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

European Statistical Forum
16th November 2021

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

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 of repeated HbA1c measures
- 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 is/was the case with many such trials, not entirely clear why values after rescue or discontinuation were excluded
- This arguably contravenes intention to treat principle
- Although not stated, the exclusion is done in order to estimate effects if you take assigned treatment
- 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?'
- The hypothetical strategy is being (implicitly) used to handle the intercurrent events (ICEs) of rescue or discontinuation

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

Here a key issue is time-varying confounding.

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

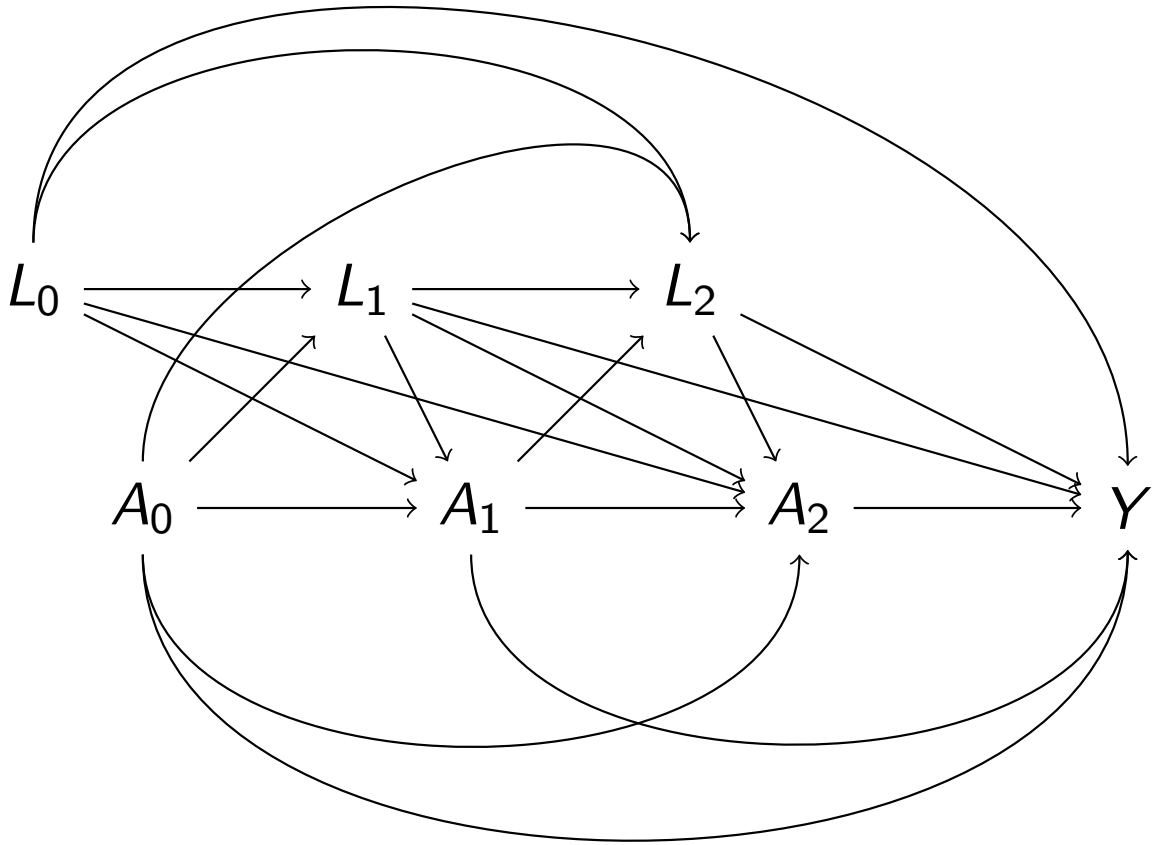
We can embed the occurrence of ICEs into this framework by treating occurrence of the ICEs as a time-varying treatment.

Notation

- Randomised treatment A_0
- Occurrence of ICE at time $t > 0$, A_t (e.g. receiving rescue treatment or discontinuation of randomised treatment)
- Outcome of interest Y (e.g. HbA1c at final time point)
- Common causes of ICEs and outcome L_t (e.g. HbA1c and FPG measured at time t)

For concreteness, in the following I will assume we have two follow-up time points at which ICE could occur.

Directed acyclic graph (DAG)



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.

Assumptions - no unmeasured confounding

G-methods assume **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.

In the diabetes trial, this means we must adjust for FPG, not just HbA1c.

Assumptions - positivity

Positivity

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

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.

This could happen with insulin rescue in diabetes trials, if patients get rescued if and only if FPG exceeds a threshold.

Estimation

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

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

I will talk mostly about G-formula, less about inverse weighting (see paper for more details).

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 scenario of interest where ICE does not occur.

