10th Data Science Round Table Meeting

Validity Assessment of Surrogate Endpoints Using a Practical Framework

2026/Jul/7

Facilitators for This Session

Facilitating today's discussion

Koji Oba	University of Tokyo
Takumi Aoki	PMDA
Kiyoshiro Tanaka	PMDA
Mizuki Yoshida	Pfizer
Yasushi Takita	Lilly

Leading an active discussion from diverse industry, regulatory, and academic perspectives

Ground Rule

Let's create a place where everyone can speak freely and have their opinions respected, regardless of gender, age, or years of experience

① Speak freely

Everyone can speak as equals, regardless of gender, age, or years of experience. Please actively share your opinions.

② Respect opinions

Don't reject opinions that are raised – first, accept them. Differing opinions are what give discussion its value.

③ Ask if you're unsure

If you're unsure about the process or content, don't hesitate to ask the facilitators or your team.

Let's work together to create a space where everyone can speak with confidence, regardless of their position

Objectives of This Workshop

① Learn the theoretical background

Without knowing the theoretical background and regulatory requirements around surrogate endpoints, you can't even discuss them with regulators. First, learn the overview.

② Surrogate EP Actually use the validity assessment framework

Surrogate EP Apply the validity assessment framework to a real case and experience conducting a validity assessment. Also learn how to gather evidence.

③ Surrogate EP Build assessment skills

Organize and share the situation using the framework, and through team discussion, develop assessment skills that hold up in regulatory negotiations.

④ Deepen dialogue across industry, regulators, and academia

From the perspectives of industry, regulators, and academia respectively, deepen understanding of diverse viewpoints and build a network for ongoing dialogue.

Timetable

① Information sharing

10:15 – 11:40

- Theory and regulatory requirements around surrogate endpoints
- Introduction to the surrogate endpoint assessment framework (REASSURE)
- Overview of each case study and how the group discussions will proceed

② Group discussion

11:40 – 15:15

- First half: Deep-dive into each case study, analyzing and discussing it using the REASSURE framework
- Second half (last 30 min): Each group presents its discussion

One presenter1 per team

③ Full presentation & sharing

15:30 – 16:45

- Overview explanation of the workshop
- Each group shares its discussion with everyone

One presenter1 per team

Information Sharing

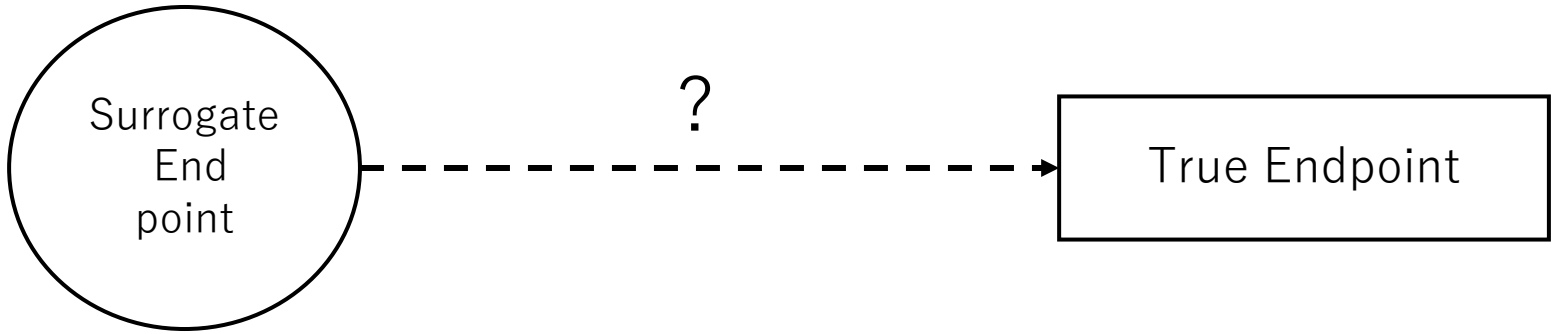
Contents

- Common framework for surrogate endpoint validity assessment overview
- The relationship between surrogate endpoints and regulation at each regulatory authority

Assessing Surrogacy ①: Does it predict prognosis? (Individual level)

- Using epidemiological research,

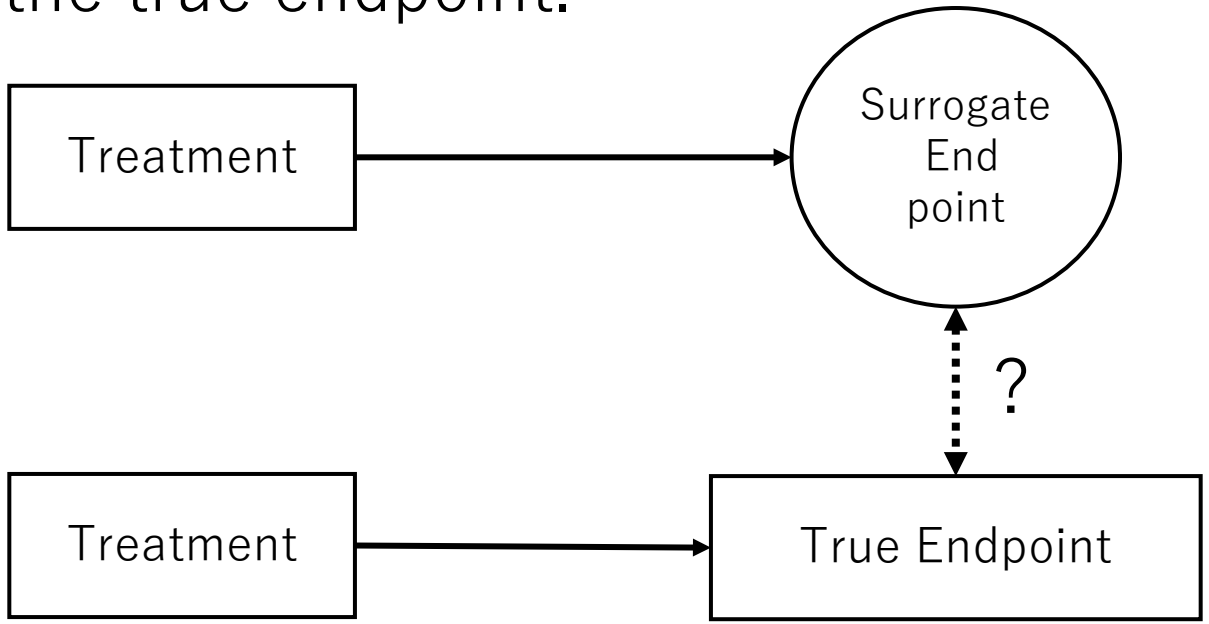
examine whether **the candidate surrogate endpoint predicts the prognosis of clinical outcomes**.



