplinary representation of participants. The meeting was conducted over two consecutive half-days (13th & 14th March 2023) as a hybrid meeting, some participants present physically and others joining virtually through an online video conference link. A specific session of the consensus meeting agenda included a session on surrogate definitions based on data identified by the scoping review and e-Delphi study.

To gauge how stakeholders judge endpoints, we asked consensus meeting participants to rate six published RCTs (chosen from our scoping and targeted review described above) with a primary outcome that could be plausibly perceived as an intermediate outcome (see e-Appendix Table 1). Participants were asked to consider whether they judged each endpoint as a surrogate endpoint (response options were ‘Yes’, ‘No’, or ‘Uncertain’) and voting collected (using [www.mentimeter.com](http://www.mentimeter.com)). To inform their decision-making, we provided links to published RCT protocols or reports and provided with a summary of reasons for and against considering each endpoint as a surrogate endpoint.

Briefly, these six primary outcomes (as stated by the trial authors) were:

- 1) “Surgical site infections (SSI) within 30 days after surgery. SSIs are classified as being superficial, deep and/or organ–space infection on the basis of validated and well-defined criteria developed by the Centers for Disease Control and Prevention.”<sup>27</sup>
- 2) “Continuous smoking abstinence rate (abstinence from conventional/combustible cigarettes during the last 4 weeks (weeks 9–12) of the treatment period of 3 months.”<sup>28</sup>
- 3) “Difference in body mass index z-scores between arms at 15 and 30 months.”<sup>29</sup>
- 4) “Duration of severe symptoms. Each symptom was scored using a six-point Likert scale, and symptoms scoring 5 or 6 were considered as severe.”<sup>30</sup>
- 5) “Spontaneous vaginal birth.”<sup>31</sup>
- 6) “Disability at 90 days, evaluated by the distribution of scores on the modified Rankin scale.”<sup>32</sup>

### A web-based survey exploring how and when to categorise intermediate outcomes in RCTs as surrogate endpoints

Following the consensus meeting, we repeated the exercise of judging the definitions of intermediate outcomes as surrogates (4th April–5th May 2023) as described above by contacting individuals who participated in the e-Delphi study or initially registered interest to participate in the e-Delphi.

### Patient and public involvement (PPI)

There was extensive PPI in all research activities described above. A PPI (DS) partner was a member of our original funding application for the CONSORT-SPIRIT surrogate project and took on the PPI lead role as a Programme Management Group member. Nineteen additional PPI partners were invited to take part in the e-Delphi survey. To facilitate their participation, we undertook a 2-h web-based briefing/learning workshop on surrogate outcomes in trials. Members of the PPI group participated individually in the e-Delphi process, with four participating in the 2-day Consensus meeting.

### Data analysis, synthesis, and presentation

The numerical results of the empirical studies were summarised quantitatively (e.g., medians, interquartile range (IQR), counts and frequencies) and tabulated or presented graphically. Free text comments received by participants are presented to illustrate key concepts.

### Role of funding

The research was funded as part of development of SPIRIT and CONSORT extensions has been funded by the UK Medical Research Council (grant number MR/V038400/1). Gary Collins was supported by Cancer Research UK (programme grant: C49297/A27294). Jane Blazeby was supported by the NIHR Bristol Biomedical Research centre. Sylwia Bujkiewicz was supported by UK Medical Research Council (MR/T025166/1) and Leicester NIHR Biomedical Research Centre. Alain Amstutz receives his salary from the Research Fund Junior Researchers of the University of Basel. Robin Christensen wants to acknowledge that Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant. CDCMF receives research productivity fellowships from the Oak Foundation (OCAY-18-774-OFIL) national council for scientific and technological development (CNPq/Brazil–Grant: 08516/

2021-4). 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.

**Results**

**Surrogate endpoint definition: scoping review**

Details of the scoping review article identification process are reported elsewhere.<sup>1</sup> The four existing surrogate definitions that were most cited (with ≥2 citations) across 32 publications identified by the scoping review were those of: Prentice (1989), Temple (1999), National Institutes of Health (NIH) Biomarkers Definitions Working Group (2001), Biomarkers, Endpoints, and other Tools (BEST) (2016) (see [Table 1](#)).<sup>6,33,34,36</sup>

Three of the four definitions emphasise the notion that a surrogate endpoint is used to substitute and predict another outcome. However, what was predicted and/or substituted for (i.e., the target outcome) differed according to definitions: “clinical endpoint”, “direct measure of how a patient feels, functions or survives”, or “patient or participant relevant final

outcome”. The Prentice definition was entirely statistical in its focus.

**Surrogate endpoint definition: e-delphi study**

A total of 219 individuals registered for the e-Delphi study, of whom 212 were deemed eligible, with 195 rating items in Round 1 and 176 in Round 2. A summary of participants who participated in the two rounds is detailed in [e-Appendix Tables 2–5](#).

Two of the definitions reached a consensus, i.e., BEST (2016) and Ciani et al. (2017). A summary of the free text comments provided by participants for the definitions is given in [Table 1](#) (see [e-Appendix](#) for details).

