

第10回データサイエンスラウンドテーブル会議

実践的フレームワークを用いた代替エンドポイントの妥当性評価

2026年7月7日

本テーマのファシリテーター

本日のディスカッションのファシリテーターを務めます

● 大庭 幸治

東京大学

● 青木 琢巳

PMDA

● 田中 清士朗

PMDA

● 吉田 瑞樹

Pfizer

● 多喜田 保志

Lilly

産官学の多様な視点から、活発な議論をリードします

Ground Rule

性別・年齢・経験年数に関係なく、自由に意見を言い、その意見が尊重される場にしましょう

● ① 自由に意見を言う

性別・年齢・経験年数に関係なく、誰もが対等に発言できます。積極的に意見を述べましょう。

● ② 意見を尊重する

出された意見は否定せず、まずは受け止めましょう。異なる意見こそ議論の価値です。

● ③ 困ったら声をかける

進め方や内容で迷ったときは、遠慮なくファシリテーターやチームに声をかけてください。

立場に関係なく、全員が安心して発言できる場を一緒に作りましょう

本ワークショップの目的

● ① 理論的背景を学ぶ

代替EPに関する理論的背景・規制要件等を知らないと、規制当局との議論すらできない。まずその概要を学ぶ。

● ② 代替EP妥当性評価フレームワークを実際に使う

代替EP妥当性評価フレームワークを実事例に適用し妥当性評価を体験する。エビデンスの集め方も学ぶ。

● ③ 代替EP評価能力を養う

フレームワークで状況を整理・共有し、チーム内議論を通じて、規制交渉の場でも通用する評価能力を身につける。

● ④ 産官学の対話を深める

産官学それぞれの立場からの考えから、多様な視点への理解を深めるとともに、継続的な対話のネットワークを築く。

タイムテーブル

● ① 情報提供

10:15 – 11:40

- 代替エンドポイントに関する理論・規制当局の要件
- 代替エンドポイント評価フレームワーク（REASSURE）の紹介
- 各事例の概要・グループディスカッションの進め方

