

データの科学
から見た

臨床研究のダイバーシティ

放射線影響研究所統計部

三角 宗近

misumi@rerf.or.jp

Munechika Misumi, Ph.D.

Department of Statistics,

Radiation Effects Research Foundation





日本糖尿病学会 COI 開示

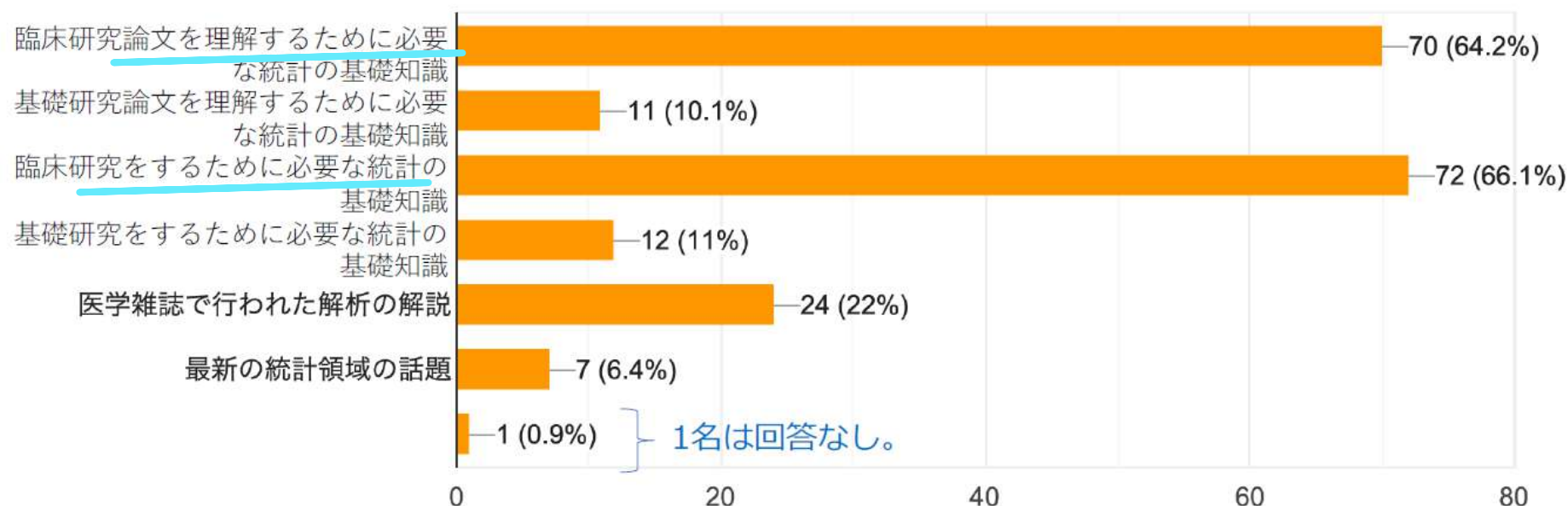
発表者名：三角 宗近

演題発表に関連し、開示すべきCOI関係にある
企業などはありません。

男女共同参画ダイバーシティ企画のワークショップでは、前回に引き続き統計の基礎に関する講演を計画しています。統計に関する事で、さらにどのような内容を希望されますか。
(回答2個まで選択可能)

男女共同参画ダイバーシティ企画のワークショップ...容を希望されますか。(回答2個まで選択可能)

109 件の回答



目的

- ・ データの科学の視点から、臨床研究のダイバーシティを紹介
- ： 臨床研究の多様性を受け入れ、理解することで、臨床研究を理解さらには実践する助けになれば

臨床研究のデザイン

- 統計学でいう「デザイン」はデータの取り方、その計画
- 臨床研究で用いられる主な研究デザイン：
 - 横断研究 (cross-sectional study)
 - 症例対象研究 (case-control study)
 - 縦断研究 (longitudinal study) 、コホート研究
 - コホート内症例対照研究 (nested case-control design)
 - ケースコホート研究

横断研究(Cross-sectional study)

- ・ ある1時点において、目的の母集団から無作為に抽出

目的：その時点における疾患の有無などの測定値をアウトカムとして、同時に得られる要因（リスク因子への曝露など）との関連を調べる

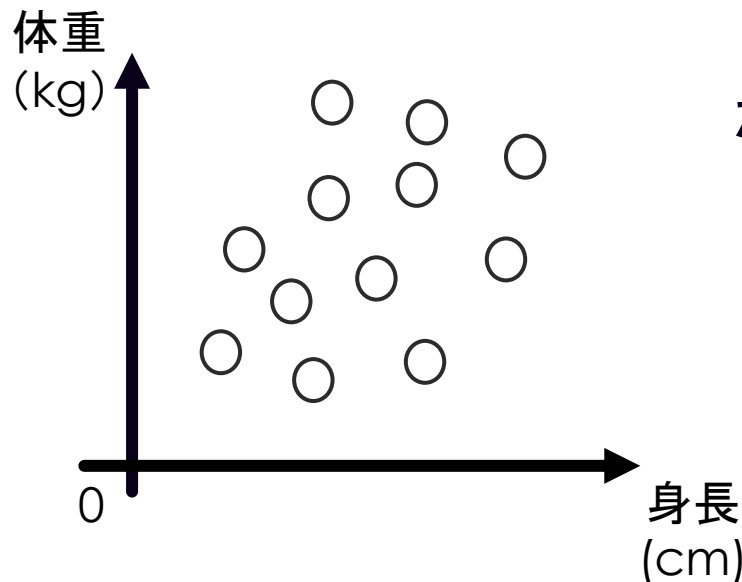
★多くの要因に関する調査を比較的短期間に低コストで行える一方、影響の時間的推移・因果関係の推論は困難

横断研究(Cross-sectional study)

- ある1時点において、目的の母集団から無作為に抽出

アウトカムは疾患の有無のような形だけとは限らない

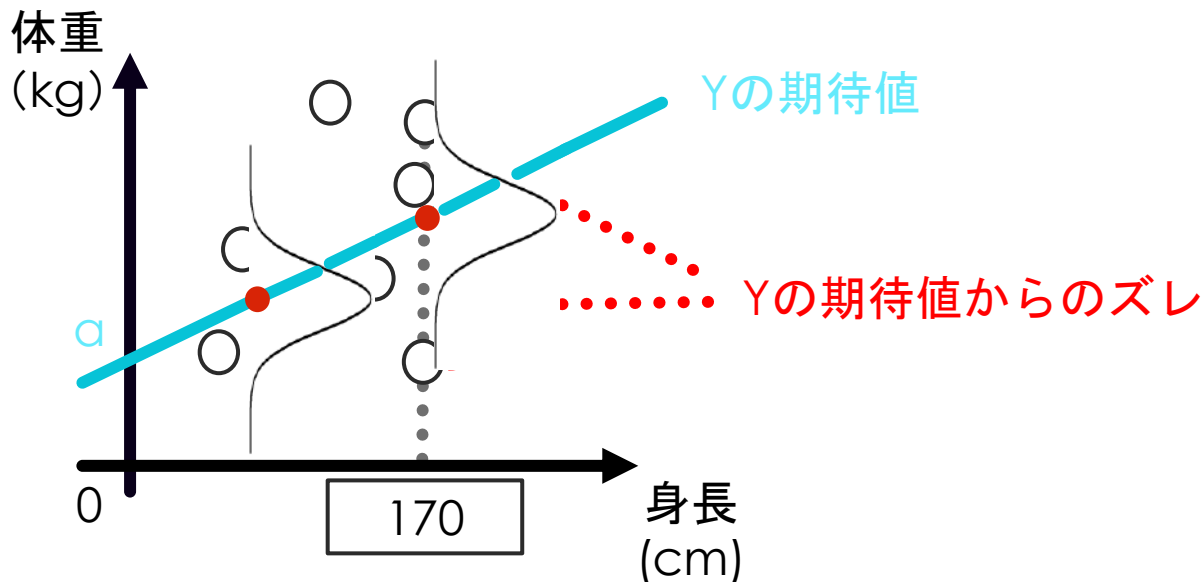
その時点で同時に得られる要因、変数間の関連を調べる



- 例えば、 Y =体重、 X_1 =身長とすると、散布図の各点がひとりの対象者

統計モデル (1)