→This means looking at the association between the surrogate endpoint and the true endpoint.

Assessing Surrogacy ②: Does it predict treatment effect? (Trial level)

- Examine whether the effect of the study treatment on the surrogate endpoint corresponds to the effect of the study treatment on the true endpoint.



→This means looking at whether the between-group difference in the surrogate endpoint is associated with the between-group difference in the true endpoint in the clinical trial.

FDA classification of surrogate endpoints and drug regulation

Classification	Detail	Approval pathway
Validated surrogate endpoint	Supported by clear biological plausibility and clinical data providing strong evidence predicting clinical benefit	Traditional approval
Reasonably likely surrogate endpoint	Supported by strong biological plausibility and/or epidemiologic evidence; likely to predict clinical benefit but without sufficient clinical data yet	Accelerated approval
Candidate surrogate endpoint	Ability to predict clinical benefit still under evaluation	Not usable

Reference: BEST (Biomarkers, EndpointS, and other Tools)Resource.

東京大学大学院
情報学環・学際情報学府
The University of Tokyo III/GSII

Introduction to the Assessment Framework (REASSURE)

Interfaculty Initiative in Information Studies,
The University of Tokyo
Koji Oba
obakoji@iii.u-tokyo.ac.jp

Recent Developments In Surrogate Endpoints

SPIRIT-Surrogate

RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

Reporting of surrogate endpoints in randomised controlled trial protocols (SPIRIT-Surrogate): extension checklist with explanation and elaboration

Anthony Muchai Manyara,[1,2] Philippa Davies,[3] Derek Stewart,[4] Christopher J Weir,[5] Amber E Young,[3] Jane Blazeby,[3,6,7] Nancy J Butcher,[8,9] Sylwia Bujkiewicz,[10] An-Wen Chan,[11,12] Dalia Dawoud,[13,14] Martin Offringa,[8,15] Mario Ouwens,[16] Asbjørn Hróbjartsson,[17,18] Alain Amstutz,[19,20,21] Luca Bertolaccini,[22] Vito Domenico Bruno,[23] Declan Devane,[24,25] Christina D C M Faria,[26] Peter B Gilbert,[27] Ray Harris,[4] Marissa Lassere,[28] Lucio Marinelli,[29,30] Sarah Markham,[4,31] John H Powers III,[32] Yousef Rezaei,[33,34,35] Laura Richert,[36] Falk Schwendicke,[37] Larisa G Tereshchenko,[38] Achilles Thoma,[39] Alparslan Turan,[40] Andrew Worrall,[4] Robin Christensen,[41] Gary S Collins,[42] Joseph S Ross,[43,44] Rod S Taylor,[1,45] Oriana Ciani[46]

For numbered affiliations see end of the article

Correspondence to: O Ciani oriana.ciani@unibocconi.it (or @OrianaCiani on Twitter; ORCID 0000-0002-3607-0508)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2024;386:e078525 http://dx.doi.org/10.1136/bmj-2023-078525

Accepted: 30 April 2024

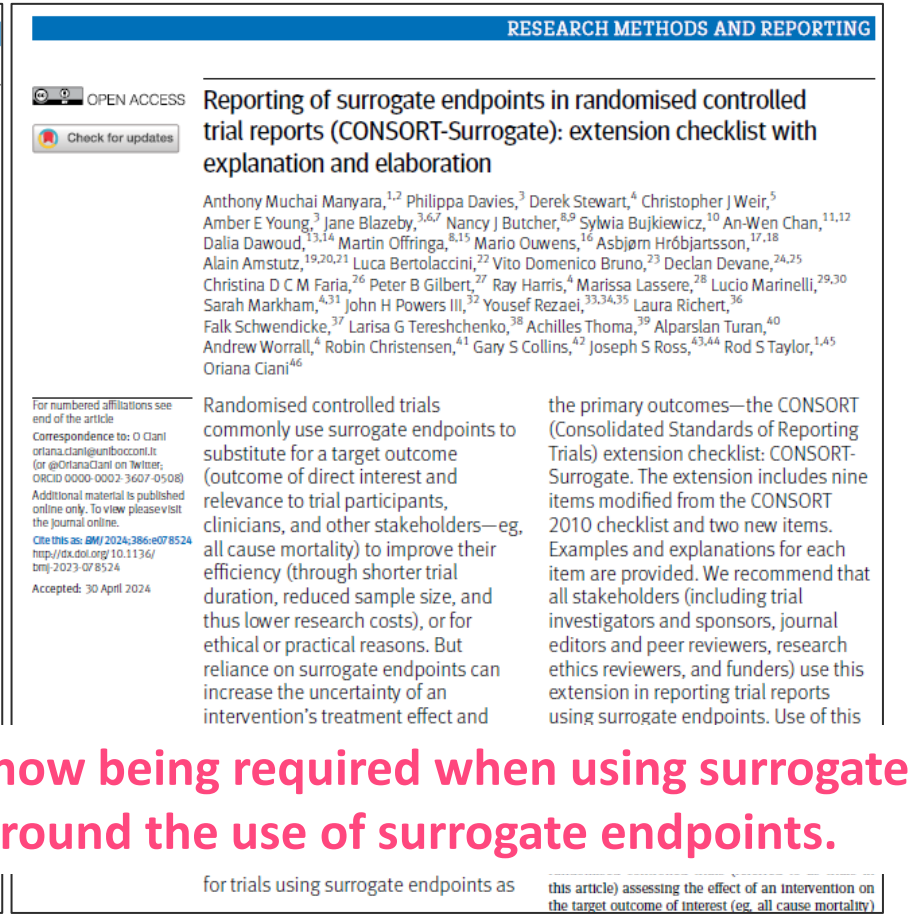
Randomised controlled trials often use surrogate endpoints to substitute for a target outcome (an outcome of direct interest and relevance to trial participants, clinicians, and other stakeholders—eg, all cause mortality) to improve efficiency (through shortened duration of follow-up, reduced sample size, and lower research costs), and for ethical or practical reasons. However, their use has a fundamental limitation in terms of uncertainty of the intervention effect on the target outcome and limited