● ② グループディスカッション

11:40 – 15:15

- 前半：各事例を深掘りし、REASSUREフレームワークに沿って分析・議論
- 後半（最後30分）：グループごとに議論内容をプレゼン

● 各チーム発表者1名

● ③ 全体発表・共有

15:30 – 16:45

- 本ワークショップの概要説明
- 各グループの議論内容を全体で共有

● 各チーム発表者1名

情報提供

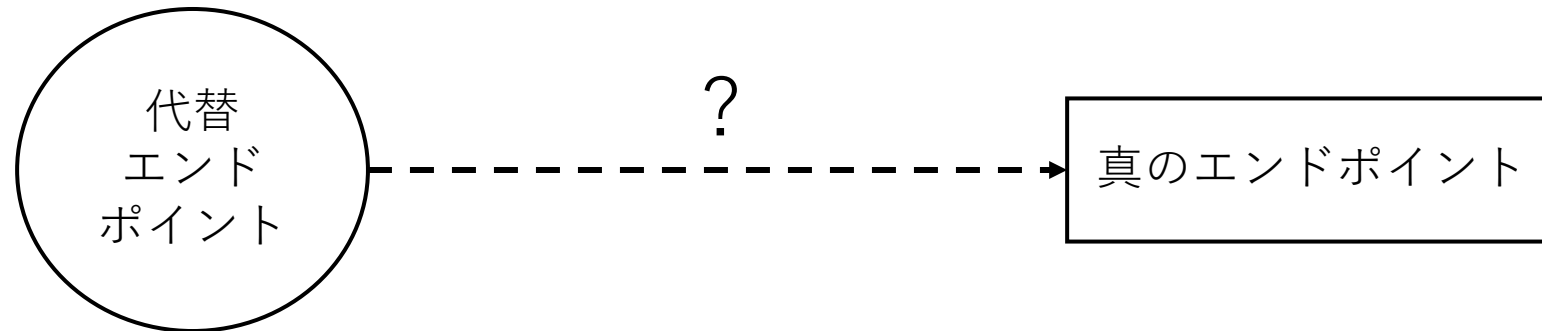
目次

- 代替エンドポイントの妥当性評価における共通の大枠
- 各規制当局における代替エンドポイントと規制の関係



代替性の検討①：予後を予測するか？（個人レベル）

- 疫学研究によって、
候補となる代替エンドポイントが**臨床結果の予後を予測するか**を検討する。

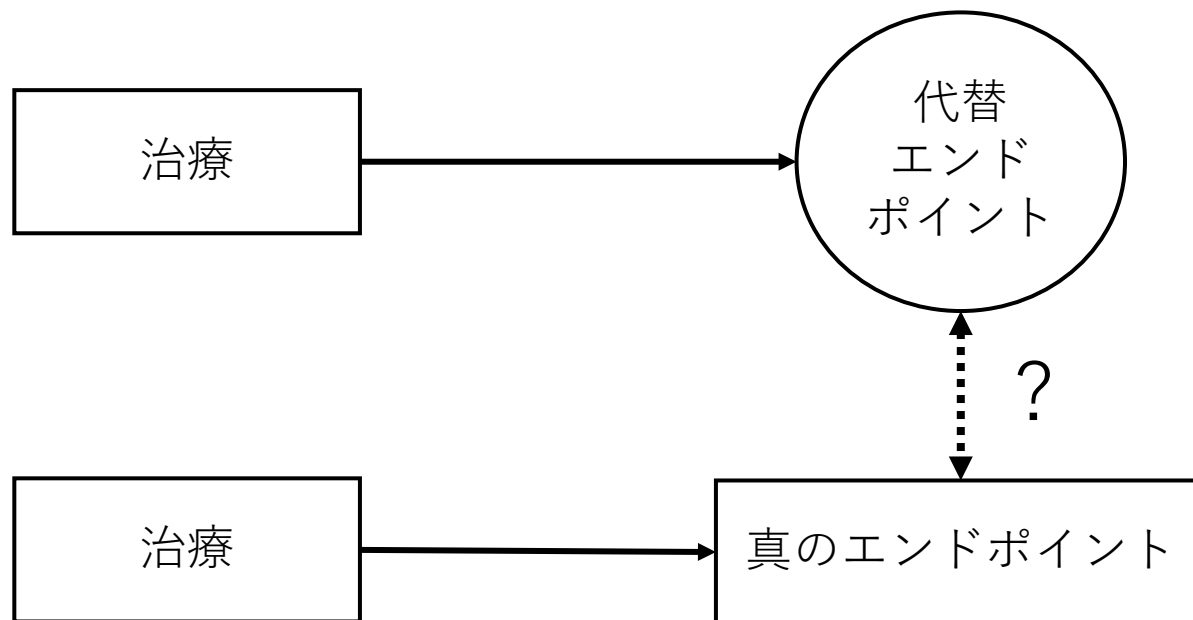


→代替エンドポイントと真のエンドポイントとの関連をみることになる。



代替性の検討②：介入効果を予測するか？（試験レベル）

- 試験治療が代替エンドポイントに及ぼす効果と、試験治療が真のエンドポイントに及ぼす効果に対応しているのか検討する。



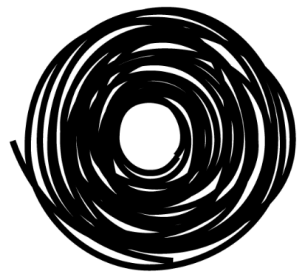
→臨床試験における代替エンドポイントの群間差と
真のエンドポイントの群間差が関連しているのかをみることになる。



FDAによる代替エンドポイントの分類と薬事規制

分類	詳細	承認経路
Validated surrogate endpoint	明確な生物学的妥当性と、臨床的效果を予測するという強力な証拠を有する臨床データによって裏付けられたもの	通常承認 (traditional approval)
Reasonably likely surrogate endpoint	強力な生物学的妥当性及び/又は疫学的根拠によって裏付けられたエンドポイントであり、臨床的效果を予測する可能性が高いが、まだ十分な臨床データがないもの	迅速承認 (accelerated approval)
Candidate surrogate endpoint	臨床的效果を予測する能力について評価中のもの	使用不可





