ISPOR Symposium:

Leveraging real-world evidence for health technology assessment – Using big data to enable patient access

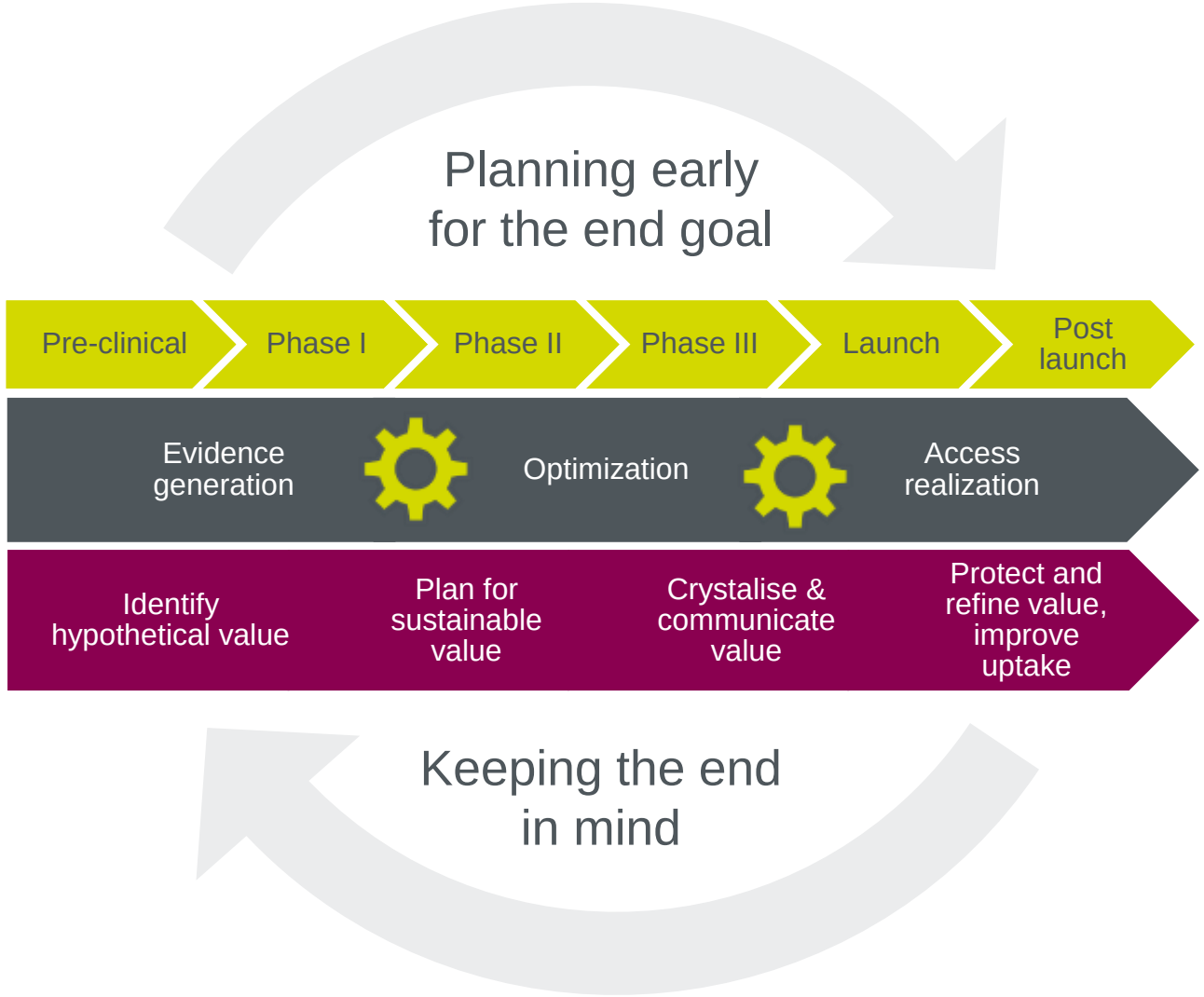
7 Nov 2022

RWE Across the Product Development Lifecycle

In 2019, 49% of FDA-approved NDAs and BLAs included RWE. In 2020, that figure jumped to 75%!

RWE included in 40% of MAAs (2018-19)

> *"Early planning with end goals in mind ensures informed and proactive decision making rather than reactive, sub-optimal responses"*



Today's Agenda

Considerations for analyzing RWD, current evidence gaps in HTA, and analytic approaches to address these gaps

Jackie Vanderpuye-Orgle, MSc, PhD
Vice President, Global Head Advanced Analytics, Parexel International, United States

Overview of NICE guidelines and other HTA agency requirements on the use of RWE

Emtiyaz Chowdhury, BSc, MA
Principal Consultant, Health Economics, Parexel International
NICE HST Committee Member, United Kingdom

Study on estimating cost-effectiveness using cancer registry data

Nicholas Latimer, MSc, PhD
Professor of Health Economics, Yorkshire Cancer Research Senior Fellow, University of Sheffield, Sheffield, UK

Using RWE to augment clinical trial data

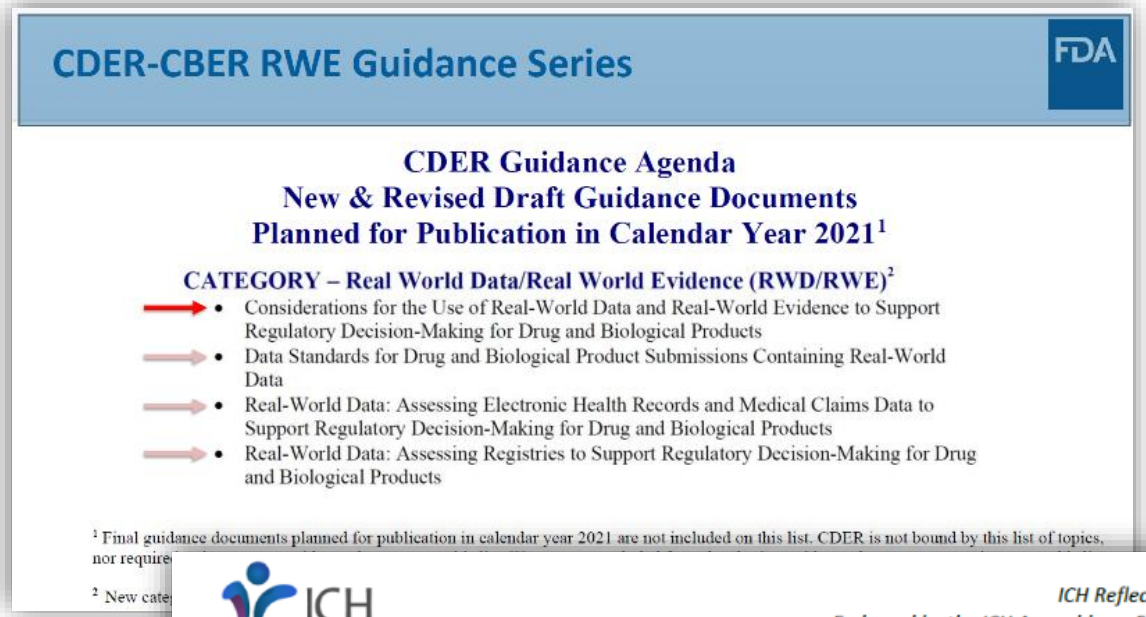
Gianluca Biao, PhD
Professor at University College London
United Kingdom

Considerations for analyzing RWD, current evidence gaps in HTA, and analytic approaches to address these gaps

Jackie Vanderpuye-Orgle

parexel

Regulatory guidelines continue to be developed to help shape the RWE landscape



CDER-CBER RWE Guidance Series

FDA

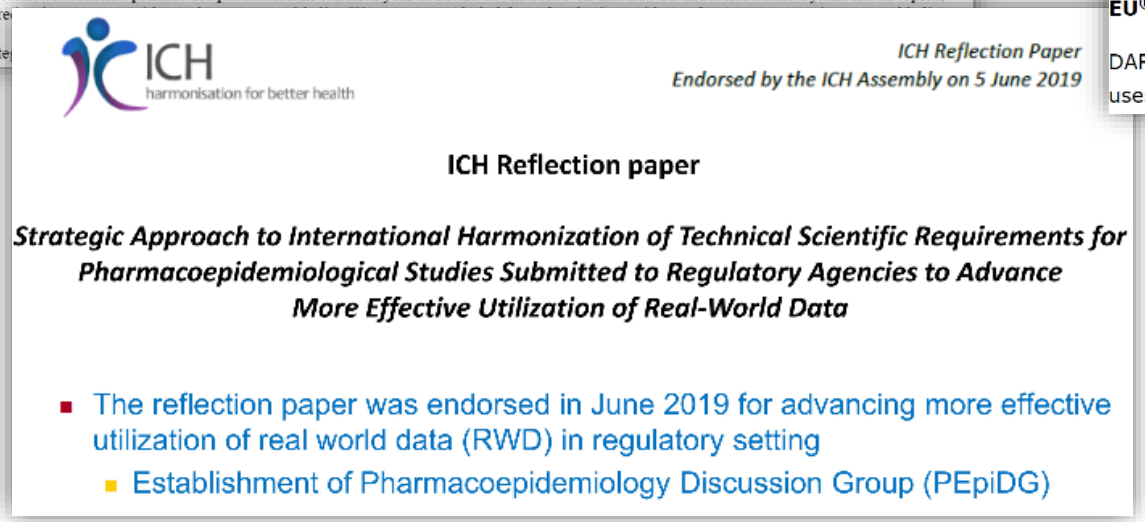
CDER Guidance Agenda
New & Revised Draft Guidance Documents
Planned for Publication in Calendar Year 2021[1]

CATEGORY – Real World Data/Real World Evidence (RWD/RWE)[2]

- Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products
- Data Standards for Drug and Biological Product Submissions Containing Real-World Data
- Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products
- Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products

[1] Final guidance documents planned for publication in calendar year 2021 are not included on this list. CDER is not bound by this list of topics, nor require

[2] New cate



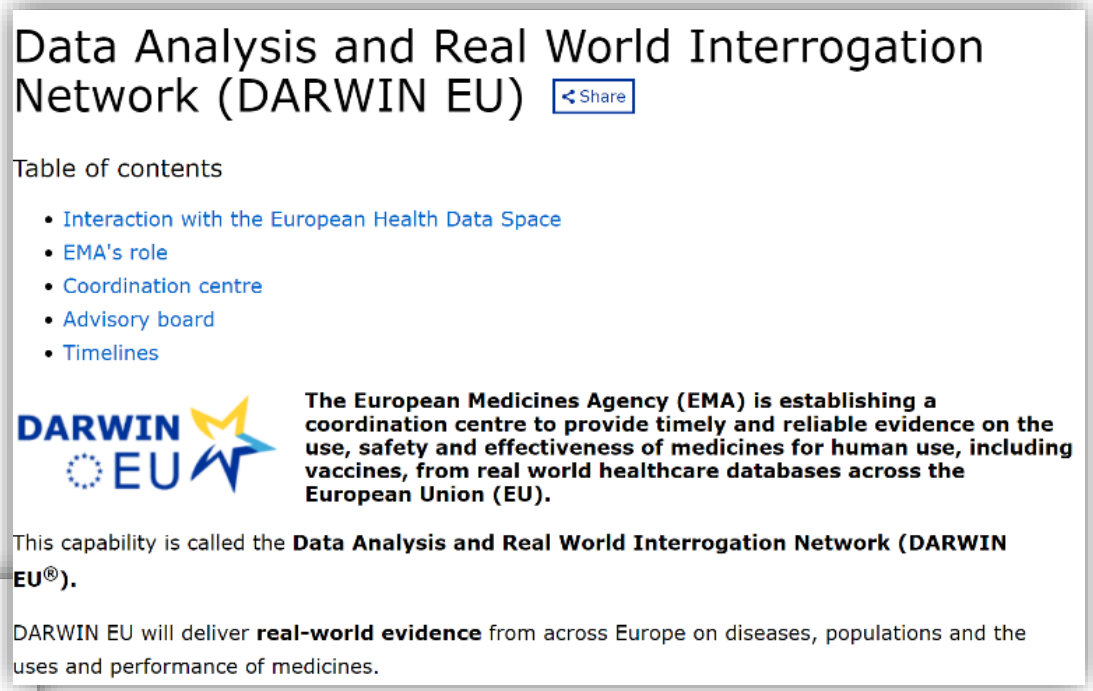
ICH harmonisation for better health

ICH Reflection Paper
Endorsed by the ICH Assembly on 5 June 2019

ICH Reflection paper

Strategic Approach to International Harmonization of Technical Scientific Requirements for Pharmacoepidemiological Studies Submitted to Regulatory Agencies to Advance More Effective Utilization of Real-World Data

- The reflection paper was endorsed in June 2019 for advancing more effective utilization of real world data (RWD) in regulatory setting
  - Establishment of Pharmacoepidemiology Discussion Group (PEpiDG)



Data Analysis and Real World Interrogation Network (DARWIN EU) Share

Table of contents

- Interaction with the European Health Data Space
- EMA's role
- Coordination centre
- Advisory board
- Timelines

DARWIN EU

The European Medicines Agency (EMA) is establishing a coordination centre to provide timely and reliable evidence on the use, safety and effectiveness of medicines for human use, including vaccines, from real world healthcare databases across the European Union (EU).

This capability is called the **Data Analysis and Real World Interrogation Network (DARWIN EU®).**

DARWIN EU will deliver **real-world evidence** from across Europe on diseases, populations and the uses and performance of medicines.

Trends in regulators' view of RWE

- Treatment for rare and/or life-threatening diseases: including synthetic / external controls
- Revised indications or drug-combination labeling
- Post-marketing evaluation: More-comprehensive assessment of safety, effectiveness of approved drugs, and to refine decision making
- Guiding clinical trial design: I/E, endpoints
- Identify the target population: Precision medicine and patient subgroups

Examples of regulatory approval based on RWE

- EMA
  - Yescarta (axicabtagene ciloleucel)
- FDA
  - Blincyto (blinatumomab)
  - Brineura (cerliponase alfa)
  - Ibrance (palbociclib)

Using real-world evidence for the purpose of product registration will require adequate communication in advance with regulatory authorities to ensure alignment on the study objectives and methodology

The Real World is messy

BUT - Adding controls and other features impacts the reality we're trying to measure

Target Trial Framework The standard for deploying RW

"The planner of a [non-randomized] study should always ask himself the question: 'How would the study be conducted if it were possible to do it by controlled experimentation' (Cochran, 1965)

American Journal of Epidemiology
© The Author 2016. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 183, No. 8
DOI: 10.1093/aje/kwv254
Advance Access publication: March 18, 2016

Practice of Epidemiology

Am J Epidemiol. 2016 Apr 15;183(8):758-64.