Interventional Trials (SPIRIT) checklist, a consensus driven reporting guideline designed for trial protocols using surrogate endpoints as the primary outcome(s). The SPIRIT-Surrogate extension includes nine items modified from the SPIRIT 2013 checklist. The guideline provides examples and explanations for each item. We recommend that all stakeholders (including trial investigators and sponsors, research ethics reviewers, funders, journal editors, and peer

items. Recommendations for

completed, and ultimately reduced research waste.

CONSORT-Surrogate



RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

Reporting of surrogate endpoints in randomised controlled trial reports (CONSORT-Surrogate): extension checklist with explanation and elaboration

Anthony Muchai Manyara,[1,2] Philippa Davies,[3] Derek Stewart,[4] Christopher J Weir,[5] Amber E Young,[3] Jane Blazeby,[3,6,7] Nancy J Butcher,[8,9] Sylwia Bujkiewicz,[10] An-Wen Chan,[11,12] Dalia Dawoud,[13,14] Martin Offringa,[8,15] Mario Ouwens,[16] Asbjørn Hróbjartsson,[17,18] Alain Amstutz,[19,20,21] Luca Bertolaccini,[22] Vito Domenico Bruno,[23] Declan Devane,[24,25] Christina D C M Faria,[26] Peter B Gilbert,[27] Ray Harris,[4] Marissa Lassere,[28] Lucio Marinelli,[29,30] Sarah Markham,[4,31] John H Powers III,[32] Yousef Rezaei,[33,34,35] Laura Richert,[36] Falk Schwendicke,[37] Larisa G Tereshchenko,[38] Achilles Thoma,[39] Alparslan Turan,[40] Andrew Worrall,[4] Robin Christensen,[41] Gary S Collins,[42] Joseph S Ross,[43,44] Rod S Taylor,[1,45] Oriana Ciani[46]

For numbered affiliations see end of the article

Correspondence to: O Ciani oriana.ciani@unibocconi.it (or @OrianaCiani on Twitter; ORCID 0000-0002-3607-0508)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2024;386:e078524 http://dx.doi.org/10.1136/bmj-2023-078524

Accepted: 30 April 2024

Randomised controlled trials commonly use surrogate endpoints to substitute for a target outcome (outcome of direct interest and relevance to trial participants, clinicians, and other stakeholders—eg, all cause mortality) to improve their efficiency (through shorter trial duration, reduced sample size, and thus lower research costs), or for ethical or practical reasons. But reliance on surrogate endpoints can increase the uncertainty of an intervention's treatment effect and

the primary outcomes—the CONSORT (Consolidated Standards of Reporting Trials) extension checklist: CONSORT-Surrogate. The extension includes nine items modified from the CONSORT 2010 checklist and two new items. Examples and explanations for each item are provided. We recommend that all stakeholders (including trial investigators and sponsors, journal editors and peer reviewers, research ethics reviewers, and funders) use this extension in reporting trial reports using surrogate endpoints. Use of this

for trials using surrogate endpoints as

this article) assessing the effect of an intervention on the target outcome of interest (eg, all cause mortality)

Even in published papers, detailed explanations are now being required when using surrogate endpoints, and various initiatives are emerging around the use of surrogate endpoints.

Manyara AM, et al. BMJ. 2024;386:e078524.
Manyara AM, et al. BMJ. 2024;386:e078525.

Extension Items

Table. Summary of the SPIRIT- and CONSORT-Surrogate Extension Items

Core SPIRIT or CONSORT item and section	Surrogate extension item[a]	Extension item applies to
Abstract and trial summary	1a. State (1) that the primary outcome is a surrogate end point and (2) the target outcome(s) whose intervention effect is being substituted for	SPIRIT and CONSORT surrogates
Introduction and background or objectives	1b. State (1) that the primary outcome is a surrogate end point and (2) the target outcome(s) whose intervention effect is being substituted for	SPIRIT and CONSORT surrogates
Outcomes	2. State the practical or scientific reason(s) for using a surrogate end point as a primary outcome	SPIRIT and CONSORT surrogates
	3. State what other surrogate end points were considered and why the current one(s) were chosen	SPIRIT-Surrogate
	4. Justification for selected surrogate end point: (1) evidence (or lack thereof) of surrogate end point validation; (2) evidence (or lack thereof) of validity being specific to the context used (eg, intervention; disease; population)	SPIRIT and CONSORT surrogates
Sample size	5. Clarify if the sample size will be or was estimated to demonstrate that a minimum effect on the surrogate end point would be predictive of a benefit on the target outcome(s)	SPIRIT and CONSORT surrogates
Ethics and patient and public engagement	6. State whether and how trial participants will be and were engaged and informed before enrollment that the trial was designed to evaluate an intervention's effect using a surrogate end point	SPIRIT and CONSORT surrogates
Results or outcomes and estimation	7. If the primary outcome is a composite outcome that includes a surrogate end point, report the intervention effect on all components	CONSORT surrogate
Harms and discussion	8. Comment on whether the trial design (including sample size and follow-up period), given the use of a surrogate end point, adequately captures the potential harms of the intervention being tested	SPIRIT and CONSORT surrogates
Statistical methods and discussion	9. State what the plans are to conduct subsequent analyses and studies to verify current findings on the target outcome(s)	SPIRIT and CONSORT surrogates
Discussion and interpretation	10. Interpretation of findings of the trial in the context of using a surrogate primary end point, including its known validity for intervention effects on the target outcome and the potential benefit-risk assessments of the tested intervention for participants	CONSORT surrogate
Dissemination and data access	11. If surrogate and target outcome data will be or were collected in the trial, state the open-access arrangements for the data for future secondary research	SPIRIT and CONSORT surrogates

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials.