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 (effects of rescue/discontinuation) on L_2 and Y.

This differs to the linear mixed model missing data analysis performed in the diabetes trial, which discards post-ICE data.

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 version of G-formula is more robust, but less efficient statistically than the first implementation.

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).

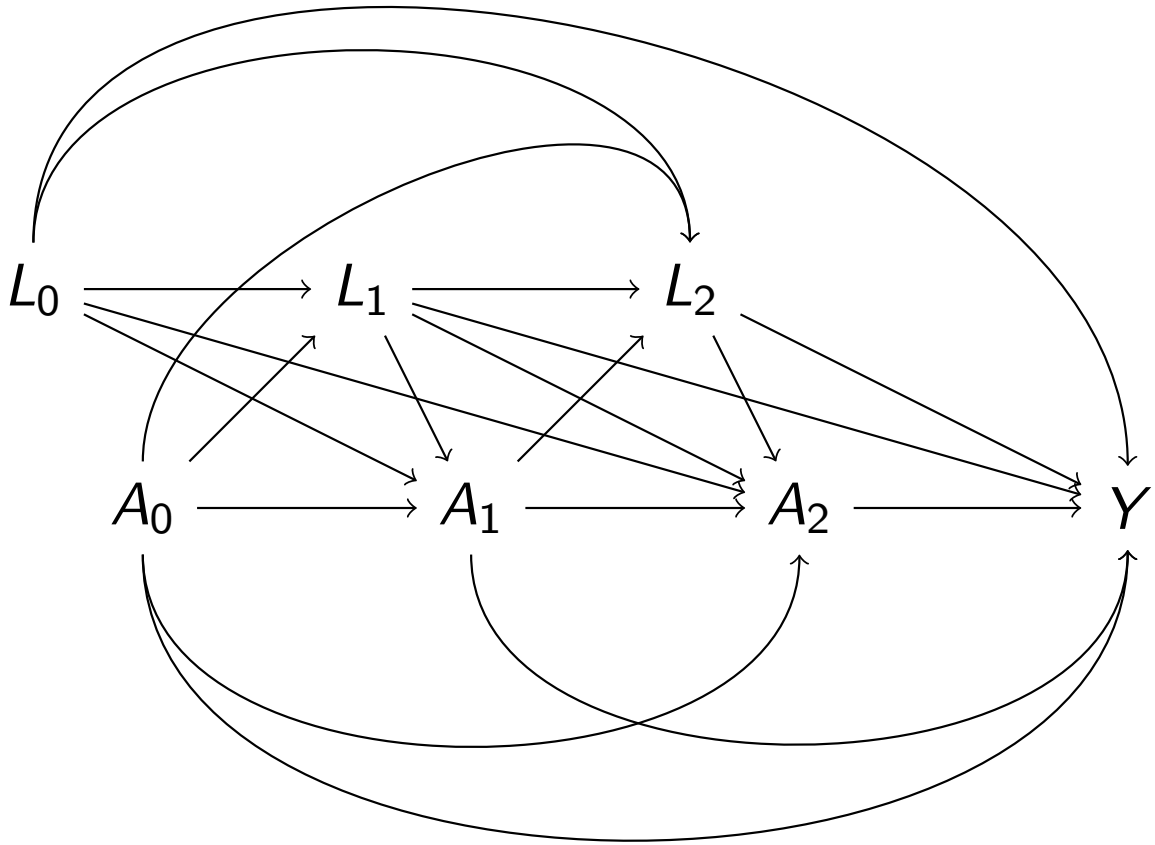
First, let's consider the MAR assumption.

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 (no rescue).

Recall our DAG from earlier...

Directed acyclic graph



Missing at random (MAR)

The DAG can be used (see paper) to show 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 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.

This conclusion is in agreement with our earlier conclusion about the no unmeasured confounding assumption.

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 from causal inference.

But need to include repeated measures of all common causes of ICE and Y, e.g. both FPG and HbA1c at each visit.

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.

Whether the extrapolation we make is reasonable will have to be considered on a case by case basis.

Outline

Conclusions

- To estimate hypothetical estimands, need to adjust for all common causes of ICE and final outcome
- Estimation of hypothetical estimands can use outcomes measured after ICE occurrence, if handled appropriately in modelling - opportunity for improved power
- Linear mixed model and multiple imputation discarding post-ICE data can be viewed as particular implementations of G-formula from causal inference
- Pre-print of paper available at `https://arxiv.org/abs/2107.04392`
- Paper includes simulation study comparing different methods.

References I

[1] M. Hernán and J. Robins.

Causal Inference: What If.

Boca Raton: Chapman & Hall/CRC, 2020.

