Hypothetical estimands in clinical trials - a unification of causal inference and missing data methods

Jonathan W. Bartlett

Department of Mathematical Sciences
University of Bath, UK
www.thestatsgeek.com

Danish Society for Biopharmaceutical Statistics
30 year anniversary
29th April 2022

Acknowledgements

This is joint work with Camila Olarte Parra (Bath) and Rhian Daniel (Cardiff).

This work was supported by a UK Medical Research Council grant (MR/T023953/1).

Context

ICH E9 estimand addendum

Causal inference

Missing data approaches

Conclusions

Outline

The context

Received: 23 April 2018 | Revised: 30 May 2018 | Accepted: 5 June 2018

DOI: 10.1111/dom.13437

WILE

ORIGINAL ARTICLE

Efficacy and safety of dapagliflozin or dapagliflozin plus saxagliptin versus glimepiride as add-on to metformin in patients with type 2 diabetes

Trial design

This trial is prototypical example of the setting we are considering.

- Randomised trial in type 2 diabetes
- $n = 939$ patients on metformin randomised 1:1:1 to receive:
  - dapa 10 mg, or
  - dapa 10 mg + saxa 5 mg, or
  - glimepiride 0 to 6 mg titrated
- Primary endpoint: change in HbA1c from baseline to 52 weeks
- HbA1c measured at week 0, 2, 4, 6, 8, 10, 12, 24, 36, 48, 52
- Open label rescue treatment with insulin possible, with initiation dependent on fasting glucose (FPG) thresholds

Trial analysis

- Primary results based on 'full analysis set'
- Only HbA1c values occurring prior to rescue treatment or discontinuation of randomised treatment used
- Linear mixed model fitted to the resulting dataset
- This was used to estimate differences in means between groups at 52 weeks
- Mixed models handle missing values assuming missing values are missing at random (MAR)

What's the estimand?

- As was historically the case with many such trials, not explicitly justified why values after rescue or discontinuation were excluded
- Only data before discontinuation or rescue are included, and an MAR analysis is performed
- As such, the implied **estimand** is something like 'what's the effect if the treatments are taken as assigned without rescue or discontinuation during 52 week follow-up?'
- Nowadays, following the ICH E9 estimand addendum, more explicit estimand specification is required

Outline

ICH E9 estimand addendum

In 2019 ICH published 'E9 (R1) addendum on estimands and sensitivity analysis in clinical trials'

It describes framework for defining clinical trial estimands

Estimand requires (according to this) specification of 5 attributes:

- the **treatments** being compared
- the **population** of patients targeted
- the **variable** to be obtained on each patient
- the strategies to handle **intercurrent events**
- the **population summary measure**, used to compare treatment groups

Intercurrent events

Intercurrent events (ICEs) are defined as:

'events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest.'

In our diabetes trial example, initiation of rescue or discontinuation or randomised treatment are ICEs.

ICH E9 intercurrent event strategies

- **treatment policy** - includes effects of ICE, by using outcomes irrespective of whether the ICE in question occurs
- **hypothetical** - what would happen in the hypothetical scenario where the ICE never occurred (e.g. withholding rescue treatment and preventing discontinuation)
- **composite variable** - occurrence of ICE included in outcome definition, e.g. a binary outcome of by 52 reduce HbA1c by x%, no need for rescue, no discontinuation of randomised treatment
- **while on treatment** - outcome value used up until time of ICE occurrence, e.g. HbA1c at 52 weeks or last one before rescue/discontinuation
- **principal stratification** - compare outcomes in strata of patients who would not experience ICE under assignment to either (or all) treatments

The diabetes trial

The diabetes trial is using the hypothetical strategy to handle initiation of rescue and discontinuation of randomised treatment.

Estimating different estimands - causal inference to the rescue?

The ICH E9 estimand addendum mostly does not mention concepts and statistical methods from causal inference.

The latter have been developed over the last 40 years, predominantly in the context of non-randomised observational studies.

We sought to deploy this existing causal inference machinery (see e.g. [1]) to the problem of **estimating hypothetical estimands** in clinical trials.

Outline

Causal inference with time-varying treatment

Causal inference methodology is well developed for estimating effects of time-varying treatments.

Here a key issue is time-varying confounding.