[a] Notes on use of extensions: Items should be used along with the core SPIRIT and CONSORT checklists; items are applicable to all intervention trials using surrogate end points as primary outcome(s). However, items could be relevant to report nonrandomized trials, observational studies, and other studies using surrogate end points; and items can be combined or reported in different sections to the ones suggested in the extension. The specific item sections are recommendations and not requirements.

Taylor RS, Manyara AM, Ross JS, Ciani O. JAMA. 2024;332(16):1340-1342.

Extension Items (Excerpt)

- Statement that the primary endpoint is a surrogate endpoint
  - Statement of the true outcome that the surrogate is meant to replace
- Statement of the reason for using a surrogate endpoint
- Justification for the selected surrogate endpoint
  - Evidence regarding surrogate endpoint validity
  - Evidence of validity for the intervention, disease, and population in question
- Sample size based on the effect on the surrogate endpoint, sufficient to predict the effect on the true endpoint
- Handling when included as one of a composite endpoint
- How adverse events are captured in the trial, etc.

21st Century Cures Act developments

Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure

Share | Post | Linkedin | Email | Print

What is the purpose of the Surrogate Endpoint Table?

FDA's surrogate endpoint table provides valuable information for drug developers on endpoints that may be considered and discussed with FDA for individual development programs. This table also fulfills a 21st Century Cures Act requirement to publish a list of "surrogate endpoints which were the basis of approval or licensure (as applicable) of a drug or a biological product" under both accelerated and traditional approval pathways.

Section 3011 of the 21st Century Cures Act established section 507 of the Federal Food, Drug, and Cosmetic Act (FD&C Act), which mandates that the FDA publish a list of "surrogate endpoints which were the basis of approval or licensure (as applicable) of a drug or biological product" under both accelerated and traditional approval provisions. The Surrogate Endpoint Table below fulfils this legislative requirement and is intended to provide valuable information for drug developers on endpoints that may be considered and discussed with FDA for individual development programs.

Content current as of:
02/28/2022

Regulated Product(s)
Drugs

https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure

Adult Surrogate Endpoint Table

Disease or Use	Patient Population	Surrogate Endpoint	Type of approval appropriate for	Drug mechanism of action
Alpha-1-antitrypsin deficiency	Patients with congenital alpha-1 antitrypsin deficiency	Plasma alpha-1 proteinase inhibitor	Traditional	Alpha-1 protease inhibitor augmentation
Acromegaly	Patients with acromegaly who don't respond to or cannot undergo other standard therapies	Serum Insulin-like growth factor-I (IGF-1)	Traditional	Growth hormone receptor antagonist
Acromegaly	Patients with acromegaly who don't respond to or cannot undergo other standard therapies	Serum growth hormone and serum insulin-like growth factor-I (IGF-1)	Traditional	Somatostatin analog
Acute Bronchospasm	Patients with acute bronchospasm associated with reversible obstructive airway disease or exercise	Forced expiratory volume in 1 second (FEV1)	Traditional	Beta-2 adrenergic agonist
Alzheimer's disease	Patients with mild cognitive impairment or mild dementia stage of Alzheimer's disease	Reduction in amyloid beta plaques	Accelerated	Monoclonal antibody
Anthrax vaccine	Persons at high risk of exposure to anthrax	Anti-protective antigen antibody response	Traditional	Induction of immunity
Anticoagulation reversal (needed due to life-threatening or uncontrolled bleeding)	Patients treated with a direct or indirect FXa inhibitor when reversal of anticoagulation is needed	Percent change in anti-FXa activity, from baseline to nadir	Accelerated	Binding and sequestering FXa inhibitors
Asthma	Patients with asthma	Forced expiratory volume in 1 second (FEV1)	Traditional	Corticosteroid; Beta-2 adrenergic agonist
Cancer: hematological malignancies	Patients with Acute Lymphoblastic Leukemia	Serum asparaginase	Traditional	Asparagine-specific enzyme
Cancer: hematological malignancies	Patients with chronic myeloid leukemia	Major hematologic response	Accelerated/Traditional§	Mechanism agnostic*

JAMA | Original Investigation

Associations Between Surrogate Markers and Clinical Outcomes for Nononcologic Chronic Disease Treatments

Joshua D. Wallach, PhD, MS; Samuel Yoon, BS, BA; Harry Doernberg, MM; Laura R. Glick, MD; Oriana Ciani, PhD; Rod S. Taylor, PhD; Maryam Mooghali, MD; Reshma Ramachandran, MD, MHS, MPP; Joseph S. Ross, MD, MHS

IMPORTANCE Surrogate markers are increasingly used as primary end points in clinical trials supporting drug approvals.

OBJECTIVE To systematically summarize the evidence from meta-analyses, systematic reviews and meta-analyses, and pooled analyses (hereafter, meta-analyses) of clinical trials examining the strength of association between treatment effects measured using surrogate markers and clinical outcomes in nononcologic chronic diseases.

DATA SOURCES The Food and Drug Administration (FDA) Adult Surrogate Endpoint Table and MEDLINE from inception to March 19, 2023.

STUDY SELECTION Three reviewers selected meta-analyses of clinical trials; meta-analyses of observational studies were excluded.

DATA EXTRACTION AND SYNTHESIS Two reviewers extracted correlation coefficients, coefficients of determination, slopes, effect estimates, or results from meta-regression analyses between surrogate markers and clinical outcomes.