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

- Framework presented and discussed at FDA sponsored workshop with Duke Margolis Center for Health Policy (Feb 2021)
- While developed for comparative observational studies, bias in an RWD-generated ECA may exist without an explicit target trial (e.g., defining 'time zero' for the beginning of patient follow-up)
- Sponsors should anticipate FDA incorporation of this framework into their upcoming 2021 RWD/E guidance and/or decision-making

- **Why is it important to emulate a randomized clinical trial?**
- **Because not doing so leads to bias**
- **Deviations from the target trial are sources of bias in observational analyses**

Source: Presentation by Miguel Hernán

- **Step 1: Specify**
- **Step 2: Emulate**

- Eligibility criteria
- Enrollment strategies
- Randomized assignment
- Start/end follow-up
- Outcomes
- Causal contrast
- Analysis plan

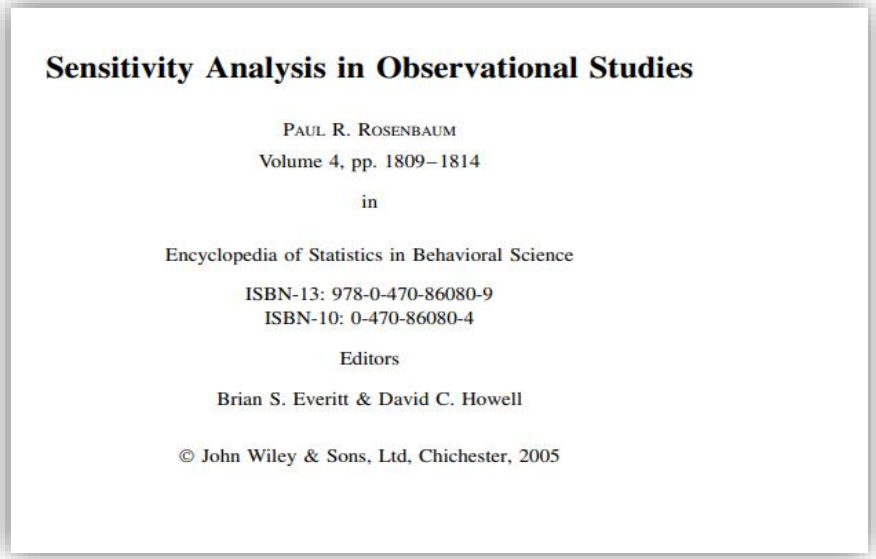
Neyman-Rubin Potential Outcomes Model

Under Neyman-Rubin, the potential outcomes that could be observed for each unit are:

- Potential outcome under treatment: the outcome that would be observed if a unit gets the treatment, Y (T = 1) = Y (1)
- Potential outcome under control: the outcome that would be observed if that unit gets the control Y (T = 0) = Y (0)

- Causal inference can be seen as a missing data problem because Yi1 and Yi0 are never both observed

Addressing unobservable characteristics and improving precision



Sensitivity Analysis in Observational Studies

PAUL R. ROSENBAUM

Volume 4, pp. 1809–1814

in

Encyclopedia of Statistics in Behavioral Science

ISBN-13: 978-0-470-86080-9
ISBN-10: 0-470-86080-4

Editors

Brian S. Everitt & David C. Howell

- There is usually the concern that some important baseline differences **were not measured**, so that individuals who appear comparable may not be
- A sensitivity analysis in an observational study addresses this possibility: it asks *what the unmeasured covariate would have to be like to alter the conclusions of the study*

Best practice recommendations are driven by two components:

- Availability of information on the unmeasured confounders, and;
- Objectives for assessing unmeasured confounding.

- Rosenbaum's Gamma (Γ) - strength of unobserved treatment assignment mechanism
- Targeted maximum likelihood estimation (TMLE)

Advanced approaches to bridging machine learning to statistical and causal inference

- **Targeted learning** is an algorithm for the construction of double-robust, semiparametric, efficient substitution estimators.
- Allows for data-adaptive estimation while supporting valid statistical inference.
- Does use the G-computation estimand (G-formula).
- Incorporates the ensemble **Super Learner**, which is a weighted composite model from a library of algorithms. Candidate learners and weights are chosen to minimize the cross-validated empirical risk loss function.

ECAs: Key decision points and considerations for establishing an external control

Statutory & FDA Criteria
- Adequate and well-controlled
- Large treatment effect
- Well known natural history of indication

Evidence Objectives
- Early specification of primary objective and endpoint is necessary to conduct feasibility assessment

Placement of Study
- In RCT protocol and SAP
- External to RCT protocol and SAP; develop a separate protocol and SAP for comparison of EC cohort to clinical trial treatment cohort

Variables
- Specify and operationally define all outcomes, endpoints, confounders
- Identify variables required for cohort balancing or comparative analyses
- Develop data abstraction rules to be used for medical informatics assessment

Confounders
- Solicit Internal Medical Expertise
- Search Literature
- Utilize Directed Acyclic Graphs
- Engage Key Opinion Leaders

Analytic Methods
- Important to decide early in decision-making process
- Cohort balancing methods
- FDA and Payer expectations

Site Selection
- All clinical trial sites
- Subset of clinical trial sites
- Non-trial sites experienced in indication of interest
- Combination of above

Patient Identification
- Site-based investigator with knowledge of prior treated patients
- Institution-wide search of EHR system by site investigator and/or or EHR systems administrator

Site & Patients Sampling
- Probability sampling
- Convenience sampling

Case Ascertainment
- Necessary for indications without an ICD-10 code
- Recommended in all other cases

Data Collection
- Data abstraction from patients' medical records
- Obtain electronic extract of identified patients' data from EHR-system

Patients
- Patients with a diagnosis in medical records
- Patients without a diagnosis in medical records but meeting diagnostic algorithm
- Combination of above

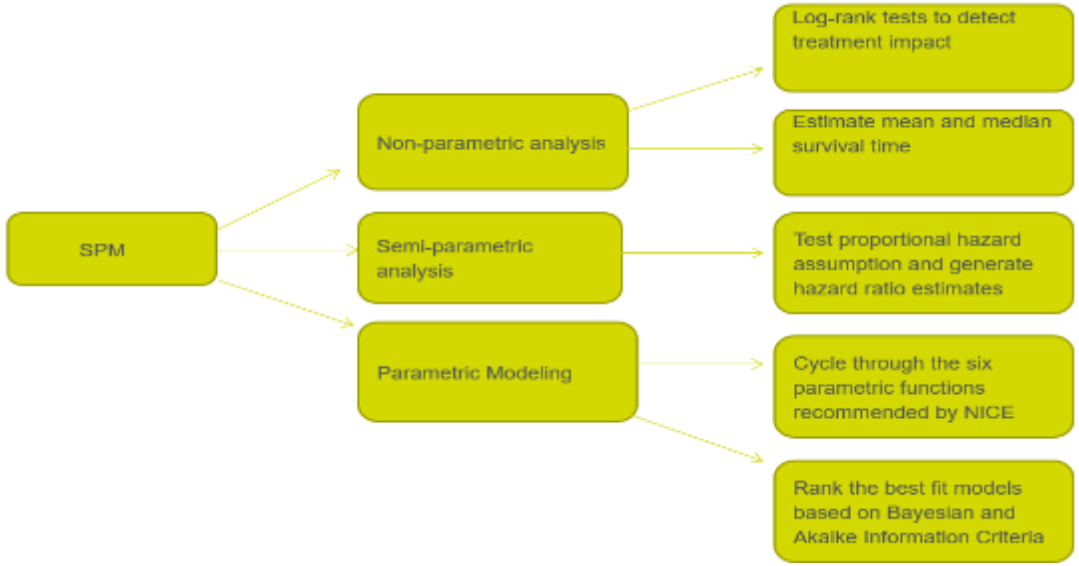
Algorithm
- Develop and implement a search algorithm to identify patients at point of EHR-administrator
- Develop and implement a search algorithm to identify patients at point of clinical trial site investigator
- Combination of above

Endpoint Adjudication
- Required if treatment response is poorly defined
- Establish Charter and process for independent medical experts

Engaging EHR Administrator
- Sponsor and/or Parexel engages EHR-administrator through site investigator
- Sponsor engages EHR-administrator directly

APMs: Using advanced parametric modelling to address data challenges in survival extrapolation

Standard parametric modeling



- SPMs may not be sufficient to model the long-term plateaus associated with certain therapies
  - Late beneficial effect of a drug/long-term survivors
  - Loss of follow-up information with increasing data maturity
  - Limited sample size

Advanced parametric modeling

Mixture Cure Fraction Models

Assumes participants who are cured/no longer at risk of disease progression have similar mortality risks to individuals of the same age, sex, and country – i.e., background mortality

Bayesian Multi-Parameter Evidence Synthesis

- Synthesizes RWD with RCT data to generate realistic long-term extrapolations and by incorporating general population/registry survival rates

Landmark Response Models

Allows extrapolations based on response to treatment, whilst selecting a landmark point to prevent misclassification of responders

ML/AI: Deploying machine-learned predictive analytics to help drive study design and value messaging

Which set of prognostic factors/patient characteristics will accurately predict outcomes?

Can we use these predictors to estimate individual patient-level probabilities for single outcomes/multiple outcomes simultaneously?

Can we identify the variables that are most informative to collect for future patients?

How do we define optimal patient response profiles to flag those who may be at increased risk for adverse events or improved outcomes?

How can we combine the above information to facilitate clinical decision making?

Applying innovative methods to assess which prognostic factors best predict outcomes to inform patient identification and stratification

Patient level data including baseline characteristics (e.g., severity, comorbidities, demographics); HCP characteristics; etc.

Predictive model to assess individual probability of outcomes (e.g., clinical outcomes, treatment discontinuation, dose titration)

Rank most informative variables predicting outcomes

Define patient response profiles for respective treatment options to inform payer decision-making and clinical practice

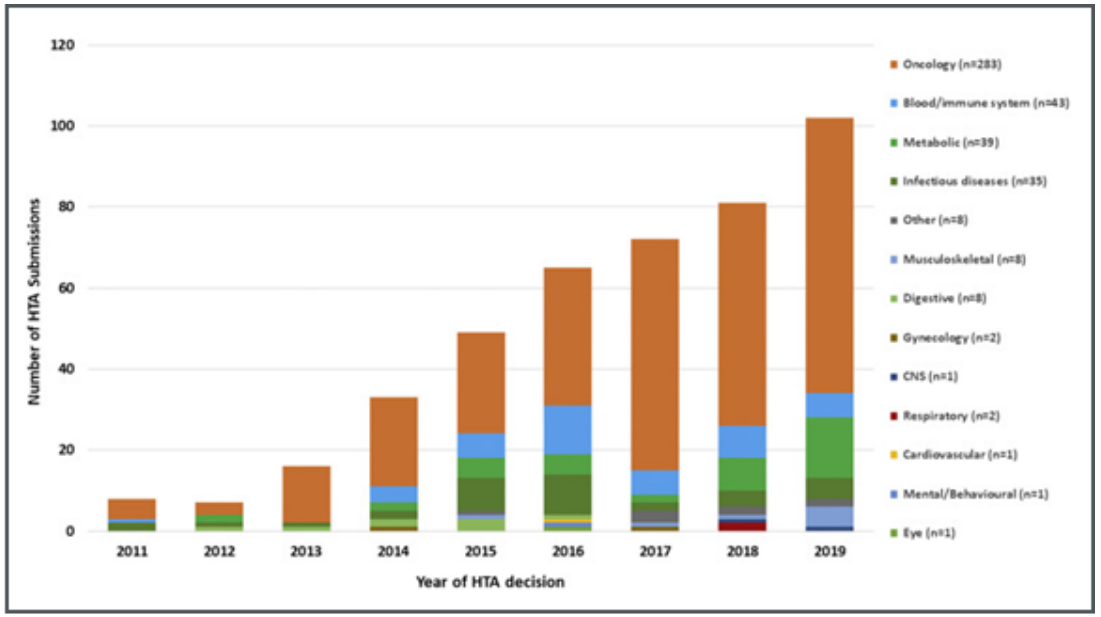
1. **Identify meaningful patient subgroups**
2. **Predict clinical outcomes**
3. **Derive of RWD-based phenotyping algorithm**
4. **Obtain rare disease predictions**
5. **Predict optimal treatment pathways**

parexel

Overview of NICE guidelines and other HTA agency requirements on the use of RWE

Emtiyaz Chowdhury

There has been an increase in the use of single arm trials to gain regulatory approval



Single arm trial submissions to HTA bodies globally (2011-2019); Patel et al., 2021

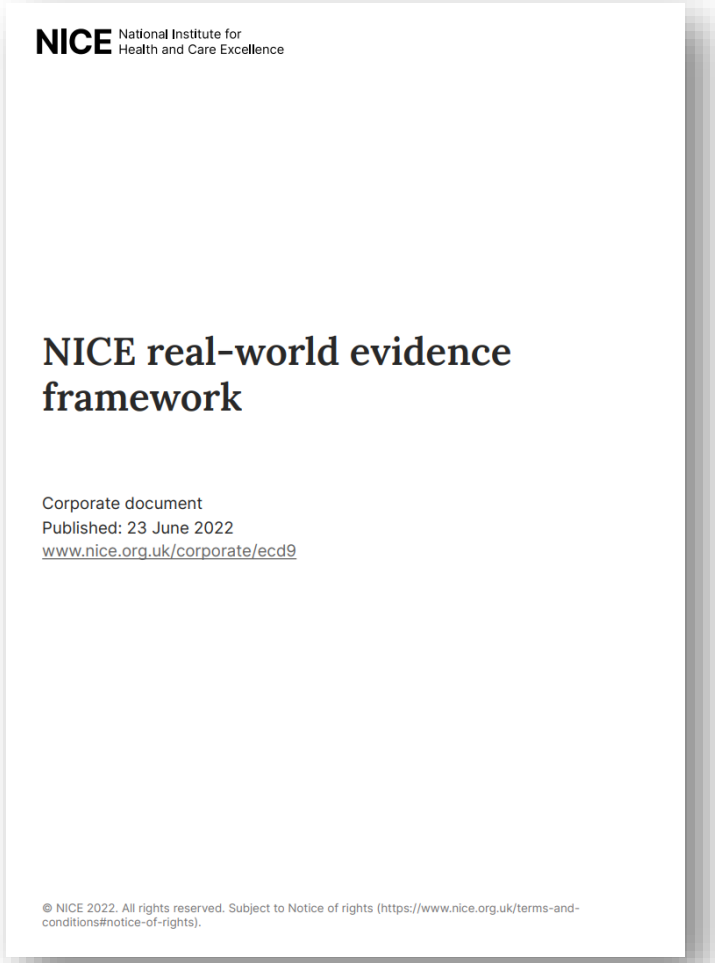
NICE published a ‘real-world evidence framework’ in June 2022

NICE’s RWE framework had two key objectives

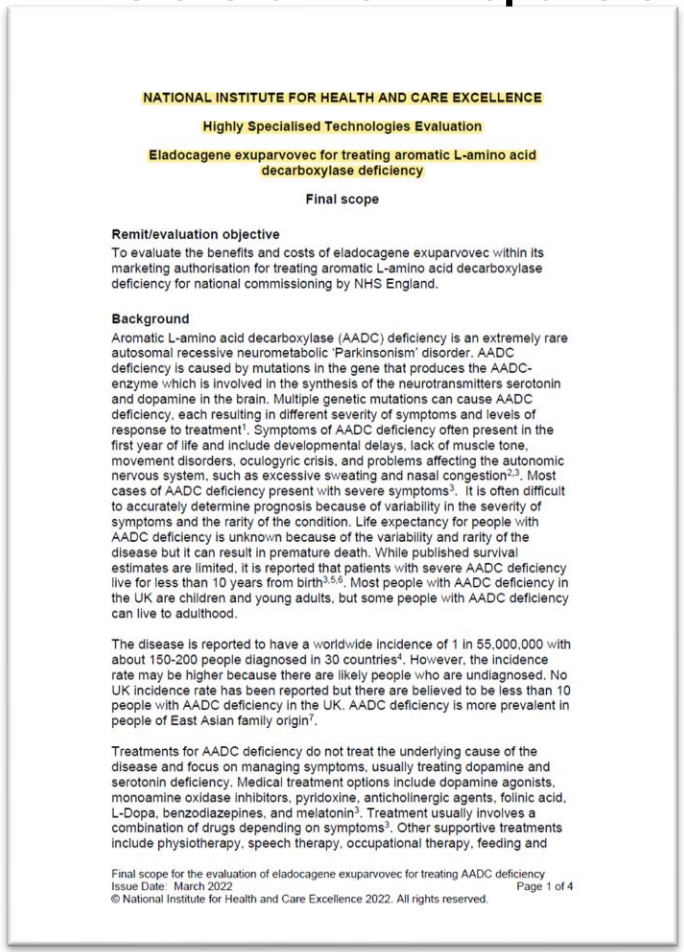
1. Identifying when RWD can be used to reduce uncertainties and improve guidance

2. Clearly describing best-practices for planning, conducting and reporting real-world evidence studies to improve the quality and transparency of evidence

The document is a living framework with at least one update expected before 2023



Randomized controlled trial are not always available or may not be sufficient to address the research question



NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Highly Specialised Technologies Evaluation

Eladocagene exuparvovec for treating aromatic L-amino acid decarboxylase deficiency

Final scope

Remit/evaluation objective

To evaluate the benefits and costs of eladocagene exuparvovec within its marketing authorisation for treating aromatic L-amino acid decarboxylase deficiency for national commissioning by NHS England.