- 生物統計で用いるモデルは、多くが線形モデル
- 対象者のYにXが関連するか調べたい場合、
 Y の平均値 $= a + cX_1$
という式を考え、集団におけるYの平均値のXとの関連を評価
★個人 (i) のデータ $Y_i = a + cX_{1i} + \text{誤差}$ と仮定
- 例) Y =体重、 X_1 =身長とすると、



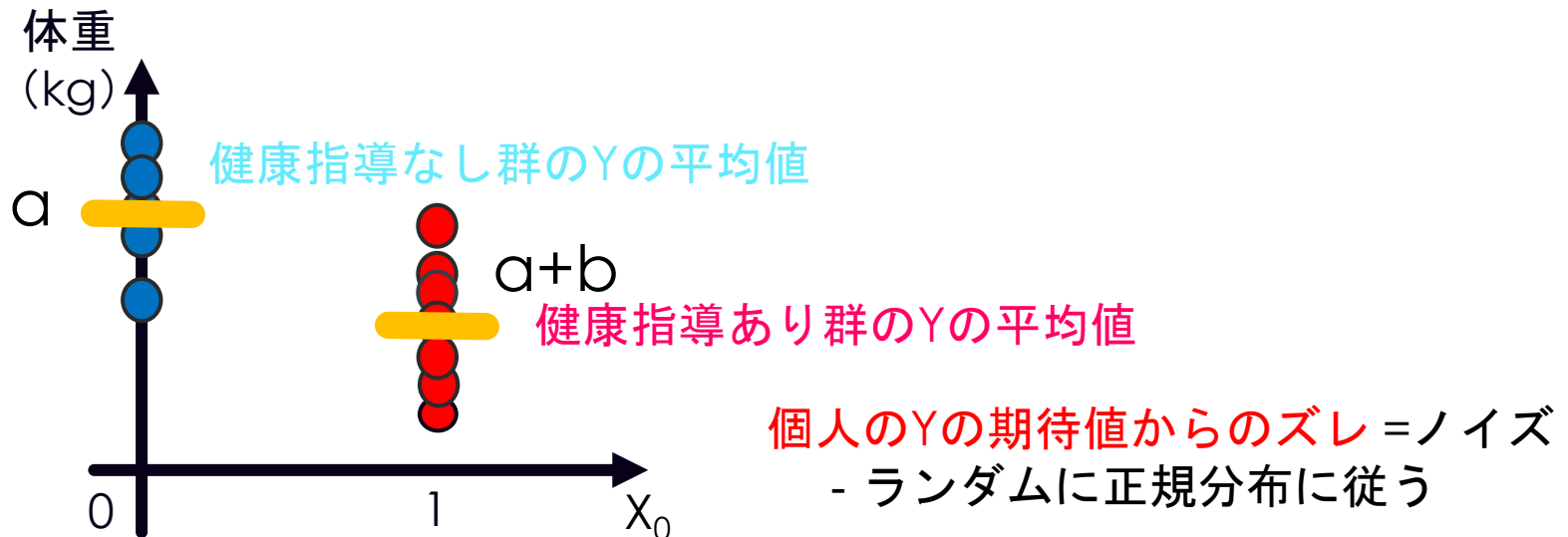
統計モデル (2)

- 生物統計で用いるモデルは、多くが線形モデル
- X (説明変数)として連続量ではなく、対照群に $X_0=0$ 、介入群に $X_0=1$ という値を付けた場合

$$Y \text{ の平均値} = a + bX_0$$

として、集団における Y の平均値と X_0 との関連を評価

- 例) X_0 =**保健指導なし(0)** **あり(1)**、 Y =3か月後の体重

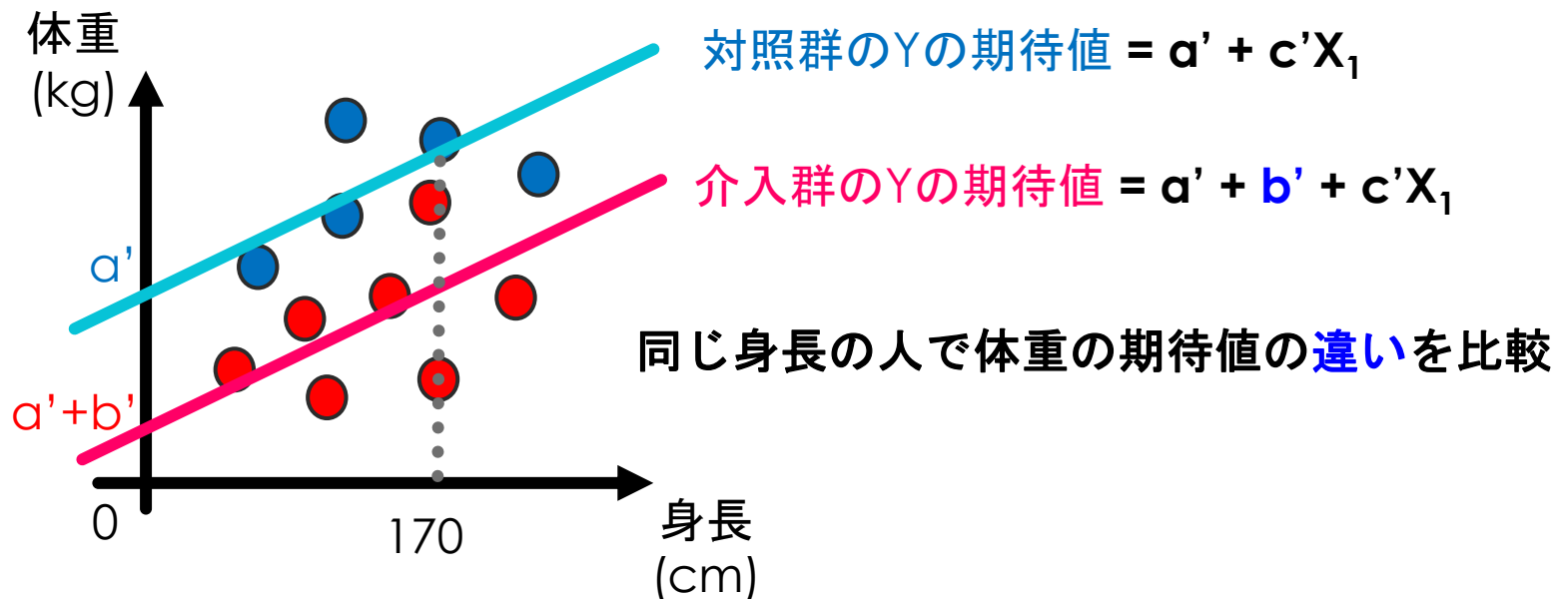


統計モデル (3)

- 介入の効果を身長の影響を調整して評価
 $X_0=0$ を介入なし、 $X_0=1$ を介入あり、 X_1 を身長として

$$Y = \alpha' + b'X_0 + c'X_1 + \text{ノイズ}$$

- 例えば、 $Y=3$ か月後の体重、 $X_0=0$ を健康指導なし、 $X_0=1$ を健康指導あり、 X_1 と身長すると、



統計モデル（4）

✓ 多くの場合、平均値にモデルを仮定

- 割合や確率を用いる場合、対象者が疾患や死亡を起こしたかどうかを2値変数（1=起こした、0=起こしていない）で収集
- 2値変数の集団での平均値をとれば、割合が得られる
割合 P を説明するモデルとして