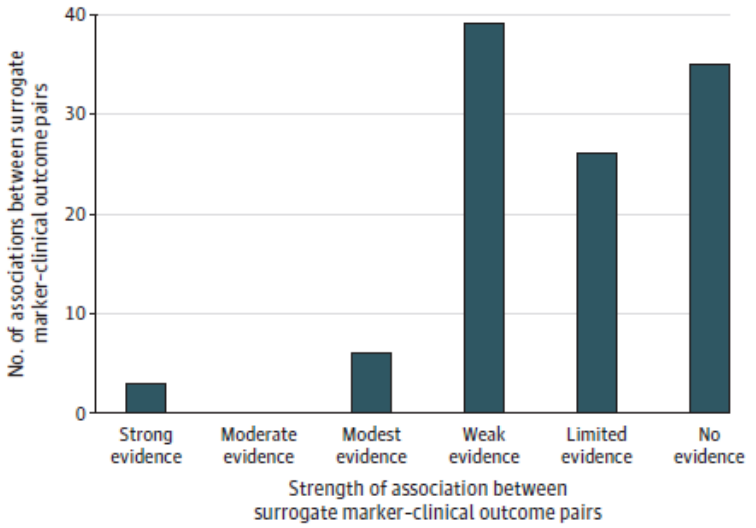
MAIN OUTCOMES AND MEASURES Correlation coefficient or coefficient of determination, when reported, was classified as high strength ($r \geq 0.85$ or $R^2 \geq 0.72$); primary findings were otherwise summarized.

RESULTS Thirty-seven surrogate markers listed in FDA's table and used as primary end points in clinical trials across 32 unique nononcologic chronic diseases were included. For 22 (59%) surrogate markers (21 chronic diseases), no eligible meta-analysis was identified. For 15 (41%) surrogate markers (14 chronic diseases), at least 1 meta-analysis was identified, 54 in total (median per surrogate marker, 2.5; IQR, 1.3-6.0); among these, median number of trials and patients meta-analyzed was 18.5 (IQR, 12.0-43.0) and 90 056 (IQR, 20 109-170 014), respectively. The 54 meta-analyses reported 109 unique surrogate marker–clinical outcome pairs: 59 (54%) reported at least 1 r or R^2, 10 (17%) of which reported at least 1 classified as high strength, whereas 50 (46%) reported slopes, effect estimates, or results of meta-regression analyses only, 26 (52%) of which reported at least 1 statistically significant result.

CONCLUSIONS AND RELEVANCE Most surrogate markers used as primary end points in clinical trials to support FDA approval of drugs treating nononcologic chronic diseases lacked high-strength evidence of associations with clinical outcomes from published meta-analyses.

Supplemental c

Figure 3. Strength of the Associations Between 109 Surrogate Marker–Clinical Outcome Pairs



Strong evidence: correlation coefficients or coefficients of determination values reported for all associations examined, and all associations classified as statistically significant and high strength according to criteria proposed by the Institute for Quality and Efficiency in Health Care ($r \geq 0.85$ or $R^2 \geq 0.72$).[14] Moderate evidence: r or R^2 values reported for all associations examined, and 1 or more (but not all) classified as statistically significant and high strength. Modest evidence: r or R^2 values reported for some associations examined, and 1 or more (but not all) classified as statistically significant and high strength. Any other r, R^2, slopes, effect estimates, or results from meta-regression analyses classified as statistically significant. Weak evidence: no r or R^2 values classified as both statistically significant and high strength, but all r, R^2, slopes, effect estimates, or results from meta-regression analyses classified as statistically significant. Limited evidence: no r or R^2 values classified as both statistically significant and high strength; some r, R^2, slopes, effect estimates, or results from meta-regression analyses classified as statistically significant and some not. No evidence: no r or R^2 values classified as statistically significant and high strength, and all r, R^2, slopes, effect estimates, or results from meta-regression analyses classified as nonstatistically significant.

FDA's Response

Surrogate Markers and Clinical Outcomes

To the Editor Dr Wallach and colleagues[1] investigated the evidence supporting surrogate end points used for drug approvals according to listings in the online US Food and Drug Administration (FDA) surrogate end points table and indicate that the FDA always relies on meta-analyses, which is untrue. The authors state that meta-analyses offered "inconsistent evidence on the strength of association" between the surrogate end points used for drug approvals and clinical outcomes for several major diseases, including HIV (plasma HIV RNA), hypercholesterolemia (low-density lipoprotein cholesterol [LDL-C]), and hypertension. This analysis incorrectly identifies the evidence the FDA used to conclude that these biomarkers can serve as surrogate end points. To assess the ability of a biomarker to predict clinical outcomes, the FDA takes into account a range of sources, including mechanistic evidence that the biomarker is on the causal pathway of disease, nonclinical models, epidemiologic data, and clinical trial data, including data from the FDA's own analyses of patient- and trial-level data to determine the quantitative association between the effect of treatment on the biomarker and the clinical outcomes.

Several examples illustrate this point. In the case of HIV disease, the FDA conducted a patient-level analysis of more than 5000 participants in multiple trials in the late 1990s that demonstrated a direct association between sustained reductions in HIV RNA and lower risks of disease progression

To assess the ability of a biomarker to predict clinical outcomes, the FDA takes into account a range of sources, including **mechanistic evidence** that the biomarker is on the **causal pathway** of disease, **nonclinical models**, **epidemiologic data**, and **clinical trial data**, including data from the FDA's own analyses of patient- and trial-level data to determine the quantitative association between the effect of treatment on the biomarker and the clinical outcomes.

There are various viewpoints in the discussion around justifying a surrogate endpoint