Time-varying confounders are variables occurring during follow up which affect both time-varying treatment and final outcome.

Correctly handling the latter requires the use of special (G-) methods, mostly developed by James Robins & coworkers.

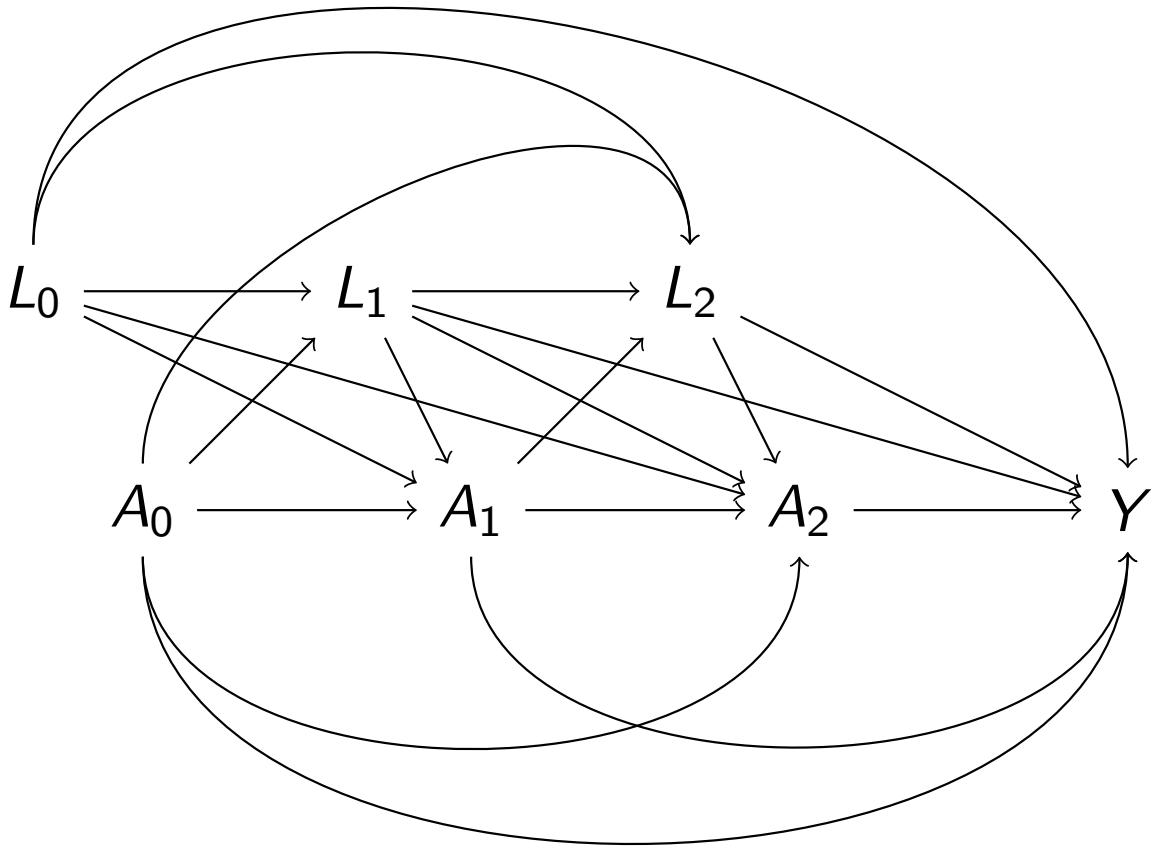
We can embed the occurrence of ICEs into this framework by treating occurrence of the ICEs as a time-varying treatment.