$$P = \beta_0 + \beta_1 X_1 + \dots$$

とすると、 P は割合なので、0以上1以下の値

→この等号が成り立つパラメター（ β_0 、 β_1 ...）が得られないこともある

ロジスティック回帰

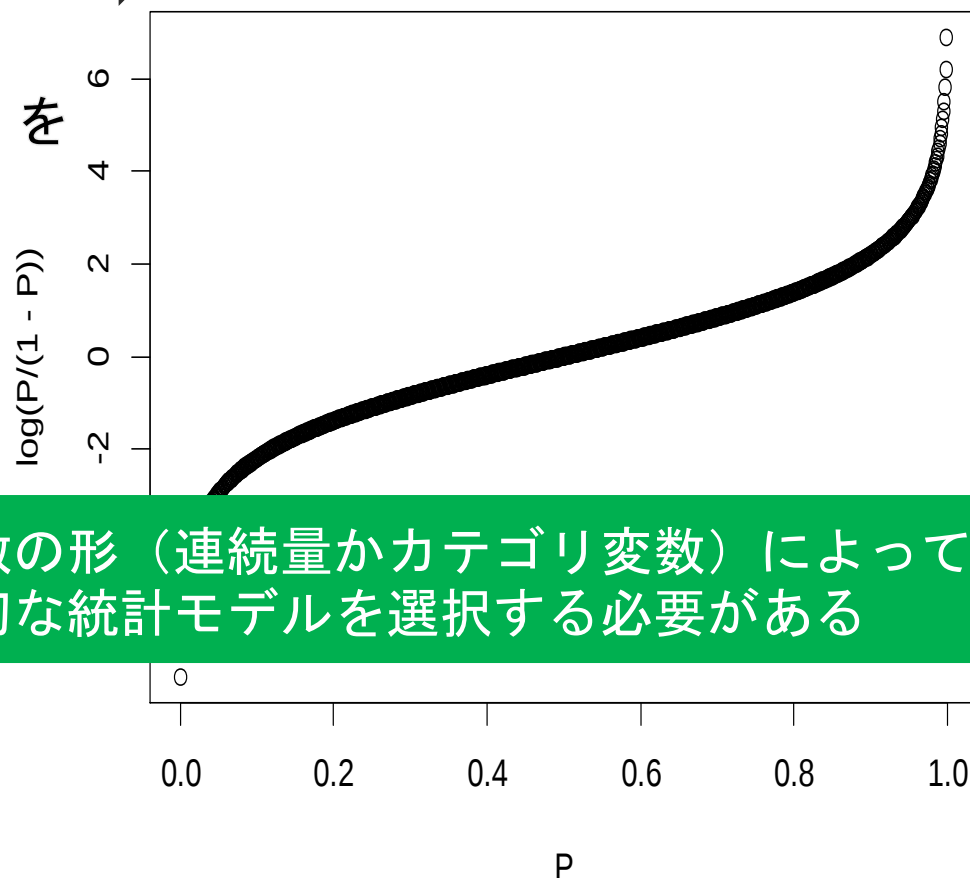
ロジット変換 $\log \frac{P}{1-P}$

→ 範囲が $(-\infty, \infty)$ へ

この変換されたもの（ロジット）を
Yとして $Y = \beta_0 + \beta_1 X_1 + \dots$
と仮定

また簡単な演算から、
以下のようにPについて
以下が得られることもわかる:

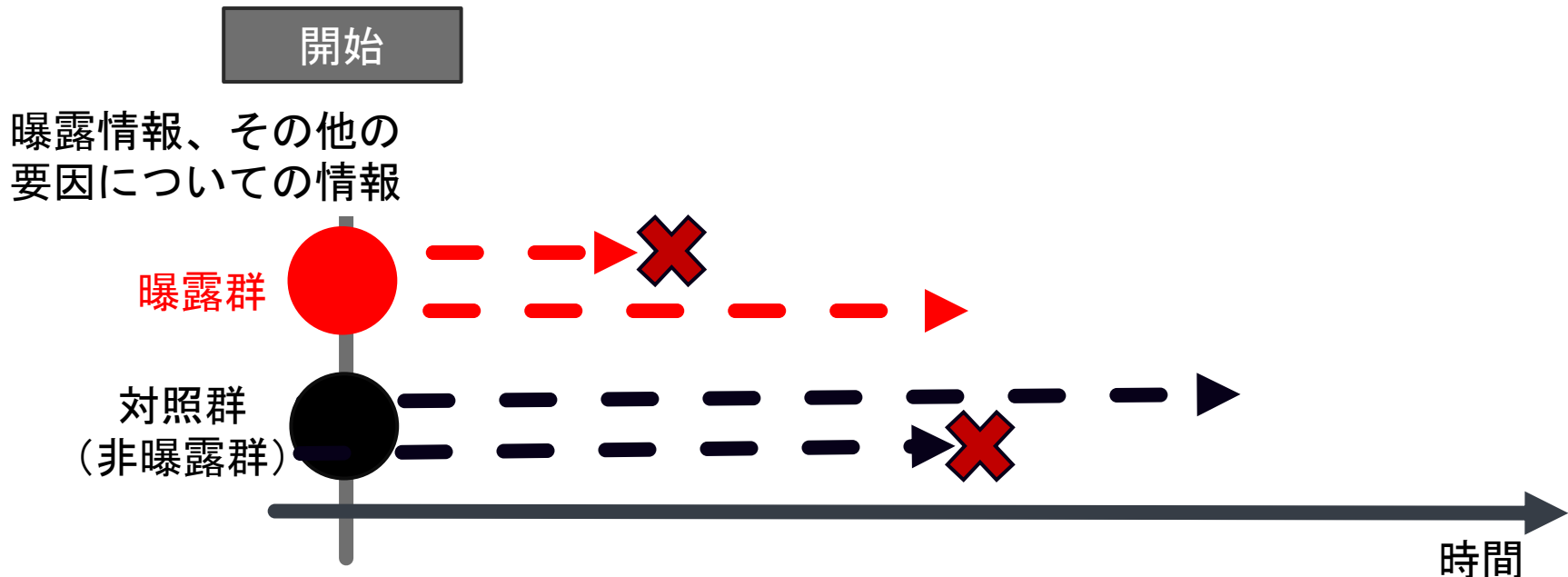
$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots}}$$



コホート研究(Cohort study)

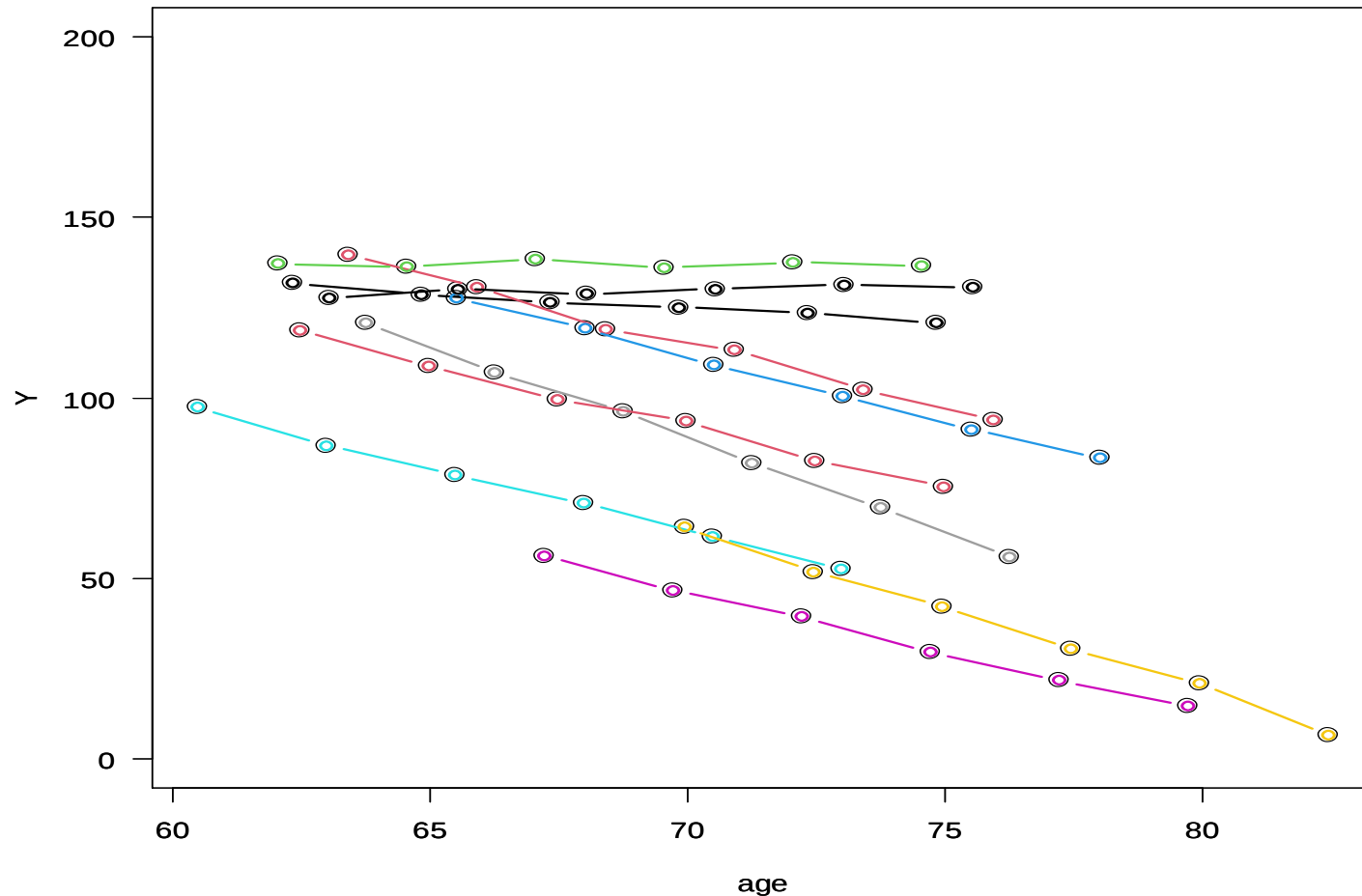
- ・ 設定した集団を一定期間追跡し、疾患の発生や死亡の発生率とリスク因子との関連を調べる