Background

Aromatic L-amino acid decarboxylase (AADC) deficiency is an extremely rare autosomal recessive neurometabolic 'Parkinsonism' disorder. AADC deficiency is caused by mutations in the gene that produces the AADC-enzyme which is involved in the synthesis of the neurotransmitters serotonin and dopamine in the brain. Multiple genetic mutations can cause AADC deficiency, each resulting in different severity of symptoms and levels of response to treatment[1]. Symptoms of AADC deficiency often present in the first year of life and include developmental delays, lack of muscle tone, movement disorders, oculogyric crisis, and problems affecting the autonomic nervous system, such as excessive sweating and nasal congestion[2,3]. Most cases of AADC deficiency present with severe symptoms[3]. It is often difficult to accurately determine prognosis because of variability in the severity of symptoms and the rarity of the condition. Life expectancy for people with AADC deficiency is unknown because of the variability and rarity of the disease but it can result in premature death. While published survival estimates are limited, it is reported that patients with severe AADC deficiency live for less than 10 years from birth[3,5,6]. Most people with AADC deficiency in the UK are children and young adults, but some people with AADC deficiency can live to adulthood.

The disease is reported to have a worldwide incidence of 1 in 55,000,000 with about 150-200 people diagnosed in 30 countries[4]. However, the incidence rate may be higher because there are likely people who are undiagnosed. No UK incidence rate has been reported but there are believed to be less than 10 people with AADC deficiency in the UK. AADC deficiency is more prevalent in people of East Asian family origin[7].

Treatments for AADC deficiency do not treat the underlying cause of the disease and focus on managing symptoms, usually treating dopamine and serotonin deficiency. Medical treatment options include dopamine agonists, monoamine oxidase inhibitors, pyridoxine, anticholinergic agents, folinic acid, L-Dopa, benzodiazepines, and melatonin[3]. Treatment usually involves a combination of drugs depending on symptoms[3]. Other supportive treatments include physiotherapy, speech therapy, occupational therapy, feeding and

Final scope for the evaluation of eladocagene exuparvovec for treating AADC deficiency
Issue Date: March 2022 Page 1 of 4

1. Randomization is considered unethical (high unmet need)
2. Patients unwilling to be allocated to one of the interventions in the trial
3. Healthcare professions unwilling to randomize patients to an intervention which they consider less effective
4. Small number eligible patients
5. Financial or technical constraints on studies
6. Not all treatment combinations (including treatment sequences) can be directly assessed

The framework does not set minimum acceptable standards, instead there is a focus on best practice (1/2)

Transparency

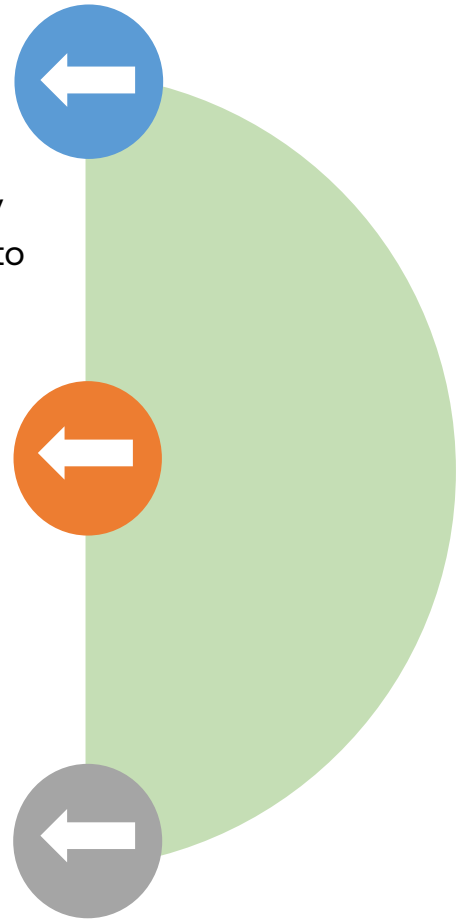
Generate evidence in a transparent way with integrity from study planning through to study conduct and reporting

Data suitability

Ensure data is trustworthy, relevant and of sufficient quality to answer research question

Methods

Use analytical methods that minimize the risk of bias and characterize uncertainty



Key messages

- Transparent reporting of data sources is essential to ensure trust in the data source and understand its fitness-for-purpose to address the research question
- Data should be of good and known provenance
  - Reporting on data sources should cover the characteristics of the data, data collection, coverage and governance
- Data fitness-for-purpose can be summarised by the data quality and relevance
  - data quality relates to the completeness and accuracy of key study variables
  - data relevance is determined by the data content, differences in patients, interventions and care settings between the data and the target population in the NHS, and characteristics of the data such as sample size and length of follow up.
- The Data Suitability Assessment Tool (DataSAT) may be used to provide consistent and structured information on data suitability
- There are reasonable trade-offs between different data sources in terms of quality, size, clinical detail and locality.
- The acceptability of a given data source may depend on the application and various contextual factors.

The framework does not set minimum acceptable standards, instead there is a focus on best practice (2/2)

Transparency

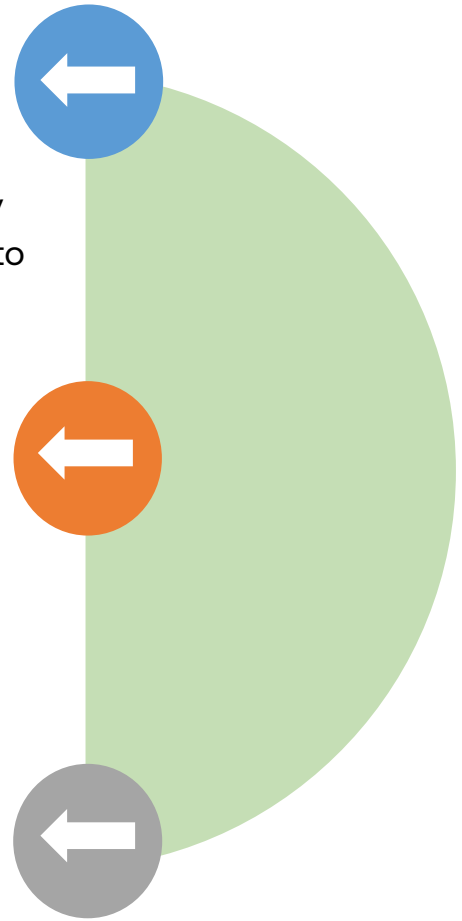
Generate evidence in a transparent way with integrity from study planning through to study conduct and reporting

Data suitability

Ensure data is trustworthy, relevant and of sufficient quality to answer research question

Methods

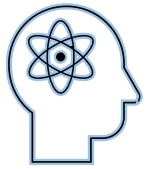
Use analytical methods that minimize the risk of bias and characterize uncertainty



Key messages

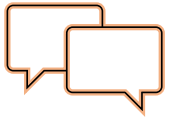
- Non-randomised studies can be used to provide evidence on comparative effects in the absence of randomised controlled trials or to complement trial evidence to answer a broader range of questions about the effects of interventions in routine settings.
- The recommendations presented here focus predominantly on cohort studies including those using real-world data to form external control arms.
- Study design
  - Design studies to emulate the preferred randomised controlled trial (target trial approach).
  - Avoid time-related biases due to differences between patient eligibility criteria being met, treatment assignment, and start of follow-up.
  - For studies using external control, select and curate data to minimise differences between data sources including availability and operational definitions of key study variables, data collection processes, patient characteristics, treatment settings, care pathways, and time periods, and consider the implications for study quality and relevance.
- Analysis
  - Identify potential confounders (including time-varying confounders) using a systematic approach and clearly articulate causal assumptions.
  - Use a statistical method that addresses confounding considering observed and unobserved confounders.
  - Consider the impact of bias from informative censoring, missing data, and measurement error and address appropriately if needed.
  - Use sensitivity and bias analysis to assess the robustness of results to main risks of bias and uncertain data curation and analysis decisions.

Manufacturers should be prepared to leverage RWE throughout the HTA process



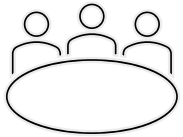
Scientific Advice

Discussions around managed access can happen early in the HTA process



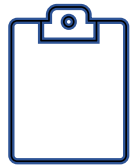
Technical Engagement

Real world evidence can be used to respond to concerns from EAG



Meeting 1

Committee concerns can be resolved by reanalyzing RWE or clinical trial data



Managed Access

Following agreement with committee, further data can be collected via MAA

Meeting 2

Early CEM is now being updated into an HTA-ready CEM

Case study [3900]: using RWE to address questions raised during the technical engagement process (1/2)

Disease, population & technology

- Cytomegalovirus (CMV) infection following solid organ or hematopoietic cell transplant (SOT/HCT) is associated with morbidity and mortality
- CMV can have a significant impact on quality of life and survival outcomes; there is a particular unmet need for patients with refractory (with or without [R/R] CMV
- Current anti-CMV drugs have a high IV burden (which in some cases require hospitalization)
- Maribavir is a novel anti-CMV drug administered orally

Issue: Time since transplant & recurrence

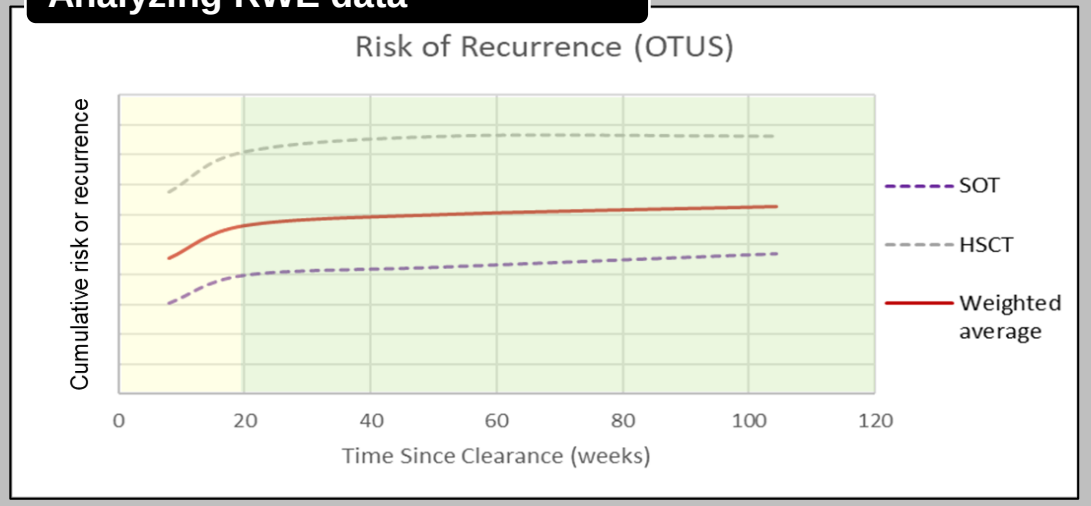
- Maribavir was superior to IAT for the primary endpoint of CMV viremia clearance at Study Week 8 (55.7% vs 23.9%, respectively).
- The EAG were of the view that time since transplant (TST) would be a prognostic factor associated with the risk of recurrence
- In response the company completed a logistic regression analysis (covariates validated with clinicians) to evidence there was no statistically significant relationship between TST and recurrence
- The logistic regression demonstrated evidence of a relationship between time since transplant (i.e., duration of time patient maintain response) and clearance

Case study [3900]: using RWE to address questions raised during the technical engagement process (2/2)

Solution: Real world evidence data, OTUS

- Company provided data from OTUS, a real-world evidence retrospective analysis of people with R/R CMV in response
- The primary objective of OTUS was to evaluate and describe clinical outcomes with current management patterns of CMV
- Results were separated for SOT and HSCT cohorts:
  - SOT included 115 patients, 58 were European that had an SOT between January 2014 and September 2021
  - HSCT included 121 patients, 39 were European that had an allogeneic HSCT from January 2017 to October 2021
- Data from OTUS supported the finding of a diminishing risk of recurrence as time since clearance increases

Analyzing RWE data



HTA agencies are demonstrating flexibility in incorporating RWE into HTA submission, manufacturers need to be responsive in supporting committees and payers reduce uncertainty

Can we estimate the comparative effectiveness of cancer treatments using UK registry data?

Dr Nicholas Latimer,

Can we estimate the comparative effectiveness of cancer treatments using UK registry data?