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

評価フレームワーク（REASSURE）の紹介

東京大学大学院 情報学環
大庭 幸治

obakoji@iii.u-tokyo.ac.jp

近年の代替エンドポイント関連の動き

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 Ouwers,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 Ciani⁴⁶

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 information on potential intervention

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 Ouwers,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 Ciani⁴⁶

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-2023-078524>
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

consensus driven reporting guideline for trials using surrogate endpoints as randomised controlled trials (referred to as trials in this article) assessing the effect of an intervention on the target outcome of interest (eg, all cause mortality)

論文上でも代替エンドポイントを利用する際に詳細な説明を求められるようになっており
代替エンドポイント使用に際して様々な取り組みが見られるようになってきている



拡張された項目

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.



拡張された項目（抜粋）

- 主要評価項目が代替エンドポイントであることの記載
 - 代替しようとする真のアウトカムについての記載
- 代替エンドポイントを利用する理由の記載
- 選択した代替エンドポイントの正当化
 - 代替エンドポイント妥当性に関するエビデンス
 - 対象の介入、疾患、集団に関する妥当性のエビデンス
- 真のエンドポイントにおける効果を予測するのに十分な代替エンドポイントにおける効果に基づくサンプルサイズ
- 複合エンドポイントの一つに含まれる場合の取り扱い
- 試験における有害事象の捉え方、など

21st Century Cures Act の動き

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



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

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*



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.

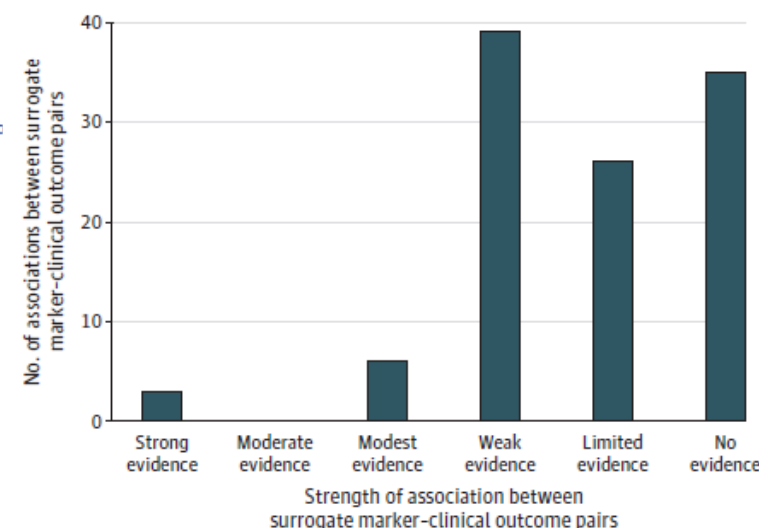
MAIN OUTCOMES AND MEASURES Correlation coefficient or coefficient of determination, when reported, was classified as high strength ($r \geq 0.85$ or $R^2 \geq 0.72$); primary findings were otherwise summarized.

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の回答

Surrogate Markers and Clinical Outcomes

To the Editor Dr Wallach and colleagues¹ 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.



代替エンドポイントの正当化の議論には、様々な観点がある

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 stable 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 baseline data?
- How specific SE is intended

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)?