目的： 関心のある処置や曝露を受けた集団と受けていない集団を合わせて追跡し、疾患や死亡などの事象(イベント)の発生率が異なるかどうかを観察する



縦断研究(longitudinal study)

- ・コホート研究のように追跡を行う中で、対象者のデータを繰り返し測定する事で、各対象者の観測値の経時的変化を観察できる

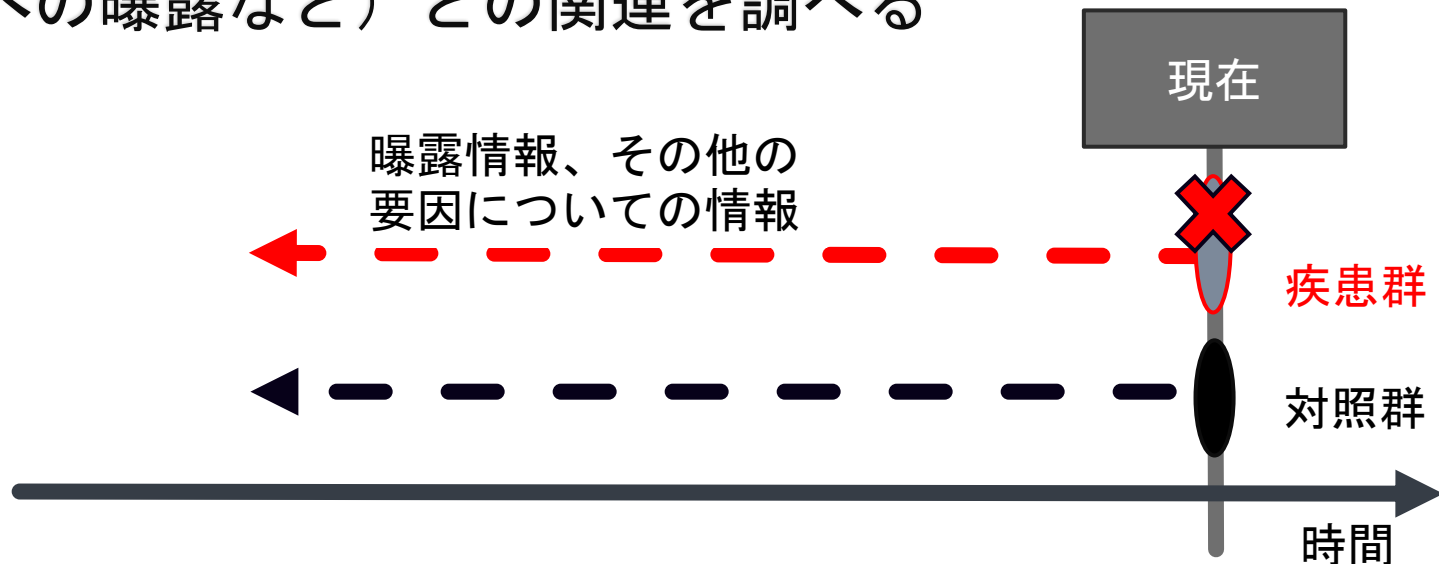


症例対照研究(Case-control study)

- ・ 1時点においてデータを収集するが、研究対象の疾患を持つ人の母集団（症例群）と持たない人の母集団（対照群）からそれぞれ対象者を抽出

後ろ向き研究

目的：その時点における疾患の有無をアウトカム、通常、その時点から過去にさかのぼって得られる要因（リスク因子への曝露など）との関連を調べる



ケース-コントロール研究

- 利点：すでに対象とする疾患を発症した人のデータがあるため、追跡する時間的・経済的コストがかからない（コントロール選択の手間は多少あるかも）
- 問題点：
 - ✓ コントロールの選択バイアス
「イベントを起こさなかった患者」の中からコントロールの対象者を選ぶため、（特に病院などでは）非ケースの患者の中から偏りなく対象者を選ぶのは容易ではない
 - ✓ 測定におけるバイアス
後ろ向きに曝露や交絡の情報を得るため、不完全な記録・記憶による誤差や、欠測が起こりうる

コントロールの選択について

- 性別、年齢を合わせてランダムに
Matched case-control study
- 曝露変数にもとづいた重みを用いた方法
:カウンターマッチング
→コントロールをケースの曝露量のカテゴリと違うカテゴリの曝露群から選択

コホート内ケースコントロール研究 (Nested case-control (NCC) study)

- ケース - コントロール研究は、問題点として挙げたバイアスについて対処できれば、比較的少ない対象者で疾患と曝露（介入）の関連の評価が可能
- コホート研究の中で、前向きにデータを集めながら、コントロールを選択すれば、少ないサンプルで曝露（介入）の効果を評価できないか？

➔ コホート内ケースコントロール研究

NCCの適用はすでに多くの臨床研究論文で見られる

Open Access

Research

BMJ Open Associations with fracture in patients with diabetes: a nested case-control study

diabetesresearchandclinicalpractice.2019.111083

BMJ Open: first published 10

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice



ELSEVIER



Issue 5, P824-834, May

Nested Case-Control Study From

Smith, MD ^{1,3} · Miriam F. Weiss, MD ^{4,5}



The association between antidepressant treatment and rates of insulin initiation in comorbid depression and type 2 diabetes: A UK electronic health record nested case-control study

Annie Jeffery ^{a,*}, Kate Walters ^b, Ian C.K. Wong ^{c,d}, David Osborn ^a, Joseph F. Hayes ^a

^a Epidemiology and Applied Clinical Research Department, Division of Psychiatry, University College London (UCL), London, United Kingdom

^b Department of Primary Care & Population Health, Institute of Epidemiology & Health, University College London (UCL), London, United Kingdom

^c Research Department of Practice and Policy, School of Pharmacy, University College London (UCL), London, United Kingdom

^d Centre for Safe Medication Practice and Research, Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

ARTICLE INFO

Keywords:

Depression
Type 2 Diabetes
Insulin therapy
Antidepressant
Long-term outcomes
Primary Care
Epidemiology

ABSTRACT

Aims: To investigate the association between antidepressant prescribing and the rate of insulin initiation in type 2 diabetes.

Methods: Using UK primary care records we completed a nested-case control study in a individuals with comorbid depression and type 2 diabetes. Cases were defined as individuals initiating insulin, controls were individuals remaining on oral antidiabetic medication. We used conditional logistic regression to estimate incident rate ratios (IRR) and the 95% confidence intervals (CI) for the association between antidepressant prescribing and initiating insulin. We adjusted for demographic characteristics, comorbidities, health service and previous medication use.

Results: We included 11,862 cases who initiated insulin, and 43,452 controls. Increased rates of insulin initiation were associated with any antidepressant prescription (IRR 3.78, 95% CI 3.53–4.04), longer (24+ months) durations of antidepressant treatment (IRR 5.61, 95% CI 5.23–6.03), and higher numbers (3+) of different antidepressant agents prescribed (IRR 5.72, 95% CI 5.25–6.24). There was no difference between recent and non-recent antidepressant prescriptions, or between different antidepressant agents.