Relationship of the SE with the Clinical Outcome

Rationale for Using the SE as a Primary Endpoint
• What is the clinical outcome the SE is proposed to predict? • What is the rationale for using an SE rather than the clinical outcome measure (e.g., feasibility, study duration, sample size, etc.)? • What evidence exists to support the relationship between the SE and the clinical outcome of interest (e.g., epidemiologic studies, randomized controlled trials, data generated from therapeutic products from the same class)? • If the SE is proposed based upon prior publications and general scientific community acceptance, discuss the current body of evidence and the use of the SE in clinical studies. • If the disease course is typically acute, what are the expected long-term sequelae and could a change in the SE also reflect these clinical outcomes (e.g., reduction in viral shedding with influenza and reduced incidence of hospitalization due to complications from illness)?
Relationship of the SE with the Causal Pathway(s)
• What is known about the causal pathway(s) for the intended disease? What is the relationship of the SE to this pathway(s)? • Does the intended disease or use have multiple causal pathways? If so, what is the evidence that the specific pathway the SE monitors is the primary pathway leading to the outcome being assessed? • If the SE is on a single causal pathway, what is the evidence that this pathway is the predominant mechanism for pathogenesis? • If the SE is not on the causal pathway, what is the rationale, the evidence, and the strength of evidence that leads to the conclusion that a change in the SE will be predictive of a change in the clinical outcome of interest?
Threshold for Change Required to Demonstrate Clinical Relevance
• How much do changes in the SE reflect changes in the clinical outcome or the probability of the clinical outcome occurring? • What is the extent and timing of change in the SE that would predict the outcome of interest? • Is the change in the SE table or does it only occur for a short time? Would timing of sample collection be feasible? • If a minimum threshold for change has been selected (both size and duration), how was this value determined (include studies conducted and data generated)? If available, is there information about the sensitivity and specificity of any measurement tools used to include?
Consistency of SE Response under Various Conditions
• How does the SE predict the clinical outcome across different subgroups of the targeted population (e.g., demographic, disease severity, co-existing conditions, concomitant medications typical of the population)? • Is there data showing that the SE predicts the clinical response similarly across relevant subgroups?
Reliably Quantifying Changes in the Clinical Outcome Before and After Treatment
• Are there scoring systems or measurement tools commonly used in clinical practice, supported by high grade clinical evidence, and recognized by clinical expert groups to assess baseline disease status and/or the effect of treatment on the clinical outcome? • How specific and sensitive is the current standard to measure the clinical endpoint that the SE is intended to predict?

Relationship of the SE with the Therapeutic Product

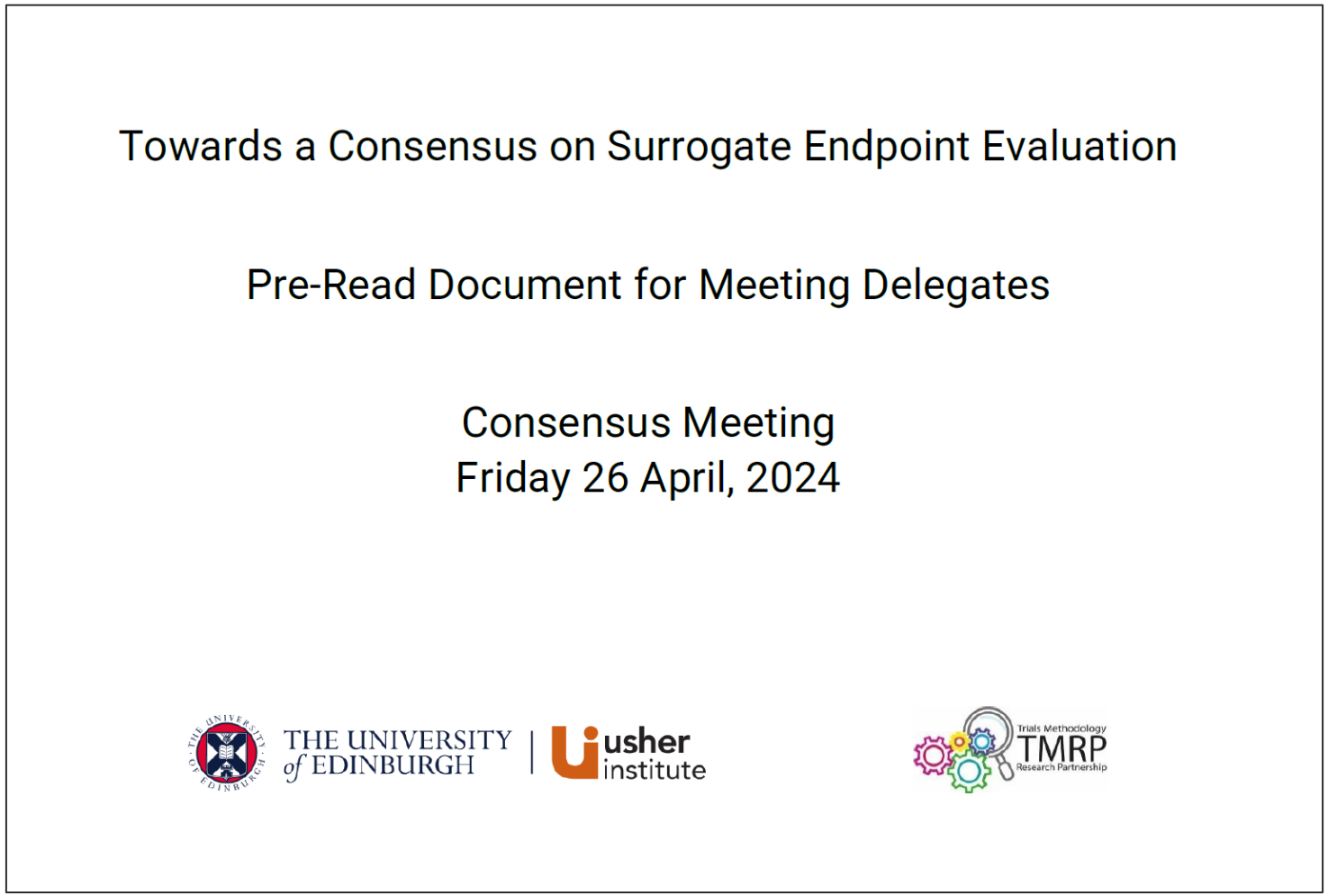
Predictive Value of Therapeutic-Induced Changes in the SE
• What is the evidence that a therapeutic-induced change in the SE will be predictive of a change in the clinical outcome (e.g., the SE is a correlate vs. the SE has actual predictive value)?
Off-Target Effects of the Therapeutic Product
• Is there evidence (identified in human or animal studies) to suggest the therapeutic product could affect off-target causal pathways, resulting in harm that may or may not be reflected by changes in the SE? • Are there any pharmacologic effects that could influence the SE but that are unrelated to modifying the disease process (e.g., renal blood flow and creatinine)?