ISPOR, Vienna, November 2022

Dr Nicholas Latimer, Professor of Health Economics, Yorkshire Cancer Research Senior Fellow, University of Sheffield, Sheffield, UK

Acknowledgements

Funders

Collaborators

- Prof Jim Chilcott
- Prof Paul Tappenden
- Prof Jon Wadsley
- Dr Peter Hall (University of Edinburgh)
- Prof Uwe Siebert (UMIT)
- Dr Rebecca Smittenaar (PHE, GRAIL Inc)
- Dr Ellie Murray (Boston University)
- **Ms Amy Chang**
- **Mrs Saleema Rex**
- And also: Mike Bradburn, Prof Ron Akehurst, Dr Robin Young, Dr Matt Winter, Dr Duncan Gillespie, Prof Janet Brown, Dr Carmel Pezaro, Prof Derek Rosario

Background

- Increased interest in “real world data”/“big data” in HTA
- Key part of this is investigating comparative effectiveness in the real world

→ **Great!...**

…but dangerous if not done correctly

→ Large proportion of HTAs are of cancer treatments. Lots of my work has been in this area, so:

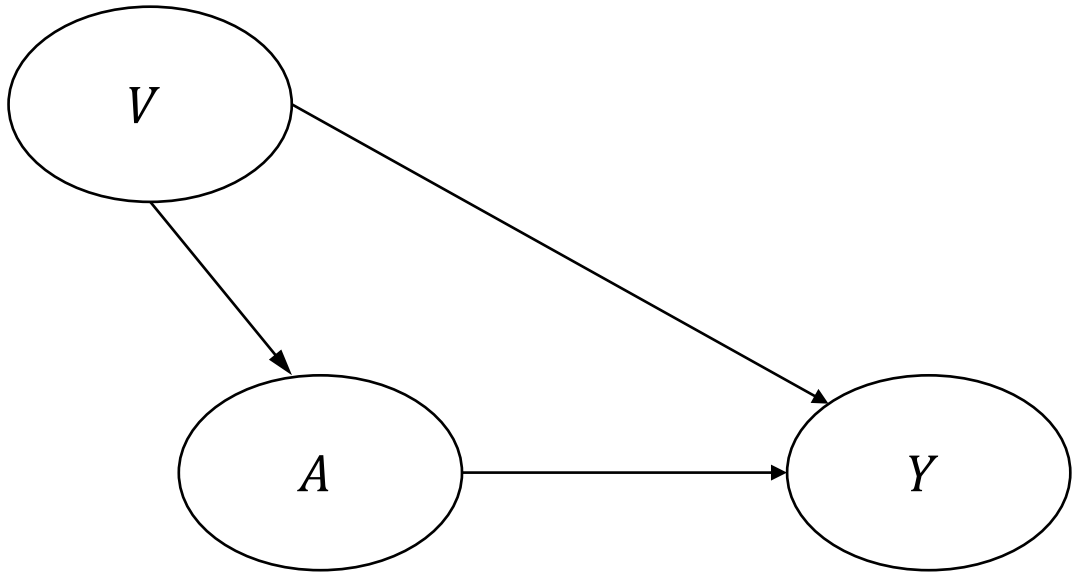
→ **Can we reliably estimate the comparative effectiveness of cancer treatments given in the NHS using English registry data?**

What are the dangers?

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation

What are the dangers?

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



Confounding by indication

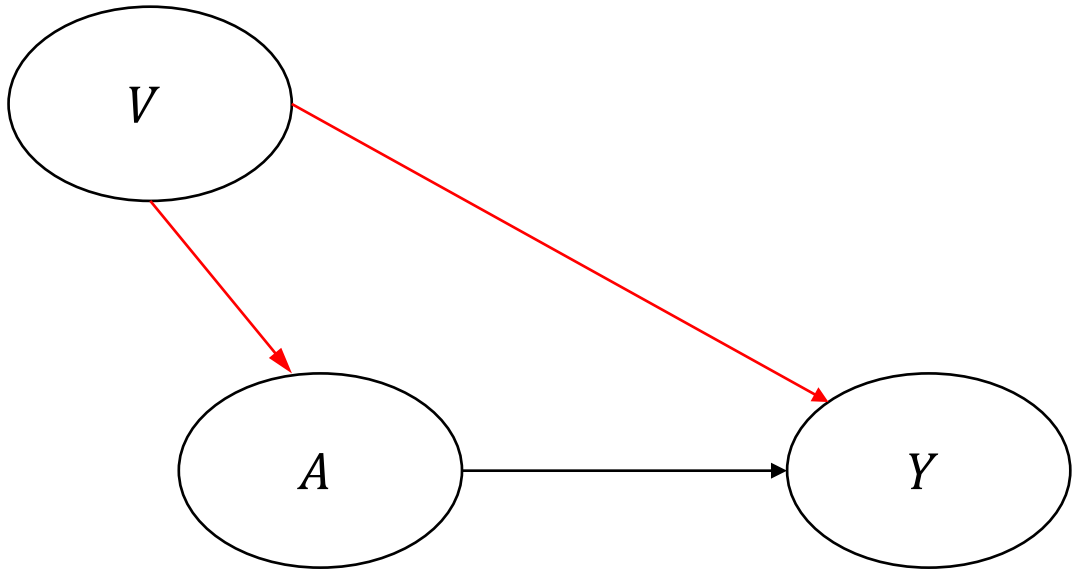
→ Imagine we want to estimate the effect of *A* on *Y*

→ Treatment received may be related to prognostic characteristics, *V*

V = prognostic variables measured at or before treatment
A = treatment indicator
Y = death indicator

What are the dangers?

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



Confounding by indication

→ Imagine we want to estimate the effect of *A* on *Y*

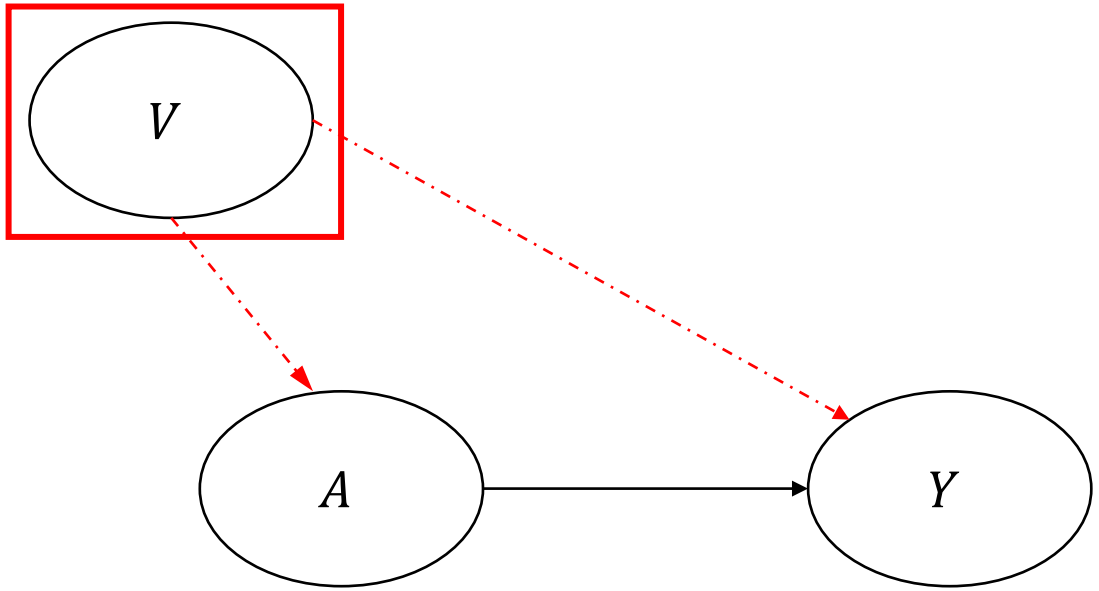
→ Treatment received may be related to prognostic characteristics, *V*

→ There is an open backdoor path between A and Y

V = prognostic variables measured at or before treatment
A = treatment indicator
Y = death indicator

What are the dangers?

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



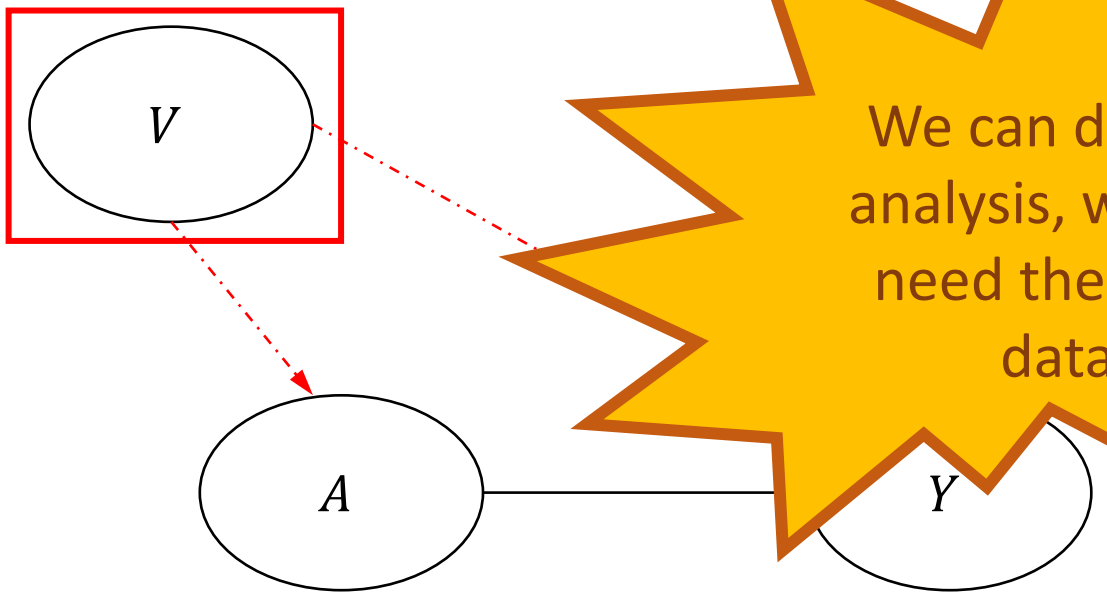
V = prognostic variables measured at or before treatment
A = treatment indicator
Y = death indicator

Confounding by indication

- → Imagine we want to estimate the effect of *A* on *Y*
- → Treatment received may be related to prognostic characteristics, *V*
- → There is an open backdoor path between *A* and *Y*
- → We need to try to block this backdoor path, by including variables *V* in our statistical model
- → For example (?):
  - → Performance status
  - → Time since diagnosis
  - → Comorbidity score

What are the dangers?

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



We can do this analysis, we just need the right data

[fo…ding by indication]

- …t to estimate the effect of *A* on *Y*
- …ent received may be related to prognostic …istics, *V*
- …s an open backdoor path between *A* and *Y*
- …e need to try to block this backdoor path, by including variables *V* in our statistical model

→ For example (?):

- → Performance status
- → Time since diagnosis
- → Comorbidity score

V = prognostic variables measured at or before treatment
A = treatment indicator
Y = death indicator

What are the dangers? (part 2)

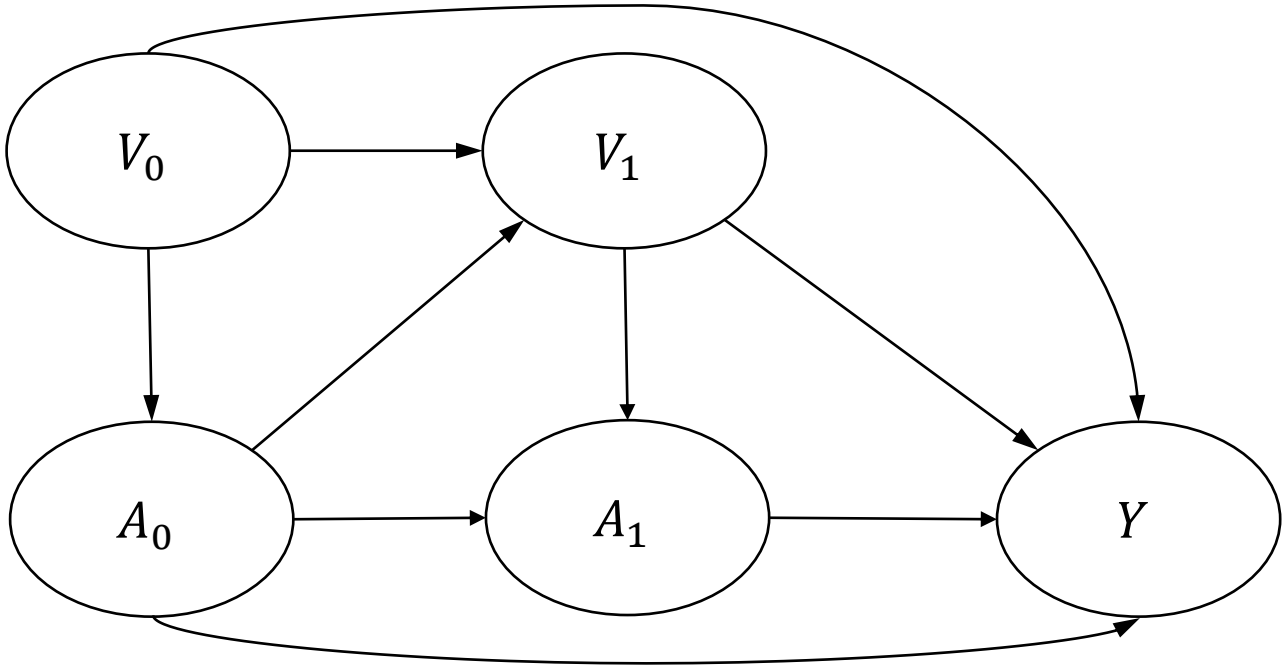
- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation

Time-dependent confounding

→ May also be important

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

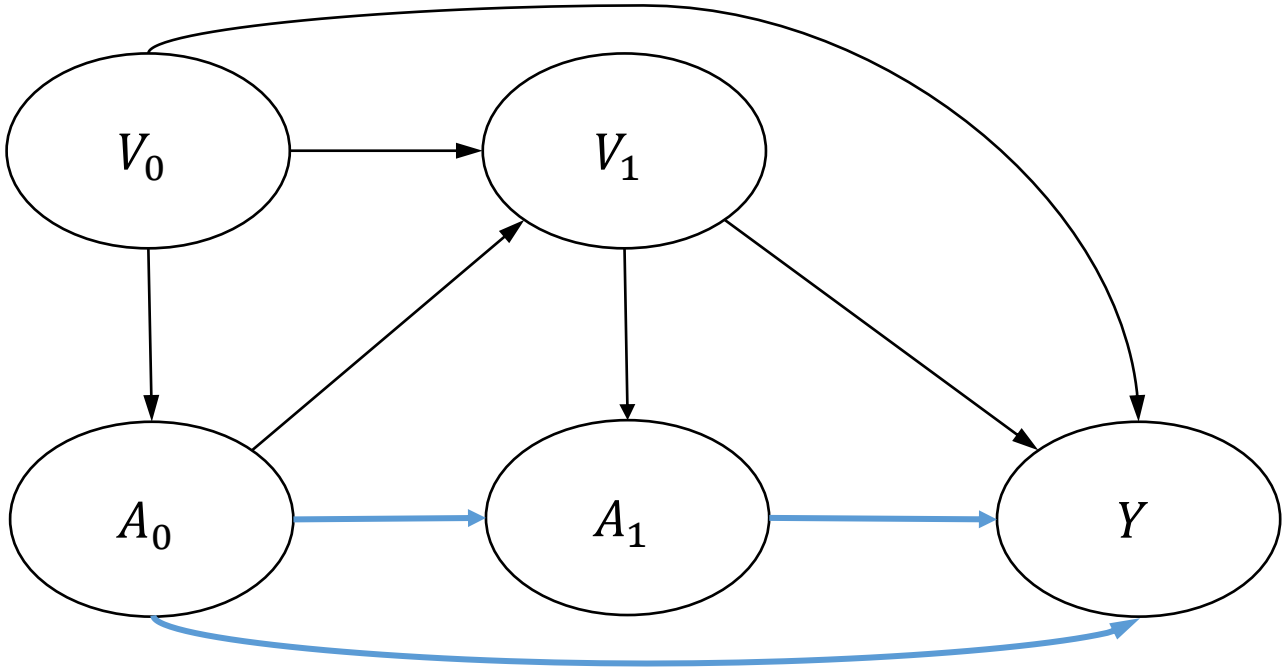
Time-dependent confounding

→ May also be important

→ Imagine we want to estimate the effect of *A* on *Y*

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

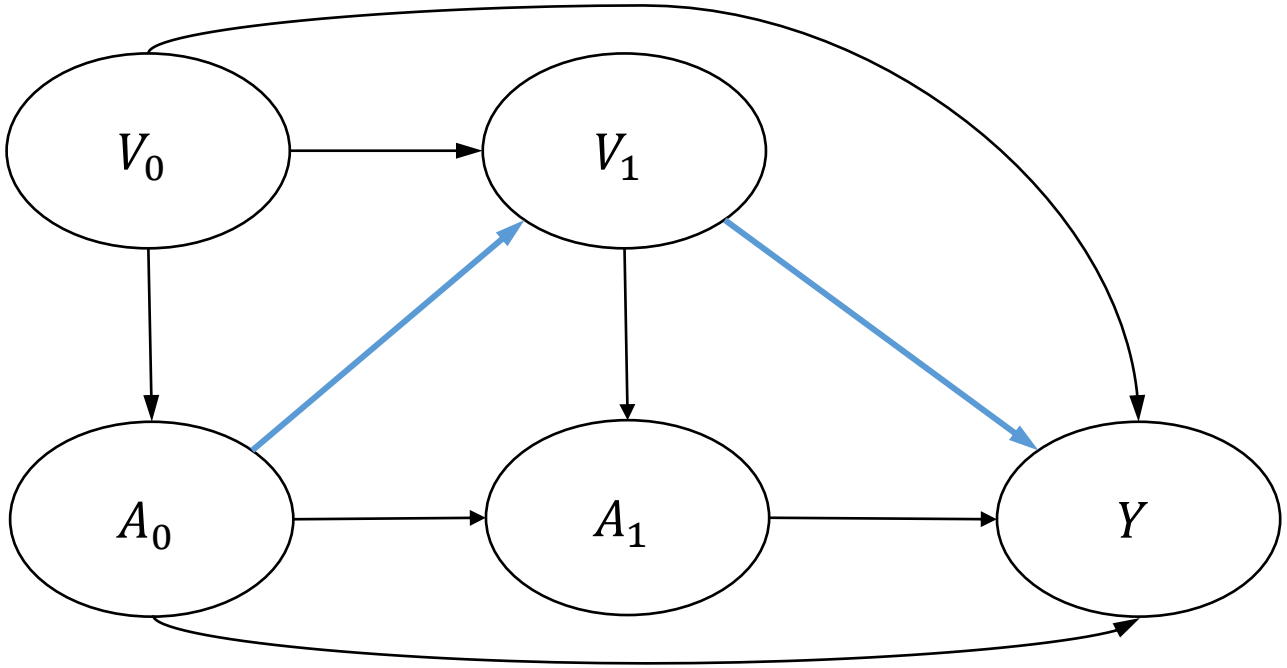
Time-dependent confounding

→ May also be important

→ Imagine we want to estimate the effect of *A* on *Y*

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

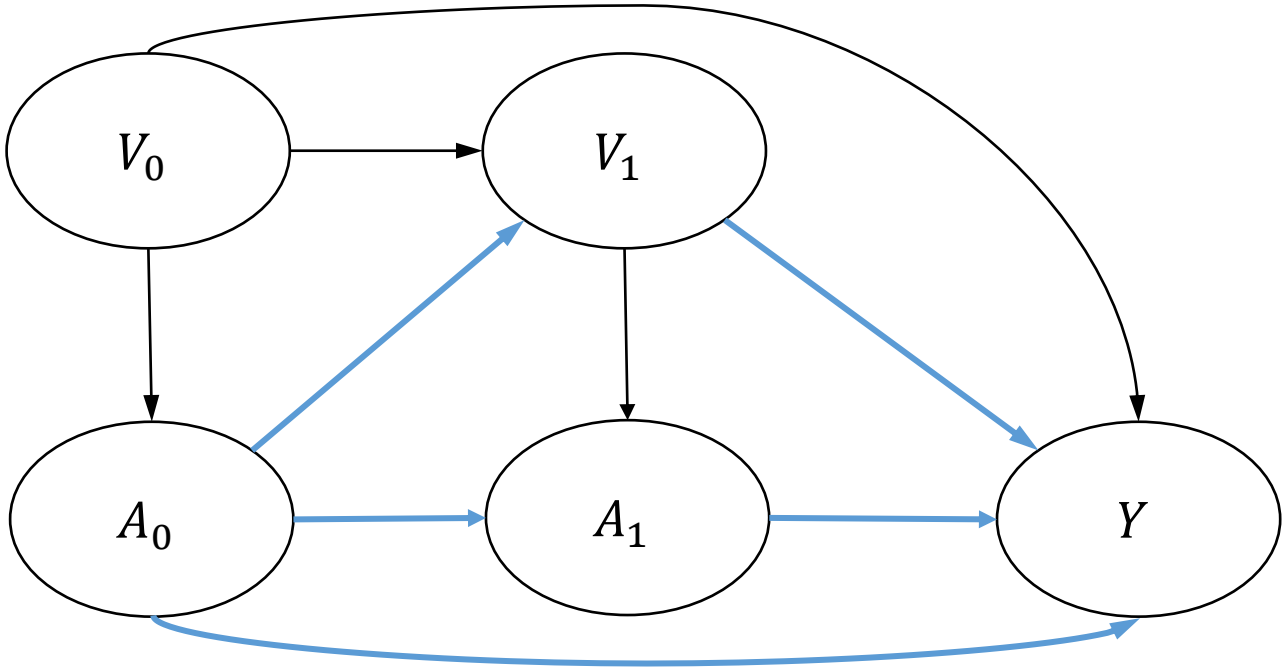
Time-dependent confounding

→ May also be important

→ Imagine we want to estimate the effect of *A* on *Y*

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

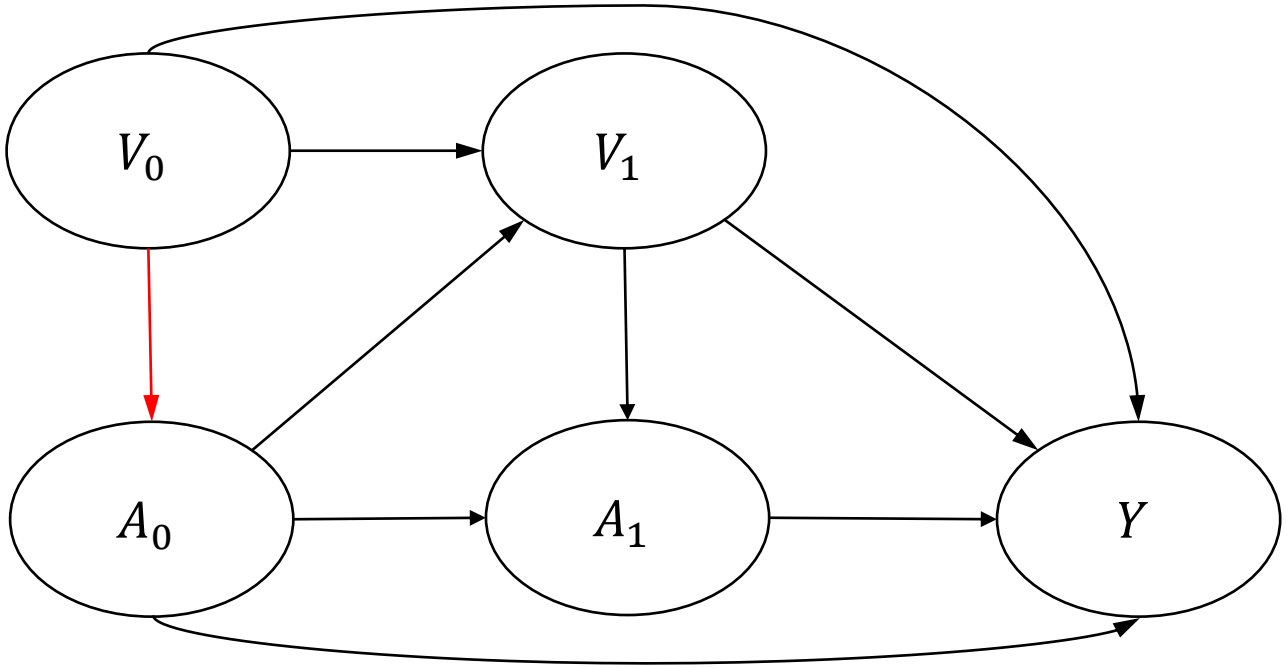
Time-dependent confounding

→ May also be important

→ Imagine we want to estimate the effect of *A* on *Y*

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

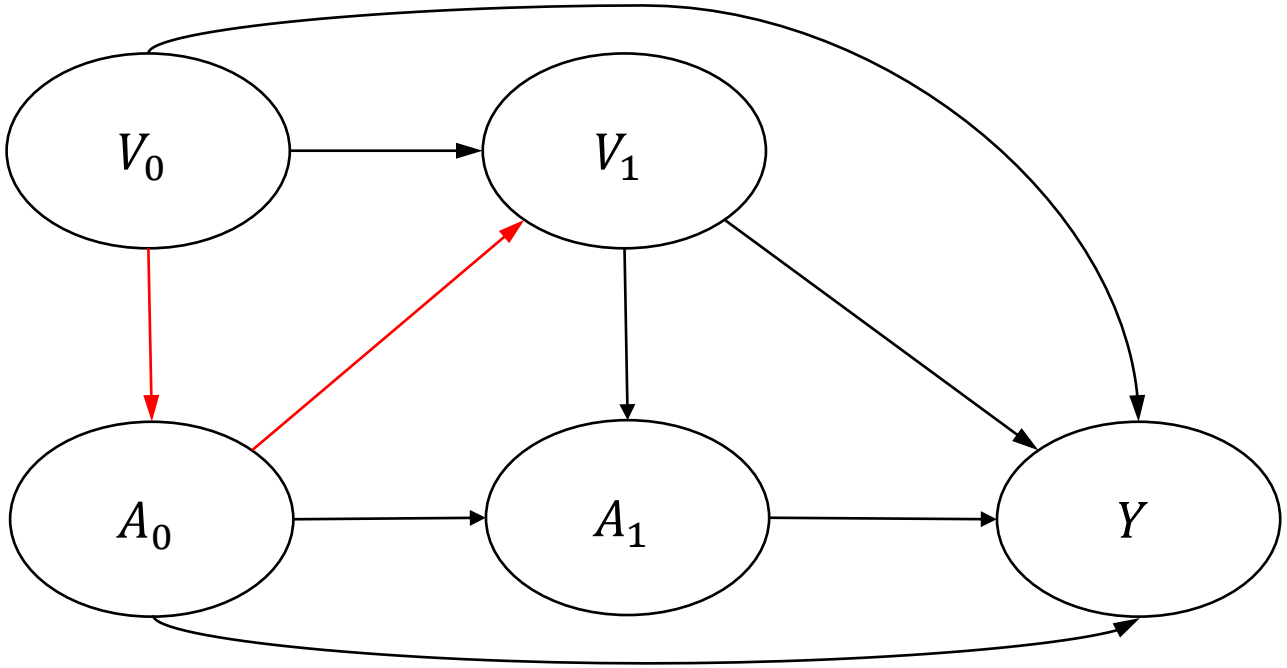
→ May also be important

→ Imagine we want to estimate the effect of *A* on *Y*

i) Treatment received (A_0) may be related to initial values of prognostic variables (V_0)

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

→ May also be important

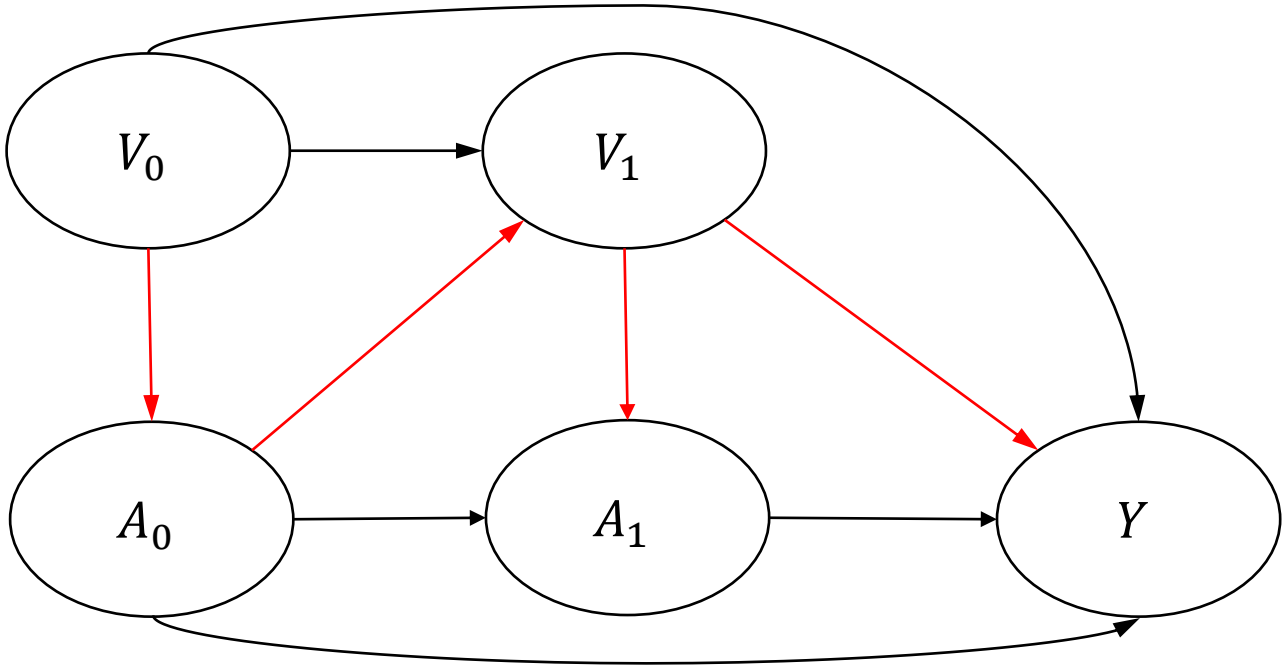
→ Imagine we want to estimate the effect of *A* on *Y*

i) Treatment received (A_0) may be related to initial values of prognostic variables (V_0)

ii) Treatment (A_0) might effect future values of prognostic variables (V_1)