Conclusions: Antidepressant prescribing was highly associated with the initiation of insulin therapy. However, this may not indicate a direct causal effect of the antidepressant medication itself, and may be a marker of more

severe depression influencing diabetes. Stored urine samples from DCCT entry and 1-9 years later when macro- or microalbuminuria occurred were measured for the

lysosomal enzyme NAG and the AGE pentosidine and AGE fluorescence. AER and adjustor variables were obtained from the DCCT



prints

< Previous article Next article >

minuria in patients with type 1 diabetes.

s Trial (DCCT).

4 cases were matched to 4 controls in a 1:1 ratio, albuminuria were matched to 110 controls in a 1:2 ratio. The first developed micro-/macroalbuminuria.

end product (AGE) fluorescence, and albumin

macroalbuminuria (AER > 300 mg/d).

Figures (1)

Figure Viewer



Article metrics

81

Citations

58

Captures

PLUMX

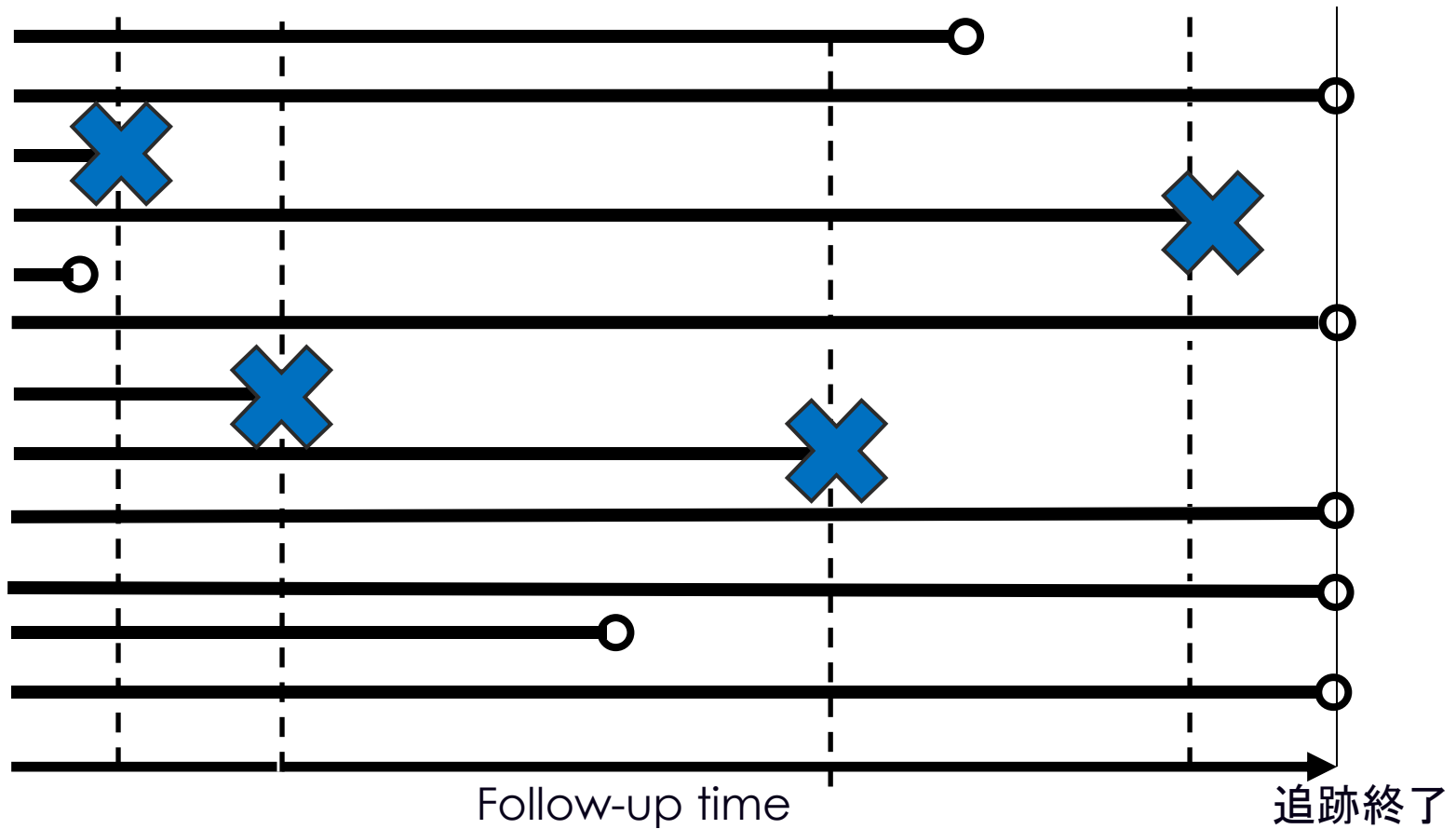
View details >

Related articles (40)

Social Isolation, Loneliness, and Risk of Microvascular Complications Among Individuals With Type 2 Diabetes Mellitus

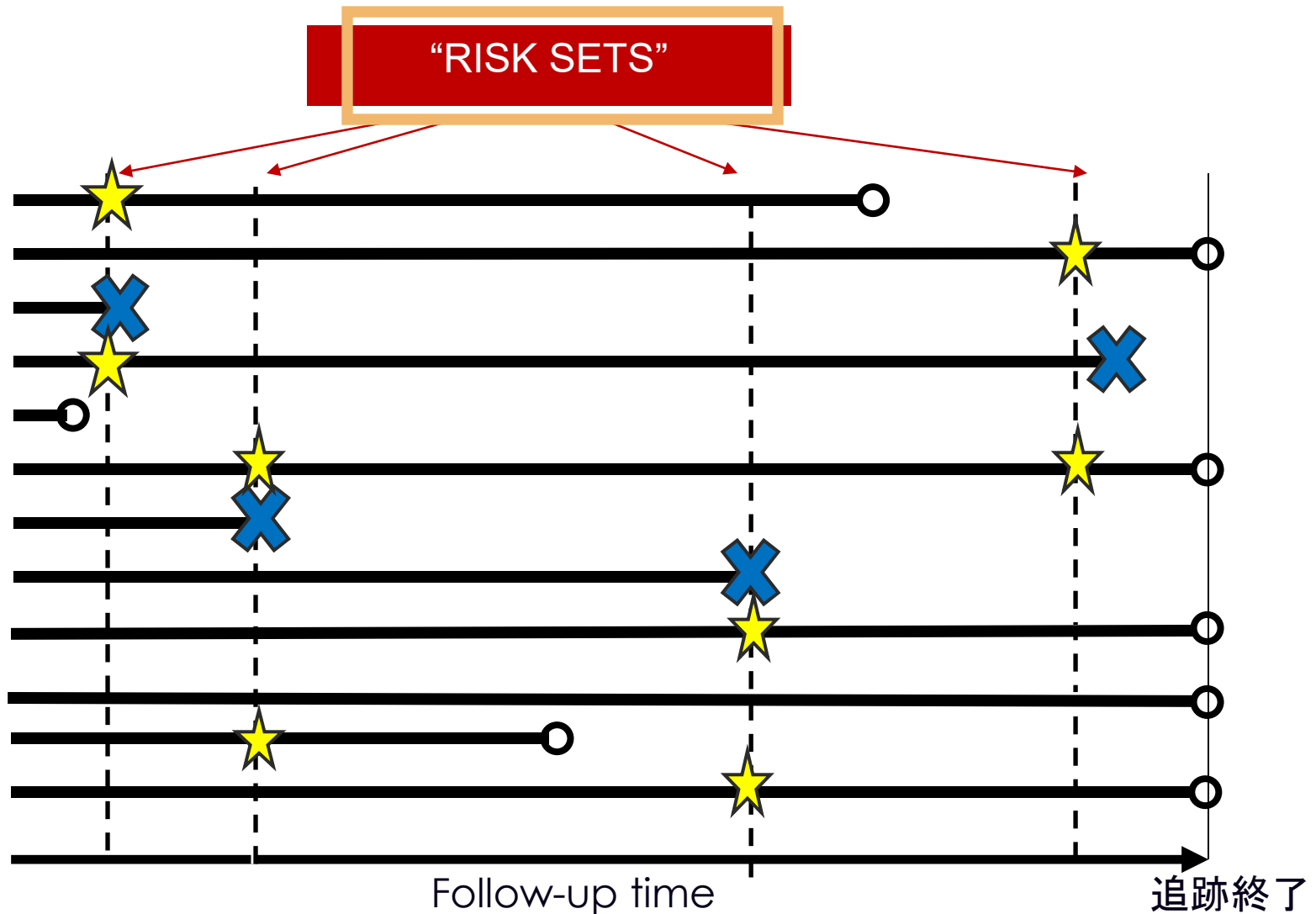
Liang et al.
American Journal of Kidney Diseases,
June 24, 2024