Reliability of the Measurement Tool(s) Used to Detect the SE

Operations Manual
• Does the operations manual include a detailed process from specimen collection to results reporting (e.g., information on sample collection and handling, sample processing, testing, results gathering, and interpretation)?
Performance Characteristics of the Measurement Tool(s)
• To what extent have the performance characteristics of the measurement tool (e.g., sensitivity, specificity, within laboratory precision and between laboratory reproducibility, and sample stability under expected handling conditions) been studied? • Is there a description of the sample type(s) that were used to generate the data for these studies included in the briefing package (e.g., prospective or retrospectively collected patient samples, left-over banked samples, samples spiked with defined levels of analyte)?

Isn't a framework that ensures transparency needed?

Efforts Toward Building a Framework



List of Participants

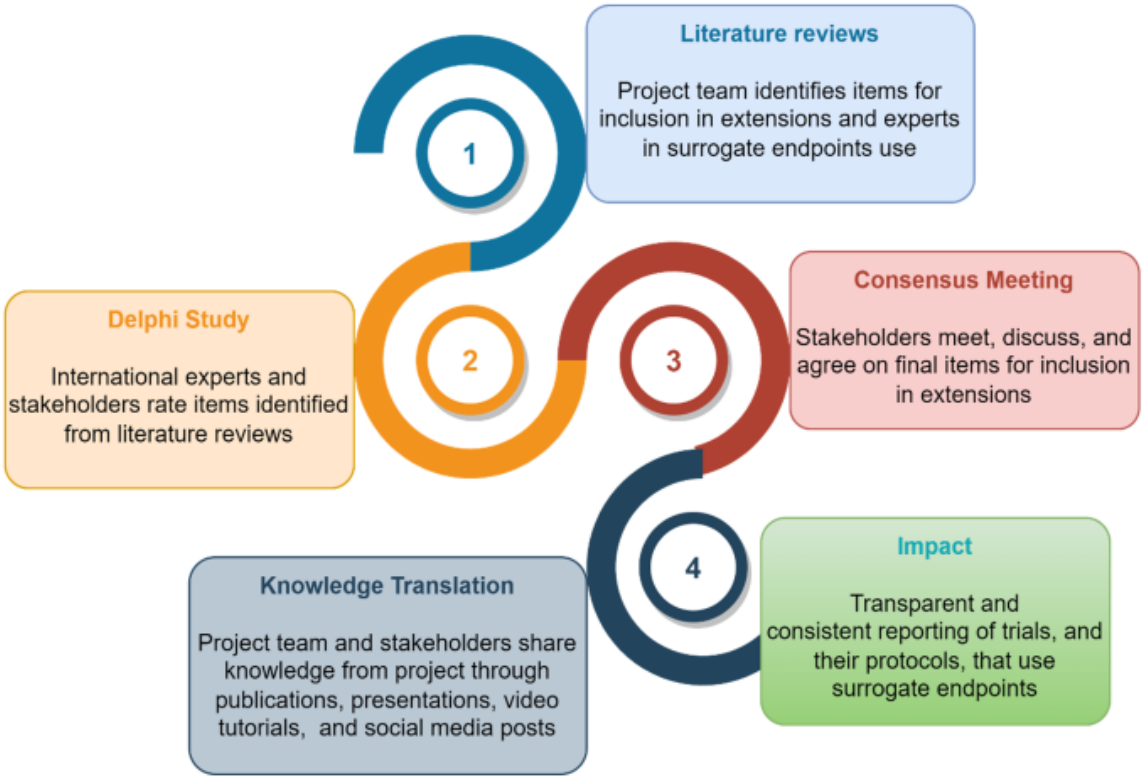
Names of Delegates and Facilitator

Name of Delegate	Institution	Country of Residence
Stuart Baker	National Cancer Institute	USA
Luca Bertolaccini	IEO, European Institute of Oncology IRCCS	Italy
Sylwia Bujkiewicz	University of Leicester	UK
Catey Bunce	The Royal Marsden NHS Foundation Trust	UK
Marc Buyse	International Drug Development Institute	Belgium
Oriana Ciani	SDA Bocconi School of Management	Italy
Jane Daniels	University of Nottingham	UK
Michael Elliott	University of Michigan	USA
Sara Gandini	IEO, European Institute of Oncology IRCCS	Italy
Tamer Gheita	Cairo University	Egypt
Peter Gilbert	Fred Hutchinson Cancer Center	USA
Laura Jaeger	U.S. Food and Drug Administration	USA
Marissa Lassere	University of New South Wales	Australia
Pok Lo (meeting chair)	The University of Edinburgh	UK
Koji Oba	The University of Tokyo	Japan
Mario Ouwens	AstraZeneca	Sweden
Emily Roberts	University of Iowa	USA
Rod Taylor	University of Glasgow	UK
Carlos Toro-Huamanchumo	Universidad San Ignacio de Loyola	Peru

Name of Facilitator	Institution	Country of Residence
Anthony Manyara	University of Bristol	UK

Consensus-Building Process

CONSORT and largely follows the same basic development process



https://www.gla.ac.uk/schools/healthwellbeing/research/mrccsosocialandpublichealthsciencesunit/programmes/complexity/methods-development/spirit-consort-surrogate/spirit-consort-surrogate-summary/

REASSURE

- Developed as a framework for judging the acceptability of surrogate endpoint use in the development of pharmaceuticals and medical devices, and in health technology assessment
  - Under submission
- Since the evaluation of each item is context-dependent, no fixed procedure is used to derive an overall assessment
  - Sub-items are evaluated qualitatively
- Use of the framework assumes collaborative work by multiple stakeholders

Key Domains of the Framework

- The framework is organized into the following 3 domains

- Domain 0: Context of use
  - Describe the disease, treatments, and the surrogate and clinical endpoints of interest.
- Domain 1: Practical Motivation
- Domain 2: Clinical and Biological Evidence
- Domain 3: Statistical Methods

Summary

- REASSURE makes it possible to organize the following points
  - Balance between time savings and increased uncertainty
  - Biological plausibility
  - RCT or RCT meta-analysis evidence
  - Clinical validity
  - Generalisability
- Evaluating each item requires expertise across multiple disciplines, including biostatistics, clinical medicine, regulatory science, and health technology assessment
  - Recommend evaluating a candidate surrogate endpoint from multiple perspectives before deciding whether to adopt it as the primary endpoint of a clinical trial

Group Discussion & Final Presentation

Group Discussion Overview

- **Surrogate EP validity assessment (filling out the REASSURE framework) → Recommendation vote**
  - Fill in each domain of the REASSURE framework based on prior information and current evidence, and based on that content, assess the validity of the surrogate EP.
  - Based on the information entered, participants vote on the following question, with each person stating their rationale.