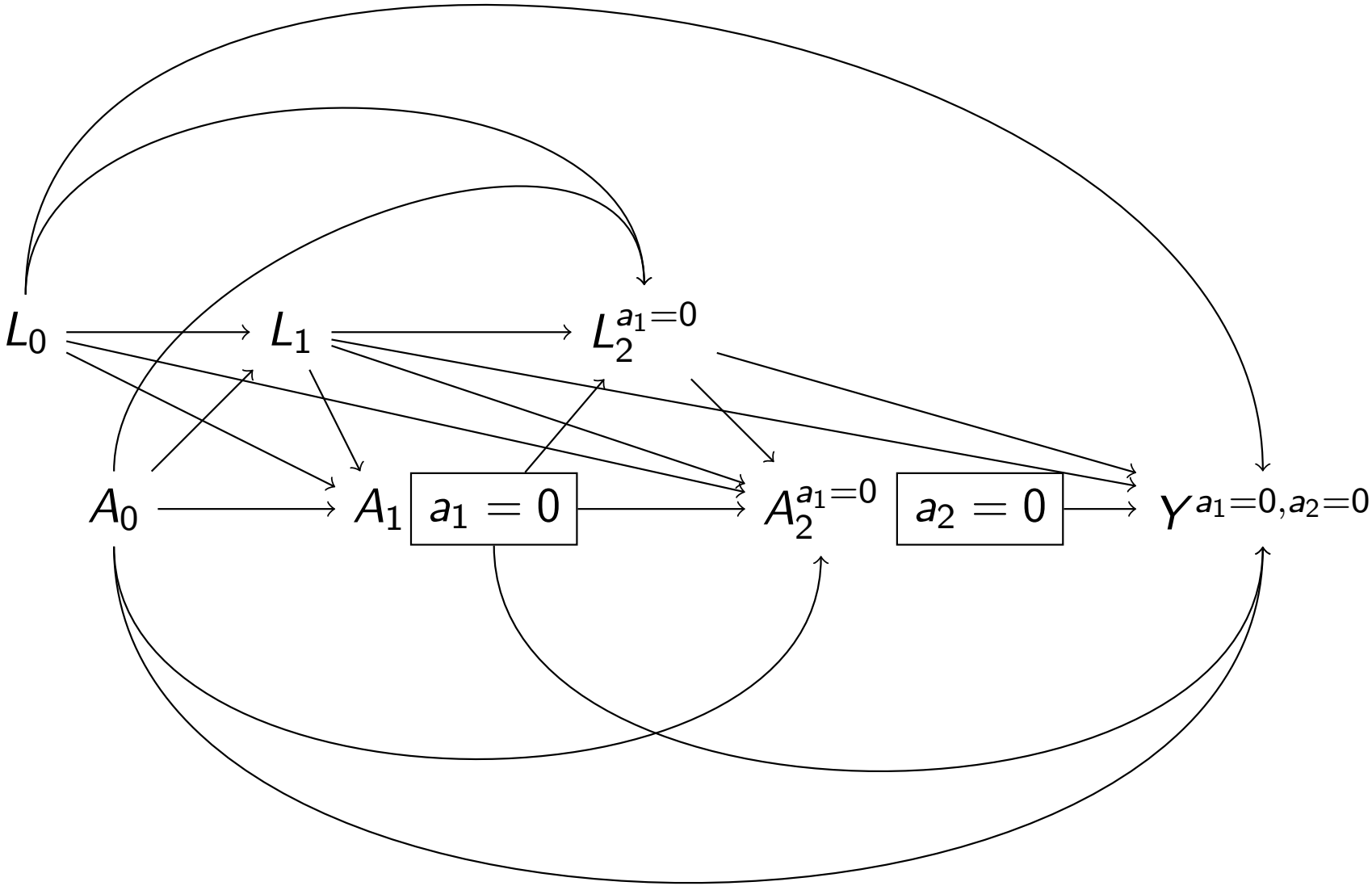
Checking MAR

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) [1].

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

SWIG for the no ICE world



Checking MAR

We have monotone missingness (by construction).

MAR means:

- $A_1 \perp\!\!\!\perp (L_2^{a_1=0}, Y^{a_1=0,a_2=0}) | A_0, L_1, L_0$
- $A_2 \perp\!\!\!\perp Y^{a_1=0,a_2=0} | A_1 = 0, A_0, L_2, L_1, L_0$

The first can be read off immediately as being true using d-separation.

For the second, in those with $A_1 = 0$, by the consistency assumption $L_2 = L_2^{a_1=0}$ and $A_2 = A_2^{a_1=0}$.

So the second part is equivalent to $A_2^{a_1=0} \perp\!\!\!\perp Y^{a_1=0,a_2=0} | A_1 = 0, A_0, L_2^{a_1=0}, L_1, L_0$, and the SWIG shows this holds.