コホート研究



コホート内ケースコントロール研究

Nested case-control study



コホート内ケースコントロール研究

- ✓ コホートの中で、サブサンプルを構成するので、コホート内ケースコントロール研究のサンプルサイズは減る→コホート全体で解析するよりも統計的検出力は下がる
- ✓ コホート内ケースコントロール研究が対象とするのは、特に、測定にコストがかかる変数をデータ解析に含める必要があるなど、制約がある場合になるべく高い検出力を保持するためのデザイン
- ✓ この名前だけど、前向き研究！
追跡の中でMatched case-control研究を行う
(マッチングによる交絡調整)

NCC 統計解析の方法

- ✓ 条件付ロジスティック回帰 (Thomas, DC (1977))

$$\Pr(Y = 1) = \frac{e^{\beta_{0i} + \beta_1 x_1 + \dots}}{1 + e^{\beta_{0i} + \beta_1 x_1 + \dots}}$$

- ✓ Matched case-control研究での多変量解析と同じ
- ✓ β_{0i} は、マッチングした組ごとに異なる切片
 - ➔ マッチした組の中の交絡因子はこの項に集約され自動的に調整される
- ✓ 推定されるのはオッズ比...だがパラメタ(β_1)の推定値は...

NCC 統計解析の方法

- ✓ Cox回帰（比例ハザードモデル）

$$h(t) = h_0(t)e^{\gamma_0 + \gamma_1 x_1 + \dots}$$

で推定されるパラメタ(γ_1)と同じなる

(Goldstein and Langholz (1992))

（もとのフルコホートにおいて比例ハザード性が成り立ち、

Cox回帰モデルが正しいモデルである場合）

- ✓ イベントが稀でなくても相対リスクの指標として解釈できる（Cox回帰では相対リスクの近似値としてハザード比を推定）

NCC での注意点

- ✓ ある時点で発生したケースに対してコントロールとしてサンプリングされた人が、のちにイベントを起こしケースとなる場合がある
 - ✓ ある時点でコントロールになった人が、別の時点で別のケースに対するコントロールになる場合がある
 - ✓ 同じ人が異なるMatchedグループに存在することになる
- ➔ (Q) 重複が起こらないように、1度サンプリングされた人はリスクセットから除外すべき？

NCC での注意点

→ (A) No.

コホートの追跡履歴からリスクセットごとに偏りなくコントロールがサンプリングされた場合にハザード比が正しく推定される

- ✓ 重複を避ける操作がサンプリングにバイアスをもたらす
- ✓ 重複が起こってもその方が妥当な（偏りのない）サンプリング

NCC のlimitation (?)

◆ アウトカムごとにコントロールを選ぶので、複数のアウトカムに関心がある場合、アウトカムごとにデザインが必要

- 全対象者（フルコホート）について高価な測定値を得るより効率的に実施できる可能性はある

しかし、（特にコホート研究を行っている施設で）もっと実用的なデザインがないか...

ケースコホート研究

- ◆ 複数のアウトカムに関心がある場合、NCCで毎回コントロールの選択が必要になる問題を克服するために考えられた (Prentice, 1986)
- 全集団からコントロールの選択を行ってケースにぶつける、というデザイン
 - ➔ まず、サブコホートを形成（ランダムサンプリング、層別ランダムサンプリング、など）
- コントロールは特定のイベントに限定されないため、1度のデザイン（サンプリング）で複数のアウトカムについて共通のコントロールとして機能する

ケースコホート研究

RESEARCH

 OPEN ACCESS



Association of plasma biomarkers of fruit and vegetable intake with incident type 2 diabetes: EPIC-InterAct case-cohort study in eight European countries

Ju-Sheng Zheng,^{1,2} Stephen J Sharp,¹ Fumiaki Imamura,¹ Rajiv Chowdhury,³ Thomas E Gundersen,⁴ Marinka Steur,¹ Iivonne Sluijs,⁵ Yvonne T van der Schouw,⁵ Antonio Agudo,⁶ Dagfinn Aune,^{7,8,9} María-Dolores Chirlaque,^{12,13} Miren I Paul W Franks,¹⁷ Guy Fagherazzi,^{16,18} Verena Katzke,²¹ Tilman Kühn,²¹ Kay Peter M Nilsson,¹⁷ Kim Overvad,^{24,25} Olov Rolandsson,²⁹ Daniel Redondo Matthias B Schulze,^{34,35,36} Annemiek Tammy Y N Tong,²⁷ Rosario Tumino,³ Adam S Butterworth,^{3,39} Elio Riboli,⁷

OPEN ACCESS Freely available online

PLOS MEDICINE

Long-Term Risk of Incident Type 2 Diabetes and Measures of Overall and Regional Obesity: The EPIC-InterAct Case-Cohort Study

The InterAct Consortium^{*}

For numbered affiliations see end of the article.

Correspondence to: N G Forouhi nita.forouhi@mrc-epid.cam.ac.uk (ORCID 0000-0002-5041-248X)

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

Cite this as: *BMJ* 2020;370:m2194 <http://dx.doi.org/10.1136/bmj.m2194>

Accepted: 2 April 2020

ABSTRACT

OBJECTIVE

To investigate the association of plasma vitamin and carotenoids, as indicators of fruit and vegetable intake, with the risk of type 2 diabetes.

DESIGN

Prospective case-cohort study.

SETTING

Populations from eight European countries

PARTICIPANTS

9754 participants with incident type 2 diabetes; a subcohort of 13 662 individuals from the Prospective Investigation into Cancer and Nutrition (EPIC) cohort of 340 234 participants; EPIC case-cohort study.

Abstract

Background: Waist circumference (WC) is a simple and reliable measure of fat distribution that may add to the prediction of type 2 diabetes (T2D), but previous studies have been too small to reliably quantify the relative and absolute risk of future diabetes by WC at different levels of body mass index (BMI).

Methods and Findings: The prospective InterAct case-cohort study was conducted in 26 centres in eight European countries and consists of 12,403 incident T2D cases and a stratified subcohort of 16,154 individuals from a total cohort of 340,234 participants with 3.99 million person-years of follow-up. We used Prentice-weighted Cox regression and random effects meta-analysis methods to estimate hazard ratios for T2D. Kaplan-Meier estimates of the cumulative incidence of T2D were calculated. BMI and WC were each independently associated with T2D, with WC being a stronger risk factor in women than in men. Risk increased across groups defined by BMI and WC; compared to low normal weight individuals (BMI 18.5–22.4 kg/m²) with a low WC (<94/80 cm in men/women), the hazard ratio of T2D was 22.0 (95% confidence interval 14.3; 33.8) in men and 31.8 (25.2; 40.2) in women with grade 2 obesity (BMI ≥ 35 kg/m²) and a high WC (>102/88 cm). Among the large group of overweight individuals, WC measurement was highly informative and facilitated the identification of a subgroup of overweight people with high WC whose 10-y T2D cumulative incidence (men, 70 per 1,000 person-years; women, 44 per 1,000 person-years) was comparable to that of the obese group (50–103 per 1,000 person-years in men and 28–74 per 1,000 person-years in women).

Conclusions: WC is independently and strongly associated with T2D, particularly in women, and should be more widely measured for risk stratification. If targeted measurement is necessary for reasons of resource scarcity, measuring WC in overweight individuals may be an effective strategy, since it identifies a high-risk subgroup of individuals who could benefit from individualised preventive action.

Please see later in the article for the Editors' Summary.