A key attribute of the NIH, BEST, and Ciani definitions identified by Delphi participants was their inclusion of the requirement for surrogate endpoint validation, i.e., the surrogate endpoint accurately predicts treatment effects on the target outcome of interest: “reliably predicts benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence”. An additional attribute identified by the Ciani definition was the explicit recognition that

Source (year)	Definition	Scoping review citations	e-Delphi rating			Summary of free-text comments	
			Median (IQR)	% of rating scores			
				1-3	4-6		7-9
Prentice (1989) <sup>33</sup>	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 endpoint.	6 (19%) records	5 (3, 6)	29.6	58.6	11.7	Complex and statistical definition with limited usability in trial design—see comments of Definition 3 in <a href="#">Appendix</a> .
Temple (1999) <sup>34</sup>	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%)	7 (5, 7)	<b>11.6</b>	31.4	57.0	Not inclusive as a surrogate endpoint extends beyond laboratory measurements and signs and their use is beyond therapeutic trials—see comments of Definition 2 in <a href="#">Appendix</a>
NIH Biomarkers Definitions Working Group (2001) <sup>35</sup>	A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint 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%)	7 (6, 7)	<b>11.0</b>	31.8	57.2	Not inclusive as surrogate endpoints extend beyond biomarkers and clinical benefit measured could still be a surrogate endpoint—see comments of Definition 1 in <a href="#">Appendix</a>
BEST (2016) <sup>6</sup>	An endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels; functions; or survives. A surrogate endpoint 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%)	8 (7, 9)	<b>0.6</b>	7.0	<b>92.4</b>	A comprehensive definition although use of ‘predict’ implies a validated surrogate endpoint—see comments of Definition 4 in <a href="#">Appendix</a>
Ciani et al. (2017) <sup>11</sup>	A biomarker or intermediate outcome used to substitute for a patient or participant relevant final outcome (i.e., severe morbidity; health related quality of life or mortality) and reliably predicts benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence	Not applicable <sup>a</sup>	8 (7, 8)	<b>2.3</b>	14.0	<b>83.6</b>	Support for inclusion of intermediate outcome in definition; however, there is limited understanding of ‘intermediate outcome’; not all trials seek to evaluate interventions based on severe morbidity, health related quality of life or death; and ‘predict’ implies a validated surrogate endpoint—see comments of Definition 5 in <a href="#">Appendix</a>
Banff Workshop (2022) <sup>25</sup>	An endpoint replacing a clinical endpoint that constitutes a basis for reliably predicting a treatment effect on the clinical endpoint in a defined context of use.	Not applicable <sup>a</sup>	7 (5.5, 8)	<b>7.8</b>	33.5	58.7	No comments received

NIH: National Institutes of Health; BEST: Biomarkers, Endpoints, and other Tool. <sup>a</sup>Not identified in the scoping review; Bold highlighted; consensus reached; Italic highlighted: consensus not reached.

**Table 1: Surrogate endpoint definitions identified by scoping review and e-Delphi rating.**



intermediate outcomes (as well as biomarkers) be considered as surrogate endpoints. Whilst positively rated in the e-Delphi study, participant free-text comments indicated that: 1) further development and clarification of the definitions were needed e.g., the need for a validated surrogate endpoint to reliably predict treatment effects on a final patient-relevant (target) outcome; and 2) the need for plain language explanations of terms such as ‘intermediate outcome’ and ‘biomarker’.

### Consensus meeting and web-based survey exploring how and when to categorise intermediate outcomes in RCTs as surrogate endpoints

During the consensus meeting, as part of the session on surrogate endpoint definition, we asked participants to judge whether each of the six primary outcomes used in recent trial protocols or reports are surrogate or not. The live meeting voting indicated mixed views about the endpoints presented, and no consensus was reached based on the thresholds that we had prespecified, i.e., none of the intermediate outcomes was considered a surrogate by more than 70% of attendees or considered a target outcome by less than 15% (see Table 2). After repeating the exercise with the participants in our e-Delphi study, out of 80 valid responses, we observed consensus on surgical site infections and spontaneous vaginal birth being target outcomes. Even in those cases, participants mentioned in free-text comments that the Center for Disease Control (CDC) definition of surgical site infection includes a composite measure of signs and symptoms and only one part of the component is perceived directly by patient-, and that the target outcomes are longer length of hospital stay, readmission and mortality for surgical site infection, or recovery time, pain, health-related quality of life for both the mother and the baby in the case of spontaneous vaginal birth (see e-Appendix Table 6). The web-survey had no consensus for the other four potential intermediate outcomes (see Fig. 1).

There was variation across stakeholders based on their primary professional role, with clinicians and regulators more likely to consider intermediate outcomes as target outcomes, in contrast to PPI

contributors and health technology assessment experts who considered them as surrogate endpoints (see Fig. 2).