Q. Would you recommend this endpoint as a surrogate endpoint?
(Answer in line with the FDA regulatory framework, choosing from the following 3 options) *

Recommend | **Recommend with conditions** | **Do not recommend**

  - Also note any points of disagreement or any comments on limitations

* This vote reflects the participants' judgment based on limited time and information within this workshop, and may not align with official regulatory positions (e.g., FDA).

Group Discussion Overview

- **Case study for each group**

Surrogate endpoint	Disease
1. eGFR Slope	Chronic kidney disease
2. MRD (measurable residual disease)	Multiple myeloma
3. NfL level	SOD1 gene mutation-associated amyotrophic lateral sclerosis (SOD1-ALS)

Surrogate Endpoint Assessment

Group 1 *: Chronic kidney disease, eGFR slope*

Disease and Endpoint Under Assessment

1 Disease / therapeutic area

[Chronic kidney disease (CKD)]

2

Surrogate endpoint under assessment

[eGFR slope]

→

3

Corresponding true endpoint

[Composite endpoint including ESKD etc. (excluding cardiovascular events)]

4 **Why a surrogate EP is needed: [The true EP takes a long time to assess, the surrogate EP can reduce this by about 3–8 years.]**

Conclusion: Should this EP be recommended as a surrogate EP?

Recommend	Recommend with conditions	Do not recommend

3 year Total slope: **Recommend**
✓ slope=-0.35 (-0.42, -0.29), R^2 = 0.97 (0.82, 1.00)

2 year Total slope: Uncertain
✓ slope=-0.27 (-0.33, -0.21), R^2 = 0.87 (0.64, 0.97)

Chronic slope: **Recommend with conditions** (judgment depends on drug characteristics; meta-analysis accounting for drug characteristics needed; used in secondary or sensitivity analyses)
✓ slope=-0.33 (-0.46, -0.20), R^2 = 0.55(0.25, 0.77)

Points of Disagreement / Limitations of the Assessment

Points of disagreement: 3-year total slope vs 2-year total slope

Reasons for support: R^2 is within an acceptable range
Reasons for uncertainty: the lower bound of the credible interval is below the acceptable threshold

Limitations of the assessment: Chronic slope assessment

It may be necessary to consider whether there was an initial dip, since the meta-analysis includes multiple interventions.

Surrogate Endpoint Assessment

Group 2 *: Multiple myeloma, MRD (measurable residual disease)*

Disease and Endpoint Under Assessment

1 Disease / therapeutic area

Multiple myeloma

- **Newly diagnosed cases (NDMM: both transplant-eligible and -ineligible)**
- **Relapsed/refractory cases (RRMM)**

2 Surrogate endpoint under assessment

At least one MRD (measurable residual disease) negative result during the assessment period (9 months or 12 months)1

→

3 Corresponding true endpoint

OS

4 Why a surrogate EP is needed: The true EP takes a long time to assess (with OS, it could take 10+ years)

Conclusion: Should this EP be recommended as a surrogate EP ?

Recommend	Recommend with conditions	Do not recommend
	5 people	2 people

- Rare disease/chronic disease with no curative treatment (high unmet medical needs)
- Biological plausibility is reasonable
  - Tumor cells undetectable at a level of 10^(-5) or higher
  - MRD lies on the causal pathway to the true endpoint
  - Residual myeloma cells -> clonal expansion -> disease progression -> death
- High result from the individual-level meta-analysis
  - Meets a pre-specified criterion (Global Odds Ratio > 3)

Points of Disagreement / Limitations of the Assessment

Points of disagreement:

Opinion against approving as a surrogate EP

- **Surrogate paradox (considering that MRD may not be an adequate surrogate EP for OS)**
- **Trial-level R^2 exceeded the pre-specified threshold of 0.5, but R^2 being around 0.6 means the evidence may not be considered strong.**

Limitations of the assessment:

- **Trial-level correlation evidence is not strong, and the number of trials was limited in diseases such as RRMM.**
  - **Additional evidence needed**
- **This endpoint has been assessed as a surrogate EP in the First line setting, and the current existence of other treatment options also needs to be considered**
- **MRD assessment timing(9 or 12 months) may vary depending on the drug used, so it cannot be judged uniformly**

Surrogate Endpoint Assessment

***Group 3** : SOD1-ALS, NfL level*

Disease and Endpoint Under Assessment

1 Disease / therapeutic area

SOD1-ALS

2

Surrogate endpoint under assessment

Plasma NfL level

→

3

Corresponding true endpoint

ALSFRS-R

4 **Why a surrogate EP is needed:** *Treatment options are limited and there is urgency*
ALSFRS-R scores show little variation,
so detecting a difference takes a long time

Conclusion: Should this EP be recommended as a surrogate EP ?

Do not recommend

- From the perspective of disease rarity and development benefits, there is motivation to use a surrogate EP.
- However, the individual-level correlation in a single trial was low (Spearman correlation coefficient r=0.07)
- Also, with only a single trial, trial-level correlation cannot be assessed

Points of Disagreement / Limitations of the Assessment

Points of disagreement

- *As a Clinical Endpoint, what is the relationship between ALS-SFRS-R and OS?Isn't something showing the degree of association with OS needed?*
- *What time point is needed? Might effects appear at a later time point?*
- *As a resource saved, the resource for measurement does not increase(would it differ in CSF?)*
- *Might the sample size decrease if the effect size is large?*

Limitations of the assessment

- *Conducting the same confirmatory trial again is not realistic.*
- *Long-term follow-up after the trial or registry data is needed*

Remaining Issues

**Given that perfect follow-up data cannot be expected for a rare disease,
is it unavoidable to judge surrogacy based on the data currently available?
(Is it realistic to require multiple trials?)**