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

→ May also be important

→ Imagine we want to estimate the effect of *A* on *Y*

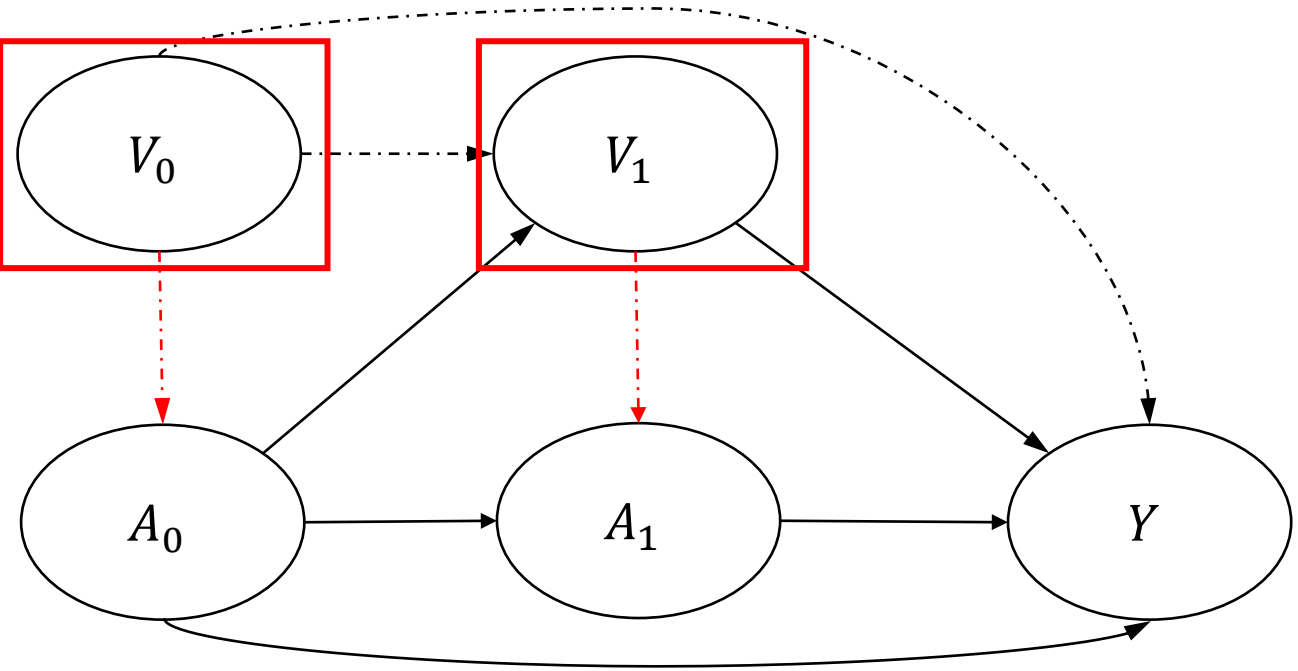
i) Treatment received (A_0) may be related to initial values of prognostic variables (V_0)

ii) Treatment (A_0) might effect future values of prognostic variables (V_1)

iii) Future values of prognostic variables (V_1) might influence future changes of treatment (A_1) (patient might start a different treatment) *and* survival

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

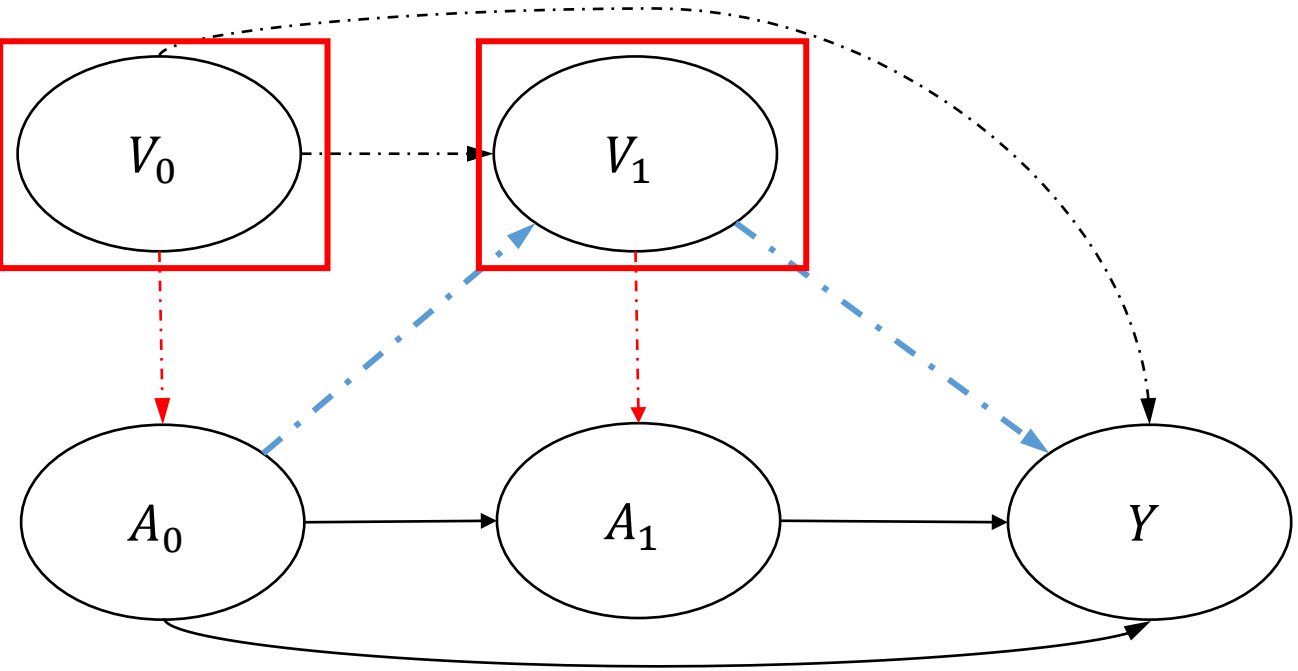
Time-dependent confounding

→ Now we have a big problem!

→ We want to block the backdoor paths between the *A*'s and *Y*…

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

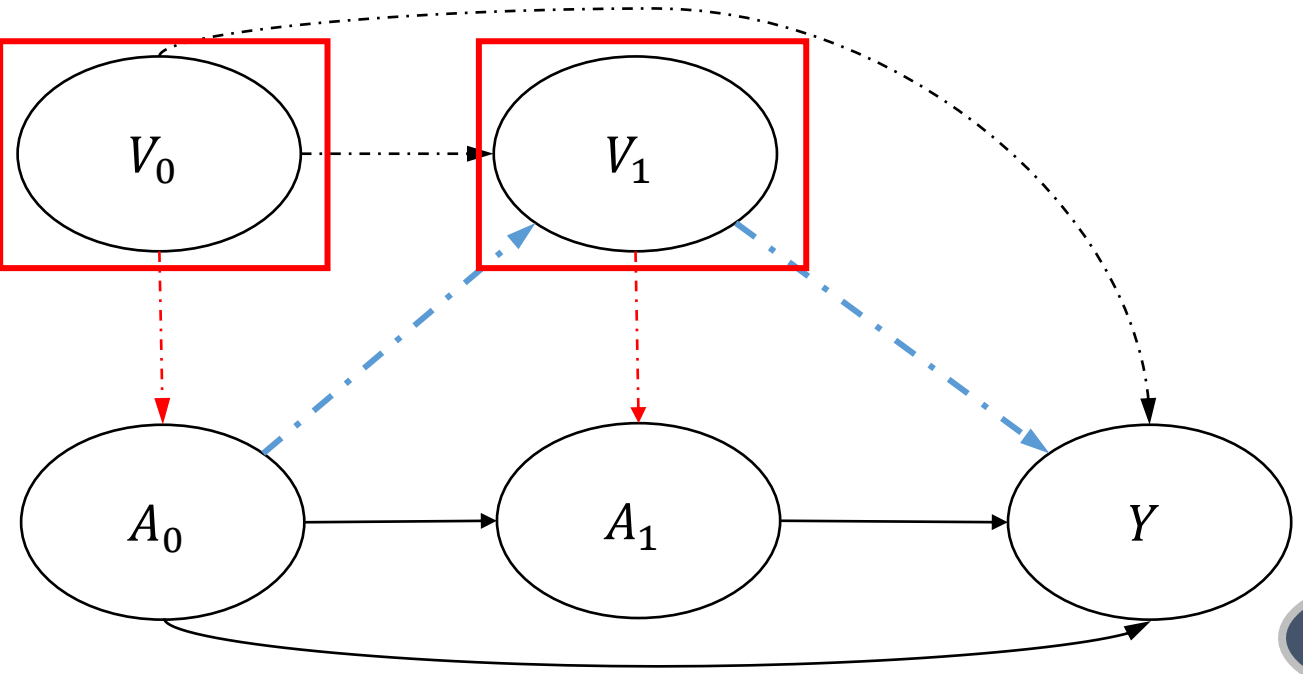
→ Now we have a big problem!

→ We want to block the backdoor paths between the *A*'s and *Y*…

→ …but if we put the *V*'s into our statistical model we end up also blocking a front door, causal path between $A_0 \rightarrow V_1 \rightarrow Y$ *!*

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

→ Now we have a big problem!

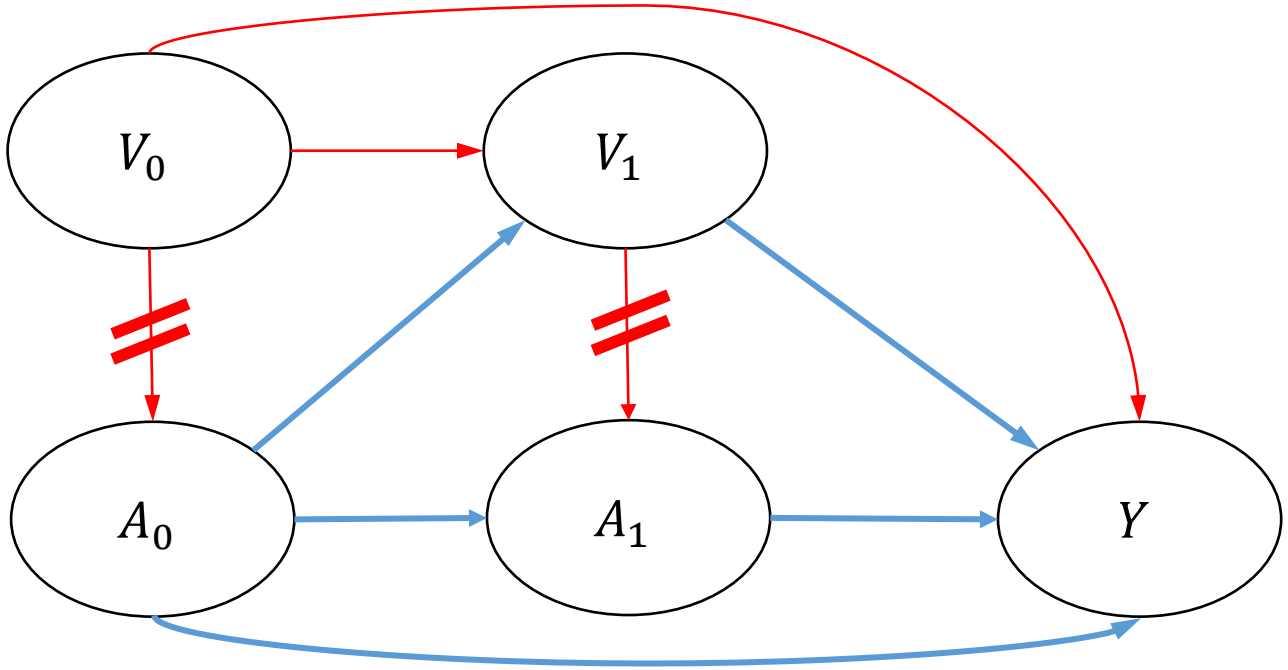
→ We want to block the backdoor paths between the *A*'s and *Y*…

→ …but if we put the *V*'s into our statistical model we end up also blocking a front door, causal path between $A_0 \rightarrow V_{\ldots} \rightarrow Y$!

If we use standard statistical models in these circumstances, our results will be biased!

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables measured over time
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

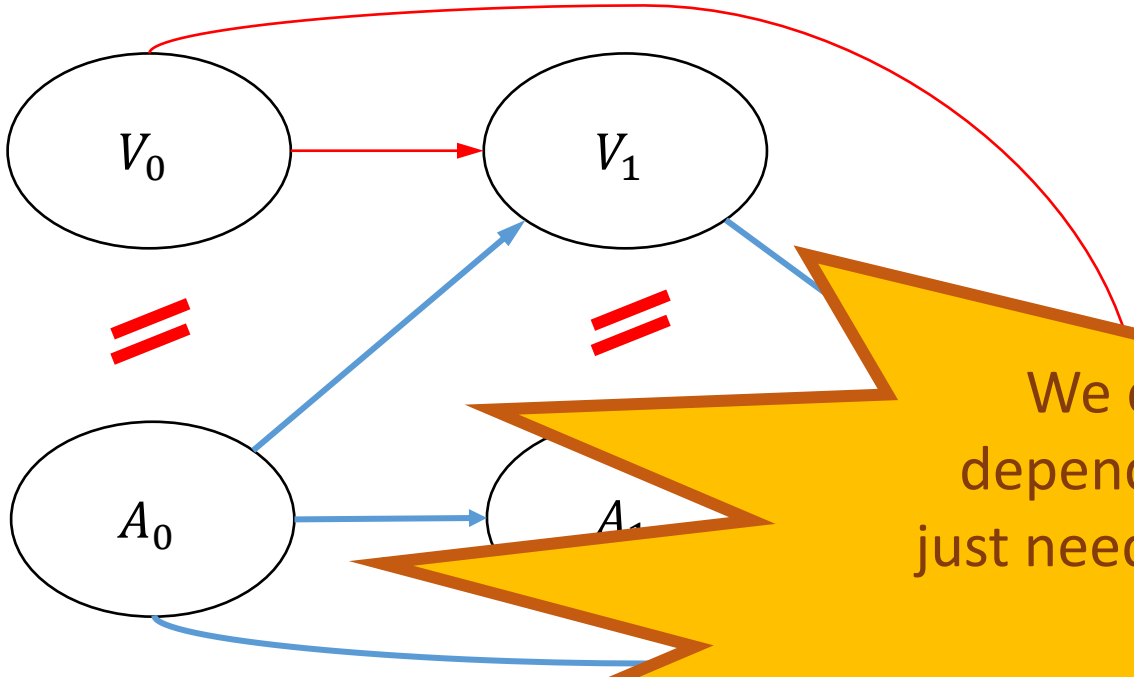
→ We need causal inference methods, like inverse probability weighting **("g-methods")**, to deal with time-dependent confounding

[essentially, these remove the arrows from V_0 and V_1 to A_0 and A_1]

→ And we need good enough data to be able to deal with all confounding variables

What are the dangers? (part 2)

- Estimating comparative effectiveness using observational data is prone to bias – due to the lack of randomisation



$V_{0.1}$ = prognostic variables m
$A_{0,1}$ = treatment indicator over time
Y = death indicator

Time-dependent confounding

→ We need causal inference methods, like inverse probability wei ing **("g-methods")**, to deal with time ent confounding

e the arrows from V_0

data to be able to
bles

We can deal with time-dependent confounding, we just need the right data and the right methods

Do these methods work?