Notation

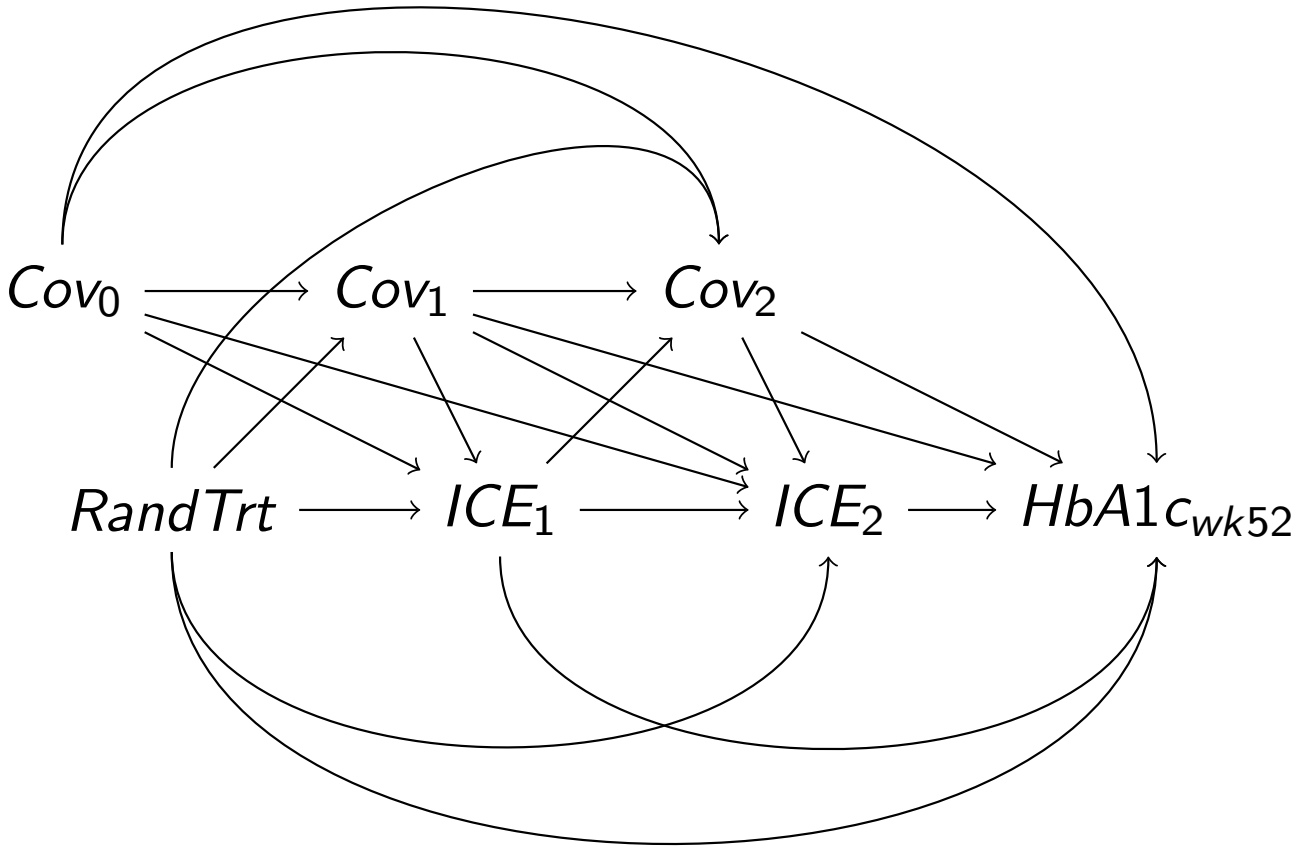
- Randomised treatment A_0
- Occurrence of ICE at follow-up time t, $A_t = 1$ if ICE occurred, and 0 otherwise
- Outcome of interest Y
- Common causes of ICEs and outcome L_t

To simplify notation and diagrams, in the following I will assume we have two follow-up time points at which ICE could occur.

Directed acyclic graph (DAG) (generic)



Directed acyclic graph (DAG) (diabetes trial)



Potential outcomes and hypothetical estimand

Let Y^{a_0,a_1,a_2} denote potential outcome if treatment A_0 is set to value a_0, ICE A_1 is set to a_1, and ICE A_2 is set to a_2.

The hypothetical estimand then targets

$$E(Y^{1,0,0}) - E(Y^{0,0,0})$$

In words: the mean difference in outcome between treatments if we prevented ICE from occurring at any time.

Identification assumptions

Consistency

Interventions (e.g. to prevent ICE) well defined so that potential outcomes are well defined and so that $Y = Y^{a_0,a_1,a_2}$ if $A_0 = a_0$, $A_1 = a_1$, $A_2 = a_2$

In words: if a patient's assigned treatment and actual ICE occurrence matches (a_0, a_1, a_2), then the outcome they actually experienced is $Y = Y^{a_0,a_1,a_2}$.

