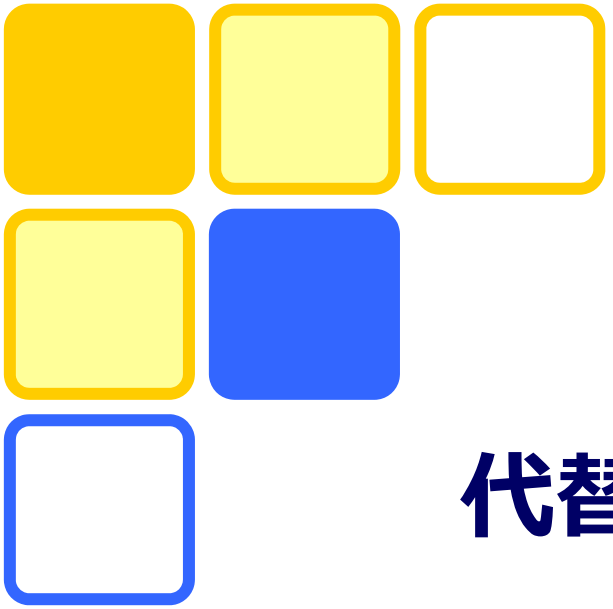
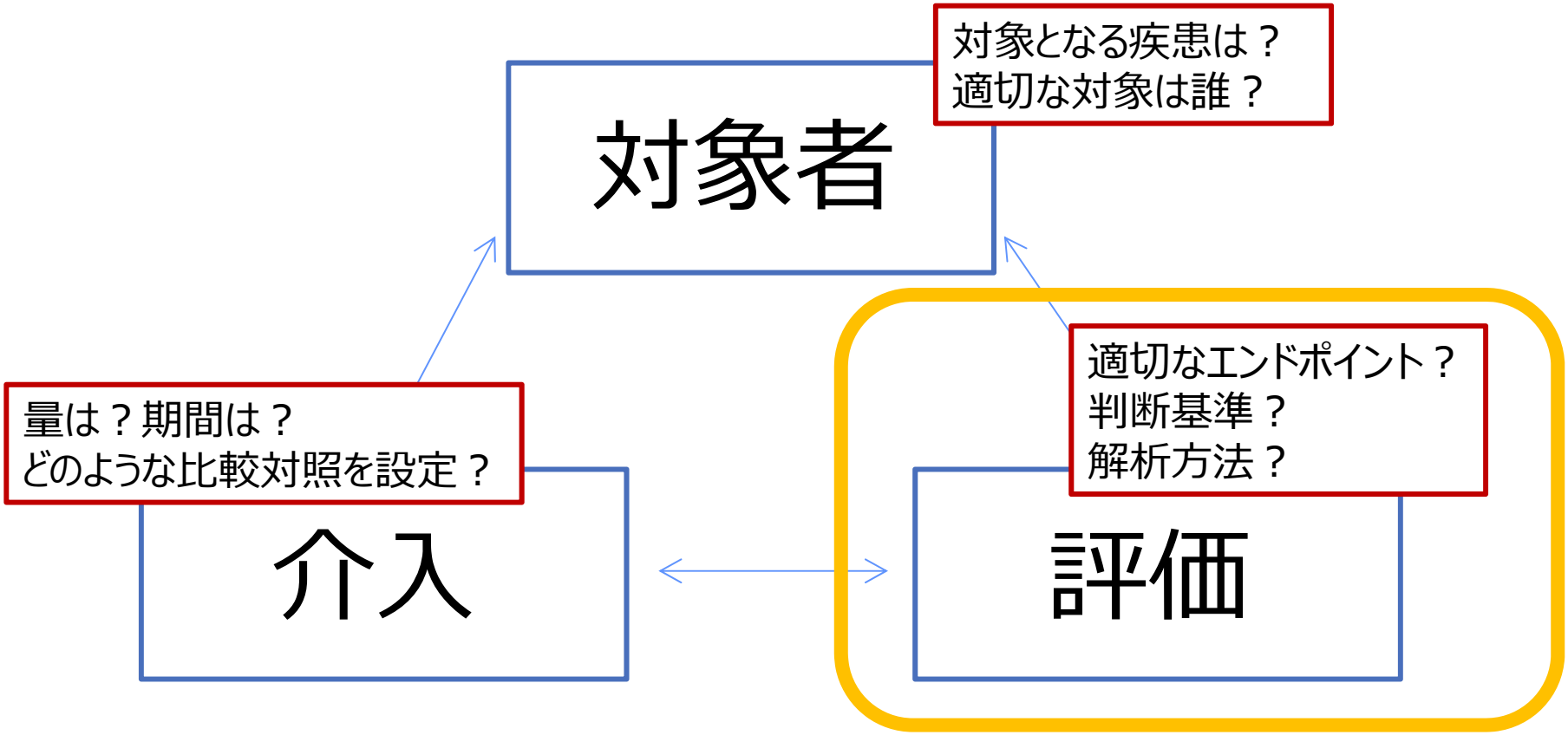


代替エンドポイントの評価と利用

東京大学大学院　情報学環
/医学系研究科　生物統計学分野（兼担）
准教授　大庭　幸治
oba@epistat.m.u-tokyo.ac.jp

営利目的でのご利用や無断転載はご遠慮ください

臨床試験デザインのエッセンス



営利目的でのご利用や無断転載はご遠慮ください

臨床試験におけるエンドポイントの重要性

- 治療法等の比較をする事を主な目的とした臨床試験では、その治療の影響を何で評価するかを事前に決定することが必要

- エンドポイント（評価項目）
  - 臨床試験の目的に関連する仮説を検証する上で臨床的に意味があり、客観的に評価できる観察・検査項目、またはそれらの合成指標
  - 目的に応じた適切なエンドポイントの選択は、臨床試験におけるデザインの中核となる

営利目的でのご利用や無断転載はご遠慮ください

良いエンドポイント?

- 良いエンドポイントの3条件
  - 治療効果を検出する感度が高いこと
    - 反応性があり評価の測定誤差が小さいこと
  - 臨床的に適切であること
  - 測定が簡便であること
    - 評価自体の分かりやすさ
    - 評価の際のバイアスの入りやすさ
    - 測定にかかる時間

これらの条件はトレードオフとなることが多い

Fleming TR, Ann Int Med 1996

営利目的でのご利用や無断転載はご遠慮ください

癌領域で用いられるエンドポイント

- あくまで個人的な印象での分類ですが…

エンドポイント	感度	適切	簡便	バイアス	時間
全生存期間 (OS)	×	○	○	○	×
無増悪生存期間 (PFS)	○	△	×	×	△
奏効率 (RR)	○	×	△	×	○
Quality of Life (QOL)	×	○	×	×	△

営利目的でのご利用や無断転載はご遠慮ください

優先されるのは臨床的適切性

- 臨床エンドポイント
  - 患者がどのように感じ (feels)、あるいは機能し、(functions) 生存しているか (survives) を反映する特性あるいは変数

    Biomarkers Definition Working Group, Clin Pharmacol Ther 2001
- 患者にもたらされる真の利益を反映する、という意味で、「真のエンドポイント」とも呼ばれる

営利目的でのご利用や無断転載はご遠慮ください

代替エンドポイント

- Surrogate endpoint (SE)
  - 臨床エンドポイントの「代わり」になる事が意図されたバイオマーカー
  - 疫学・治療学・病態生理学または他の科学的根拠に基づき、臨床上の便益・害の有無を予測することが期待されるもの

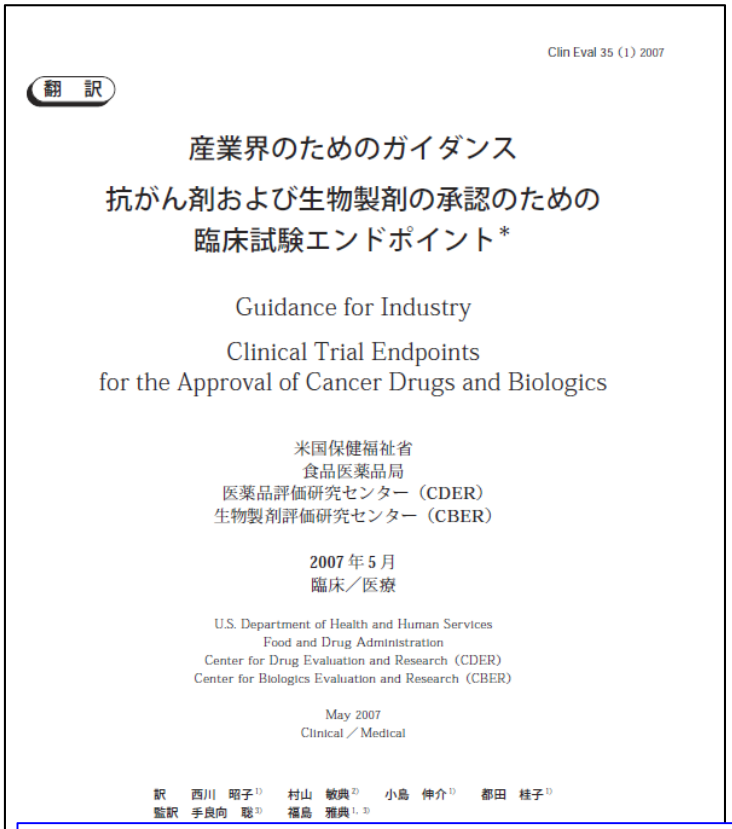
    Biomarkers Definition Working Group, Clin Pharmacol Ther 2001

- 臨床的な検証のレベルにより、validated SE, reasonably likely SE, candidate SE に分類

  FDA-NIH Biomarker Working Group, BEST (Biomarker, EndpointS, and other Tools) Resource [Internet], 2018

営利目的でのご利用や無断転載はご遠慮ください

癌エンドポイントに関するFDAガイダンス



Clin Eval 35 (1) 2007

翻　訳

産業界のためのガイダンス

抗がん剤および生物製剤の承認のための
臨床試験エンドポイント*

Guidance for Industry

Clinical Trial Endpoints
for the Approval of Cancer Drugs and Biologics

米国保健福祉省
食品医薬品局
医薬品評価研究センター（CDER）
生物製剤評価研究センター（CBER）

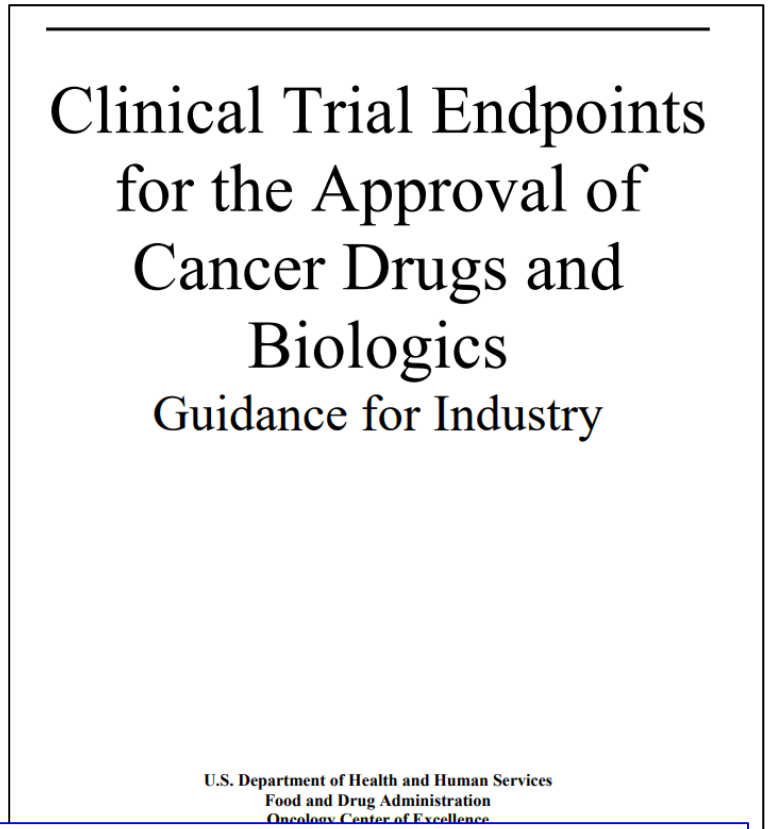
2007年5月
臨床／医療

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2007
Clinical／Medical

訳　西川　昭子[1)]　村山　敏典[2)]　小島　伸介[1)]　都田　桂子[1)]
監訳　手良向　聡[3)]　福島　雅典[1, 3)]

Updated



Clinical Trial Endpoints
for the Approval of
Cancer Drugs and
Biologics

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Oncology Center of Excellence

Surrogate endpoints that have undergone this extensive testing are called validated surrogate endpoints and these are accepted by the FDA as evidence of benefit. Between 2010 and 2012, the FDA approved 45 percent of new drugs based on a surrogate endpoint.

営利目的でのご利用や無断転載はご遠慮ください

本日のお話

- 代替エンドポイントの性能評価とは？
  - 過去の失敗事例の紹介
  - 代替エンドポイントの性能評価のための統計的基準
- 代替エンドポイントを評価した事例紹介
  - GASTRIC研究
- 最近の話題とまとめ

営利目的でのご利用や無断転載はご遠慮ください

本日のお話

- 代替エンドポイントの性能評価とは？
  - 過去の失敗事例の紹介
  - 代替エンドポイントの性能評価のための統計的基準
- 代替エンドポイントを評価した事例紹介
  - GASTRIC研究
- 最近の話題とまとめ

営利目的でのご利用や無断転載はご遠慮ください

代替エンドポイントの議論のきっかけ

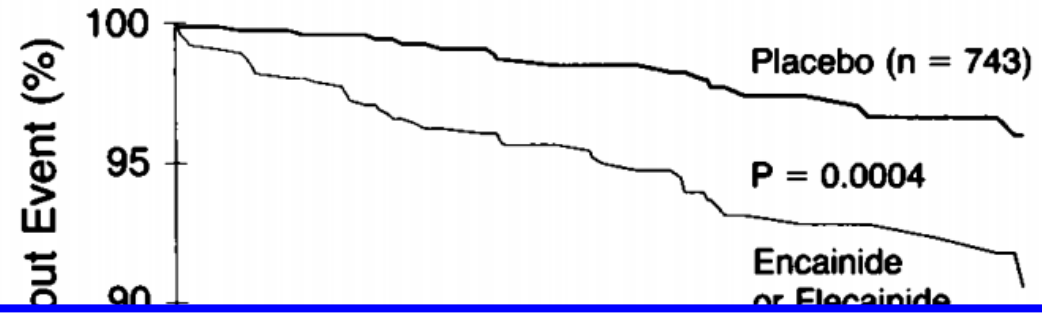
- **CAST**: cardiac arrhythmia suppression trial
- 試験背景
  - 心筋梗塞慢性期に心室性期外収縮が多発すると予後が悪い
  - 抗不整脈薬で心室性期外収縮を抑制できる
  - CASTの実施
- 目的
  - 抗不整脈薬が、心筋梗塞後の心停止や突然死を減少させる、という仮説の検証
    - 多くの人は仮説が証明されるだろうと考えた

The Cardiac Arrhythmia Suppression Trial, NEJM 1989

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

CAST試験の結果



- 結果：新たに狭心症、心筋梗塞、心不全の増加は群間で違いなかったものの、10ヶ月後突然死発症割合が抗不整脈薬群で増加
  - 1989年4月、試験の途中中止

Placebo	743	632	516	412	292	201
Active drug	755	631	507	392	286	198

Figure 1. Actuarial Probabilities of Freedom from Death or Cardiac Arrest Due to Arrhythmia in 1498 Patients Receiving Encainide or Flecainide or Corresponding Placebo.