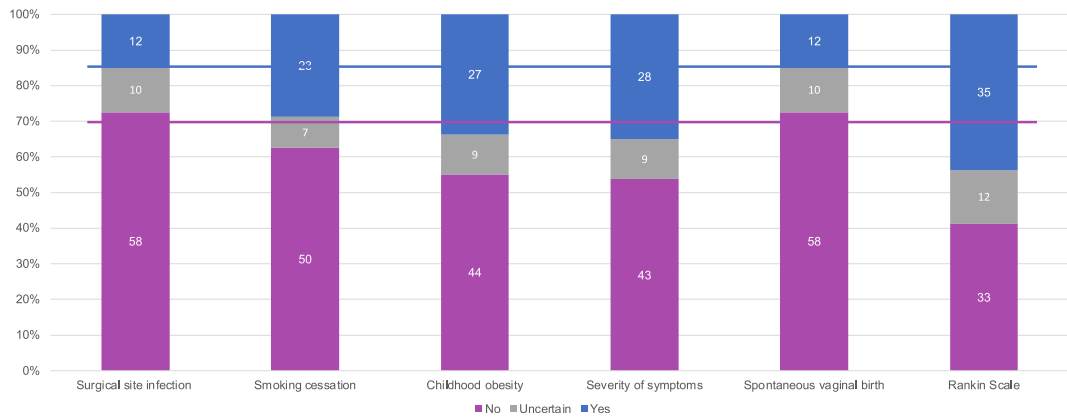
Based on consensus meeting discussions, free-text comments from the web survey, and our reflections, we identified four criteria for classifying intermediate outcomes as surrogate or target endpoints: 1) type of measurement; 2) whether the outcome is perceived by patients (patient reported) or represents health benefits per se; 3) intervention intent or trial hypothesis; and 4) association with target outcomes. Table 3 summarises these criteria along with counterarguments or further considerations for each. We illustrate these criteria using obesity as an intermediate outcome. Use of these criteria results in different conditionally dependent conclusions on an endpoint being a surrogate or a target outcome. For example, the measurement criterion posits that obesity can be a target outcome if a valid patient-reported outcome (PRO) instrument is used (e.g., obesity-related quality of life instrument<sup>35</sup>) and a surrogate endpoint if body mass index (BMI) is used.

### Discussion

Evidence for the benefits and harms of interventions should come from high-quality RCTs that directly assess treatment effects on a target outcome of interest to patients, clinicians, and other stakeholders, such as all-cause mortality.<sup>37</sup> Evaluating effects on such target outcomes can require trials with large sample sizes, long follow-up times, and high costs. While accepting treatment effects on a biomarker as a replacement for a target outcome has become increasingly commonplace in the regulatory setting,<sup>7-9</sup> there has been much less consideration for the use and implications of intermediate outcomes as surrogate endpoints in the wider context of healthcare interventional trials.<sup>11-13</sup> As part of the development of the SPIRIT and CONSORT extensions for the reporting of RCTs with a primary surrogate endpoint,<sup>16</sup> we undertook four linked empirical studies (a scoping review, e-Delphi study, consensus meeting, and a web survey) to inform the development of a definitional framework for the improved design, reporting, and interpretation of trials using surrogate endpoints.

Do you consider this a surrogate endpoint?	No (%)	Yes (%)	Uncertain (%)
Surgical site infection	44%	30%	26%
Smoking cessation	24%	72%	4%
Childhood obesity	36%	46%	18%
Severity of symptoms	63%	19%	19%
Spontaneous vaginal birth	61%	25%	14%
Rankin Scale	50%	42%	8%

Table 2: Classification of intermediate outcomes as surrogate endpoints during the consensus meeting (28 participants).