With $a_1 = a_2 = 0$, $Y = Y^{a_0,a_1,a_2} = Y^{a_0,0,0}$ is the outcome in the hypothetical trial where we set initial treatment to level a_0 and (somehow) prevent subsequent ICEs from occurring.

Identification assumptions

Conditional exchangeability (no unmeasured confounding)

$$Y^{a_0,0,0} \perp\!\!\!\perp A_1 | A_0 = a_0, L_0, L_1$$
$$Y^{a_0,0,0} \perp\!\!\!\perp A_2 | A_0 = a_0, A_1 = 0, L_0, L_1, L_2$$

for $a_0 = 0$ and $a_0 = 1$.

This holds under the DAG shown previously.

We need to measure (and use in the analysis) all **common causes of ICE and outcome** Y.

Identification assumptions

Positivity

$$P(A_1 = 0|A_0 = a_0, L_0, L_1) > 0$$
$$P(A_2 = 0|A_0 = a_0, A_1 = 0, L_0, L_1) > 0$$

for $a_0 = 0$ and $a_0 = 1$.

At all values of L_0 and L_1 which can occur, there is a non-zero probability of the ICE A_1 **not occurring** (similar for A_2).

This would be violated if rescue treatment A_1 is initiated deterministically based on L_1.

Short aside - is all this complication really needed?!

Are all these technicalities really needed?

They're not needed (apparently) when I do my MMRM MAR analysis.

As we will discuss further, there is really a 1-1 connection between the 'causal inference' and 'missing data approaches':

- Well defined potential outcomes = Well defined hypothetical missing values
- Conditional exchangeability = Missing at random

So they are there (and needed) in the missing data approach, but not as explicitly characterised typically (particularly the well defined potential outcome and consistency aspects).

Estimation methods from 'causal inference'

To estimate $E(Y^{0,0,0})$ and $E(Y^{1,0,0})$, we can use:

- G-formula
- Inverse probability of treatment weighting (here 'treatment' is A_0, A_1, A_2)

G-formula

To estimate $E(Y^{0,0,0})$:

1. specify and fit models for
   - $f(L_1|A_0, L_0)$
   - $f(L_2|A_0, A_1, L_0, L_1)$
   - $f(Y|A_0, A_1, A_2, L_0, L_1, L_2)$
2. for every patient
   - simulate L_1^* from $f(L_1|A_0 = 0, L_0)$
   - simulate L_2^* from $f(L_2|A_0 = 0, A_1 = 0, L_0, L_1^*)$
   - simulate Y^* from $f(Y|A_0 = 0, A_1 = 0, A_2 = 0, L_0, L_1^*, L_2^*)$
   - calculate mean of Y^* across patients

For $E(Y^{1,0,0})$ replace $A_0 = 0$ with $A_0 = 1$ in the second part.

G-formula intuition and points to note

G-formula can be viewed as simulating/imputing longitudinal history (L_1, L_2, Y) for every patient under the hypothetical scenarios of interest where ICE does not occur.

The preceding G-formula implementation exploits randomisation which means $A_0 \perp\!\!\!\perp L_0$, so that a model for L_0 is not needed.

Observations of L_2 and Y after occurrence of ICE in the real trial are (by default) not excluded from the model fitting process.

But this requires us to model the effects of ICE occurrence (i.e. of A_1 and A_2) on L_2 and Y.

This differs to the linear mixed model missing data analysis performed.

Excluding data after ICE

In fact, since for the hypothetical estimand we are only interested in no ICE potential outcomes, we can avoid modelling effects of ICE A_1 on L_2 and Y and A_2 on Y.

We can specify models for:

- $f(L_1|A_0, L_0)$
- $f(L_2|A_0, A_1 = 0, L_0, L_1)$
- $f(Y|A_0, A_1 = 0, A_2 = 0, L_0, L_1, L_2)$

since these are all we need for step 2.

This variation of G-formula makes fewer modelling assumptions, but ignores some of the observed data.

It is more robust, but less efficient statistically than the first implementation.