The Cardiac Arrhythmia Suppression Trial, NEJM 1991

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

代替エンドポイントの利用による失敗例

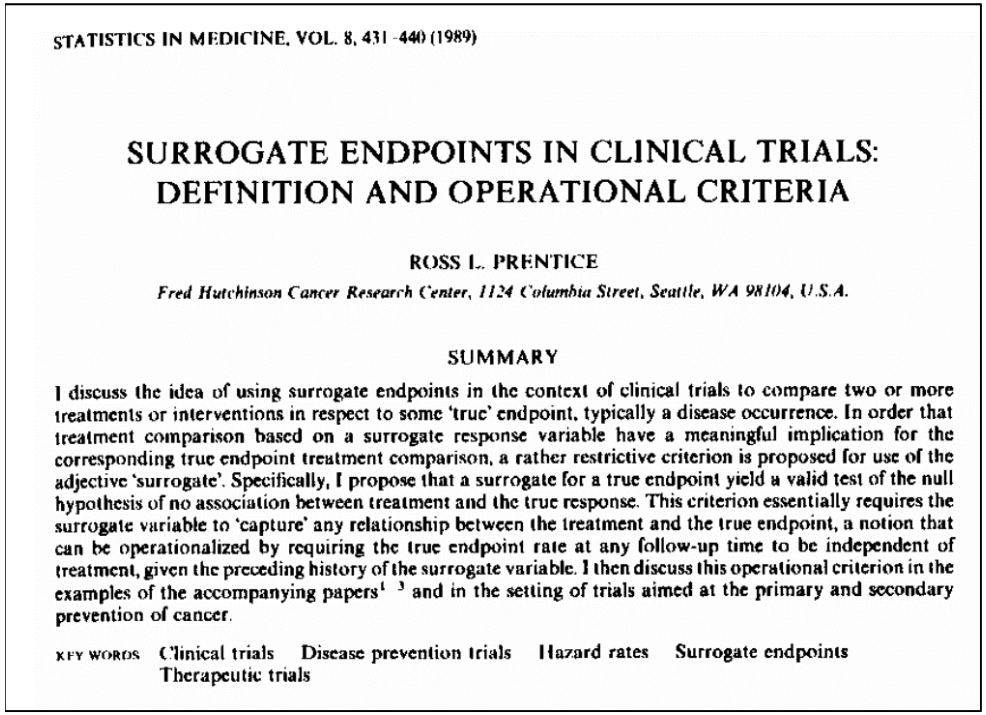
疾患	治療	効果		試験または解析
		代替エンドポイント	臨床エンドポイント	
心筋梗塞後	抗不整脈薬	心室性不整脈の減少	急死の増加	CAST
心房細動	キニジン (Quinidine)	洞調律の維持 (1年)	死亡率の増加	メタ分析
うっ血性心不全	ミルリノン/フロセキナン/エポプロステノール	改善された心拍出量/増加した運動耐性	死亡率の増加	PROMISE PROFILE FIRST
閉経後女性の心疾患	ホルモン補充療法	血清リポタンパク質レベルに対する好ましい効果	増加した冠状動脈性心疾患/心筋梗塞	HERS WHIT PEPI
心臓病	コレステロール低下剤	コレステロール値の低下	死亡率の増加	WHO Gordon meta-analysis
骨粗鬆症	ナトリウムフッ化物	骨密度の増加	骨折発生率の増加	
HIV	ジドブジン (Zidovudine)	CD4 +細胞数の低下	日和見感染を減少の失敗	British-French Concord Trial
正常血圧患者	緑内障の管理	眼内圧を下げる	長期視野損失の治療効果の欠損	

Shi Q et al., Int J Clin Oncol 2009

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

代替エンドポイントの代替性評価



STATISTICS IN MEDICINE, VOL. 8, 431–440 (1989)

SURROGATE ENDPOINTS IN CLINICAL TRIALS: DEFINITION AND OPERATIONAL CRITERIA

ROSS L. PRENTICE

Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104, U.S.A.

SUMMARY

I discuss the idea of using surrogate endpoints in the context of clinical trials to compare two or more treatments or interventions in respect to some 'true' endpoint, typically a disease occurrence. In order that treatment comparison based on a surrogate response variable have a meaningful implication for the corresponding true endpoint treatment comparison, a rather restrictive criterion is proposed for use of the adjective 'surrogate'. Specifically, I propose that a surrogate for a true endpoint yield a valid test of the null hypothesis of no association between treatment and the true response. This criterion essentially requires the surrogate variable to 'capture' any relationship between the treatment and the true endpoint, a notion that can be operationalized by requiring the true endpoint rate at any follow-up time to be independent of treatment, given the preceding history of the surrogate variable. I then discuss this operational criterion in the examples of the accompanying papers[1–3] and in the setting of trials aimed at the primary and secondary prevention of cancer.

KEY WORDS Clinical trials Disease prevention trials Hazard rates Surrogate endpoints Therapeutic trials

Prentice RL., Statist Med 1989

Surrogate End Points in Clinical Trials: Are We Being Misled?

Thomas R. Fleming, PhD, and David L. DeMets, PhD

Phase 3 clinical trials, which evaluate the effect that new interventions have on the clinical outcomes of particular relevance to the patient (such as death, loss of vision, or other major symptomatic event), often require many participants to be followed for a long time. There has recently been great interest in using surrogate end points, such as tumor shrinkage or changes in cholesterol level, blood pressure, CD4 cell count, or other laboratory measures, to reduce the cost and duration of clinical trials. In theory, for a surrogate end point to be an effective substitute for the clinical outcome, effects of the intervention on the surrogate must reliably predict the overall effect on the clinical outcome. In practice, this requirement frequently fails. Among several explanations for this failure is the possibility that the disease process could affect the clinical outcome through several causal pathways that are not mediated through the surrogate, with the intervention's effect on these pathways differing from its effect on the surrogate. Even more likely, the intervention might also affect the clinical outcome by unintended, unanticipated, and unrecognized mechanisms of action that operate independently of the disease process. We use examples from several disease areas to illustrate how surrogate end points have been misleading about the actual effects that treatments have on the health of patients.

Surrogate end points can be useful in phase 2 screening trials for identifying whether a new intervention is biologically active and for guiding decisions about whether the intervention is promising enough to justify a large definitive trial with clinically meaningful outcomes. In definitive phase 3 trials, except for rare circumstances in which the validity of the surrogate end point has already been rigorously established, the primary end point should be the true clinical outcome.

Clinical trials are the standard scientific method for evaluating a new biological agent, drug, device, or procedure for the prevention or treatment of disease in humans. The phase 3 trial is designed to evaluate a new agent's clinical benefit and possible side effects; as such, it is considered to be the definitive test of the agent's usefulness (1–3). For phase 3 trials, the primary end point should be a clinical event relevant to the patient, that is, the event of which the patient is aware and wants to avoid. Examples are death, loss of vision, symptomatic events of the acquired immunodeficiency syndrome (AIDS), the need for ventilatory support, and other events causing a reduction in quality of life. Trials with these clinical outcomes often have a long duration and are expensive. As a consequence, there has recently been great interest in the development of alternative outcomes, or surrogate end points, to reduce the cost and shorten the duration of phase 3 trials (4–17). As defined by Temple (13),

> a surrogate endpoint of a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.

Examples of surrogate end points are increased CD4 cell counts or decreased viral load measures for trials of therapy for human immunodeficiency virus (HIV) infection or AIDS, suppression of ventricular arrhythmias or reduction in cholesterol level

Fleming TR et al., Ann Int Med 1996

営利目的でのご利用や無断転載はご遠慮ください

相関関係≠代替性

- A correlate does not make a surrogate
- 単純に、代替エンドポイントについて（期外収縮）
  - 真のエンドポイントと相関している（突然死が多い）
  - 代替エンドポイントに効果がある（期外収縮抑制）
  - ~~適切な代替エンドポイント~~
  - ~~臨床試験で、代替エンドポイントを利用可能~~
- 代替エンドポイントと真のエンドポイント間の相関だけでは、代替エンドポイントの妥当性の条件としては不十分

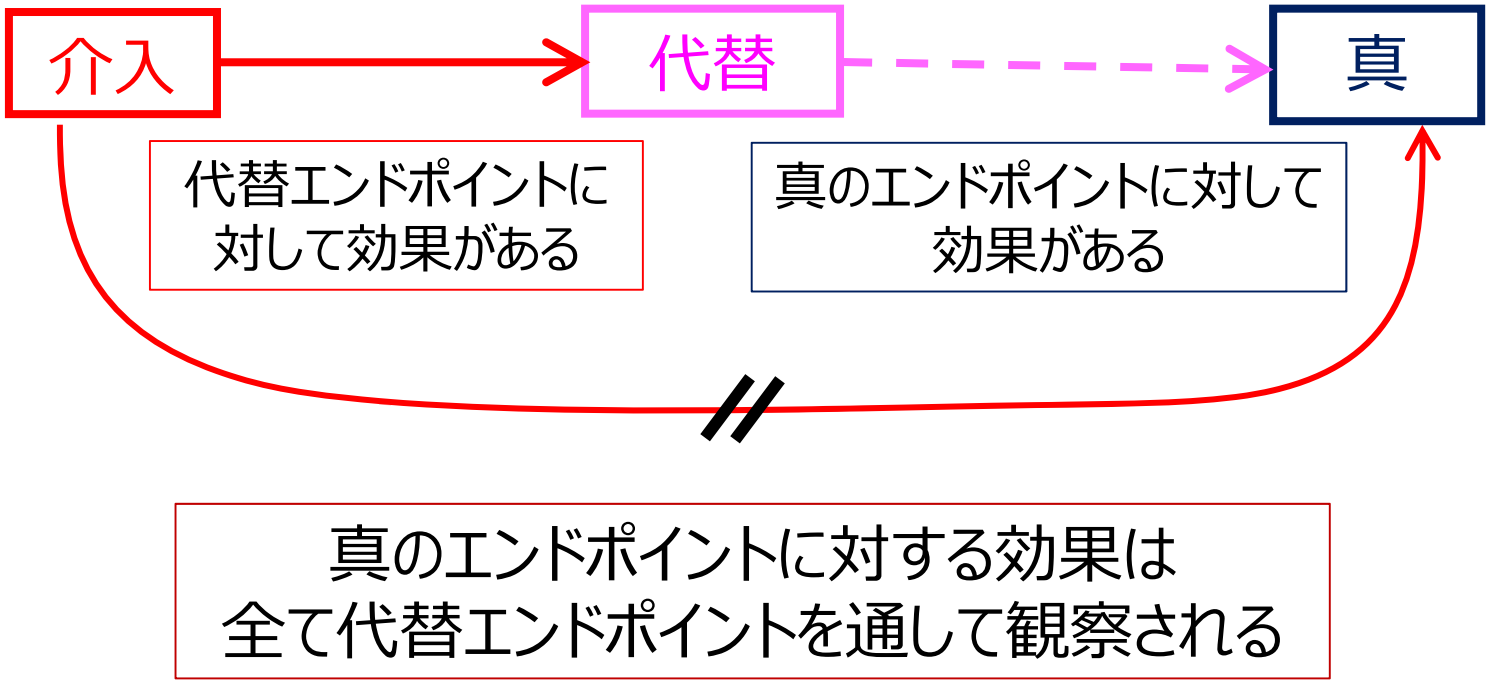
Fleming TR et al., Ann Int Med 1996

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

理想的な代替エンドポイント

- Prenticeの基準（代表的な統計的基準）

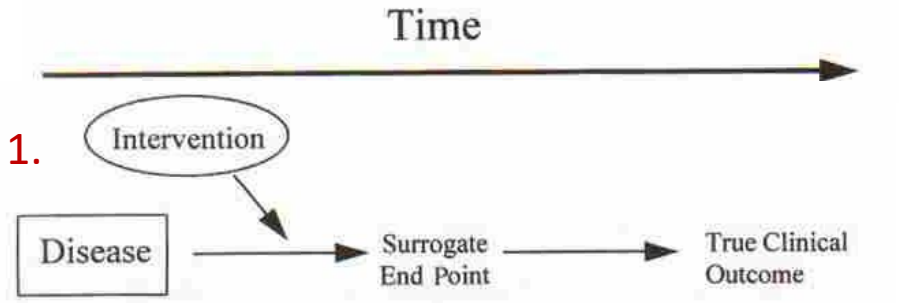


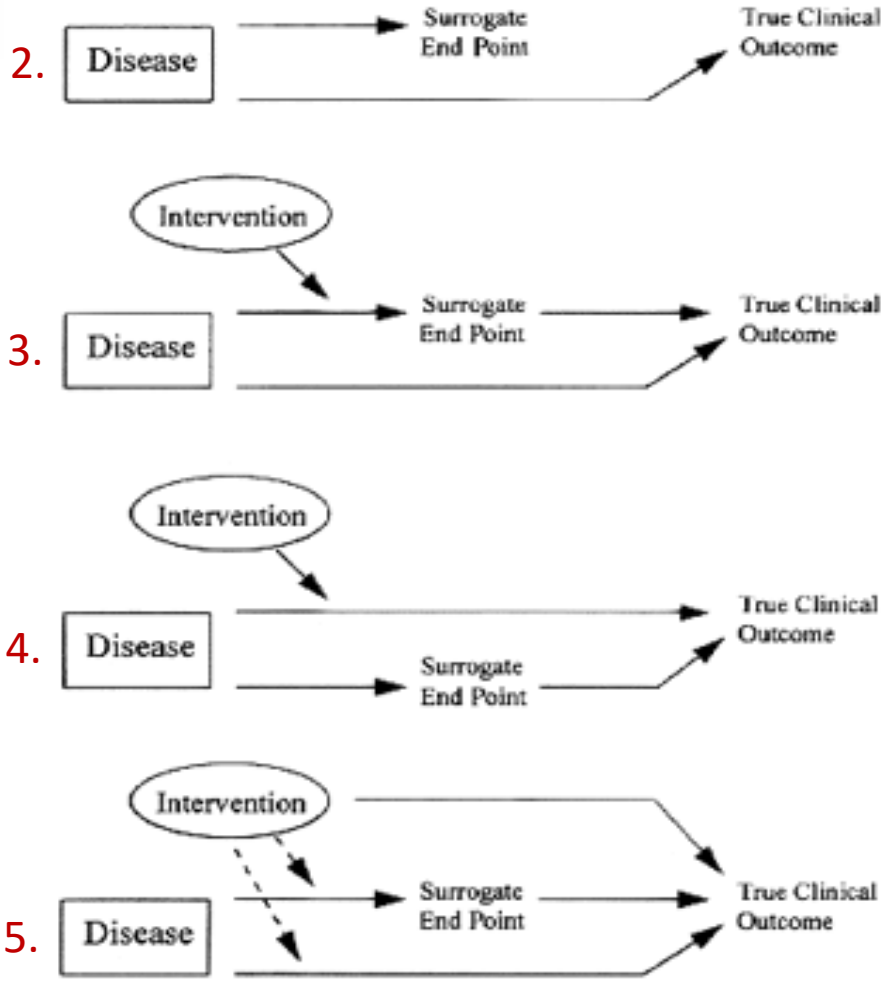
Prentice RL, Stat Med 1989

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

複雑な疾患発症過程と介入





1~5はいずれも代替と真のエンドポイント間に相関が生じる状況

1. 理想的な状況
2. 疾患→代替の過程と疾患→真のエンドポイントの過程が違う
3. 複数の経路のうち，代替を経由する発症過程以外の経路がある
4. 代替が介入の影響を受けない、もしくは感度が悪い
5. 介入が、疾患から結果への発症過程以外の経路で影響を与える

Fleming TR et al., Ann Int Med 1996

営利目的でのご利用や無断転載はご遠慮ください

代替ポイントの妥当性評価

- Prenticeの基準は厳しすぎる
- 理想的には、疾患発症や介入メカニズムに関する因果経路を十分に把握しておくことが重要
  - 疾患発症経路は非常に複雑であり、完全な把握・評価は、実際には非常に困難

営利目的でのご利用や無断転載はご遠慮ください

国際的な統計ガイドラインでの基準

- ICH E9ガイドラインでは、代替性の証拠の強さは以下の3つに依存している
- 代替変数と臨床的結果の関連の生物学的合理性
  - 重要だが限界あり
- 代替変数が臨床的結果の予後を予測する上で有益であると疫学研究によって示されていること
- 試験治療の代替変数に対する効果が臨床的効果と対応しているという臨床試験の結果
  - ➡ メタアナリシスアプローチの基づく評価

ICH E9ガイドライン「臨床試験のための統計的原則」について, 1998

営利目的でのご利用や無断転載はご遠慮ください

メタアナリシス

From: *Medical Research Library of Brooklyn - http://library.downstate.edu/ebm/2100.htm*