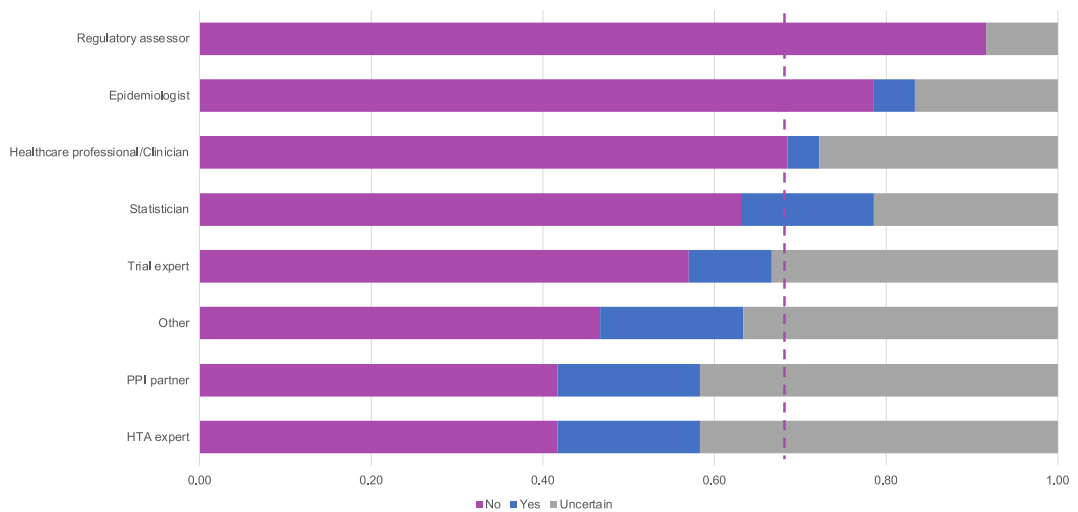


**Fig. 1: Classification of intermediate outcomes as surrogate endpoints by participants in the web-survey (80 participants).** Solid lines indicate 70% and 15% frequency according to criteria for consensus. Rankin Scale was collapsed to 'yes' for the following response categories: symptoms, symptoms and mild disability, and symptoms and all forms of disability. See e-Appendix Table 1 for details of the 6 intermediate outcomes assessed in web-survey.

Our e-Delphi study showed good support for the 2016 BEST (Biomarkers, Endpoints, and other Tools) definition from the NIH-FDA Working Group that biomarkers (e.g., brain amyloid plaque protein, blood pressure, tumour response) can act as surrogate endpoints in interventional trials by acting as a substitute and predictor of treatment effects on target outcome(s) for how a patient feels, functions, and survives.<sup>6</sup> The BEST definition reflects regulatory practice over the last three decades, as more than half of recent new drug and biologic approvals by US FDA, European Medicine Agency (EMA), Pharmaceuticals and Medical Devices Agency in Japan are based on trial evidence of treatment effects on a biomarker endpoint as an acceptable

surrogate endpoint.<sup>7-9</sup> Support of this definition did not reach consensus thresholds in our consensus meeting. Furthermore, both our scoping review and e-Delphi study demonstrated an increasing recognition that this surrogate endpoint definitional framework needs to be extended to capture a broader definition that includes intermediate outcomes in certain contexts.

Intermediate outcomes are typically measured more proximally in time to the target outcome(s) and may have potential direct relevance to patients, clinicians, and other stakeholders. However, as with biomarkers, a common underlying surrogacy rationale for intermediate outcomes in intervention trials is that they act as a substitute and predictor of treatment effects on a target



**Fig. 2: Classification of intermediate outcomes as surrogate endpoints by professional role (80 participants).** \*Dashed line indicates frequency of "No" across the whole sample. Trial experts include trial investigators, methodologists, and managers. See e-Appendix Table 1 for details of the 6 intermediate outcomes assessed in e-survey.

Criteria	Type of measurement	Outcome is perceived/has health benefits	Intervention intent or hypothesis	Association with target outcome(s)
Explanation	<ul style="list-style-type: none"> <li>■ Considers a target outcome as what an individual feels, functions, or how long they survive.</li> <li>■ When an outcome is directly reported by individual, e.g., a patient reported outcome measure, it can be considered a target outcome.</li> <li>■ When outcome is indirectly assessed or inferred from a measure, e.g., signs and clinician assessed scales, it can be considered a surrogate endpoint</li> </ul>	<ul style="list-style-type: none"> <li>■ If outcome is perceived (e.g., felt, reported by) by individual, then it is not considered a surrogate outcome.</li> <li>■ If the outcome has any benefits including short-term health and non-health benefits, then it could be considered target outcome.</li> </ul>	<ul style="list-style-type: none"> <li>■ Consideration of whether an outcome is a surrogate endpoint based on whether the trial aimed at using an outcome as a replacement or predictor of a target outcome(s).</li> </ul>	<ul style="list-style-type: none"> <li>■ Outcomes that have strong association with target outcomes are validated surrogate endpoints.<sup>3</sup></li> <li>■ In health technology assessment, the target outcome is often ultimately an assessment of quality adjusted life years (QALYs).</li> </ul>
Counter arguments or further considerations on criteria	<ul style="list-style-type: none"> <li>■ Duration of measurement is important: patient reported outcomes for a short period should be considered a surrogate endpoint for the outcome at longer duration.</li> <li>■ Death is not patient reported yet a definitive target outcome.</li> <li>■ A patient reported outcome measure that is not validated can be regarded as a surrogate endpoint.</li> </ul>	<ul style="list-style-type: none"> <li>■ Consideration of perception of in the surrogate endpoint/target outcome categorisation assumes only biomarkers (which are not perceived) to be the only surrogate endpoint.</li> <li>■ Being perceived by individual can be considered lack of outcome blinding.</li> <li>■ Duration of health benefits is important: short-term benefits can be surrogates for longer term benefits.</li> </ul>	<ul style="list-style-type: none"> <li>■ Explicit recognition of target outcomes rarely happens in trial hypothesis statements.</li> </ul>	<ul style="list-style-type: none"> <li>■ Target outcomes are not always the only important outcomes for stakeholders.</li> </ul>
Application of the four criteria to the intermediate outcome of obesity/weight loss measures				
Argument	Obesity measured using validated patient reported outcome would be considered a target outcome. If inferred from anthropometric measures, such as body mass index, then consider as a surrogate endpoint.	Most people can feel weight gain/loss and participants can enrol for weight loss interventions for non-health outcomes including body attractiveness and preventing a wardrobe change hence a trial using weight loss as primary outcome has used target outcome.	View of obesity as a disease means that trials that evaluating management of obesity can be regarded as using a target outcome.	Evaluation of weight loss interventions often judged on how well they result in health benefit (e.g., QALYs) based on impact on various diseases and conditions contributed to by obesity.
Counterargument	Patient reported outcomes in the short term could be considered surrogate endpoints for target outcomes such as severe morbidity and death.	Weight gain/loss is still a surrogate endpoint that is not blinded. Furthermore, even with non-health outcomes duration of weight loss is important: sustained weight loss/control would be the target to see benefits in body attractiveness.	Obesity is a global health and health system challenge due to its association with many diseases and complications and hence weight loss is a surrogate endpoint for these other health impacts.	Obesity-related diseases and complications are not the only considerations for people who engage in weight loss.
<p><sup>a</sup>For details of surrogate validation methodology see. <a href="#">8,18-23</a></p>				
<p><b>Table 3: Criteria in considering a trial intermediate outcome as a surrogate endpoint or target outcome.</b></p>				