透明性を担保したフレームワークの構築が必要では？



フレームワークづくりに向けた取り組み

Towards a Consensus on Surrogate Endpoint Evaluation

Pre-Read Document for Meeting Delegates

Consensus Meeting

Friday 26 April, 2024



THE UNIVERSITY
of EDINBURGH

Usher
institute



参加者一覧

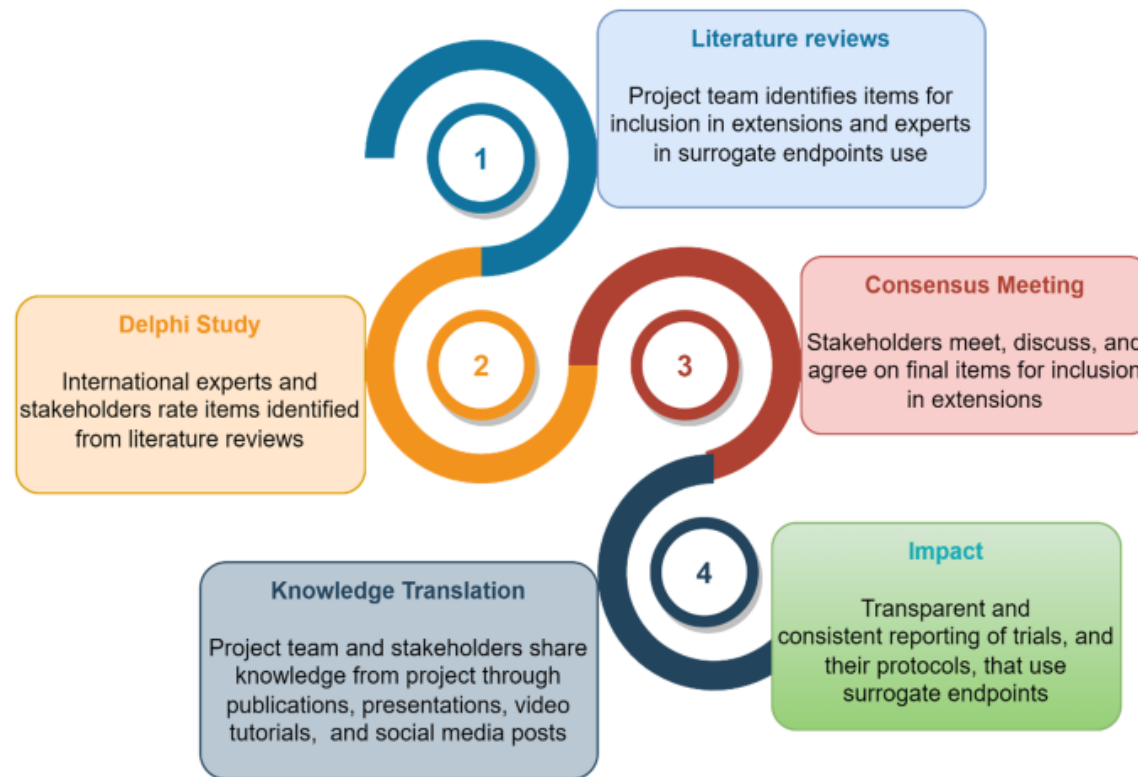
Names of Delegates and Facilitator

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 Ouwers	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

コンセンサス構築の流れ

CONSORTなどの構築プロセスと基本的には同じ



REASSURE

- 医薬品および医療機器の開発、医療技術評価における代替エンドポイントの使用の許容性を判断するための指針として開発
 - 投稿中
- 各項目の評価は文脈に依存するため、決まった総合評価のスコアリング手順は行わない
 - サブ項目は定性的に評価
- フレームワークの利用は、多くの関係者による共同作業を想定



フレームワークを構成するドメイン

- 以下の3つのドメインでフレームワークを整理
- ドメイン 0: 使用状況
 - 対象としている疾患、医薬品、想定する代替エンドポイントや臨床エンドポイントに関する記載
- ドメイン 1: 実務的動機
- ドメイン 2: 臨床的・生物学的エビデンス
- ドメイン 3: 統計手法



ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン0：使用状況（Context of Use）				
治療（Treatment）	0.1a	当該治療の適応および対象集団を記述する	該当なし	
	0.1b	当該治療のクラス（分類）を記述する	該当なし	
	0.1c	当該治療のラインを記述する	該当なし	
エンドポイント（Endpoints）	0.2a	代替エンドポイント（surrogate endpoint）を記述する	該当なし†	
	0.2b	臨床エンドポイント（clinical endpoint）を記述する	該当なし†	
試験フェーズ（Phase）	0.3	当該代替エンドポイントが用いられ得る臨床試験フェーズを記述する	該当なし	