営利目的でのご利用や無断転載はご遠慮ください

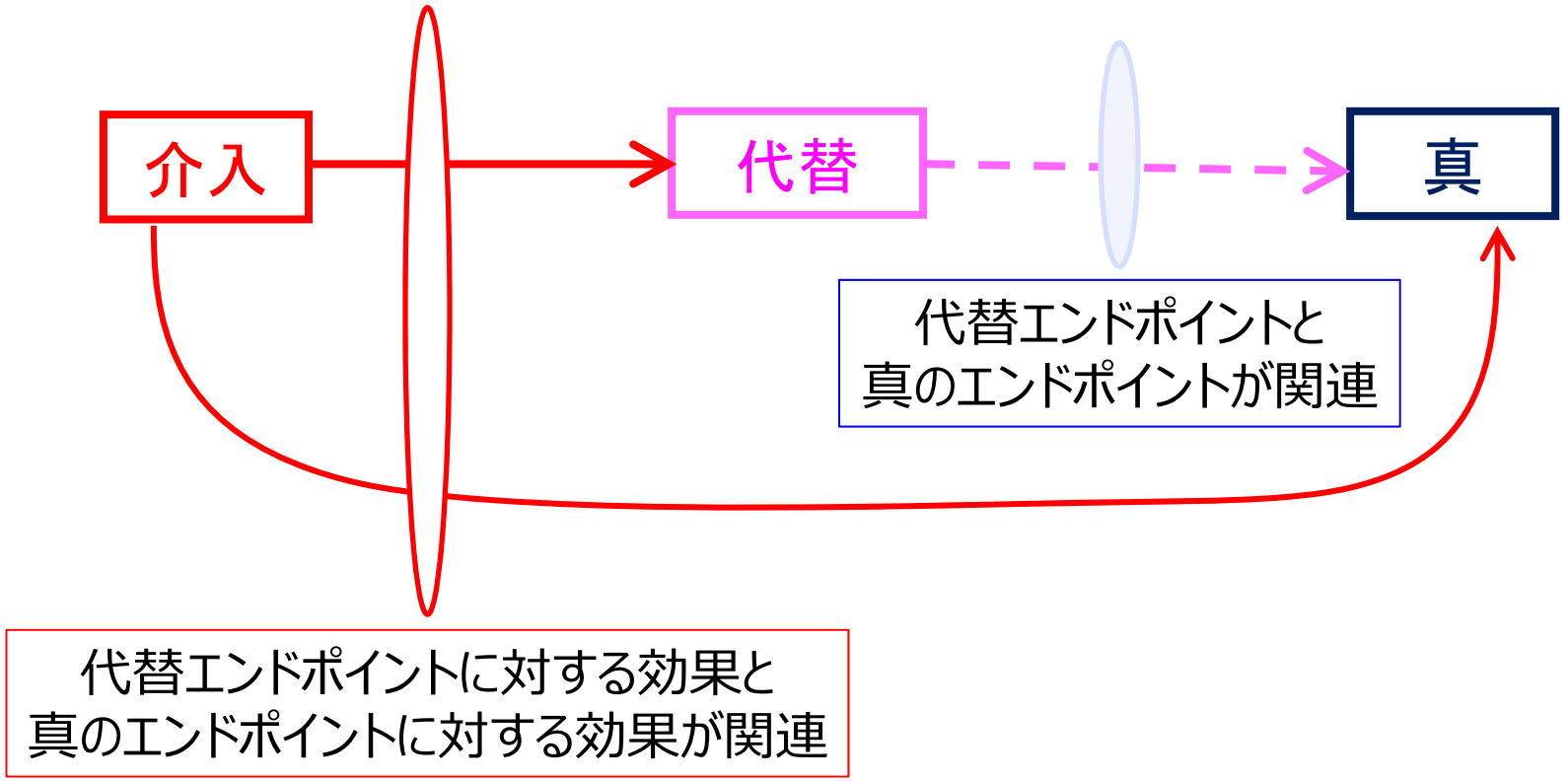
メタアナリシスの役割

- 多数のデータに基づく治療効果の検出
  - 推定精度、検出力の向上
  - 一般化可能性
- 異質性の検討
  - サブグループ解析
  - 予後規定因子に関する解析
- 代替エンドポイントの妥当性評価
  - Individual-Patient-Data (IPD) の利用

営利目的でのご利用や無断転載はご遠慮ください

メタアナリシスによる妥当性評価

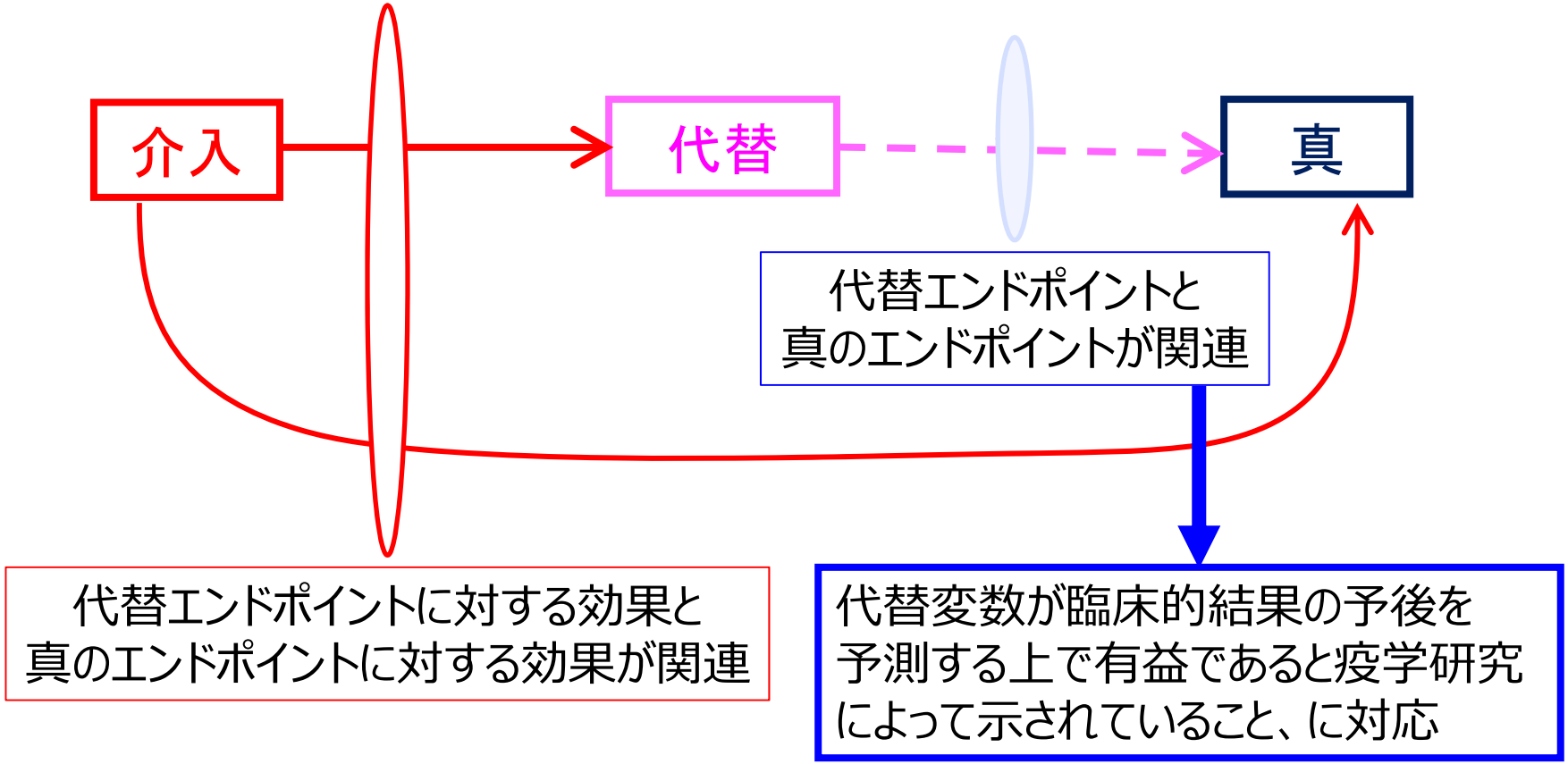
- 介入効果が完全に代替エンドポイントを経由する事は稀



営利目的でのご利用や無断転載はご遠慮ください

メタアナリシスによる妥当性評価

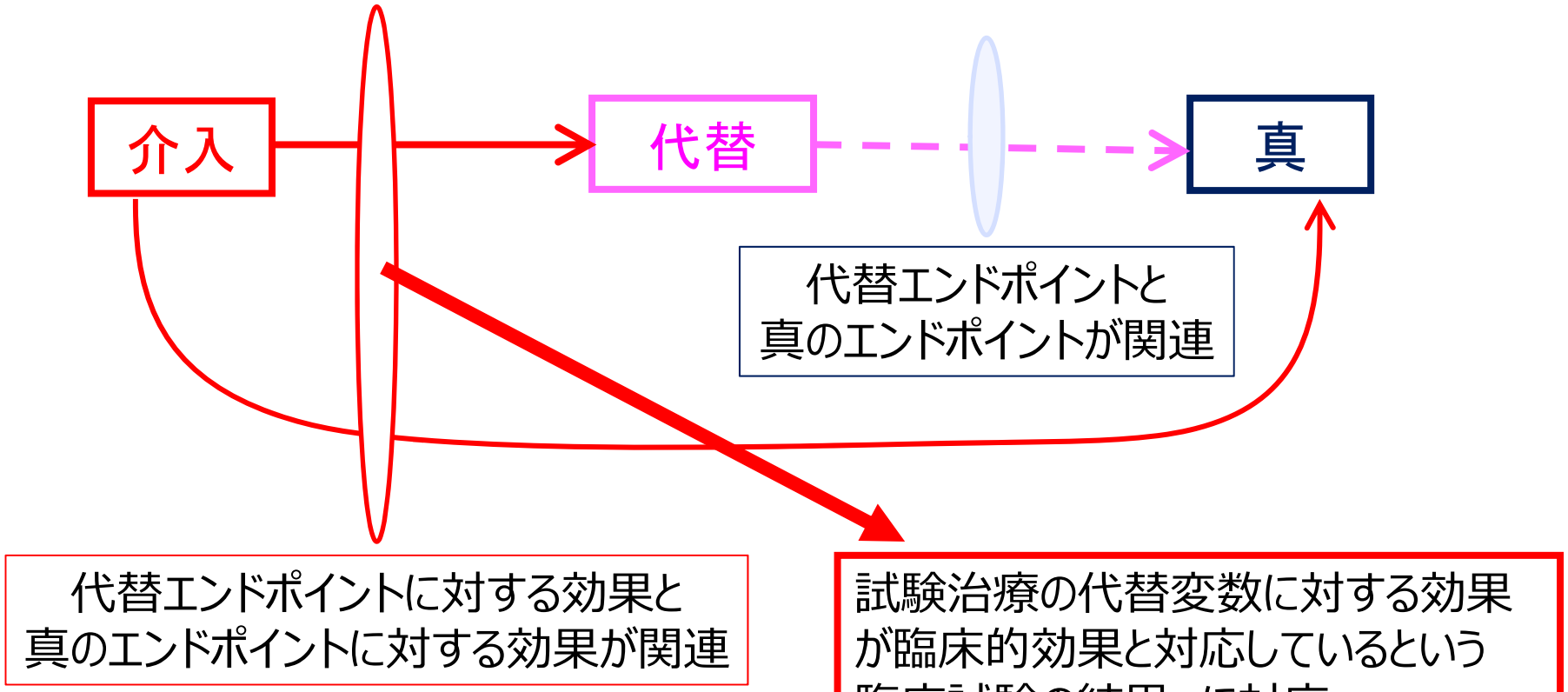
- 介入効果が完全に代替エンドポイントを経由する事は稀



営利目的でのご利用や無断転載はご遠慮ください

メタアナリシスによる妥当性評価

- 介入効果が完全に代替エンドポイントを経由する事は稀



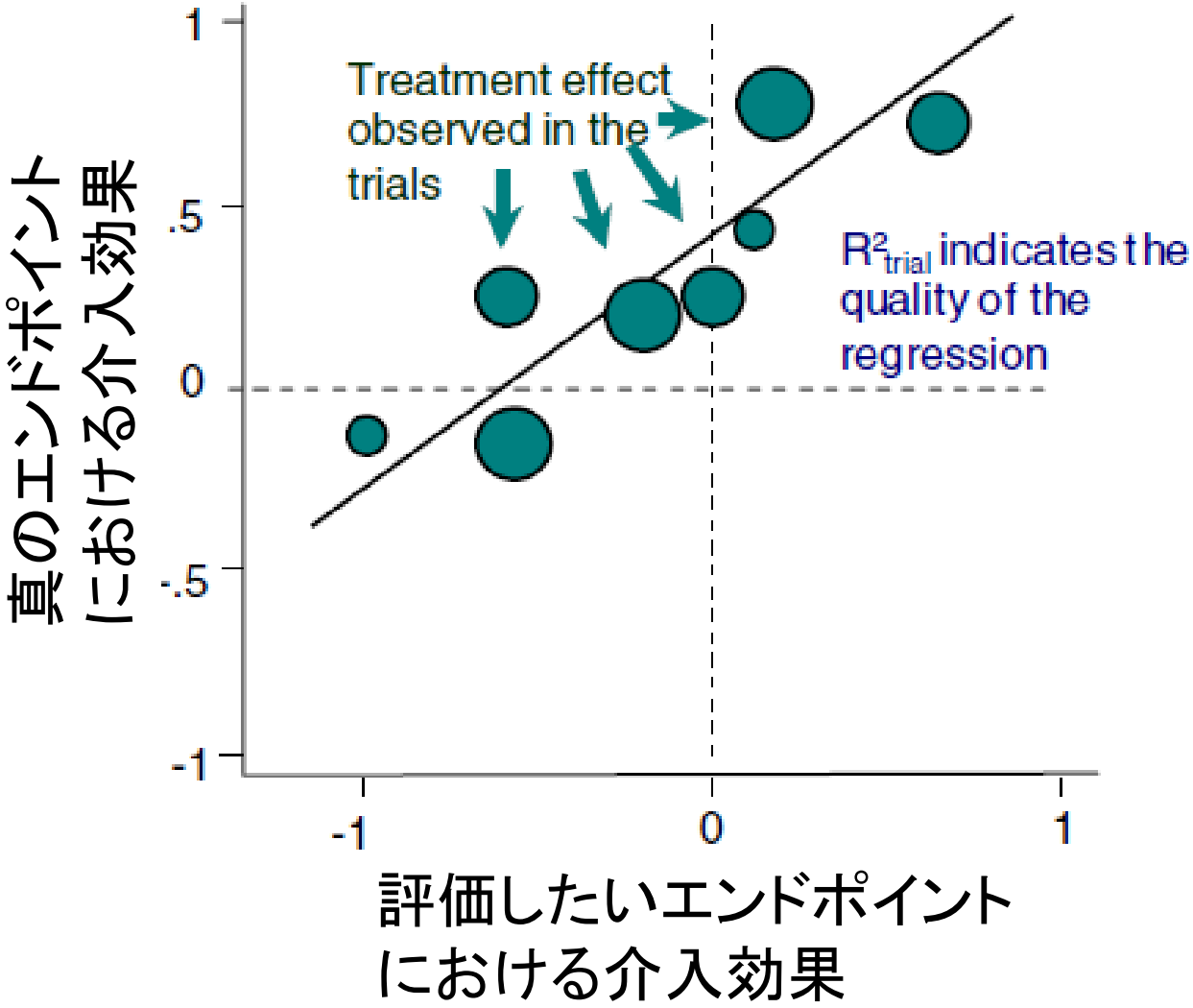
営利目的でのご利用や無断転載はご遠慮ください

メタアナリシスを用いた2つの評価指標

- 個人レベルの代替性 (R^2_{indiv})
  - 真のエンドポイントと代替エンドポイントの間の決定係数（または、相関係数）
  - 実際の患者さんの予後を予測する上で有用
- 試験レベルの代替性 (R^2_{trial})
  - 真のエンドポイントにおける治療効果と代替エンドポイントにおける治療効果の間の決定係数(または相関係数)
  - 新しい試験における真のエンドポイントでの治療効果の予測性を評価する上で有用