Inverse probability of treatment weighting (IPTW)

To estimate $E(Y^{0,0,0})$ (similarly for $E(Y^{1,0,0})$):

1. Fit logistic regressions for
   - $P(A_1 = 0|A_0, L_0, L_1)$
   - $P(A_2 = 0|A_0, A_1, L_0, L_1, L_2)$
2. Calculate weights
   $W_i = \frac{1}{P(A_1=0|A_0=0,L_{i0},L_{i1})P(A_2=0|A_0=0,A_1=0,L_{i0},L_{i1},L_{i2})}$
3. Estimate $E(Y^{0,0,0})$ by

$$\frac{\sum_{i=1}^{n} I(A_0 = 0, A_1 = 0, A_2 = 0) W_i Y_i}{\sum_{i=1}^{n} I(A_0 = 0, A_1 = 0, A_2 = 0) W_i}$$

IPTW intuition and points to note

This is a weighted average of the outcomes among those on treatment 0 who in the real trial did not experience the ICE.

Again, we can consider alternative implementations: only fit models for

- $P(A_1 = 0|A_0, L_0, L_1)$
- $P(A_2 = 0|A_0, A_1 = 0, L_0, L_1, L_2)$

Outline

Missing data approaches

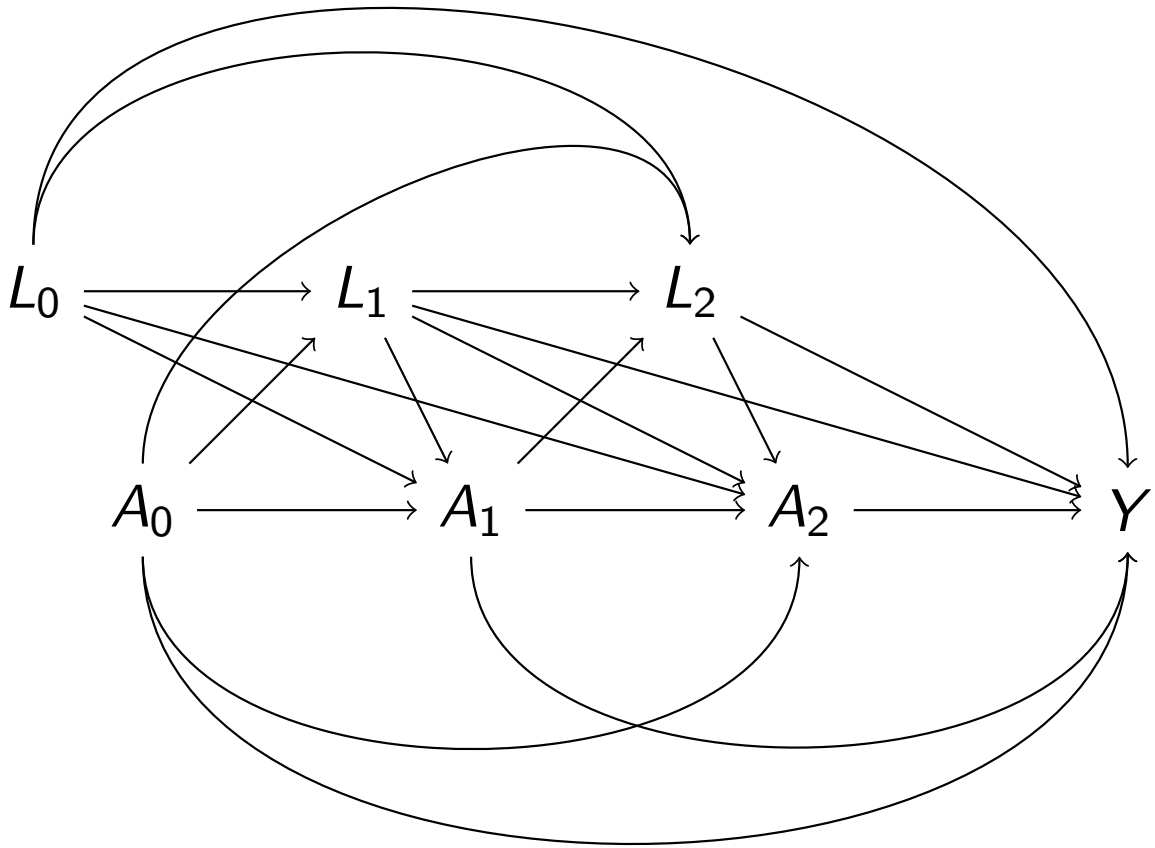
Recall the standard approach excludes data on HbA1c after ICE occurs, and fits a linear mixed model to repeated measures of HbA1c assuming missing values are missing at random (MAR).