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン1：実務的動機（Practical Motivation）				
疾患の有病率*	1.1	疾患の有病率*	希少／一般的	
臨床的ニーズ*	1.2	臨床的ニーズ*	緊急性が高い または未充足 ／いずれでもない	
節減される時間	1.3	臨床エンドポイントの代わりに 代替エンドポイントを用いること による時間の節減（臨床エン ドポイントに対する予測治療効 果の不確実性増加との対比にお ける）	意義あり／意 義なし／判定 不能	
節減されるその他の リソース*	1.4	臨床エンドポイントの代わりに 代替エンドポイントを用いること によるリソースの節減（臨床 エンドポイントに対する予測治 療効果の不確実性増加との対比 における）*	意義あり／意 義なし／判定 不能	

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン2：臨床的・生物学的エビデンス				
メタアナリシス**	2.1a	件数	複数／単一／なし	
	2.1b	各メタアナリシスの質（バイアスリスク）	バイアスリスク評価に関するガイダンスを参照†	
	2.1c	サンプルサイズ（各メタアナリシスに組み入れられた試験数）	該当なし†	
	2.1d	代替エンドポイントと臨床エンドポイントとの個人レベル関連*	全メタアナリシスにわたる関連の一致性：高／中／低／該当なし。すべての数値および指標を記載する†	
	2.1e	代替エンドポイントへの治療効果と臨床エンドポイントへの治療効果との試験レベル関連	全メタアナリシスにわたる関連の一致性：高／中／低／該当なし。すべての数値および指標を記載する†	
メタアナリシス：総合結論		代替エンドポイントを支持／代替エンドポイントに反対／不確定／メタアナリシスなし		

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン2：臨床的・生物学的エビデンス				
ランダム化比較試験 (RCT) **	2.2a	件数	複数／単一／なし	
	2.2b	各試験の質（バイアスリスク）	バイアスリスク評価に関するガイダンスを参照†	
	2.2c	サンプルサイズ（各試験の参加者数）	該当なし†	
	2.2d	代替エンドポイントと臨床エンドポイントとの個人レベル関連*	全試験にわたる関連の 一貫性：高／中／低／ 該当なし。すべての数 値および指標を記載す る†。	
	2.2e	代替エンドポイントへの治療効果と臨床エンドポイントへの治療効果との試験レベル関連	全試験にわたる関連の 一貫性：高／中／低／ 該当なし。すべての数 値および指標を記載す る†。	
ランダム化比較試験：総合結論		代替エンドポイントを支持／代替エンドポイントに反対／不確定／ランダム化比較試験なし		

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン2：臨床的・生物学的エビデンス				時間的制約のため、 DSRT 10 当日の議論のScope外とする。
観察研究*	2.3a	件数	複数／単一／なし	
	2.3b	各研究の質（バイアスリスク）	バイアスリスク評価に関するガイダンスを参照†	
	2.3c	サンプルサイズ（各研究の参加者数）	該当なし†	
	2.3d	代替エンドポイントと臨床エンドポイントとの個人レベル関連*	全研究にわたる関連の 一貫性：高／中／低／ 該当なし。すべての数 値および指標を記載す る†。	
	2.3e	代替エンドポイントへの治療効果と臨床エンドポイントへの治療効果との研究レベル関連	全研究にわたる関連の 一貫性：高／中／低／ 該当なし。すべての数 値および指標を記載す る†。	
観察研究：総合結論		代替エンドポイントを支持／代替エンドポイントに反対／不確定／観察研究なし		

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン2：臨床的・生物学的エビデンス				
生物学的妥当性 (Biological plausibility)	2.4	推奨される生物学的妥当性の検討要因†	良好な理解／妥当な理解／限定的な理解	
臨床的妥当性 (Clinical plausibility)	2.5a	因果性：疾患プロセスから臨床エンドポイントに至る経路における代替エンドポイントの位置づけ	因果経路上に直接位置する／複数ある因果経路のうち少なくとも1つの経路上に位置するが、すべてではない／因果経路と良好に相関するが、因果経路上にはない／因果経路は不明または不確実	
	2.5b	推奨される臨床的妥当性の検討要因*†	該当なし	
一般化可能性 (Generalisability)	2.6a	治療クラス	他に複数の治療クラスへ一般化可能／他に1つの治療クラスへ一般化可能／一般化可能性なし	
	2.6b	リスク集団	他に複数のリスク集団へ一般化可能／他に1つのリスク集団へ一般化可能／一般化可能性なし	