営利目的でのご利用や無断転載はご遠慮ください

複数の試験を用いた効果の予測

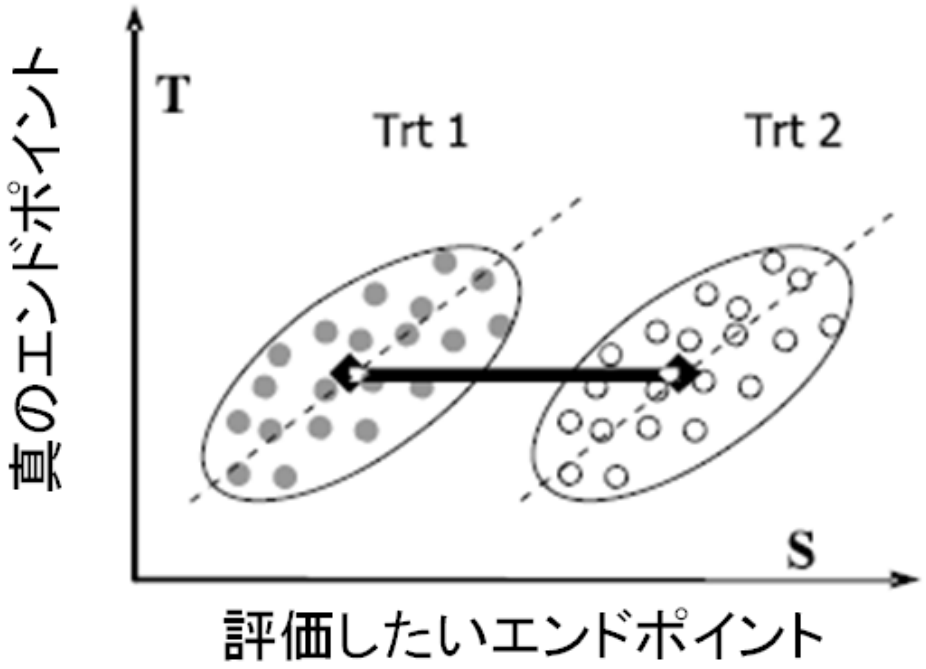


営利目的でのご利用や無断転載はご遠慮ください

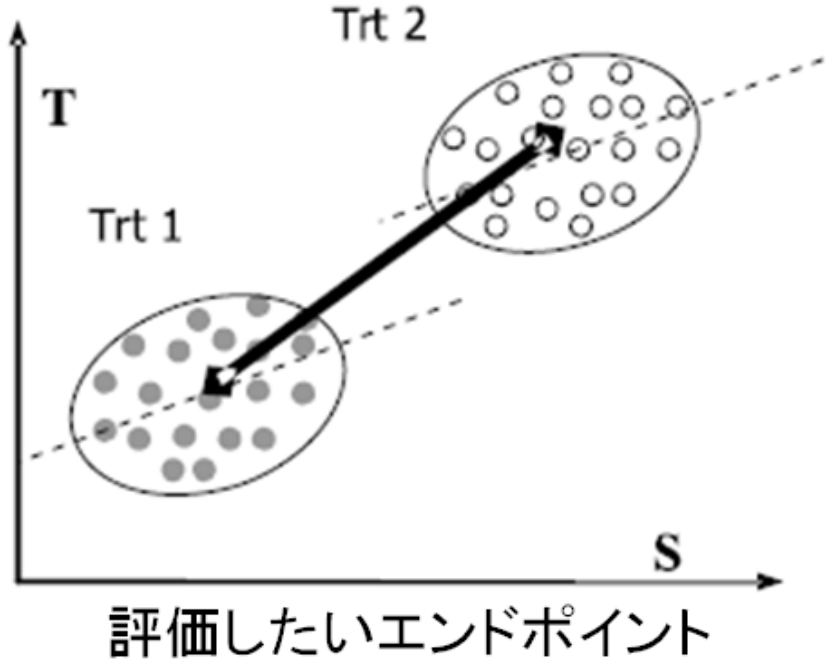
個人レベルと試験レベルの代替性は必ずしも一貫しない

- 両者が十分に高ければ、代替性の証拠は強いが…

個人レベルの代替性は高い
試験レベルの代替性は低い



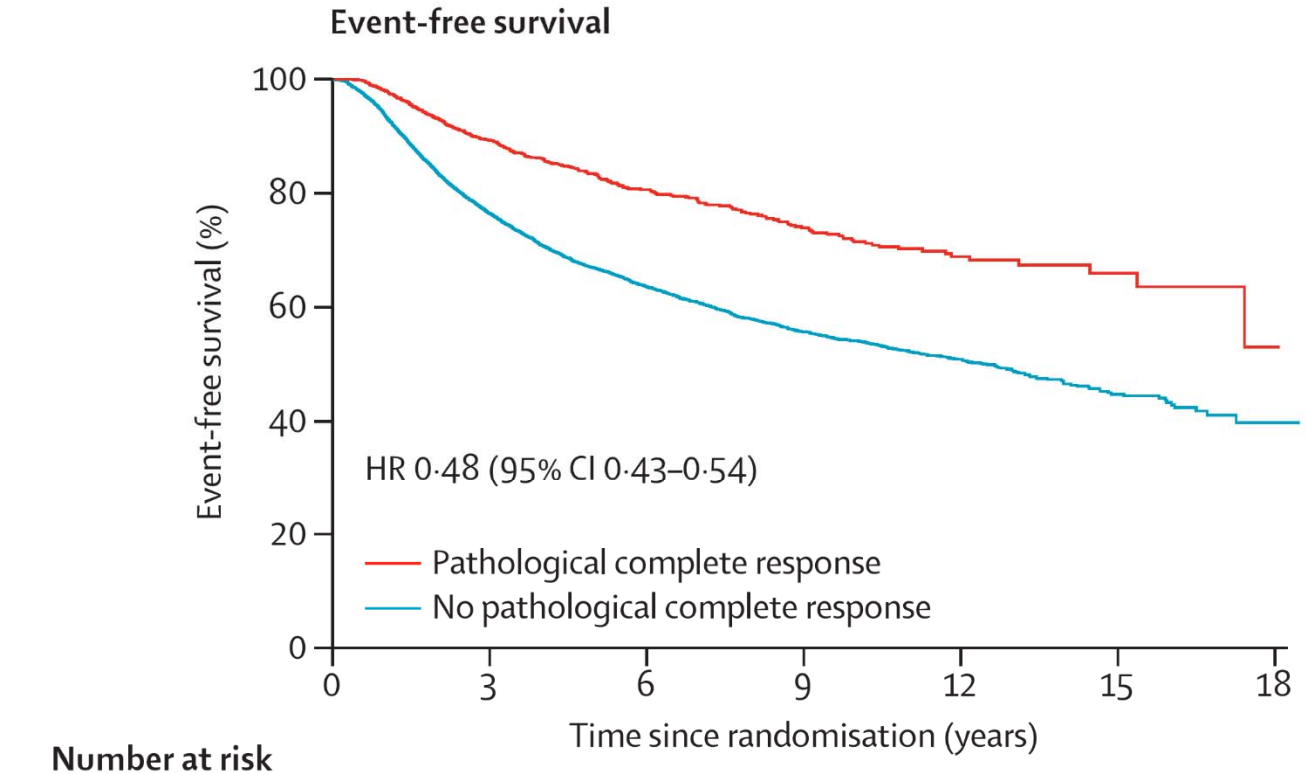
個人レベルの代替性は低い
試験レベルの代替性は高い



Collette L et al, Eur J Cancer 2006

営利目的でのご利用や無断転載はご遠慮ください

乳癌における術前化学療法のpCR



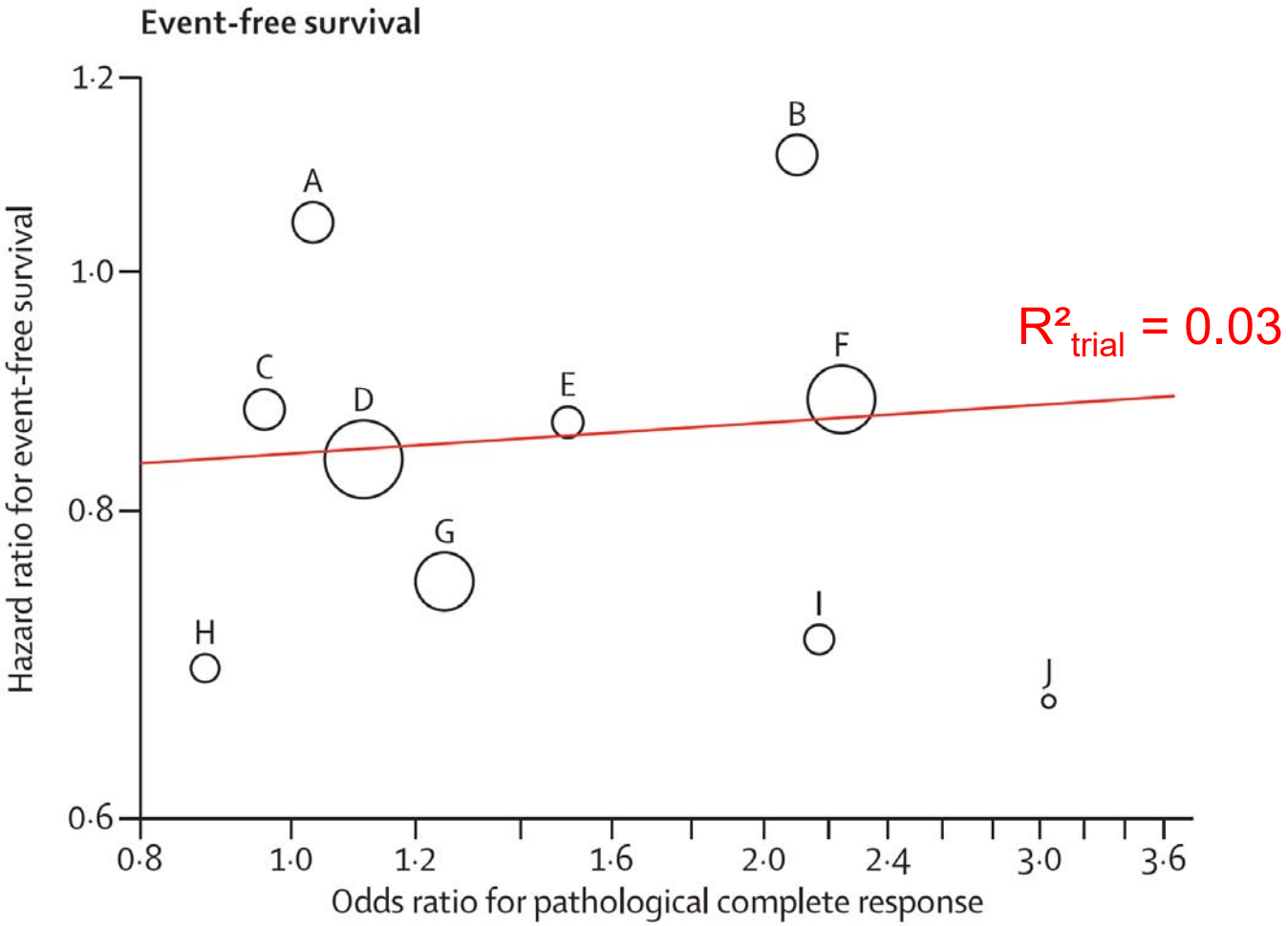
Number at risk							
Pathological complete response	2131	1513	583	337	124	35	2
No pathological complete response	9824	6169	2674	1523	525	165	1

Cortazar et al, Lancet 2014.

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

pCRでの効果がEFSと関連しない…



Cortazar et al, Lancet 2014.

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

PDUFA IV下でのFDAにおける議論

- 代替エンドポイントを利用したい場合に、事前面談で検討すべき事項：大カテゴリとしては以下の3点
- Relationship of the SE with the Clinical Outcome
- Relationship of the SE with the Therapeutic Product
- Reliability of the Measurement Tool(s) Used to Detect the SE

https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM614581.pdf

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

本日のお話

- 代替エンドポイントの性能評価とは？
  - 過去の失敗事例の紹介
  - 代替エンドポイントの性能評価のための統計的基準
- 代替エンドポイントを評価した事例紹介
  - GASTRIC研究
- 最近の話題とまとめ

営利目的でのご利用や無断転載はご遠慮ください

GASTRIC研究

営利目的でのご利用や無断転載はご遠慮ください

GASTRIC研究

- The **G**lobal **A**dvanced/Adjuvant **S**tomach **T**umor **R**esearch through **I**nternational **C**ollaboration (GASTRIC)
- 根治術後胃がんに対して術後化学療法を評価したランダム化比較試験、または、進行／再発胃がんを対象とした化学療法を評価したランダム化比較試験の個人データに基づくメタアナリシス（IPDメタアナリシス）を行うアカデミア主導の国際プロジェクト
  - 2006～2013年：1st Round
  - 2014年～現在：2nd Round実施中

営利目的でのご利用や無断転載はご遠慮ください

GASTRIC 1st Roundの概要

- 2005年以前に実施されたランダム化比較試験を用いて、主に2つの目的としてIPDメタアナリシスを実施

- 化学療法の治療効果に関するメタアナリシス
  - 特に胃がん術後補助化学療法の有効性
- 進行胃がん・術後胃がんを対象とした臨床試験における全生存(overall survival, OS)に対する代替性の検討
  - 無増悪生存期間（PFS）
  - 無病生存期間（DFS）

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

GASTRIC研究と共同研究者

- **Steering Committee**
  - Y Bang
  - H Bleiberg
  - T Burzykowski
  - M Buyse
  - C Delbaldo
  - S Michiels
  - S Morita
  - K Oba
  - Y Ohashi
  - X Paoletti
  - C Pozzo
  - JP Pignon
  - P Rougier
  - J Sakamoto
  - D Sargent
  - M Sasako
  - E Van Cutsem

- **Advanced setting**
  - J Ajani
  - N Boku
  - O Bouche
  - J Buckner
  - C Coombes
  - S Cullinan
  - M Dank
  - B Glimelius
  - R Hawkins
  - W Koizumi
  - M Moehler
  - Y Nio
  - A Ohtsu

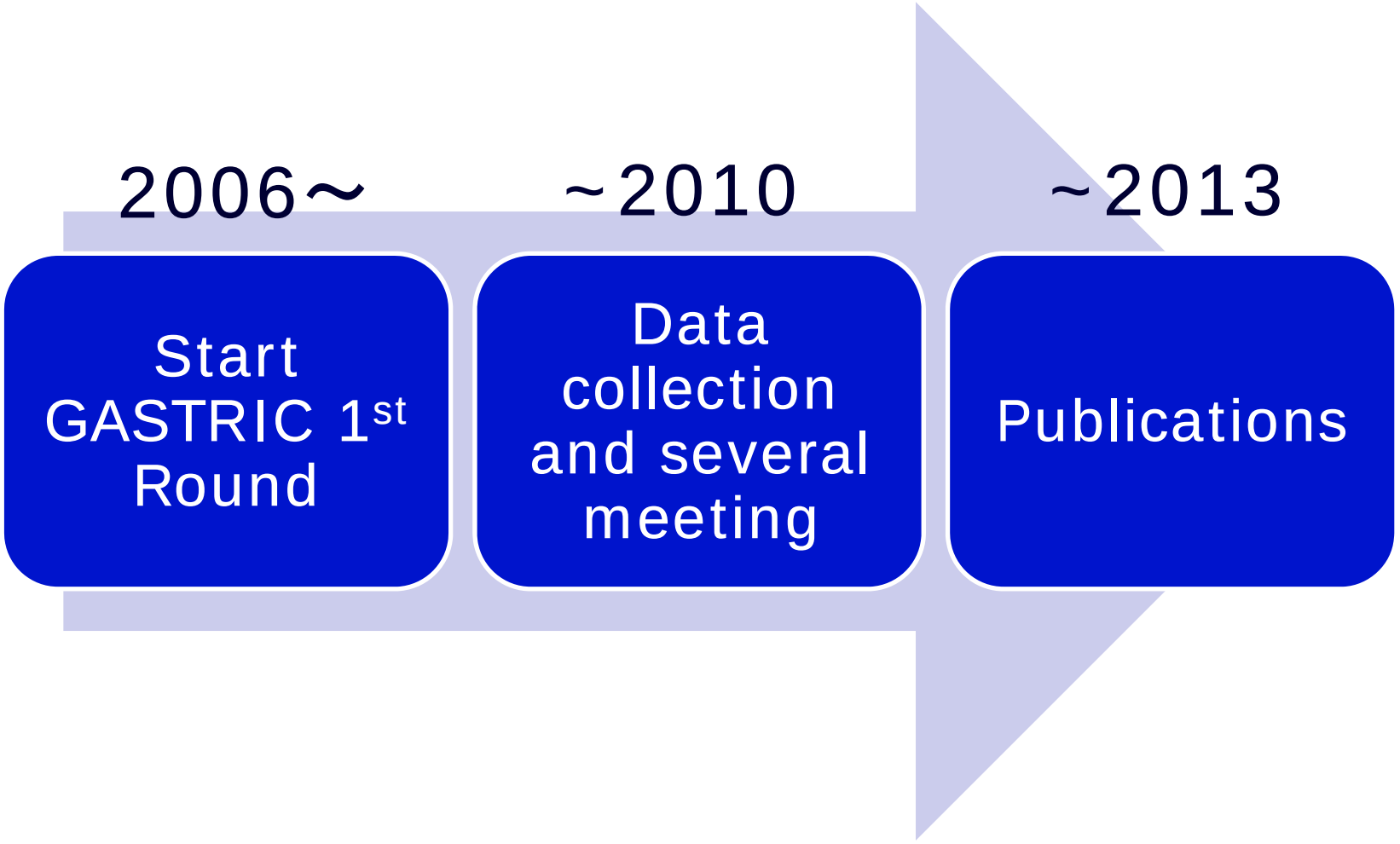
- **Adjuvant setting**
  - S Alberts
  - E Bajetta
  - J Benedetti
  - N Boku
  - O Bouche
  - RC Coombes
  - J Grau
  - J Krook
  - T Kinoshita
  - M Lise
  - JS Mac Donald
  - T Nakajima
  - A Nashimoto
  - G Nelson