Citation: The InterAct Consortium (2012) Long-Term Risk of Incident Type 2 Diabetes and Measures of Overall and Regional Obesity: The EPIC-InterAct Case-Cohort Study. *PLoS Med* 9(6): e1001230. doi:10.1371/journal.pmed.1001230

Case-cohort study

Full
cohort



⋮

Case-cohort study

select subjects for the sub-cohort



Case-cohort study

select subjects for the sub-cohort



Case-cohort study

select subjects for the sub-cohort



⋮

Case-cohort study

selected sub-cohort

○

○

○

○

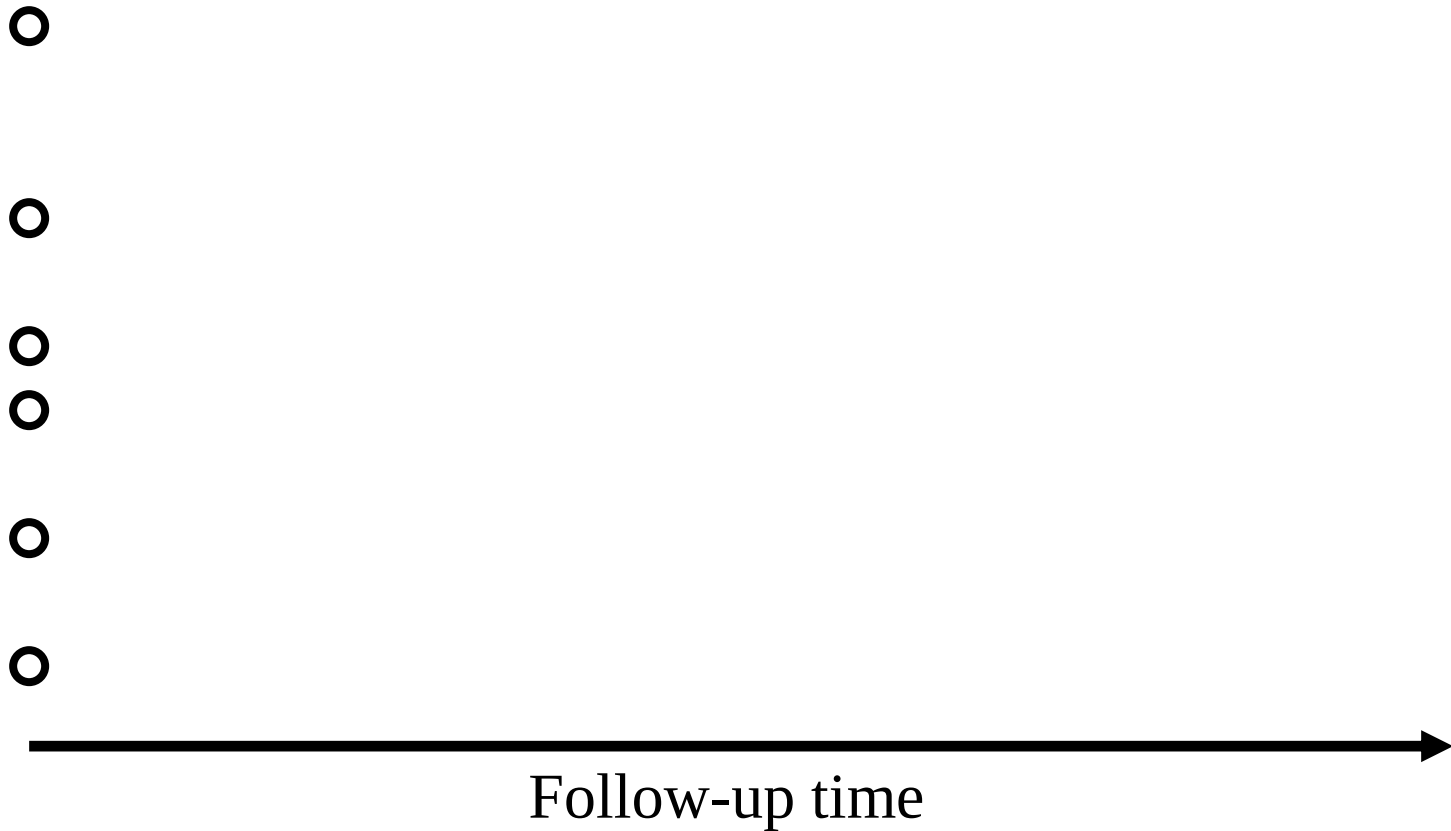
○

○

⋮

repeat