outcome. For example, in the public health trial setting, a primary endpoint of body mass index (BMI) for dietary and physical activity promotion interventions in children could be interpreted as an intermediate outcome and surrogate endpoint for health improvement, including morbidity and mortality associated with the risk of future development of cardiometabolic disease.<sup>38</sup> Importantly, our web survey showed a lack of consensus across stakeholders (including patients, clinicians, researchers, and health technology assessment experts) on acceptance of whether intermediate outcomes used as primary outcomes were surrogate endpoints or target outcomes. Patients and health technology assessment experts appeared more likely to consider intermediate outcomes as surrogate endpoints, whereas this was less likely for clinicians and regulators.

Based on this study's findings we propose a framework for the broader definition and interpretation of surrogate endpoints in interventional trials (see [Fig. 3](#)) along with criteria that can aid stakeholders to decide on

whether trial endpoints are surrogate endpoints or target outcomes (see [Table 3](#)). Teams designing healthcare trials may find these criteria helpful in considering the choice of their primary outcome(s). Misalignment in what is regarded as a surrogate endpoint or target outcome between different stakeholders may result in suboptimal utilisation of trial findings to inform clinical decision making and healthcare policy, leading to unrealistic expectations among patients and clinicians and wastage of research.<sup>39</sup> Trial teams should consider what criteria allow for achieving the trial objective and lead to optimal use of trial findings by other researchers. Additionally, considering what is a surrogate endpoint, or a target outcome could benefit from consultation with stakeholders including regulators, health technology assessment experts, clinicians, patients, and the wider public.

Few previously published studies have sought to formally assess stakeholders' views on the use of surrogate endpoints.<sup>40,41</sup> Schievink et al. conducted an

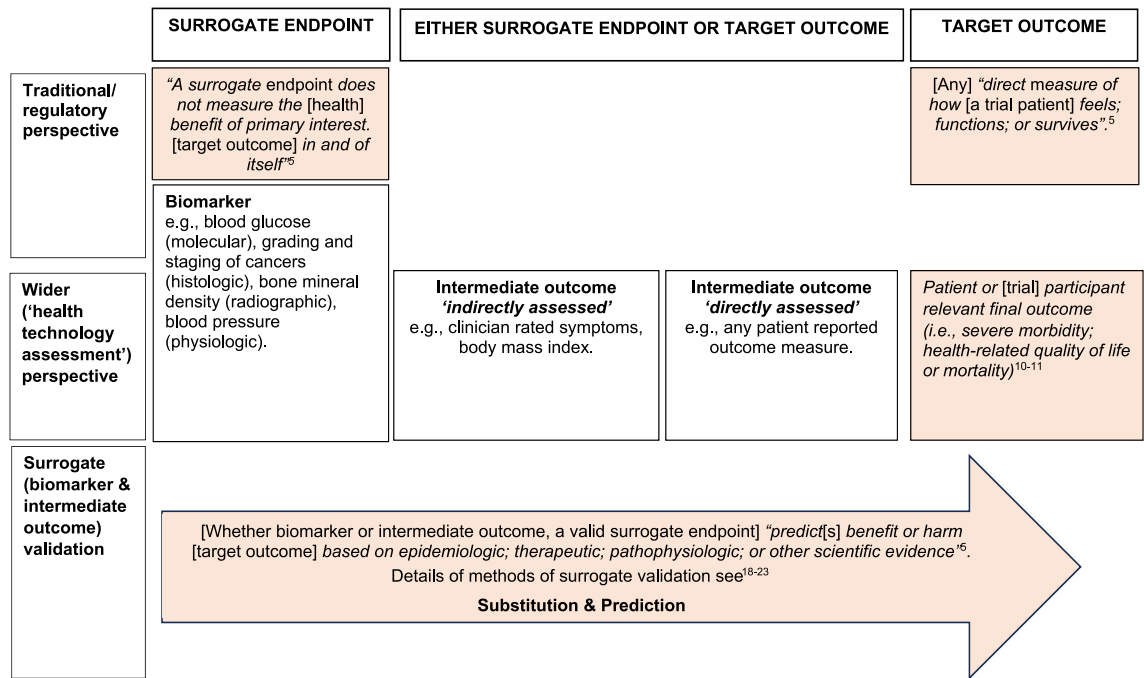


Fig. 3: Proposed framework for the definition and interpretation of surrogate endpoints in interventional trials.