- **Secretariats**
  - K Oba
  - X Paoletti
  - M Buyse
  - S Michiels

←半分は生物統計家、半分は臨床家

営利目的でのご利用や無断転載はご遠慮ください

1st Roundの流れ



営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

JAMA®

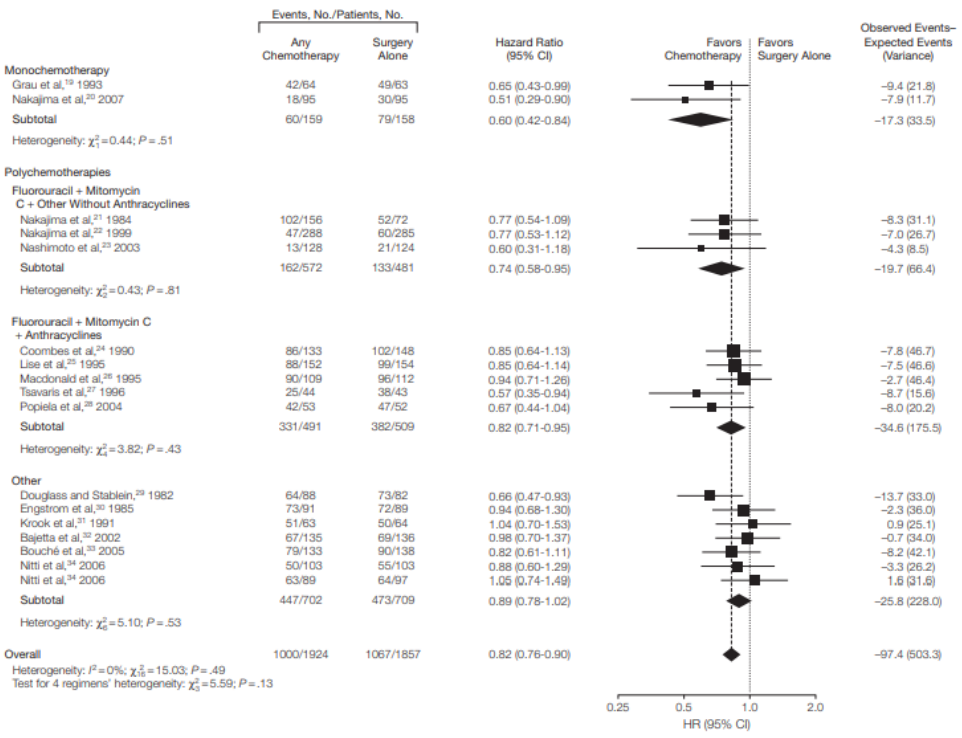
Online article and related content current as of May 4, 2010.

Benefit of Adjuvant Chemotherapy for Resectable Gastric Cancer: A Meta-analysis

The GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group

JAMA. 2010;303(17):1729-1737 (doi:10.1001/jama.2010.534)

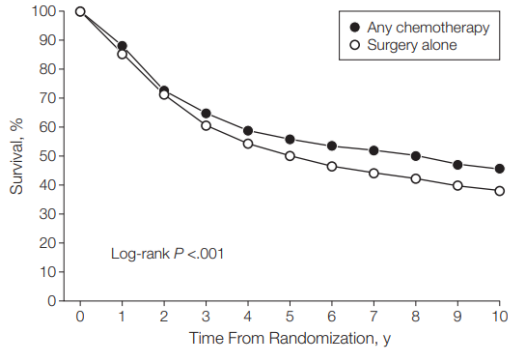
Figure 2. Individual Trial and Overall Hazard Ratio for Overall Survival When Comparing Any Adjuvant Chemotherapy vs Surgery Alone



	Events, No./Patients, No.					
	Any Chemotherapy	Surgery Alone	Hazard Ratio (95% CI)	Favors Chemotherapy	Favors Surgery Alone	Observed Events–Expected Events (Variance)
Monochemotherapy						
Grau et al,[19] 1993	42/64	49/63	0.65 (0.43-0.99)			−9.4 (21.8)
Nakajima et al,[20] 2007	18/95	30/95	0.51 (0.29-0.90)			−7.9 (11.7)
Subtotal	60/159	79/158	0.60 (0.42-0.84)			−17.3 (33.5)
Heterogeneity: $\chi^2_1=0.44$; $P=.51$						
Polychemotherapies						
Fluorouracil + Mitomycin C + Other Without Anthracyclines						
Nakajima et al,[21] 1984	102/156	52/72	0.77 (0.54-1.09)			−8.3 (31.1)
Nakajima et al,[22] 1999	47/288	60/285	0.77 (0.53-1.12)			−7.0 (26.7)
Nashimoto et al,[23] 2003	13/128	21/124	0.60 (0.31-1.18)			−4.3 (8.5)
Subtotal	162/572	133/481	0.74 (0.58-0.95)			−19.7 (66.4)
Heterogeneity: $\chi^2_2=0.43$; $P=.81$						
Fluorouracil + Mitomycin C + Anthracyclines						
Coombes et al,[24] 1990	86/133	102/148	0.85 (0.64-1.13)			−7.8 (46.7)
Lise et al,[25] 1995	88/152	99/154	0.85 (0.64-1.14)			−7.5 (46.6)
Macdonald et al,[26] 1995	90/109	96/112	0.94 (0.71-1.26)			−2.7 (46.4)
Tsavaris et al,[27] 1996	25/44	38/43	0.57 (0.35-0.94)			−8.7 (15.6)
Popiela et al,[28] 2004	42/53	47/52	0.67 (0.44-1.04)			−8.0 (20.2)
Subtotal	331/491	382/509	0.82 (0.71-0.95)			−34.6 (175.5)
Heterogeneity: $\chi^2_4=3.82$; $P=.43$						
Other						
Douglass and Stablein,[29] 1982	64/88	73/82	0.66 (0.47-0.93)			−13.7 (33.0)
Engstrom et al,[30] 1985	73/91	72/89	0.94 (0.68-1.30)			−2.3 (36.0)
Krook et al,[31] 1991	51/63	50/64	1.04 (0.70-1.53)			0.9 (25.1)
Bajetta et al,[32] 2002	67/135	69/136	0.98 (0.70-1.37)			−0.7 (34.0)
Bouché et al,[33] 2005	79/133	90/138	0.82 (0.61-1.11)			−8.2 (42.1)
Nitti et al,[34] 2006	50/103	55/103	0.88 (0.60-1.29)			−3.3 (26.2)
Nitti et al,[34] 2006	63/89	64/97	1.05 (0.74-1.49)			1.6 (31.6)
Subtotal	447/702	473/709	0.89 (0.78-1.02)			−25.8 (228.0)
Heterogeneity: $\chi^2_6=5.10$; $P=.53$						
Overall	1000/1924	1067/1857	0.82 (0.76-0.90)			−97.4 (503.3)
Heterogeneity: $I^2=0\%$; $\chi^2_{16}=15.03$; $P=.49$						
Test for 4 regimens' heterogeneity: $\chi^2_3=5.59$; $P=.13$						

The inverse of the variance of observed events minus expected events measures the weight of each trial in the analysis. *P* values are from *P*-for-effect modification testing for heterogeneity within or across the groups of regimens. The sizes of data markers are proportional to the number of deaths in the trials. CI indicates confidence interval; HR, hazard ratio.

Figure 3. Overall Survival Estimate After Any Chemotherapy or Surgery Alone Truncated at 10 Years



No. at risk											
Any chemotherapy	1924	1688	1385	1217	1080	929	709	526	390	297	243
Surgery alone	1857	1568	1300	1092	952	782	583	407	267	172	138

The estimates of the survival curves use an actuarial approach as described in the Methods.

GASTRIC, JAMA 2010.

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

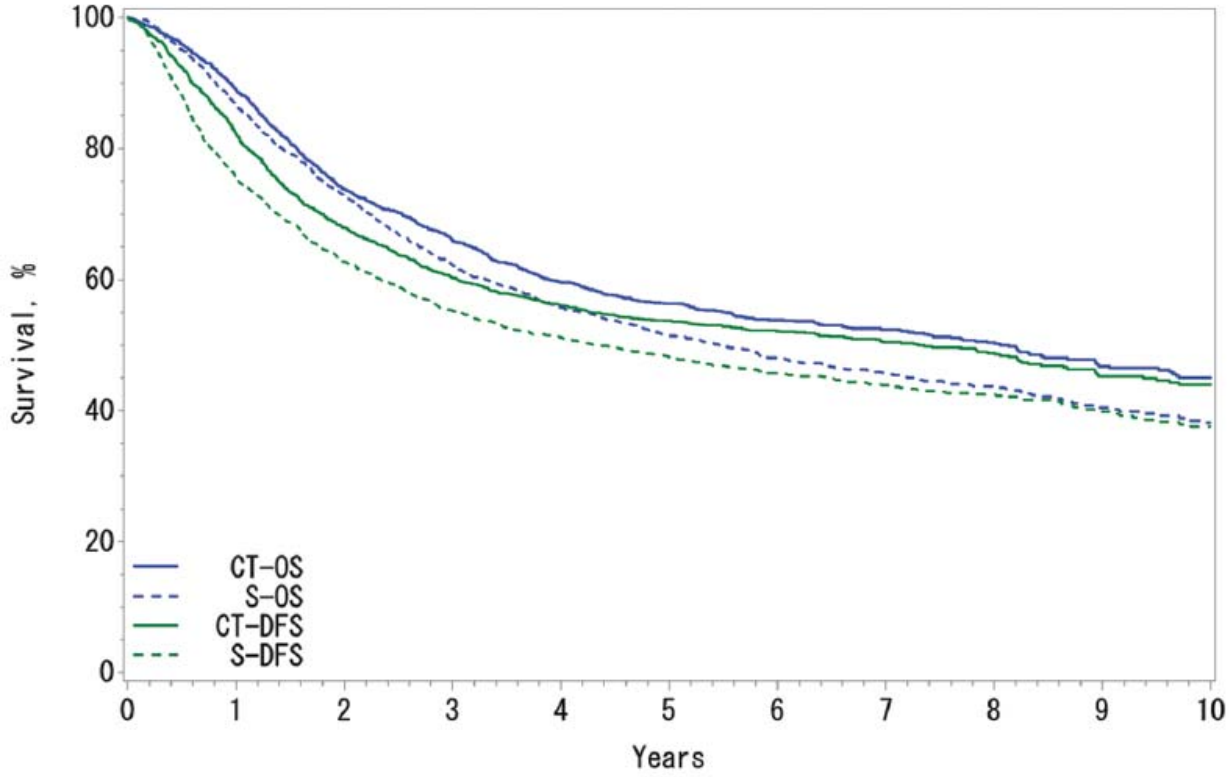
DFS、PFSに関する代替性の検討

- Setting：根治術後胃癌
  - 真のエンドポイント：OS、代替エンドポイント：DFS
  - 14試験、3288名の個人レベルのデータ
    - Surgery alone vs Adjuvant CT
- Setting：進行／再発胃癌
  - 真のエンドポイント：OS、代替エンドポイント：PFS
  - 20試験、4069名の個人レベルのデータ
    - Experimental CT vs Control CT

Oba K et al, JNCI 2013.
Paoletti X et al, JNCI 2013.

営利目的でのご利用や無断転載はご遠慮ください

カプランマイヤー曲線: DFS and OS

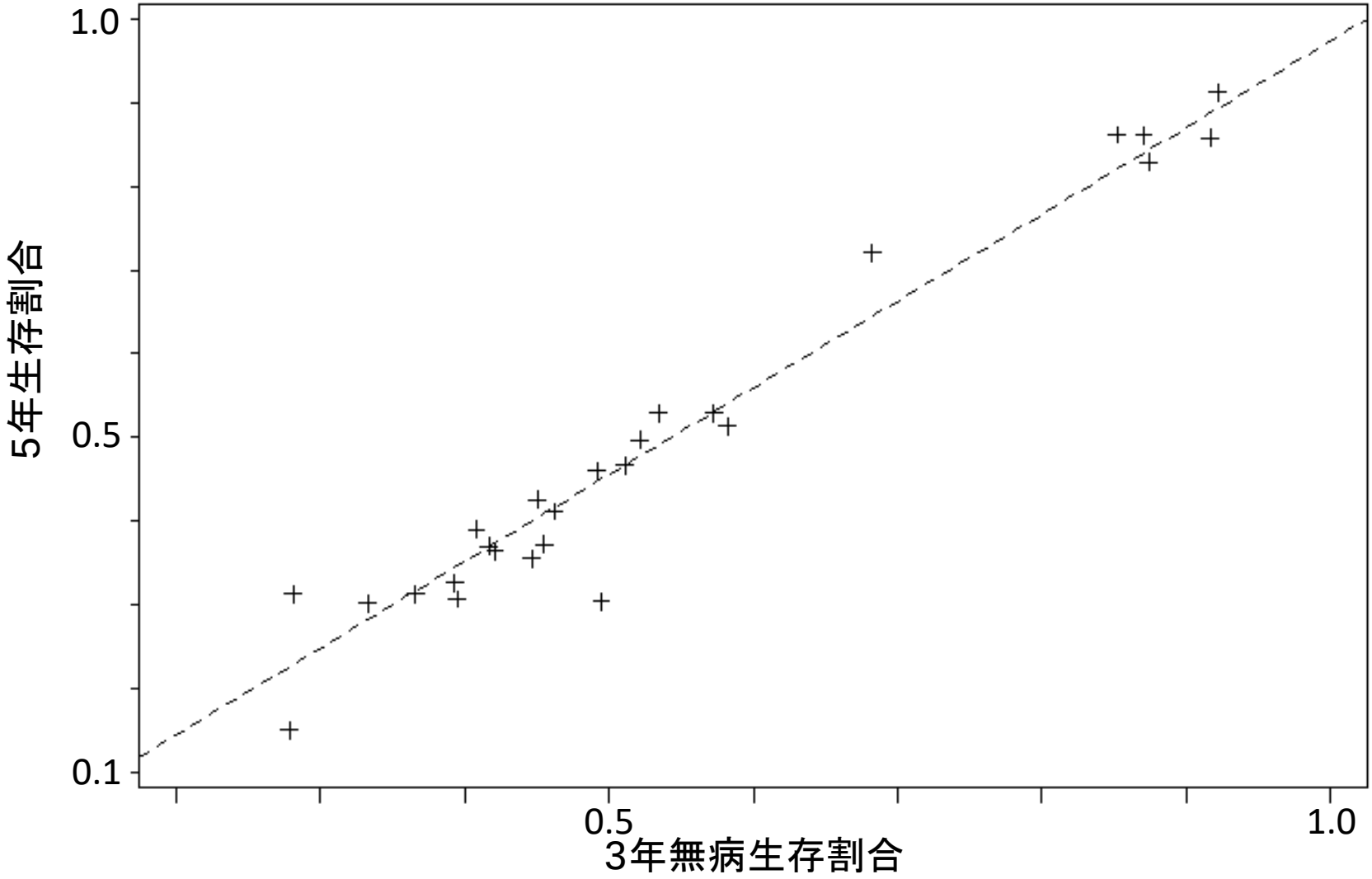


CT-OS	1634	1448	1193	1050	927	781	570	395	270	180	133
S-OS	1654	1421	1194	1009	875	718	531	363	228	135	102
CT-DFS	1634	1338	1096	956	867	738	545	374	253	164	124
S-DFS	1654	1242	1027	897	801	666	496	341	214	129	97