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

ドメイン / 項目	項目番号	サブ項目	評価グレード	【 記 入 欄 】
ドメイン3：統計手法（Statistical Methods）				
個人レベル解析*	3.1	手法の選択	該当なし†	
試験レベル解析	3.2a	R二乗（R-squared）	該当なし†	
	3.2b	Surrogate Threshold Effect*	該当なし†	
	3.2c	手法の選択	該当なし†	
閾値設定（Thresholding）	3.3	R二乗（R-squared）*	該当なし†	
付与すべき総合グレード：当該代替エンドポイントは、(i) 医薬品または医療機器の開発、および／または (ii) 当該医薬品または医療機器の医療技術評価（HTA）への使用として受容可能か？ 状況依存性（context）の重要性および柔軟性に鑑み、定型的な総合グレードの組合せは提示しない。ただし上記の問いに対するいかなる回答も、ドメイン0、1、2および3のそれぞれから得られたエビデンスの包括的統合により裏付けられなければならない。				

※翻訳は今回のデータサイエンスラウンドテーブル用にファシリテーターが行ったため、正式なものではありません

総合判定（Overall Grade）

問い：当該代替エンドポイントは、(i) 医薬品または医療機器の開発、および／または (ii) 医療技術評価（HTA）への使用として受容可能か？

注記：定型的な総合グレードの組合せは提示しない。ドメイン0、1、2および3のそれぞれから得られたエビデンスの包括的統合により裏付けること。

【記入欄】 総合判定・根拠・コメント（ドメイン0～3の統合）

まとめ

- REASSurEにより、以下のポイントの整理につなげることを可能とする
 - 時間短縮と不確実性増加のバランス
 - 生物学的妥当性
 - RCTまたはRCTメタアナリシスによるエビデンス
 - 臨床的妥当性
 - 一般化可能性（Generalisability）
- 各項目を評価するには、生物統計学、臨床医学、規制科学、医療技術評価、など複数分野の専門知識が必要
 - 多面的な視点から候補代替エンドポイントを評価したうえで、臨床試験の主要評価項目として採用すべきかを判断することを推奨



グループディスカッション & 最終発表

グループディスカッション概要

● 代替EP妥当性評価（フレームワーク記入） → 推奨投票

- 事前情報および現在のエビデンスに基づき、フレームワークの各ドメインを埋め、その内容に基づき、代替EPの妥当性評価を行う。
- 記入した情報に基づき、以下の問いについて投票し、各自その根拠を述べる

Q. あなたはこのエンドポイントを代替エンドポイントとして推奨するか
(回答はFDAの規制フレームワークに準拠し、以下の3種類から選択)

推奨

条件付き推奨

推奨しない

- 意見の相違があった点、または限界に関する意見があれば併せて記述する

グループディスカッション概要

● 各グループの事例

代替エンドポイント	疾患
1. eGFR Slope	慢性腎臓病
2. MRD（微小残存病変）	多発性骨髄腫
3. NfL濃度	SOD1遺伝子変異を伴う筋萎縮性側索硬化症（SOD1-ALS）

代替エンドポイント評価

グループ1 : 慢性腎臓病, eGFR slope

対象疾患・評価エンドポイント

1 / 5

1 対象疾患・治療領域

[慢性腎臓病（CKD）]

2

評価対象の代替エンドポイント

[eGFR slope]



3

対応する真のエンドポイント

[ESKD等（心血管イベントは除く）の複合エンドポイント]

4

なぜ代替EPが必要とされているか： [真のEPは評価に時間がかかる、代替EPで3~8年程度削減可能。]

結論：このEPを代替EPとして推奨するか

2 / 5

推奨する

条件付きで推奨

推奨しない

3 year Total slope：推奨する

✓ slope=-0.35 (-0.42, -0.29)、 $R^2 = 0.97$ (0.82, 1.00)

2 year Total slope：不確定

✓ slope=-0.27 (-0.33, -0.21)、 $R^2 = 0.87$ (0.64, 0.97)