online questionnaire survey in 74 stakeholders (including representatives from drug regulators, health technology assessment agencies, drug industry, and academia) to assess their perception of the acceptability of surrogate endpoints.<sup>40</sup> The study authors reported that the survey respondents generally supported the use of surrogate endpoints in drug approval; however there was a recommendation for prioritising surrogates supported by formal scientific evidence of their validation for treatment effects on outcomes. In contrast to the current study, this previous survey was restricted to only one specific clinical area (cardiorenal disease) and only considered biomarkers (e.g., blood pressure, glycosylated haemoglobin/HbA1c). Another interview-based study with different stakeholder representatives (i.e., healthcare professionals, payers or representatives of health technology assessment bodies, regulators, statisticians, health economists, health technology manufacturers, journal editors), in accord with the findings of the current study, reported a wide variation in the acceptance of the use of surrogate endpoints across the different stakeholder groups.<sup>41</sup>

To our knowledge, our study represents the most comprehensive analysis of definitional frameworks for using surrogate endpoints in the context of interventional trials. However, we acknowledge our four linked studies each have limitations. Our scoping review excluded literature published in languages other than English albeit we included documents in English from regions where the primary language is not English

(Europe, Asia, and South America). Consistent with previous surveys of surrogate trials, we chose six major general medical journals for our targeted review because they are likely to widely used in clinical and policy decision making.<sup>14</sup> We searched these journals over the last 5-year period (2017–2022) in order to reflect contemporary trial practice. However, the experience of these journals of the use of surrogate endpoints in trials may not reflect that of discipline specific journals oriented to a particular specialty. Whilst we sought to recruit a wide range of participants to our e-Delphi, surveys, and consensus meeting with members of the wider public, clinicians, researchers from academia and healthcare industry, regulators, and health technology assessment experts, our sample was not a random one with participants responding to various expressions of interest that we posted and advertised. We therefore cannot claim our study samples are fully representative of either these stakeholder groups individually or the wider stakeholder community. Relatedly, and despite our best efforts, the sample size of some the stakeholder subgroups was relatively small; therefore, caution is needed in comparing our results across these subgroup populations. Our survey of intermediate outcomes was limited to a sample of only six trials and may, again, lack generalisability.

The judicious use of surrogate endpoints in interventional trials provides an important opportunity to expedite access to new innovative treatments for

conditions of high unmet health need, such as early Alzheimer's disease, where the outcomes of ultimate interest for patients, families, clinicians, and policy makers are difficult to collect (e.g., due to longer follow-up and larger sample size requirements). Furthermore, surrogate endpoints in small feasibility and pilot trials can provide early signals of biological activity to guide the design and investment in a fully powered randomised trial to examine impacts on a target endpoint(s).<sup>42</sup> However, a societal consequence of this reliance on surrogate endpoints is the increase of the uncertainty of intervention's 'true value' (including clinical efficacy/effectiveness, safety, and cost-effectiveness). Observed changes in a surrogate endpoint may fail to reliably predict the efficacy of an intervention in terms of its actual impact on patient relevant target outcomes. A meta-epidemiological study has shown that the treatment effects from RCTs with surrogate primary endpoints on average overpredict effects by more than 40% compared with trials based on primary final patient-relevant endpoints, such as mortality.<sup>43</sup> More concerning, improvement in a surrogate endpoint may not be related to the harms associated with a therapy.<sup>44</sup> For example, the diabetes drug rosiglitazone, approved by the US FDA and EMA in 1999/2000 based on phase I-III clinical trials showing improved levels of blood glucose and HbA1c (surrogate endpoints) was later withdrawn in 2010 when found to increase risk for heart failure hospitalisation and myocardial infarction.<sup>45</sup> To minimise such uncertainty, it has been widely recommended that trials are limited to validated surrogate endpoints where there is statistical evidence that the treatment effect on the surrogate endpoint is strongly predictive of the treatment effect on the target/final participant-relevant outcome(s) including safety.<sup>8,18</sup> Methods for the statistical validation of surrogates and examples of multidimensional frameworks evaluating the level of evidence for surrogate endpoint validity are described elsewhere.<sup>19–23</sup>

In conclusion, our results have important implications for the future use of surrogate endpoints in the design, interpretation, and reporting of interventional trials. Determining what is (and what is not) a surrogate endpoint can be challenging and dependent on the framing within trials. We therefore propose a definitional framework that extends the scope of surrogate endpoints in interventional trials from a focus on only biomarkers ('the traditional drug regulatory perspective') to include intermediate outcomes in certain contexts that include measures of function or symptoms. This framework will directly inform the SPIRIT and CONSORT extensions for the reporting of trials using surrogate outcomes currently in development.<sup>14</sup> It provides a decision tool to help those involved in the design, reporting, and interpretation of trials to weigh up their judgement of trial evidence based on surrogate endpoints.