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

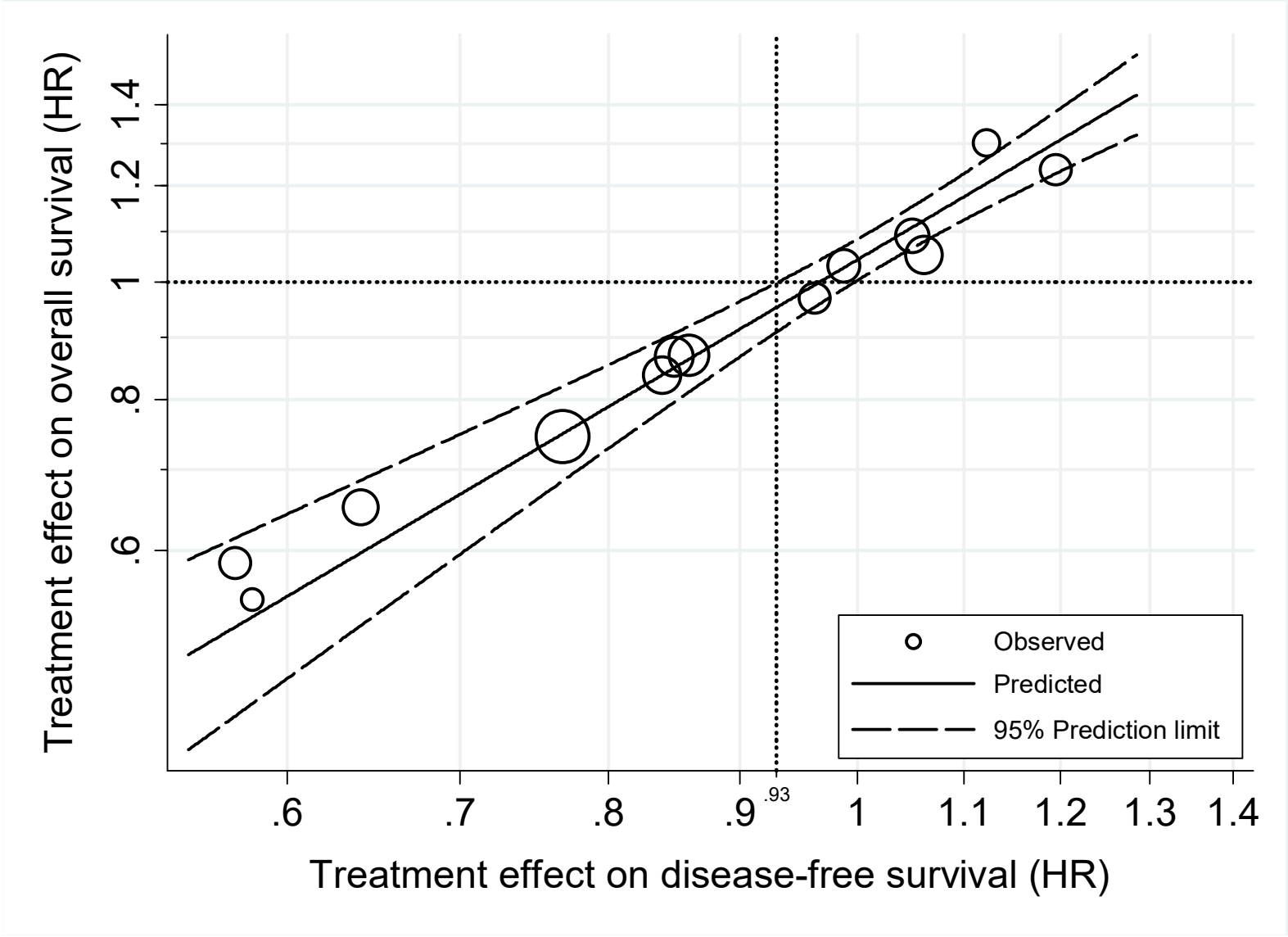
群別 3年DFS vs 5年OS



営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

治療効果間の関連性



営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

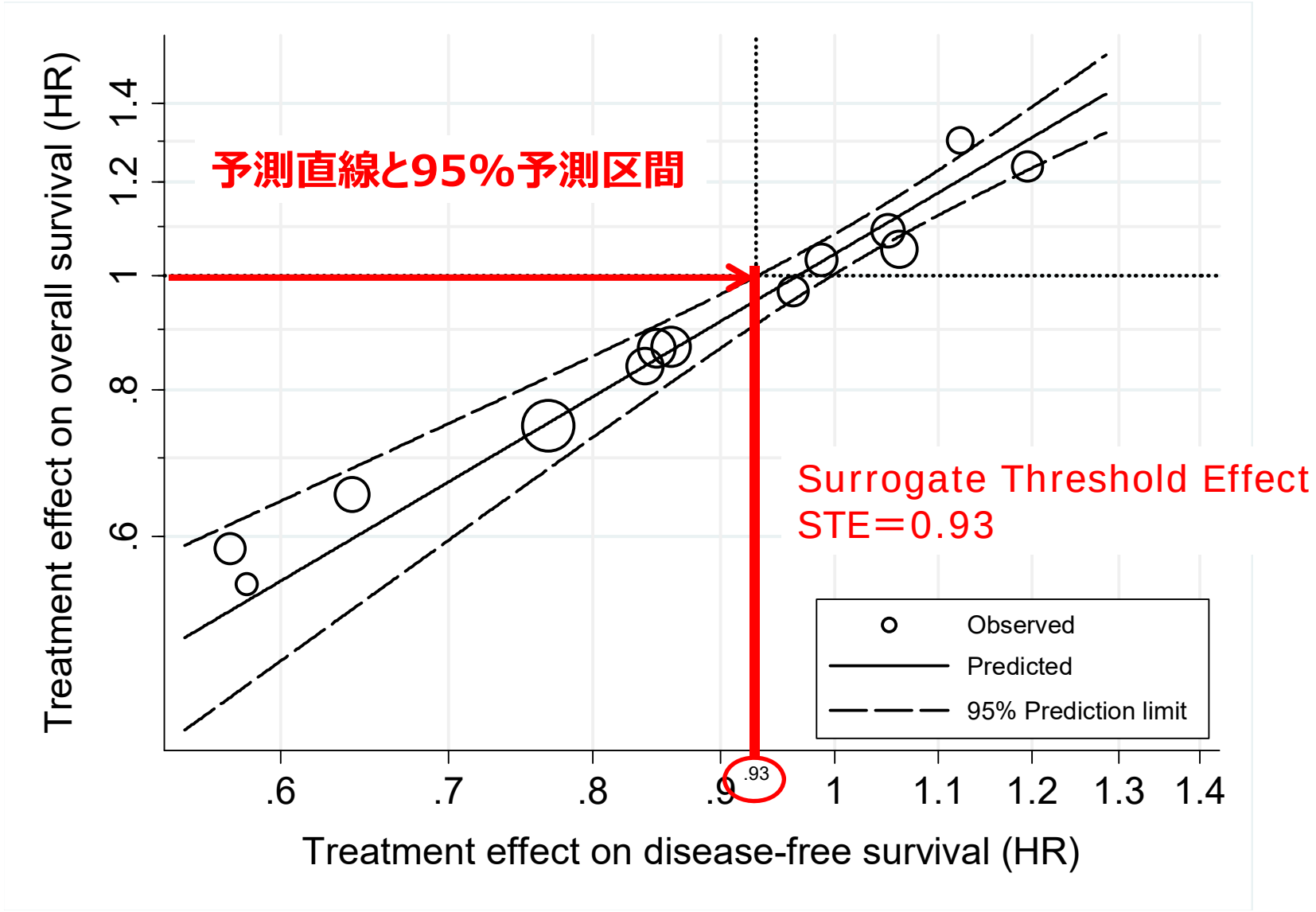
DFSの代替性

- 個人レベル（エンドポイント間）の代替性
  - 順位相関係数： 0.976 (95%CI 0.966-0.987)
- 試験レベル（治療効果間）の代替性
  - $R^2 \approx 1$ (95% CI = 0.99997-1.0000)
  - $\ln(HR_{OS}) = 0.047 + 1.239 \times \ln(HR_{DFS})$

- 個人レベルの代替性、試験レベルの代替性、ともに高い代替性を示し、代替性の証拠は強いと考えた

営利目的でのご利用や無断転載はご遠慮ください

DFSで超えるべきハードルの目安



営利目的でのご利用や無断転載はご遠慮ください

追加試験を用いた予測性のバリデーション

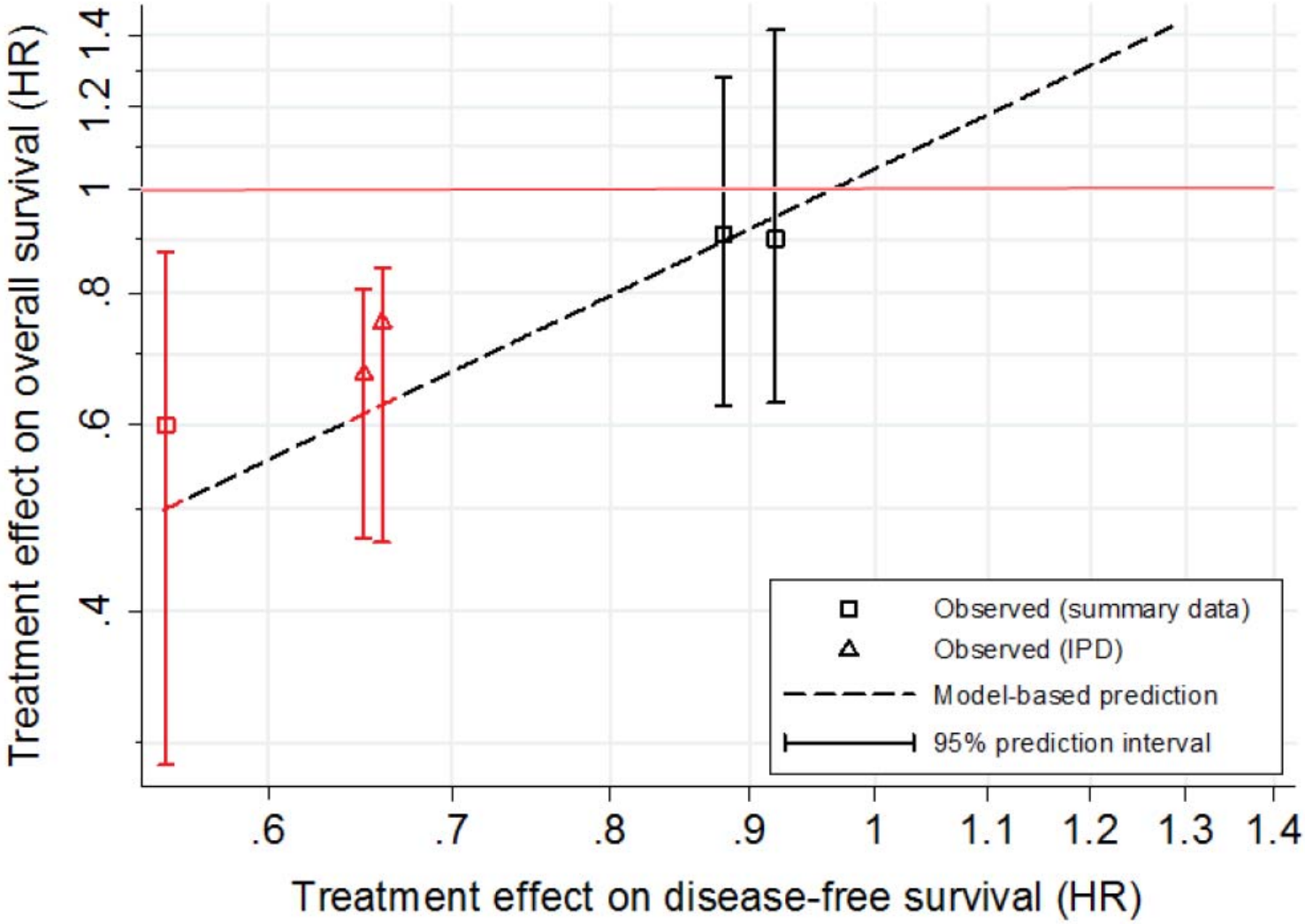
- データ収集後、実施された試験（バリデーション試験）を用いてDFSによるOSの予測性を評価
  - 2試験（ACTS-GC、INT-1018）は個人データ
  - 3試験は文献データ

Trial	Observed HR_{DFS}	Observed HR_{OS}
Cirera	0.55 [0.36, 0.85]*	0.60 [0.39, 0.93]*
ACTS-GC	0.65 [0.54, 0.79]*	0.67 [0.54, 0.83]*
INT-1018	0.66 [0.53, 0.82]*	0.75 [0.61, 0.92]*
GOIM-9602	0.88 [0.66, 1.17]	0.91 [0.69, 1.21]
GOIRC	0.92 [0.66, 1.27]	0.90 [0.64, 1.26]

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

追加試験を用いた予測性のバリデーション



営利目的でのご利用や無断転載はご遠慮ください

仮想的に時点を区切った場合の代替性

	2y DFS/5y OS	3y DFS/5y OS	4y DFS/5y OS	All
Events	1054/1371	1279/1371	1390/1371	1619/1561
個人レベルρ	0.953 (0.926, 0.980)	0.957 (0.934, 0.979)	0.961 (0.941, 0.980)	0.976 (0.965, 0.987)
試験レベル R^2	0.745 (0.505, 0.984)	0.898 (0.792, 1.00)	0.940 (0.876, 1.00)	0.978 (0.954, 1.00)
STE (HR)	undefined	0.66	0.87	0.93

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

Disease-Free Survival as a Surrogate for Overall Survival in Adjuvant Trials of Gastric Cancer: A Meta-Analysis

Koji Oba, Xavier Paoletti, Steven Alberts, Yung-Jue Bang, Jacqueline Benedetti, Harry Bleiberg, Paul Catalano, Florian Lordick, Stefan Michiels, Satoshi Morita, Yasuo Ohashi, Jean-pierre Pignon, Philippe Rougier, Mitsuru Sasako, Junichi Sakamoto, Daniel Sargent, Kohei Shitara, Eric Van Cutsem, Marc Buyse, Tomasz Burzykowski; on behalf of the GASTRIC group

Manuscript received February 12, 2013; revised July 25, 2013; accepted July 25, 2013.

Correspondence to: Koji Oba, PhD, Translational Research and Clinical Trial Center, Hokkaido University Hospital, Kita 14, Nishi 5, Kita-ku, Sapporo, Hokkaido 0608648, Japan (e-mail: k.oba@huhp.hokudai.ac.jp).

Background In investigations of the effectiveness of surgery and adjuvant chemotherapy for gastric cancers, overall survival (OS) is considered the gold standard endpoint. However, the disadvantage of using OS as the endpoint is that it requires an extended follow-up period. We sought to investigate whether disease-free survival (DFS) is a valid surrogate for OS in trials of adjuvant chemotherapy for gastric cancer.

Methods The GASTRIC group initiated a meta-analysis of individual patient data collected in randomized clinical trials comparing adjuvant chemotherapy vs surgery alone for patients with curatively resected gastric cancer. Surrogacy of DFS was assessed through the correlation between the endpoints as well as through the correlation between the treatment effects on the endpoints. External validation of the prediction based on DFS was also evaluated.

Results Individual patient data from 14 randomized clinical trials that included a total of 3288 patients were analyzed. The rank correlation coefficient between DFS and OS was 0.974 (95% confidence interval [CI] = 0.971 to 0.976). The coefficient of determination between the treatment effects on DFS and on OS was as high as 0.964 (95% CI = 0.926 to 1.000), and the surrogate threshold effect based on adjusted regression analysis was 0.92. In external validation, the six hazard ratios for OS predicted according to DFS were in very good agreement with those actually observed for OS.