→ Can investigate using trial replication techniques (Hernan, Robins. *Am J Epidemiol* 2016;183(8):758–764)

Zidovudine for HIV-positive men (Hernan, Brumback, Robins, Epidemiology 2000;11(5):561-70)

- Yes! Showed that "simple" methods gave biased results, but appropriate g-methods worked well

RCT Duplicate (Franklin, Patorno, Desai *et al*. Circulation. 2020 Dec 17)

- Replicate 30 completed RCTs using US claims data, use appropriate analyses and compare results
- So far, 8/10 RWE emulations resulted in treatment effect estimates within 95% CI of corresponding RCT

SEER-Medicare linked data analysis (Petito, Garcia-Albeniz, Logan *et al*. JAMA Netw Open 2020;Mar 2;3(3))

- Trial replications for colorectal cancer and pancreatic cancer, using SEER-Medicare linked data
- Simple methods gave biased results, but g-methods worked well

Analysis with Scottish cancer registry data (Gray, Marti, Bewster *et al*. J Clin Epidemiol 2019;109:125-132)

- Real world effectiveness of chemo for breast cancer, using registry data linked to hospital data
- Some methods produced results similar to RCTs, some methods did not

The University Of Sheffield.

yorkshire cancer research

Saving Yorkshire Lives

Can we do it with English cancer registry data?

The datasets – are the data sufficient?

NCRAS (National Cancer Registration and Analysis Service)

- Cancer registry
- Systemic anti-cancer therapy (SACT)
- Hospital Episodes Statistics

The datasets

NCRAS

- **Cancer registry**
- Systemic anti-cancer therapy (SACT)
- Hospital Episodes Statistics

- Age
- Sex
- Diagnosis date
- Tumour site, grade, size, histology, stage
- Charlson co-morbidity (2 and 6 year)
- Geographical info inc. deprivation measures
- Date of death

The datasets

NCRAS

- Cancer registry
- **Systemic anti-cancer therapy (SACT)**
- Hospital Episodes Statistics

- Height
- Weight
- Performance Status (ECOG) (over time)
- Co-morbidity indicator
- Treatment details (including treatment outcomes, treatment changes)

"SACT… is the world's first comprehensive database, allowing us to understand treatment patterns and outcomes on a national scale"

The datasets

NCRAS

- Cancer registry
- Systemic anti-cancer therapy (SACT)
- **Hospital Episodes Statistics**

- Episode information for:
  - Admitted care
  - Outpatient episodes
  - Accident & Emergency

The datasets

NCRAS

- Cancer registry
- Systemic anti-cancer therapy (SACT)
- Hospital Episodes Statistics

Seems like reasonably good breadth of data on potentially prognostic variables

But is this enough? ("no unmeasured confounding")

Are the data good enough quality / accurate?

Work plan

Plan is to emulate previously completed RCTs using Hernan and Robins "Target Trial" framework **(similar to RCT Duplicate)**

1. Emulate RCT inclusion criteria / treatment strategies as far as possible
2. Conduct analyses that control for confounding
3. Compare results to RCT results → are they similar?

Work plan

Plan is to emulate previously completed RCTs using Hernan and Robins "Target Trial" framework **(similar to RCT Duplicate)**

1. Emulate RCT inclusion criteria / treatment strategies as far as possible
2. Conduct analyses that control for confounding
3. Compare results to RCT results → are they similar?

Cancer	Trial	Details	Researcher
Pancreatic	ESPAC-4[6]	Gemcitabine vs gem + capecitabine	Nick
Pancreatic	ACCORD[7]	FOLFIRINOX vs gemcitabine	Nick
Pancreatic	CRUK-GEM-CAP[8]	Gemcitabine vs gem + capecitabine	Nick
Pancreatic	MPACT[9]	Gemcitabine vs gem + nab-paclitaxel	Nick
Lung	LUX-Lung[10]	Afatinib vs gefitinib	Saleema
Lung	Keynote-024[11]	Pembrolizumab vs chemotherapy	Saleema
Breast	TNT[12]	Carboplatin vs docetaxel	Saleema
Prostate	Khalaf *et al.* 2019[13]	Enzalutamide and abiraterone sequencing study, using Flatiron and NCRAS data	Amy
Renal Cell	RECORD-3[14]	Sunitinib followed by everolimus sequencing study	Amy

Discussion

- We should try to make use real world data (we collect it already)
- Important for patients, clinicians, and healthcare decision-making / resource use
- **But are the data good enough?**

→ We will see!

Using RWE to augment clinical trial data

Gianluca Biao

Leveraging real-world evidence for health technology assessment - Using big data to enable patient access

Gianluca Baio

Department of Statistical Science | University College London

g.baio@ucl.ac.uk
https://gianluca.statistica.it/
https://egon.stats.ucl.ac.uk/research/statistics-health-economics/
https://github.com/giabaio
https://github.com/StatisticsHealthEconomics
@gianlubaio

ISPOR Europe 2022, Vienna (Austria)

7 November 2022

UCL Statistical Science

What are we talking about?...

What are we talking about?...

... Well, there are many problems!

Data

1. We may (or may not!) access individual level data for "our" trial, but not for the competitors'
2. The trial data have a very limited follow up, which implies large amount of censoring
   - This is often OK(-ish!) for "medical stats" analysis. But HORRIBLE for economic evaluation! ⇒ Extrapolation
3. Often the data are manipulated by the stats team within the sponsor and the economic modellers only get summaries/estimates
   - It is ALWAYS good to leave things to statisticians. But the modellers can (should?!) be statisticians too, so they could handle the data!...

What are we talking about?...

... Well, there are many problems!

Data

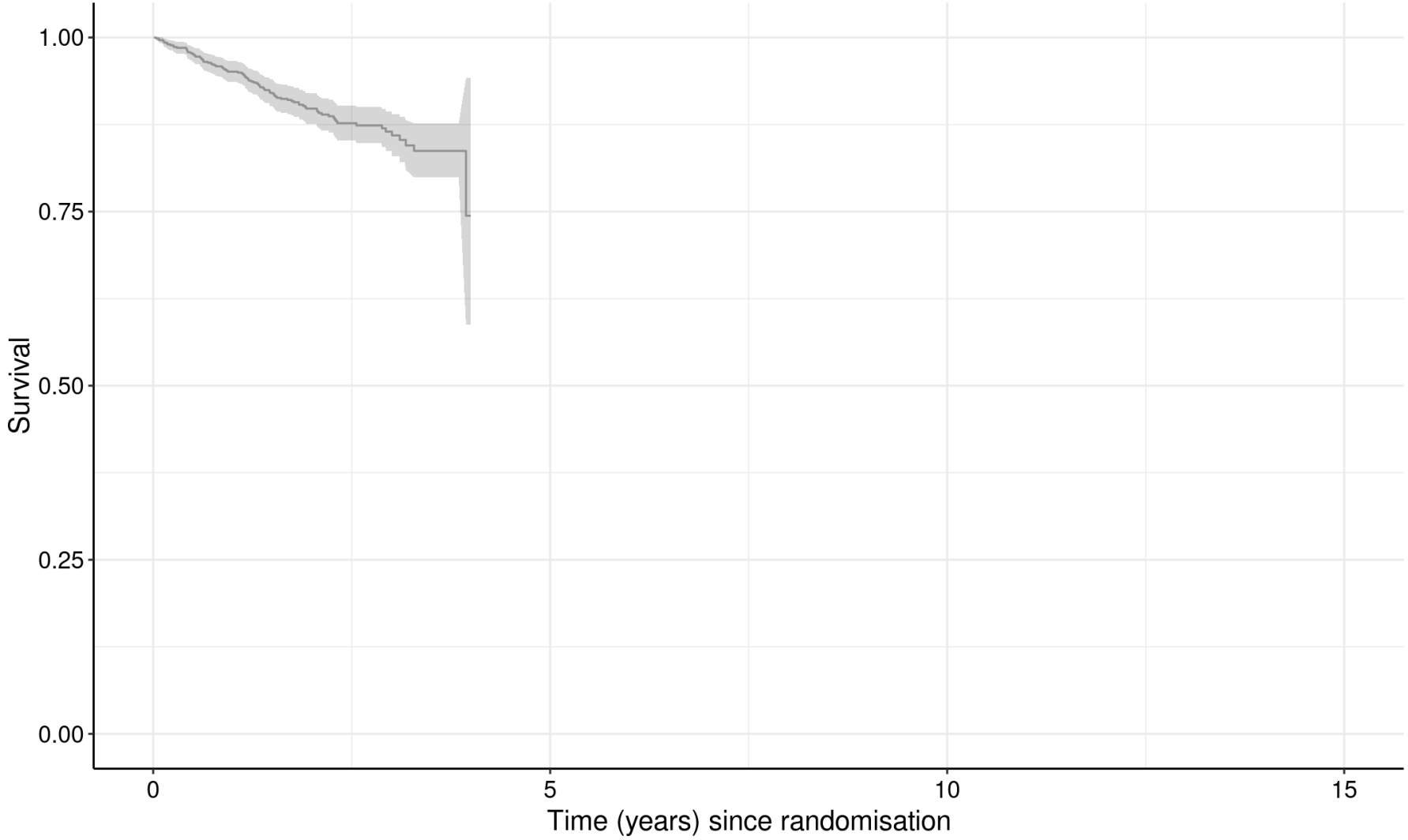
1. We may (or may not!) access individual level data for "our" trial, but not for the competitors'
2. The trial data have a very limited follow up, which implies large amount of censoring
   - This is often OK(-ish!) for "medical stats" analysis. But HORRIBLE for economic evaluation! ⇒ Extrapolation
3. Often the data are manipulated by the stats team within the sponsor and the economic modellers only get summaries/estimates
   - It is ALWAYS good to leave things to statisticians. But the modellers can (should?!) be statisticians too, so they could handle the data!...

Models

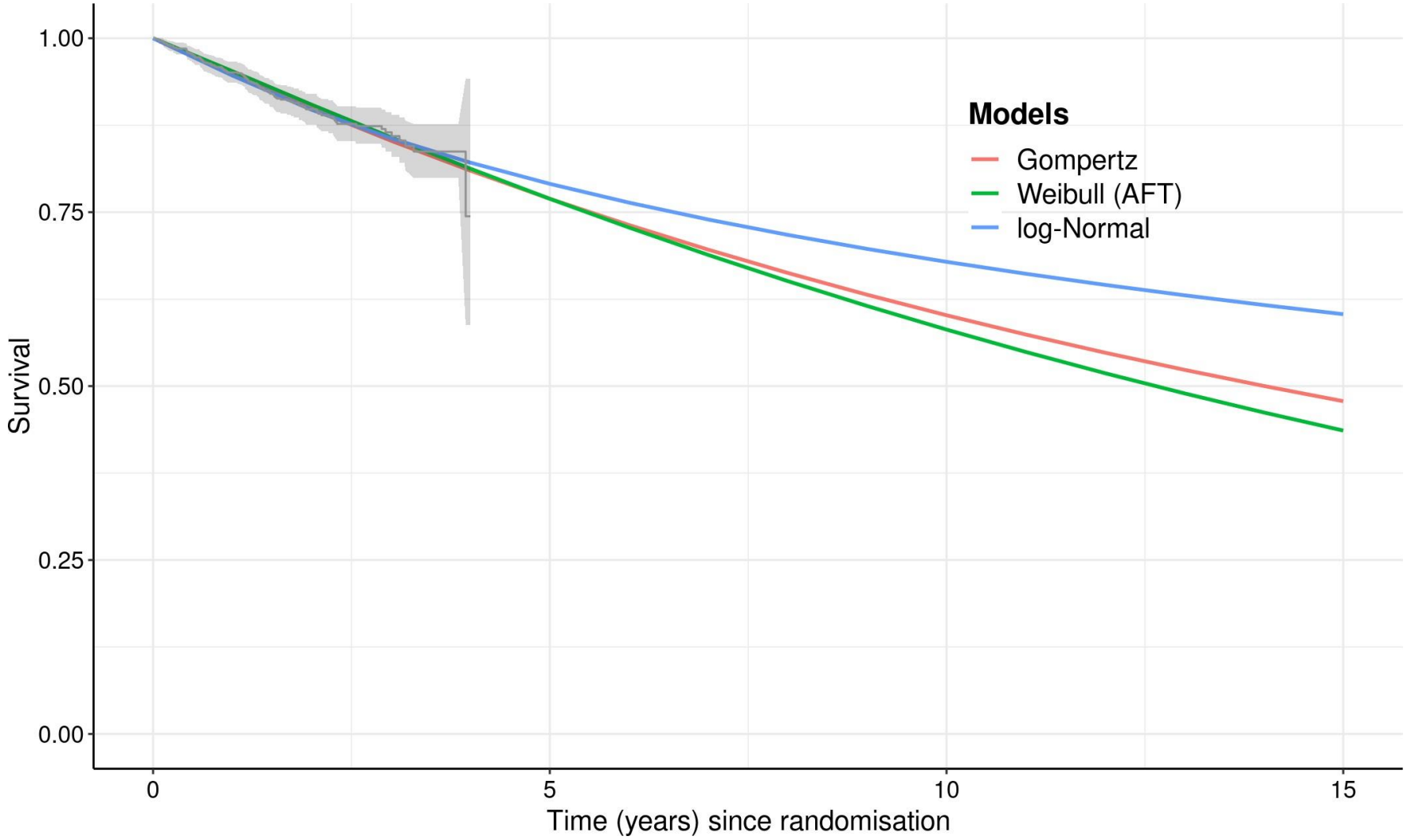
1. Which model is the "best fit" - how to judge that?
2. Is modelling even enough? (How to make the most of "external data")
3. Should you be Bayesians about this?
   - (Spoiler alert: the answer is *always* Yes!...)

Example

Observed data (NICE TA 174)



Parametric fitting/extrapolation



"Blended" survival curves

Che et al (2022)

Consider two separate process

"Blended" survival curves

doi Che et al (2022)

Consider two separate process

1. Driven *exclusively* by the observed data
   - Similar to a "standard" HTA analysis - use this to estimate $S_{obs}(t \mid \theta_{obs})$
   - Main objective: produce the best fit possible to the *observed* information
   - NB: Unlike in a "standard" modelling exercise where the issue of overfitting is potentially critical, achieving a very close approximation to the observed dynamics has much less important implications in the case of blending

"Blended" survival curves

doi Che et al (2022)

Consider two separate process

1. Driven *exclusively* by the observed data
   - Similar to a "standard" HTA analysis - use this to estimate $S_{obs}(t \mid \theta_{obs})$
   - Main objective: produce the best fit possible to the *observed* information
   - NB: Unlike in a "standard" modelling exercise where the issue of overfitting is potentially critical, achieving a very close approximation to the observed dynamics has much less important implications in the case of blending

2. **"External" process**
   - Used to derive a separate survival curve, $S_{ext}(t \mid \theta_{ext})$ to describe the ***long-term*** estimate for the survival probabilities
   - Could use "hard" evidence (eg RWE/registries/cohort studies/etc)...
   - ...Or, purely subjective knwoledge elicited from experts (or both!)