#### Contributors

OC, AMM, RST, CJW, PD, and DS conceptualised the programme of research and obtained research funding. All authors contributed either to the e-Delphi study, consensus meeting, or e-survey. The manuscript was drafted by OC, AMM, and RST. All authors reviewed the manuscript and approved the final version.

#### Data sharing statement

All data for this study is available upon request to the corresponding author.

#### Declaration of interests

Sylwia Bujkiewicz is a member of the NICE Decision Support Unit (DSU) and NICE Guidelines Technical Support Unit (TSU), has served as a paid consultant, providing methodological advice, to NICE, Roche, IQVIA and RTI Health Solutions, received payments for educational events from NICE and Roche and has received research funding from European Federation of Pharmaceutical Industries & Associations (EPIA) and Johnson & Johnson. Mario Ouwers 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 (NEST), from the Food and Drug Administration for the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) 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. Alain Amstutz and Yousef Rezaei are Associate Editors at BMC Trials. Robin Christensen is a founding member of the OMERACT Technical Advisory Group which might be perceived as a possible conflict of interest. Ray Harris has shares in Johnson and Johnson. John Powers has been a consultant for Adaptive Phase, Arrevas, Atheln, Bavaria Nordic, Cellularity, Eicos, Evofem, Eyecheck, Gilead, GSK, Mustang, OPKO, Otsuka, Resolve, Romark, Spine-BioPharma, UTILITY, Vir. Achilles Thoma received royalties from Springer Publishing for his book "Evidence Based Surgery: A Guide to Understanding and Interpreting the Surgical Literature", 2019.

#### Acknowledgements

We are indebted to Professor Amber Young (University of Bristol) for her contribution to the planning and conduct of the SPIRIT|CONSORT-SURROGATE project and who sadly passed away in September 2022. The SPIRIT|CONSORT-SURROGATE Project team and Consensus Group are listed as an e-appendix. We acknowledge all professional organisations and networks who helped in Delphi survey participant mobilisation and all who took part in the Delphi and/or definitions survey; we list these organisations/networks and all participants in the e-Appendix.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2023.102283>.

#### References

- 1 Manyara AM, Davies P, Stewart D, et al. Definitions, acceptability, limitations, and guidance in the use and reporting of surrogate endpoints in trials: a scoping review. *J Clin Epidemiol.* 2023;160:83–99.
- 2 Walsh S, Merrick R, Milne R, Brayne C. Aducanumab for Alzheimer's disease? *BMJ.* 2021;374:n1682.
- 3 Walsh S, Merrick R, Richard E, Nurock S, Brayne C. Lecanemab for Alzheimer's disease. *BMJ.* 2022;379:o3010.