Chronic solope：条件付きで推奨（薬剤特性によって判断、薬剤特性を考慮したメタ解析が必要、副次や感度分析等で使用）

✓ slope=-0.33 (-0.46, -0.20)、 $R^2 = 0.55$ (0.25, 0.77)

意見の相違・評価の限界

4 / 5

意見の相違： [3年総スロープ vs 2年総スロープ]

支持する理由： R^2 が許容範囲

不確定な理由： 信用区間の下限が許容以下

評価の限界： [Chronic slopeの評価]

メタ解析に複数の介入が含まれており、initial dipの有無を考慮することが必要かもしれない。

代替エンドポイント評価

グループ2: 多発性骨髄腫, MRD (微小残存病変)

対象疾患・評価エンドポイント

1 / 5

1 対象疾患・治療領域

多発性骨髄腫

- 新規診断例（NDMM：移植適格・不適格両者）
- 再発/難治症例（RRMM）

2

評価対象の代替エンドポイント

評価期間中（9か月 or 12ヵ月）少なくとも1回でもMRD（微小残存病変）陰性



3

対応する真のエンドポイント

OS

4

なぜ代替EPが必要とされているか：真のEPは評価に時間がかかる（OSだと10年以上かかる可能性も）

結論：このEPを代替EPとして推奨するか

2 / 5

推奨する

条件付きで推奨

推奨しない

5名

2名

- 希少疾患/慢性疾患であり、根治治療が無い（Unmet medical needsが高い）
- 生物学的妥当性が合理的である
 - 腫瘍細胞が $10^{(-5)}$ 以上のレベルで検出限界以下
 - MRDが真のエンドポイントへの因果経路上にある
 - 骨髄腫細胞の残存 -> クローンの増殖 -> 疾患進行 -> 死亡
- 個人レベルのメタ解析の結果が高い
 - 事前規定された基準（Global Odds Ratio > 3）を満たす

意見の相違：

代替EPとして承認しないの意見

- サロゲートパラドックス（MRDがOSの適切な代替EPとなっていないことを考慮）
- 試験レベルの R^2 が事前規定された基準0.5を超えたが、 R^2 が0.6前後であり、エビデンスが高いとは言えないのではないかという意見あり。

評価の限界：

- 試験レベル相関のエビデンスが高くなく、RRMMなど疾患においては試験数が限られた。
 - 追加のエビデンスが必要
- 本エンドポイントはFirst lineでの代替EPとしての評価であり、他の治療オプションが存在する現状も考慮する必要がある
- MRDの評価時期(9か月または12か月)は使用する薬剤によって適切な時期が異なる可能性があり、一律に判断できない

代替エンドポイント評価

グループ3 : *SOD1-ALS, NfL* 濃度

対象疾患・評価エンドポイント

1 / 5

1 対象疾患・治療領域

SOD1-ALS

2

評価対象の代替エンドポイント
血漿中NfL濃度



3

対応する真のエンドポイント
ALSFRS-R

4 なぜ代替EPが必要とされているか：治療の選択肢が限られており、緊急性を要する
ALSFRS-Rスコアは変動が小さく、差を検出するのに長い時間が必要

結論：このEPを代替EPとして推奨するか

2 / 5

推奨しない

- ・ 疾患の希少性、開発上のメリットの観点からは代替EPを使う動機がある。
- ・ 一方、単試験での個人アウトカムの相関が低かった（スピアマン相関係数 $r=0.07$ ）
- ・ また、単試験のみで、試験間相関を評価することができない

意見の相違

- Clinical Endpoint としてALS-SFRS-RとOSの関係は？OSとの関連の程度を示すものが必要ではないか
- どの時点の情報が必要か？より長い時点では効果がでるのでは？
- 節約するリソースとして、測定のリソースは増えない(髄液中だと変わる？)
- 効果量が大きいのであればサンプルサイズが減る可能性がある？

評価の限界

- 同じ検証的試験を行うのは現実的ではない。
- 試験後の長期的な追跡やレジストリのデータが必要

希少疾患において、完璧な追跡データが望めない以上、
今あるデータで代替性を判断することはやむを得ないか？
(複数の試験を要求することは現実的か？)