First, let's consider the MAR assumption.

The 'full data' of interest here are the potential outcomes $Y^{A_0,0,0}$ - i.e. for each patient what their outcome would have been under no ICE.

Recall our DAG from earlier...

Directed acyclic graph (DAG)



Checking MAR

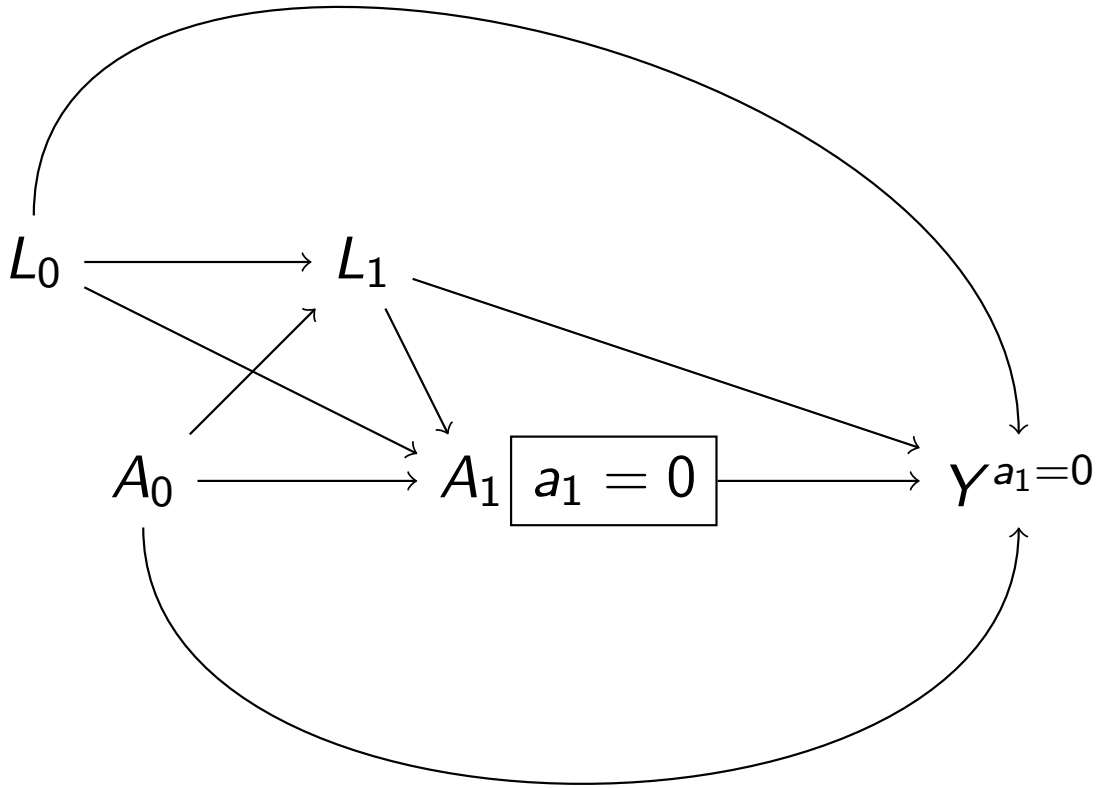
The DAG contains the actual outcomes, not the potential outcomes of interest $Y^{A_0,0,0}$.

We make use **single-world intervention graphs** (SWIG) [3].

This shows what happens in the hypothetical world of interest under no ICE.

We will consider the even simpler case with just one intermediate follow-up visit...

SWIG for the no ICE world



Checking MAR

MAR means probability of being missing (here indicated by A_1) is independent of the partially observed variable ($Y^{a_1=0}$) conditional on the fully observed variables (A_0, L_0, L_1).

This conditional independence holds in this SWIG by d-separation - there are no open paths from A_1 to $Y^{a_1=0}$.

For the case with more time points, which is slightly more involved, see our paper.

Implications

The MAR assumption for the partially observed no ICE outcomes is correct **if** we condition on all the common causes (i.e. L_1 and L_2) of ICE occurrence and Y.

In the diabetes trial, the mixed model is fitted to repeated measurements of HbA1c.