- 4 Rosenthal MB. Novel Alzheimer disease treatments and reconsideration of us pharmaceutical reimbursement policy. *JAMA*. 2023;330:50. <https://doi.org/10.1001/jama.2023.11702>.
- 5 Manyara AM, Ciani O, Taylor RS. Reply to commentary by Cummings (2022): surrogate endpoints extend beyond biomarkers. *Alzheimers Dement*. 2022;8:e12344.
- 6 BEST (Biomarkers, EndpointS, and other Tools) resource FDA-NIH biomarker working group. Food and Drug Administration (US); Bethesda (MD); National Institutes of Health (US); Silver Spring (MD); 2016. <https://www.ncbi.nlm.nih.gov/books/NBK326791/>.
- 7 Mitra-Majumdar M, Gunter SJ, Kesselheim AS, et al. Analysis of supportive evidence for US Food and drug administration approvals of novel drugs in 2020. *JAMA Netw Open*. 2022;5:e2212454.
- 8 Schuster Bruce C, Brhlikova P, Heath J, McGettigan P. The use of validated and nonvalidated surrogate endpoints in two European Medicines Agency expedited approval pathways: a cross-sectional study of products authorised 2011-2018. *PLoS Med*. 2019;16:e1002873.
- 9 Maeda H, Shingai R, Takeda K, Hara A, Murai Y, Ofuchi M. Assessment of surrogate end point trends in clinical trials to approve oncology drugs from 2001 to 2020 in Japan. *JAMA Netw Open*. 2023;6:e238875.
- 10 O'Bryant SE, Waring SC, Cullum CM, et al. Staging dementia using clinical Dementia rating scale sum of Boxes scores: a Texas alzheimer's research consortium study. *Arch Neurol*. 2008;65:1091-1095.
- 11 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;20:487-495.
- 12 Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Use of surrogate end points in healthcare policy: a proposal for adoption of a validation framework. *Nat Rev Drug Discov*. 2016;15:516.
- 13 European Network for Health Technology Assessment (EUNetHTA). *D4.4 outcomes*; 2023. <https://www.eunetha.eu/wp-content/uploads/2023/01/EUNetHTA-21-D4.4-practical-guideline-on-Endpoints-v1.0.pdf>.
- 14 la Cour JL, Brok J, Göttsche PC. Inconsistent reporting of surrogate outcomes in randomised clinical trials: cohort study. *BMJ*. 2010;341:c3653.
- 15 Walia A, Haslam A, Prasad V. FDA validation of surrogate endpoints in oncology: 2005-2022. *J Cancer Policy*. 2022;34:100364.
- 16 Manyara AM, Davies P, Stewart D, 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:e064304.
- 17 Manyara AM, Davies P, Stewart D, 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:e062798.
- 18 Hey SP, Kesselheim AS, Patel P, Mehrotra P, Powers JH 3rd. 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;180:131-138.
- 19 Xie W, Halabi S, Tierney JF, et al. A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. *JNCI Cancer Spectr*. 2019;3:pkz002.
- 20 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;43:272989X231162852.
- 21 Weir CJ, Taylor RS. Informed decision-making: statistical methodology for surrogacy evaluation and its role in licensing and reimbursement assessments. *Pharmaceut Stat*. 2022;21:740-756.
- 22 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 Stats. *Methods Med Res*. 2008;17:303-340.
- 23 Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (Bio-Surrogate) evaluation schema (BSES). *BMC Med Res Methodol*. 2012;12:27.
- 24 Ritchie J, Lewis J, Elam G. *Designing and selecting samples*. Qualitative research methods; 2003:77-108.
- 25 Banff International Research Station. *Statistical challenges in the identification, validation, and use of surrogate markers workshop, final report*. Oaxaca, Mexico; 2022. <https://www.birs.ca/cmoworkshops/2022/22w5184/report22w5184.pdf>.
- 26 Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7:e1000217.
- 27 Vignaud M, Paugam-Burtz C, Garot M, et al. Comparison of intravenous versus combined oral and intravenous antimicrobial prophylaxis (COMBINE) for the prevention of surgical site infection in elective colorectal surgery: study protocol for a multicentre, double-blind, randomised controlled clinical trial. *BMJ Open*. 2018;8:e020254.
- 28 Berlin I, Dautzenberg B, Lehmann B, et al. Randomised, placebo-controlled, double-blind, double-dummy, multicentre trial comparing electronic cigarettes with nicotine to varenicline and to electronic cigarettes without nicotine: the ECSSMOKE trial protocol. *BMJ Open*. 2019;9:e028832.
- 29 Adab P, Pallan MJ, Lancashire ER, et al. Effectiveness of a childhood obesity prevention programme delivered through schools, targeting 6 and 7 year olds: cluster randomised controlled trial (WAVES study). *BMJ*. 2018;360:k211.
- 30 Llor C, Moragas A, Bayona C, et al. The STOP-AB trial protocol: efficacy and safety of discontinuing patient antibiotic treatment when physicians no longer consider it necessary. *BMJ Open*. 2017;7:e015814.
- 31 Epidural and Position Trial Collaborative Group. Upright versus lying down position in second stage of labour in nulliparous women with low dose epidural: BUMPES randomised controlled trial. *BMJ*. 2017;359:j4471.
- 32 Martins SO, Mont'Alverne F, Rebelo LC, et al. Thrombectomy for stroke in the public health care system of Brazil. *N Engl J Med*. 2020;382:2316-2326.
- 33 Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989;8:431-440.
- 34 Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69:89-95.
- 35 Kolotkin RL, Williams VSL, Ervin CM, et al. Validation of a new measure of quality of life in obesity trials: impact of weight on quality of life-lite clinical trials version. *Clin Obes*. 2019;9:e12310.
- 36 Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA*. 1999;282:790-795.
- 37 Ciani O, Manyara AM, Taylor RS. Surrogate endpoints in trials-a call for better reporting. *BMJ*. 2022;378:o1912.
- 38 Brown T, Moore TH, Hooper L, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev*. 2019;7:CD001871.
- 39 Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383:101-104.
- 40 Schievink B, Lambers Heerspink H, Leufkens H, De Zeeuw D, Hoekman J. The use of surrogate endpoints in regulating medicines for cardio-renal disease: opinions of stakeholders. *PLoS One*. 2014;9:e108722.
- 41 Ciani O, Grigore B, Taylor RS. Development of a framework and decision tool for the evaluation of health technologies based on surrogate endpoint evidence. *Health Econ*. 2022;31(Suppl 1):44-72.
- 42 Campbell MJ, Lancaster GA, Eldridge SM. A randomised controlled trial is not a pilot trial simply because it uses a surrogate endpoint. *Pilot Feasibility Stud*. 2018;4:130.
- 43 Ciani O, Buyse M, Garside R, 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.
- 44 Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med*. 1996;125:605-613.
- 45 Cohen D. Rosiglitazone: what went wrong? *BMJ*. 2010;341:c4848.