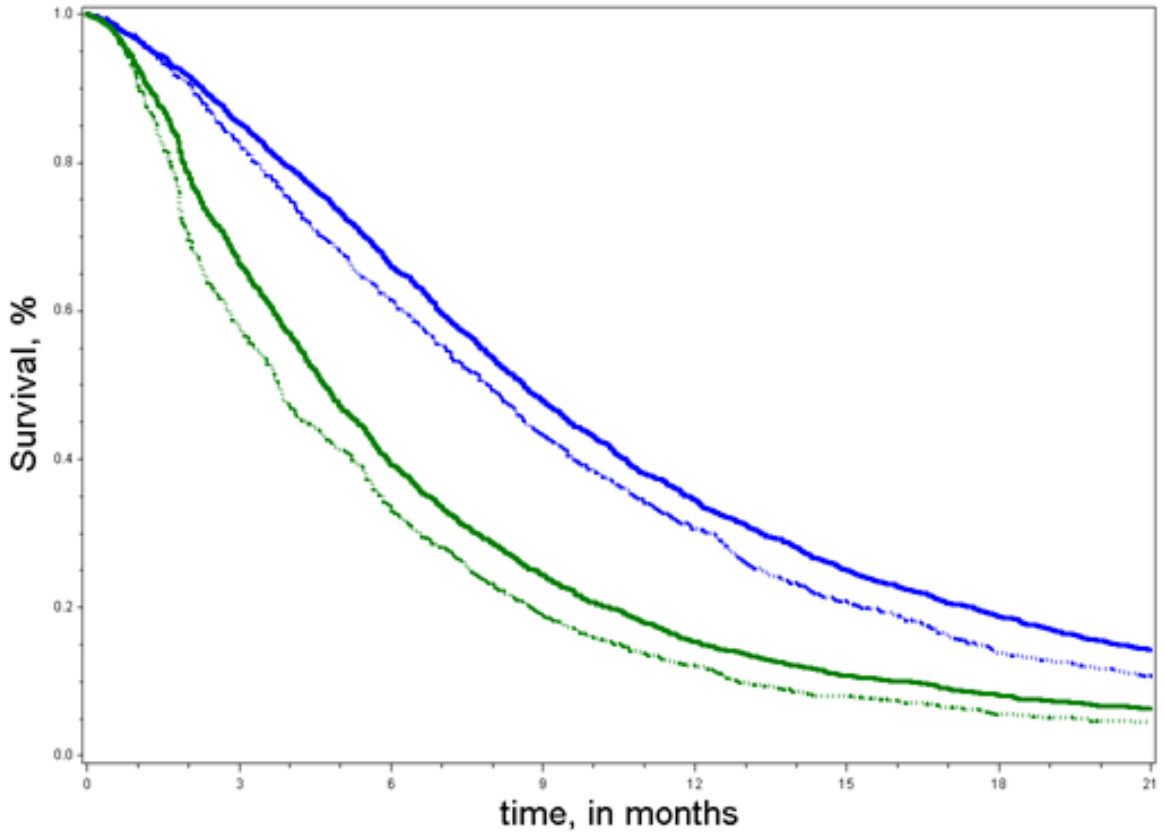
Conclusions DFS is an acceptable surrogate for OS in trials of cytotoxic agents for gastric cancer in the adjuvant setting.

J Natl Cancer Inst;2013;105:1600–1607

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

カプランマイヤー曲線：PFS vs OS

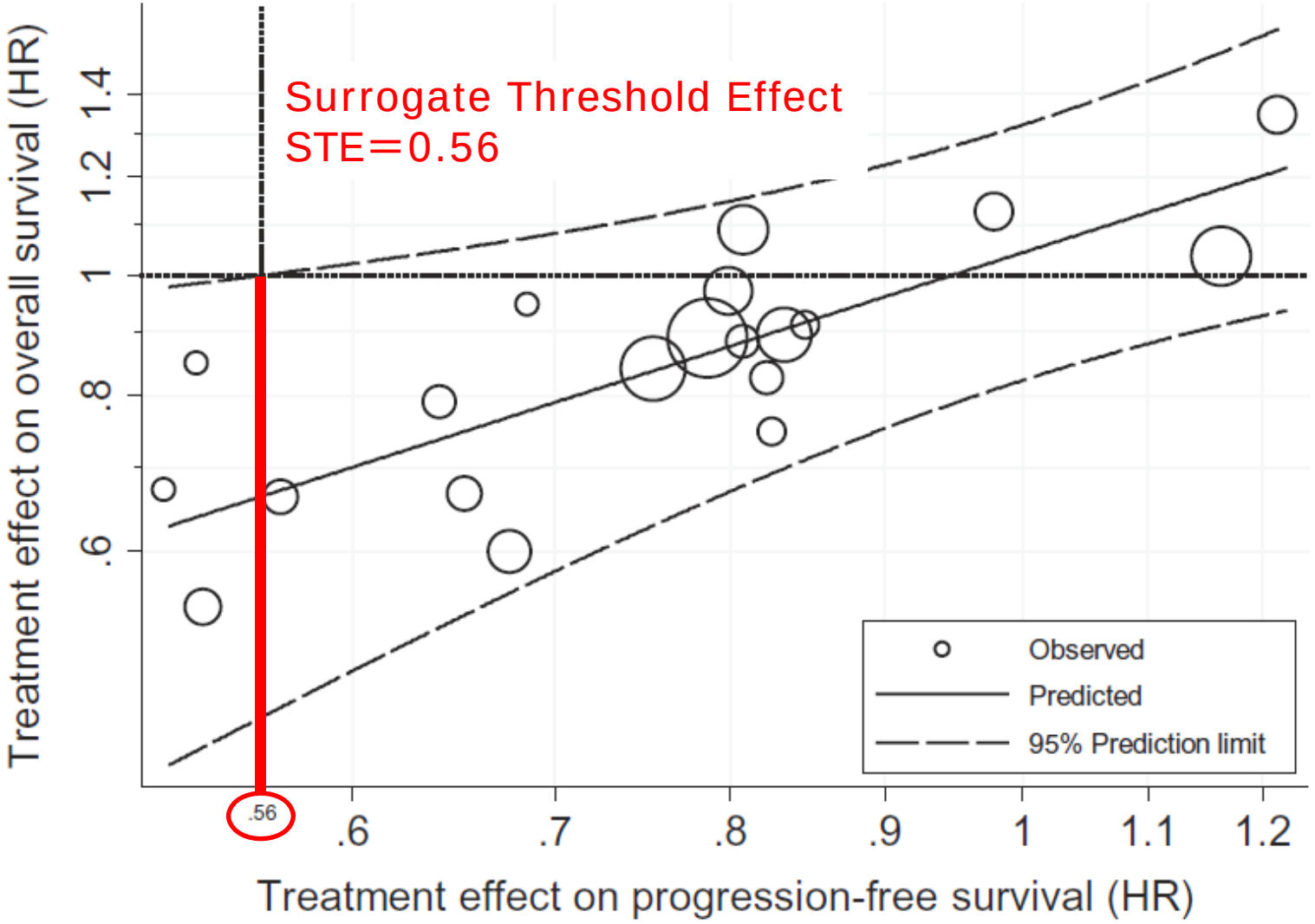


Exp－OS	2477	2093	1602	1150	814	556	405	286
Ctrl－OS	1737	1420	1042	726	498	316	197	139
Exp－PFS	2404	1599	934	571	353	232	168	118
Ctrl－PFS	1669	962	545	306	187	112	69	49

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

治療効果間の関連性



営利目的でのご利用や無断転載はご遠慮ください

PFSの代替性

- 個人レベル（エンドポイント間）の代替性
  - 順位相関係数： 0.853 (95%CI 0.852-0.854)
- 試験レベル（治療効果間）の代替性
  - R^2 = 0.61 (95% CI = 0.04-1.00)
  - $\ln(HR_{OS}) = 0.042 + 0.779 \times \ln(HR_{DFS})$

- 個人レベルの代替性は高いが、試験レベルの代替性が中程度であり、代替性の証拠としては不十分

営利目的でのご利用や無断転載はご遠慮ください

追加試験を用いた予測性のバリデーション

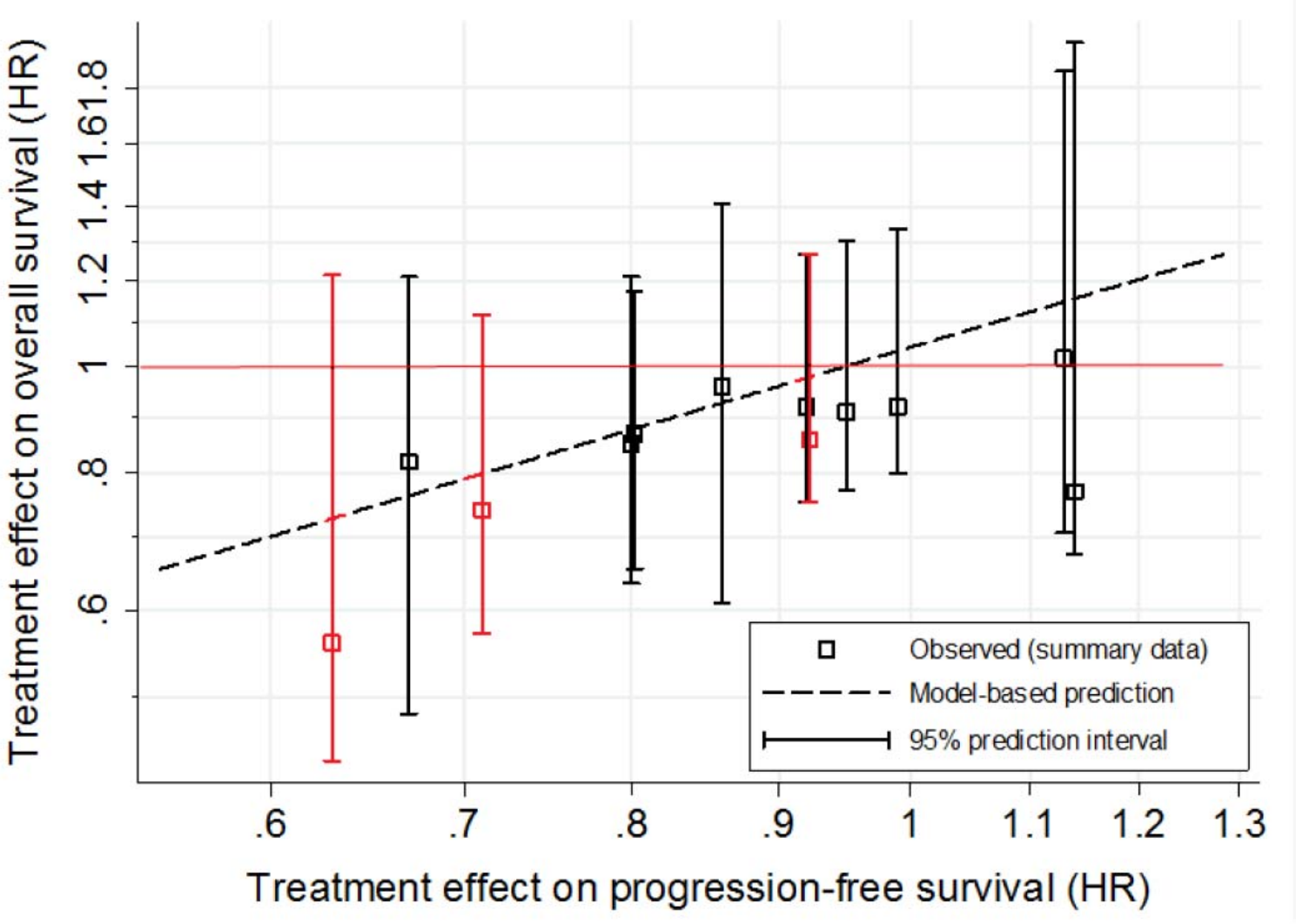
- データ収集後、実施された試験（バリデーション試験）を用いてPFSによるOSの予測性を評価

Trial	Observed HR_{PFS}	Observed HR_{OS}	Predicted HR_{OS}
Jeung	0.63 [0.38, 1.05]	0.56 [0.35, 0.88]*	0.73 [0.46, 1.04]
AIO	0.67 [0.43, 1.04]	0.82 [0.47, 1.45]	0.76 [0.53, 1.07]
ToGA	0.71 [0.59, 0.85]*	0.74 [0.60, 0.91]*	0.80 [0.58, 1.09]
AVAGAST	0.80 [0.68, 0.93]*	0.87 [0.73, 1.03]	0.88 [0.76, 1.14]
Kang	0.80 [0.63, 1.03]	0.85 [0.64, 1.13]	0.88 [0.76, 1.14]
Park	0.86 [0.54, 1.37]	0.96 [0.60, 1.52]	0.93 [0.71, 1.18]
REAL(a)	0.92 [0.80, 1.04]	0.92 [0.80, 1.10]	0.98 [0.77, 1.22]
REAL(b)	0.92 [0.81, 1.05]	0.86 [0.80, 0.99]*	0.98 [0.77, 1.22]
Ross	0.95 [0.80, 1.08]	0.91 [0.76, 1.04]	1.00 [0.79, 1.29]
FLAGS	0.99 [0.86, 1.14]	0.92 [0.80, 1.05]	1.03 [0.81, 1.31]
Rao	1.13 [0.63, 2.01]	1.02 [0.61, 1.70]	1.14 [0.89, 1.46]
Moehler	1.14 [0.59, 2.21]	0.77 [0.51, 1.17]	1.15 [0.90, 1.48]

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

追加試験を用いた予測性のバリデーション



営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

本日のお話

- 代替エンドポイントの性能評価とは？
  - 過去の失敗事例の紹介
  - 代替エンドポイントの性能評価のための統計的基準
- 代替エンドポイントを評価した事例紹介
  - GASTRIC研究
- 最近の話題とまとめ

営利目的でのご利用や無断転載はご遠慮ください

代替性のレベルに基づく分類

- **Validated Surrogate Endpoint**
  - An endpoint supported by a clear mechanistic rationale and clinical data providing strong evidence that an effect on the surrogate endpoint predicts a specific clinical benefit.
  - …can be used to support marketing approval of a medical or tobacco product in a defined context without the need for additional studies to demonstrate the clinical benefit directly
- **Reasonably Likely Surrogate Endpoint**
  - An endpoint supported by strong mechanistic and/or epidemiologic rationale such that an effect on the surrogate endpoint is expected to be correlated …, but without sufficient clinical data to show that it is a validated surrogate endpoint.
  - …may be used for accelerated approval for drugs and potentially also for approval or clearance of medical devices.

FDA-NIH Biomarker Working Group, BEST (Biomarker, EndpointS, and other Tools) Resource [Internet], 2018

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

どの程度、関連の強さが求められる？

- IQWIG (Institute for Quality and Efficiency in Health Care) ガイドラインによると
- **Validity proven**
  - 相関係数の95%信頼区間下限が0.85以上
- **Unclear validity**
  - R < 0.85 to >0.7
- **Proven lack of validity**
  - 相関係数の95%信頼区間上限が0.7以下

Kemp R et al, BMC Med 2017.