But fasting plasma glucose (FPG) measurements were used to determine eligibility for rescue treatment.

Conditioning/adjusting for longitudinal measurements of HbA1c is is not sufficient for MAR to hold.

Missing data estimation approaches

There are various 'missing data' approaches, assuming MAR:

- observed data likelihood methods (e.g. linear mixed models)
- multiple imputation
- inverse probability of missingness weighting

Equivalence of G-formula and linear mixed models

We show (see paper) that linear mixed models (discarding post ICE data) correspond to a particular version of G-formula.

Intuitively this is because:

- the (mean of the) simulated outcomes from G-formula for those with no ICE match the observed mean from those with no ICE used to fit the models
- the simulated outcomes from G-formula for those who did experience ICE are the same as the MAR implied predicted outcomes from the mixed model

But for MAR to hold, as previously mentioned, need to adjust for common causes of ICE and Y

Doing this in practice (i.e. with software) with mixed models is possible, but gets trickier to implement.

Multiple imputation

If based on same data and model assumptions, multiple imputation and mixed models are (essentially) equivalent.

Therefore, multiple imputation (discarding post ICE data) also corresponds to a particular version of G-formula from causal inference.

Software for multiple imputation makes it easier to adjust for full set of time-varying confounders L which affect ICE occurrence and final outcome Y.

Multiple imputation may therefore be an attractive approach for adjusting for full set of variables affecting ICE occurrence and final outcome Y.

IPW methods and positivity violations

Inverse probability of treatment weighting (from 'causal inference') is the same as inverse probability of missingness weighting (from 'missing data').

If the positivity assumption is violated, IPW estimators fail.

G-formula and missing data likelihood methods (mixed models and multiple imputation) can still give consistent estimates.

But their consistency relies on modelling assumptions that cannot be checked from the observed data.

We are extrapolating beyond the data.

Due to the popularity of likelihood/imputation approaches in trials, in the trials arena this issue I think has not been given so much attention...

Practicalities

Implementations of G-formula and IPW both assume there are no missing 'actual' data.

In practice we often have some (often not too much) intermediate missing data.

This is the case in the diabetes trial, which are currently analysing using these approaches.

To handle this when applying G-formula and IPW, we are first using MI to handle missing actual data.

Outline

Conclusions - 1

- Formalism of causal inference framework is useful for clarifying assumptions we must make when estimating hypothetical estimands
- In particular, important to think about feasibility, even in theory, of how one would prevent the ICEs occurring that you deal with by the hypothetical strategy
- To estimate hypothetical estimands, need to adjust for all common causes of ICE (c.f. missingness) and final outcome / common
- Estimation of hypothetical estimands can via G-formula use outcomes measured after ICE occurrence, if handled appropriately in modelling
- Using outcomes after ICE offer opportunity for improved power, but complicates modelling

Conclusions - 2

- Linear mixed model and multiple imputation discarding post-ICE data (the standard approaches used now) can be viewed as particular implementations of G-formula from causal inference
- Pre-print of our paper available at `https://arxiv.org/abs/2107.04392` [2], shortly to be published in Statistics in Biopharmaceutical Research
- Paper includes simulation study comparing different methods

Ongoing and future work

- We are currently analysing the aforementioned diabetes trial to compare the approaches discussed here, with a paper to follow soon
- Further work needed to explore whether in practice precision gains through exploiting post ICE data would be sufficiently large to negate increased concerns about model misspecification
- Hypothetical estimand as defined here is a controlled direct effect of treatment on outcome
- An argument against it is it prevents all ICEs and this is scenario is never realistic in practice
- Other types of direct effects could be of interest and may be preferable, such as natural direct effects or interventional direct effects, which are also investigating

References I

[1] M. Hernán and J. Robins.

Causal Inference: What If.

Boca Raton: Chapman & Hall/CRC, 2020.

[2] C. Olarte Parra, R. M. Daniel, and J. W. Bartlett.

Hypothetical estimands in clinical trials: a unification of causal inference and missing data methods.

arXiv preprint arXiv:2107.04392, 2021.

[3] T. S. Richardson and J. M. Robins.

Single world intervention graphs: a primer.

In *Second UAI workshop on causal structure learning, Bellevue, Washington*. Citeseer, 2013.