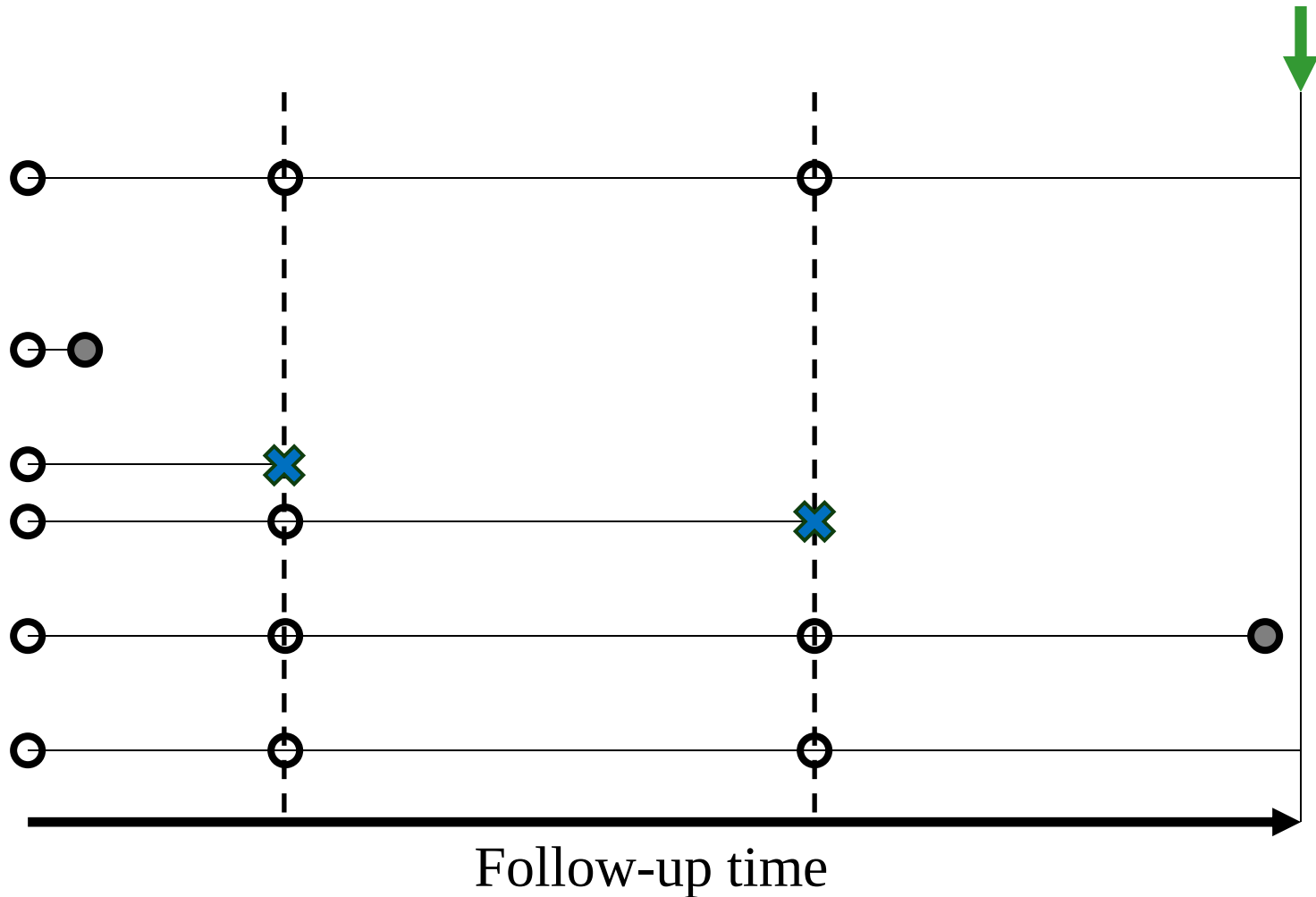
Case-cohort study



Case-cohort study

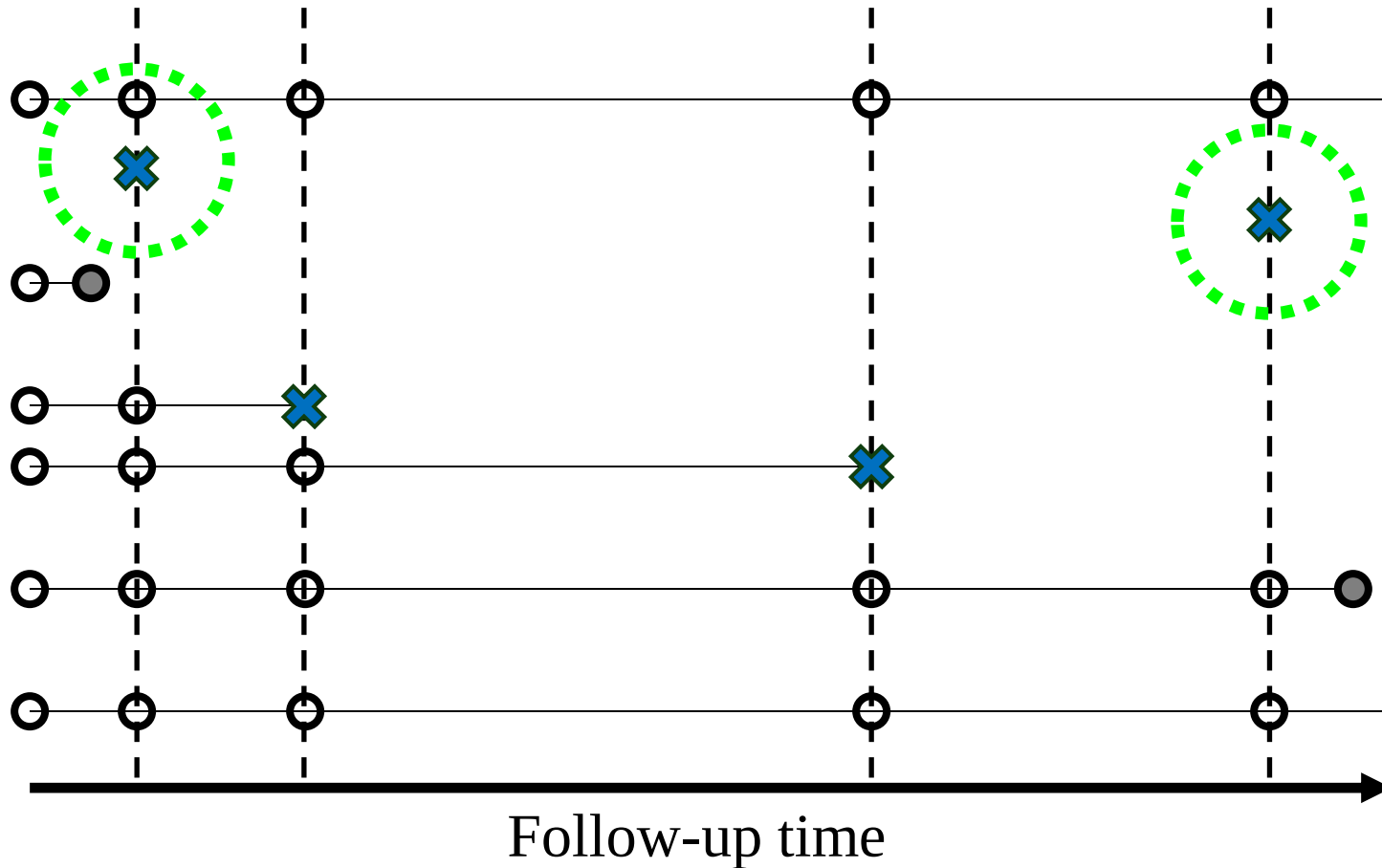
● = 脱落

End of
follow-up



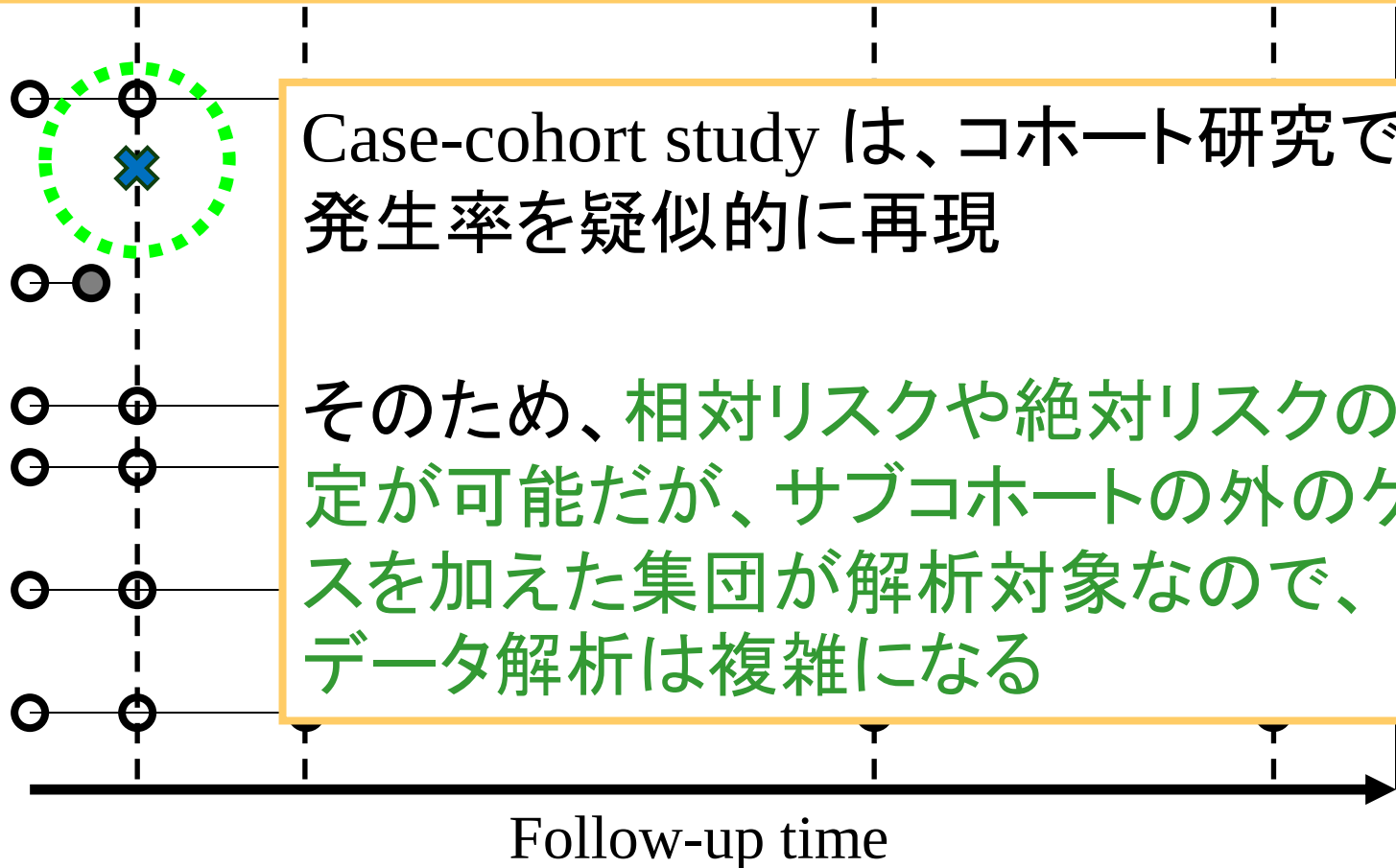
Case-cohort study

サブコホート外のイベントも解析対象に加える



Case-cohort study

サブコホートの外で起こったイベントは他のリスクセットに寄与していないことに注意



生存時間解析

- Cox回帰（比例ハザードモデル）で解析する...が、Case-cohort研究の解析データは、コホートからのランダムサンプルではない
- ➡ 重みづけを行うなど、通常のCox回帰とは少し異なる方法論が必要

まとめ

- データ分析は、
 - ✓ どのようにデータが取られたか（研究デザイン）
 - ✓ どのようなデータが取られたか（データの種類）による多様な方法が存在
- データの重要性が認識されるようになって、臨床研究でも生物統計家の必要性が定着してきた
- 臨床研究もチーム戦：専門性のダイバーシティを促進することが成功の鍵

ご清聴ありがとう
ございました