営利目的でのご利用や無断転載はご遠慮ください

ICRweb臨床研究入門 http://www.icrweb.jp

代替エンドポイントリストの提供

- FDAが既に薬剤開発で利用された代替エンドポイント
- ガイダンスや他文書で受け入れを示唆したエンドポイント

	A	B	C	D	E
1	**Disease or Use**	**Patient Population**	**Surrogate endpoint**	**Type of approval appropriate for**	**Drug mechanism of action**
14	Cancer: hematological malignancies	Patients with Acute Lymphoblastic Leukemia	Serum asparaginase	Traditional	Asparagine-specific enzyme
15	Cancer: hematological malignancies	Patients with diffuse large B-cell lymphoma	Event-free survival (EFS) ×	Traditional	Mechanism agnostic*
16	Cancer: hematological malignancies	Patients with chronic myeloid leukemia; hypereosinophilic syndrome/chronic eosinophilic leukemia	Major hematologic response	Accelerated/Traditional §	Mechanism agnostic*
17	Cancer: hematological malignancies	Patients with acute myeloid leukemia and acute lymphoblastic leukemia	Durable complete remission rate	Accelerated/Traditional §	Mechanism agnostic*
18	Cancer: hematological malignancies	Patients with acute lymphoblastic leukemia; myelodysplastic/myeloproliferative diseases; chronic myeloid leukemia	Major hematologic response and cytogenic response	Accelerated/Traditional §	Mechanism agnostic*
19	Cancer: hematological malignancies	Patients with B-cell precursor acute lymphoblastic leukemia in first or second complete remission	Minimal residual disease response rate	Accelerated	Mechanism agnostic*
20	Cancer: hematological malignancies	Patients with T-cell lymphoma; mantle cell lymphoma; classical hodgkin lymphoma; anaplastic large cell lymphoma and mycosis fungoides; non-hodgkin's lymphoma; multiple myeloma; chronic myeloid leukemia; acute lymphoblastic leukemia; small lymphocytic lymphoma; Waldenström's macroglobulinemia; marginal zone lymphoma	Durable objective overall response rate (ORR)	Accelerated/Traditional §	Mechanism agnostic*

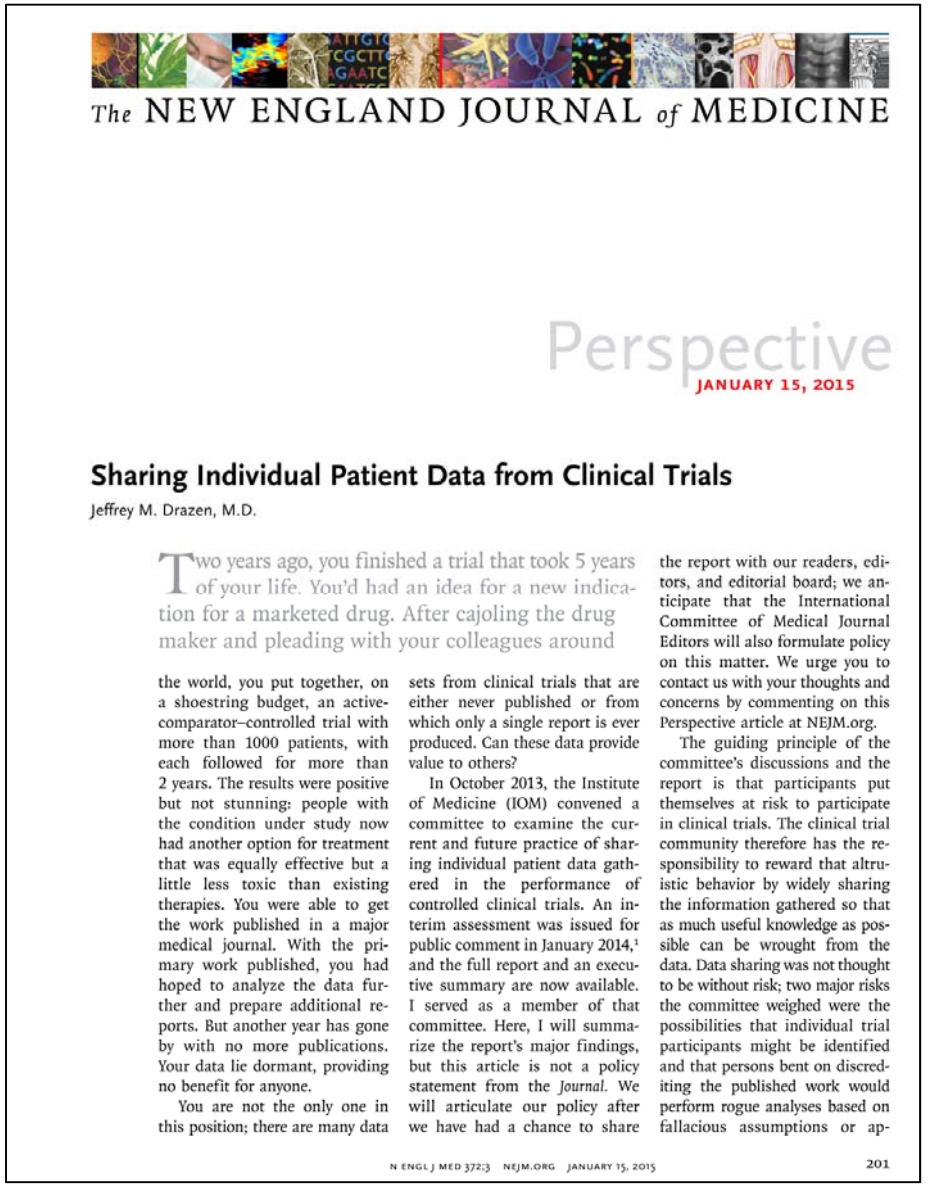
営利目的でのご利用や無断転載はご遠慮ください

代替性は状況に依存する

- 状況として代表的なもの
  - 疾患
  - 使用する薬剤の種類
  - 場（医療環境、人種、地域、など）
- 確立された関連性があってもバリデーションや、代替性の再評価が必要
  - **データベースの構築の重要性**

営利目的でのご利用や無断転載はご遠慮ください

データ公開に関する海外での動き



The NEW ENGLAND JOURNAL of MEDICINE

Perspective

JANUARY 15, 2015

Sharing Individual Patient Data from Clinical Trials

Jeffrey M. Drazen, M.D.

Two years ago, you finished a trial that took 5 years of your life. You'd had an idea for a new indication for a marketed drug. After cajoling the drug maker and pleading with your colleagues around the world, you put together, on a shoestring budget, an active-comparator–controlled trial with more than 1000 patients, with each followed for more than 2 years. The results were positive but not stunning: people with the condition under study now had another option for treatment that was equally effective but a little less toxic than existing therapies. You were able to get the work published in a major medical journal. With the primary work published, you had hoped to analyze the data further and prepare additional reports. But another year has gone by with no more publications. Your data lie dormant, providing no benefit for anyone.

You are not the only one in this position; there are many data sets from clinical trials that are either never published or from which only a single report is ever produced. Can these data provide value to others?

In October 2013, the Institute of Medicine (IOM) convened a committee to examine the current and future practice of sharing individual patient data gathered in the performance of controlled clinical trials. An interim assessment was issued for public comment in January 2014,[1] and the full report and an executive summary are now available. I served as a member of that committee. Here, I will summarize the report's major findings, but this article is not a policy statement from the *Journal*. We will articulate our policy after we have had a chance to share the report with our readers, editors, and editorial board; we anticipate that the International Committee of Medical Journal Editors will also formulate policy on this matter. We urge you to contact us with your thoughts and concerns by commenting on this Perspective article at NEJM.org.

The guiding principle of the committee's discussions and the report is that participants put themselves at risk to participate in clinical trials. The clinical trial community therefore has the responsibility to reward that altruistic behavior by widely sharing the information gathered so that as much useful knowledge as possible can be wrought from the data. Data sharing was not thought to be without risk; two major risks the committee weighed were the possibilities that individual trial participants might be identified and that persons bent on discrediting the published work would perform rogue analyses based on fallacious assumptions or ap-

EDITORIAL

Data Sharing Statements for Clinical Trials
A Requirement of the International Committee of Medical Journal Editors

Darren B. Taichman, MD, PhD; Peush Sahni, MB, BS, MS, PhD; Anja Pinborg, MD; Larry Peiperl, MD; Christine Laine, MD, MPH; Astrid James, MB, BS; Sung-Tae Hong, MD, PhD; Abraham Haileamlak, MD; Laragh Gollogly, MD, MPH; Fiona Godlee, FRCP; Frank A. Frizelle, MB, ChB, FRACS; Fernando Florenzano, MD; Jeffrey M. Drazen, MD; Howard Bauchner, MD; Christopher Baethge, MD; Joyce Backus, MSLS

The International Committee of Medical Journal Editors (ICMJE) believes there is an ethical obligation to responsibly share data generated by interventional clinical trials because trial participants have put themselves at risk. In January 2016 we published a proposal aimed at helping to create an environment in which the sharing of deidentified individual participant data becomes the norm. In response to our request for feedback we received many comments from individuals and groups.[1] Some applauded the proposal while others expressed disappointment it did not more quickly create a commitment to data sharing. Many raised valid concerns regarding the feasibility of the proposed requirements, the necessary resources, the real or perceived risks to trial participants, and the need to protect the interests of patients and researchers.

It is encouraging that data sharing is already occurring in some settings. Over the past year, however, we have learned that the challenges are substantial and the requisite mechanisms are not in place to mandate universal data sharing at this time. Although many issues must be addressed for data sharing to become the norm, we remain committed to this goal.

Therefore, ICMJE will require the following as conditions of consideration for publication of a clinical trial report in our member journals:

1. As of July 1, 2018, manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement as described below.
2. Clinical trials that begin enrolling participants on or after January 1, 2019, must include a data sharing plan in the trial's registration. The ICMJE's policy regarding trial registration is explained at http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html. If the data sharing plan changes after registration this should be reflected in the statement submitted and published with the manuscript and updated in the registry record.

Table. Examples of Data Sharing Statements That Fulfill These ICMJE Requirements[a]

	Example 1	Example 2	Example 3	Example 4
Will individual participant data be available (including data dictionaries)?	Yes	Yes	Yes	No
What data in particular will be shared?	All of the individual participant data collected during the trial, after deidentification	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices)	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures and appendices)	Not available
What other documents will be available?	Study protocol, statistical analysis plan, informed consent form, clinical study report, analytic code	Study protocol, statistical analysis plan, analytic code	Study protocol	Not available
When will data be available (start and end dates)?	Immediately following publication; no end date	Beginning 3 months and ending 5 years following article publication	Beginning 9 months and ending 36 months following article publication	Not applicable
With whom?	Anyone who wishes to access the data	Researchers who provide a methodologically sound proposal	Investigators whose proposed use of the data has been approved by an independent review committee ("learned intermediary") identified for this purpose	Not applicable
For what types of analyses?	Any purpose	To achieve aims in the approved proposal	For individual participant data meta-analysis	Not applicable
By what mechanism will data be made available?	Data are available indefinitely at (*link to be included*)	Proposals should be directed to xxx@yyy. To gain access, data requestors will need to sign a data access agreement. Data are available for 5 years at a third-party website (*link to be included*).	Proposals may be submitted up to 36 months following article publication. After 36 months the data will be available in our university's data warehouse but without investigator support other than deposited metadata. Information regarding submitting proposals and accessing data may be found at (*link to be provided*).	Not applicable

[a] These examples are meant to illustrate a range of, but not all, data sharing options.

営利目的でのご利用や無断転載はご遠慮ください

長期追跡の重要性

- 代替エンドポイントで承認された場合、長期の安全性の懸念がないか、という点には注意が必要

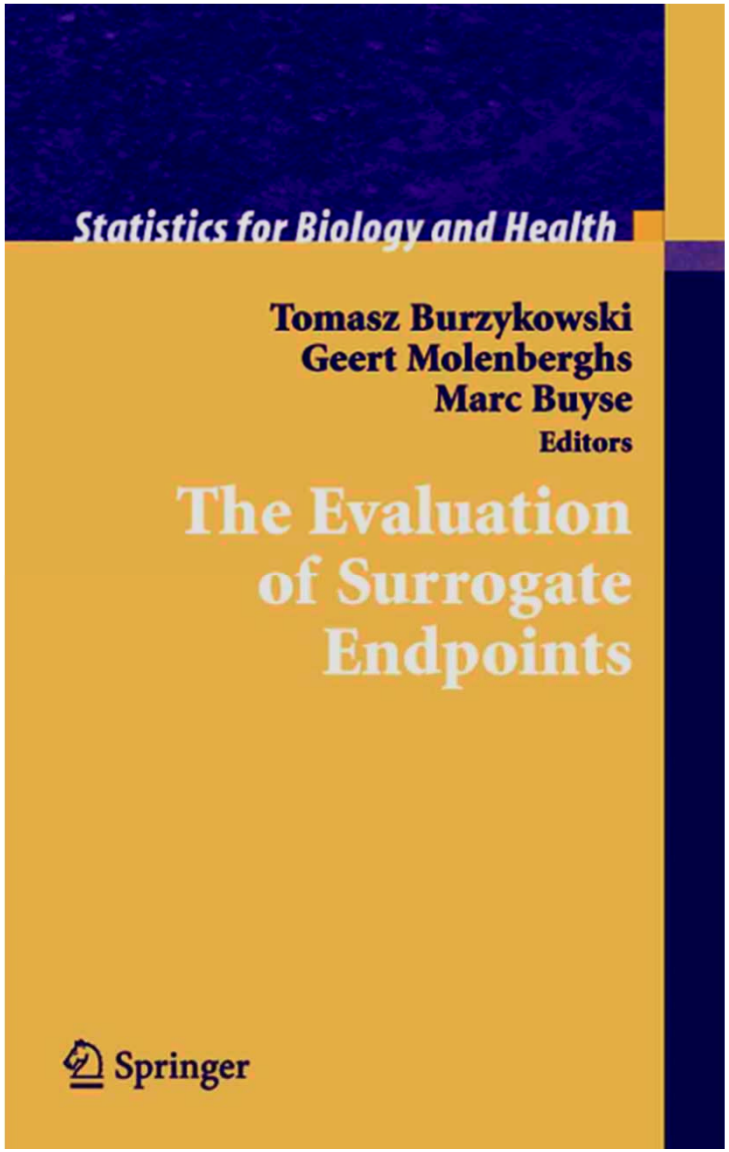
営利目的でのご利用や無断転載はご遠慮ください

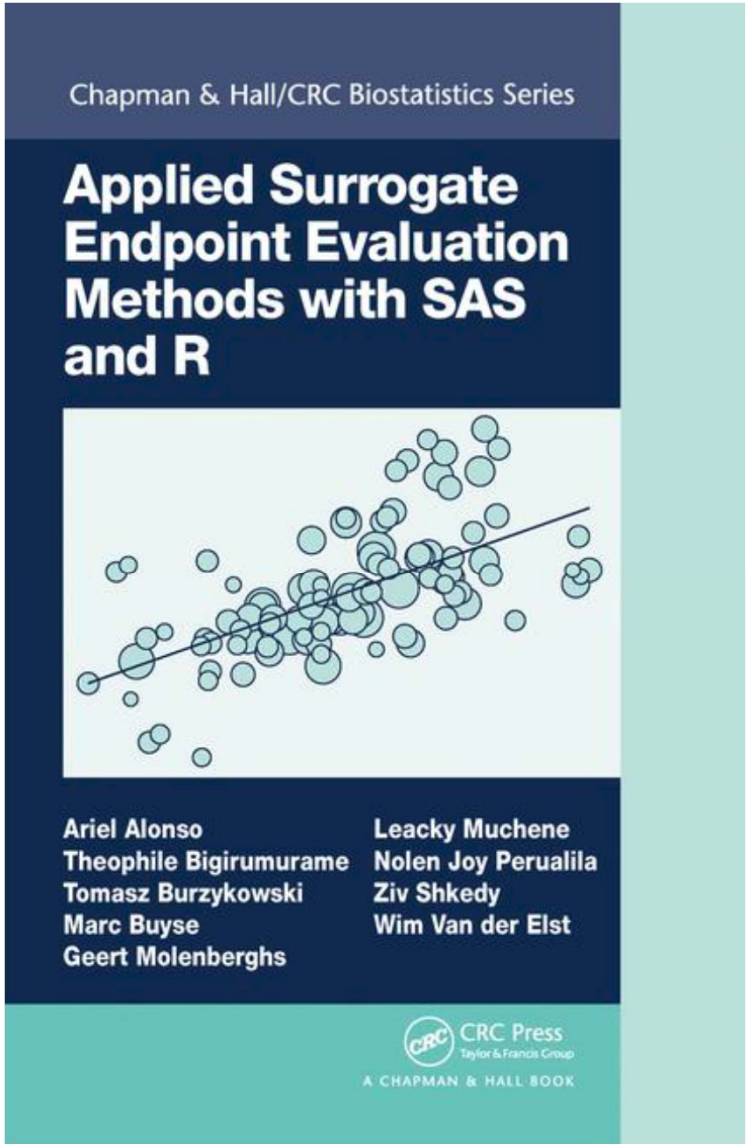
まとめ

- 臨床試験においてエンドポイントの設定は重要
- 代替エンドポイントを利用するためには、その妥当性をデータから評価しておくことが必要
  - 個人レベルの代替性、試験レベルの代替性
  - 個人データに基づくランダム化比較試験のメタアナリシスがゴールドスタンダード
- 代替エンドポイントの妥当性は、患者・場・治療状況に依存するため継続的な評価が必要
  - 継続的な評価を可能とするためのデータベース構築は重要

営利目的でのご利用や無断転載はご遠慮ください

統計手法について詳細に学びたい方へ





営利目的でのご利用や無断転載はご遠慮ください