R Code (for the paper)

R package blendR

Example - Blended survival curves

Combine the two processes to obtain

$$S_{ble}(t \mid \theta) = S_{obs}(t \mid \theta_{obs})^{1-\pi(t;\alpha,\beta,a,b)} \times S_{ext}(t \mid \theta_{ext})^{\pi(t;\alpha,\beta,a,b)}$$

where:

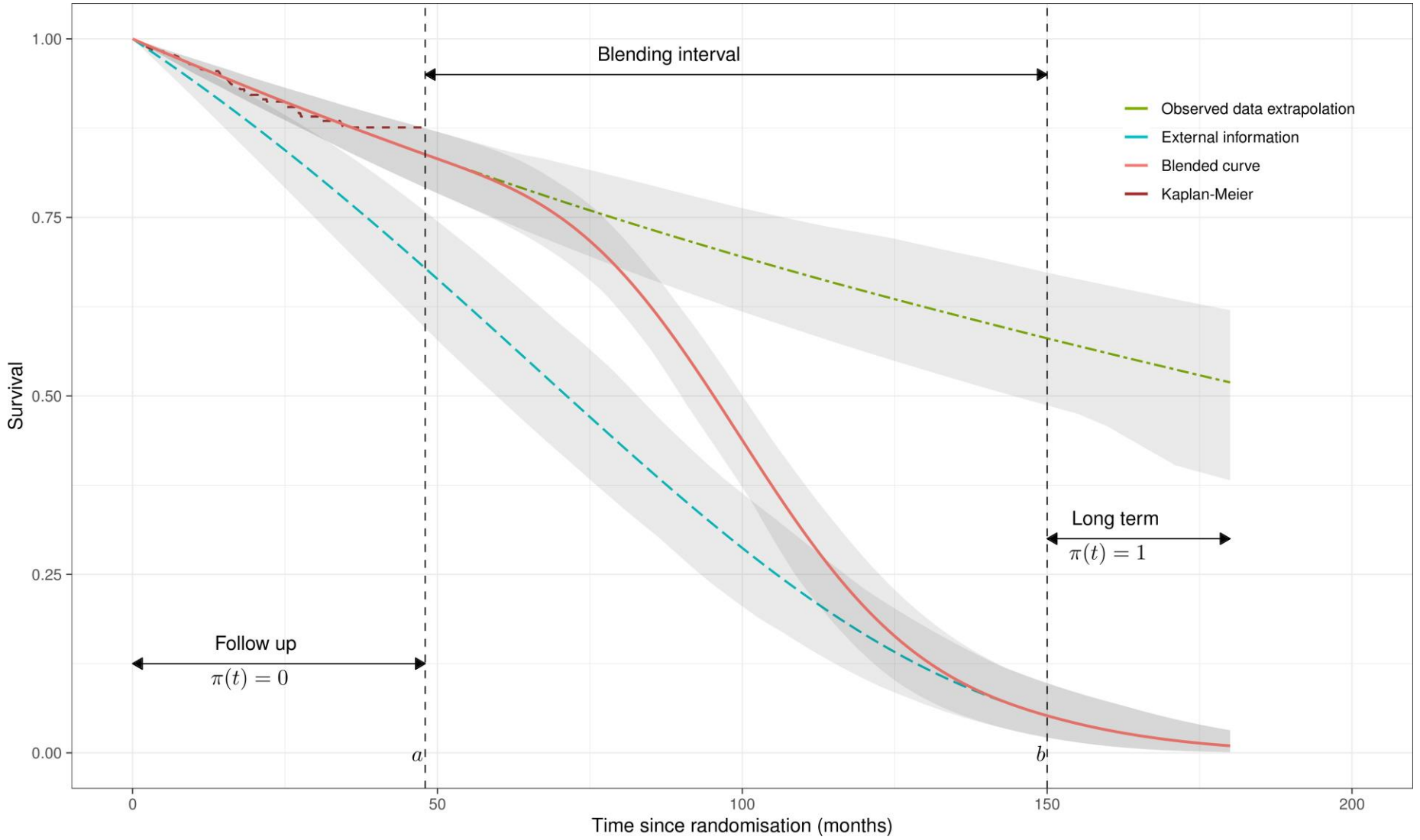
- $\theta = \{\theta_{obs}, \theta_{ext}, \alpha, \beta, a, b\}$ is the vector of model parameters
- $\pi(t;\alpha,\beta,a,b) = \Pr\left(T \leq \frac{t-a}{b-a} \mid \alpha,\beta\right) = F_{\text{Beta}}\left(\frac{t-a}{b-a} \mid \alpha,\beta\right)$ is a weight function controlling the extent to which $S_{obs}(\cdot)$ and $S_{ext}(\cdot)$ are blended together
- $t \in [0, T^*]$, is the interval of times over which we want to perform our evaluation

NB: This is *not* the same as a "mixture cure model"!

- In MCM, one mixed survival curve (cured vs non cured individuals)
- In BSC, short- vs long-term processed modelled explicitly

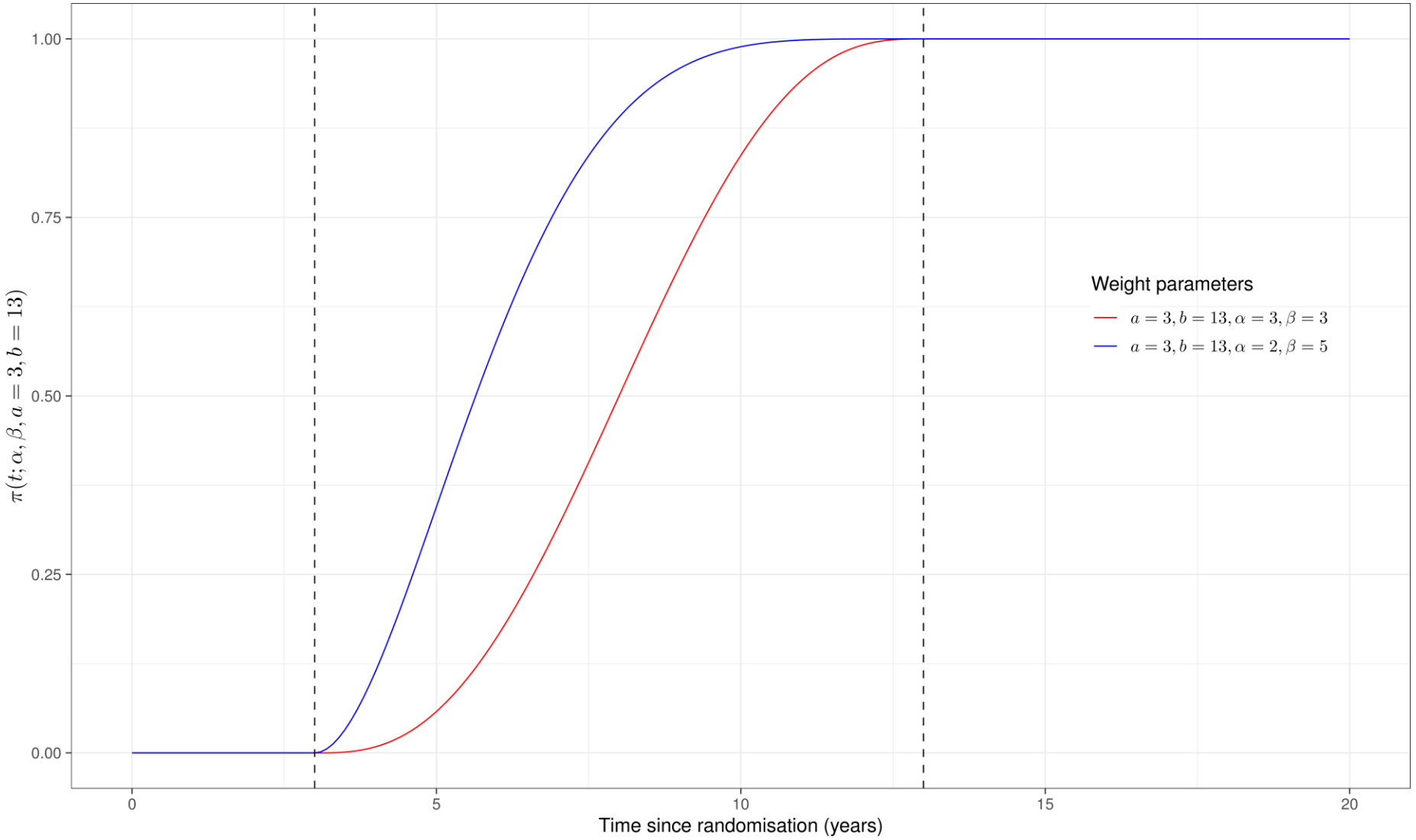
Example - Blended survival curves

Stats | Graphical representation | Weights | What do the weights do?...



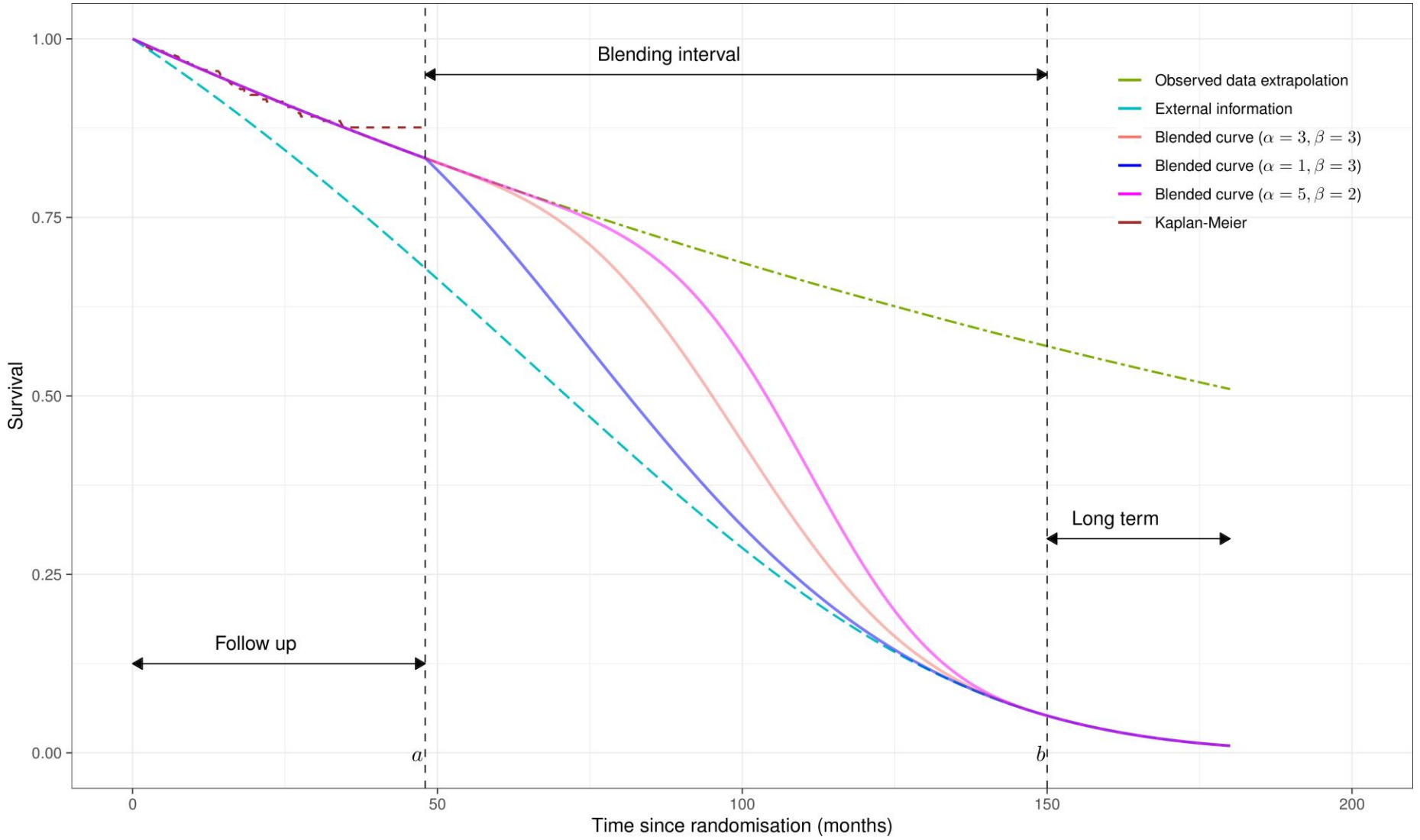
Example - Blended survival curves

Stats | Graphical representation | Weights | What do the weights do?...



Example - Blended survival curves

Stats | Graphical representation | Weights | What do the weights do?...



Comments

- The main point of the "blending" procedure is to recognise that, sometimes (often...), the observed data are just not good enough to simultaneously
  1. Provide the best fit to the observed data
  2. Provide a reasonable extrapolation for the long-term survival
- Instead, we let the observed data tell us about the short-term survival and some external information tell us something about the long-term survival

Comments

- The main point of the "blending" procedure is to recognise that, sometimes (often...), the observed data are just not good enough to simultaneously
  1. Provide the best fit to the observed data
  2. Provide a reasonable extrapolation for the long-term survival
- Instead, we let the observed data tell us about the short-term survival and some external information tell us something about the long-term survival

- When external data/RWE are available, they should be leveraged
  - BCS allows to do this in a relatively straightforward way
  - The "heavy-lifting" is done by the weight function that determines how the sources are blended together
  - This is based on (possibly untestable, but certainly open/upfront!) assumptions

Comments

- The main point of the "blending" procedure is to recognise that, sometimes (often...), the observed data are just not good enough to simultaneously
  1. Provide the best fit to the observed data
  2. Provide a reasonable extrapolation for the long-term survival
- Instead, we let the observed data tell us about the short-term survival and some external information tell us something about the long-term survival

- When external data/RWE are available, they should be leveraged
  - BCS allows to do this in a relatively straightforward way
  - The "heavy-lifting" is done by the weight function that determines how the sources are blended together
  - This is based on (possibly untestable, but certainly open/upfront!) assumptions

- This combination of difference sources of evidence is naturally Bayesian
  - Ultimately, we don't really care about the two components - rather we want to fully characterise the uncertainty in the blended curve
  - ... But to get that is simple algebra to combine the posterior distributions for $S_{obs}(t \mid \theta_{obs})$ and $S_{ext}(t \mid \theta_{ext})$


PBS NEWS HOUR
Thank you, all!