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

Jonathan W. Bartlett

Department of Medical Statistics
LSHTM
https://thestatsgeek.com

Australian Trials Methodology Conference 2023
17th October 2023

Acknowledgements

This is joint work with Camila Olarte Parra (LSHTM), Rhian Daniel (Cardiff) and David Wright (AstraZeneca)

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

Motivation

ICH E9 estimand addendum

Causal inference and G-formula

Missing data approaches

Conclusions

Outline

Illustrative motivating trial

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 [7] 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 (so-called MMRM) 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?

- The estimand was not stated explicitly (the trial pre-dated the ICH E9 estimand addendum)
- Exclusion of HbA1c after rescue or discontinuation + MAR assumption implies a hypothetical estimand
- Namely - what would have happened had rescue not been made available and discontinuation had not been permitted
- Important questions about relevance of such an estimand, and indeed whether it is sufficiently well-defined - how would we completely prevent discontinuation?
- I shall not dwell further on this important aspect...

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.

Treatment policy and hypothetical '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)

Addendum also describes composite, while-on-treatment and principal stratification strategies.

The diabetes trial

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

Although note this was not stated explictly - the diabetes paper pre-dates the ICH E9 estimand addendum.

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. [3]) 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.

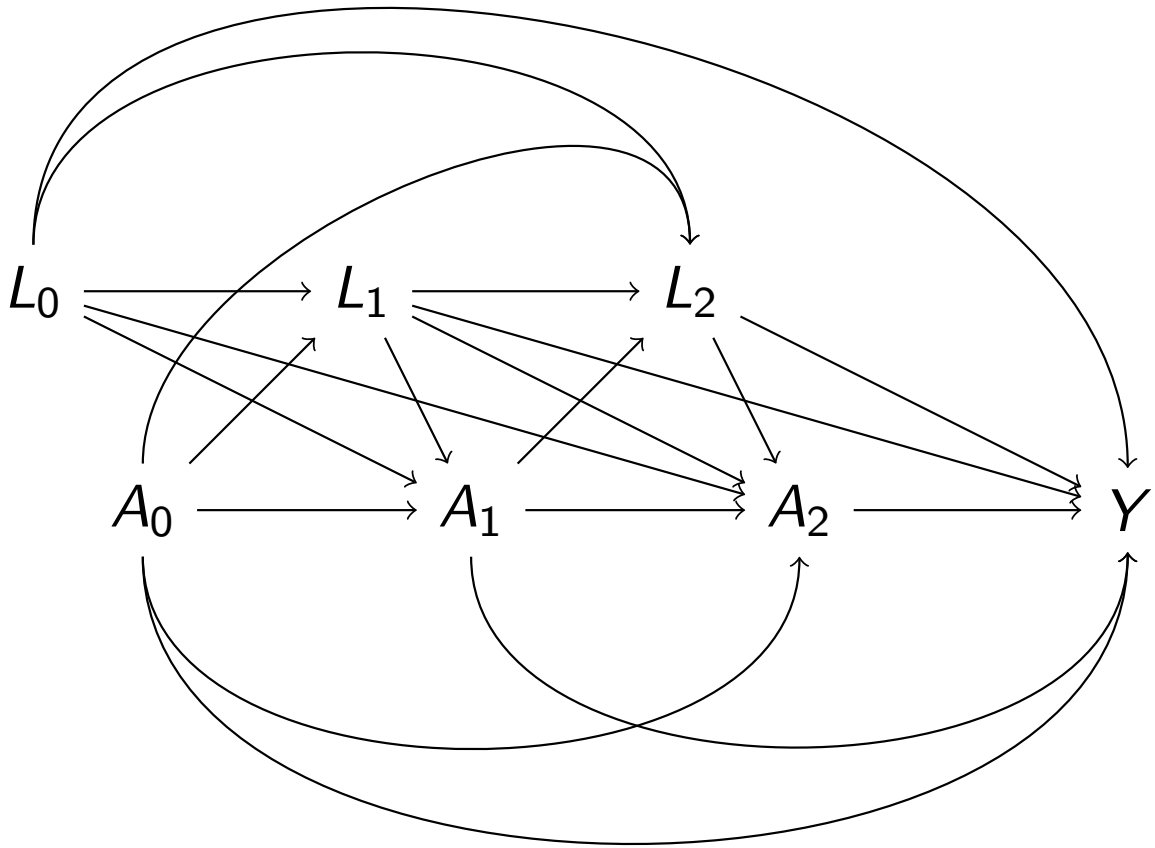
We can embed the occurrence of ICEs into this framework by treating occurrence of the ICEs as a time-varying treatment.

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 fasting plasma glucose (FPG) measured at time t)

For simplicity, in the following I will assume we have just 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 (a) 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 - consistency

Consistency (not usual 'statistical consistency')

Interventions to prevent ICE are well defined so that $Y = Y^{a_0,a_1,a_2}$ if $A_0 = a_0$, $A_1 = a_1$, $A_2 = a_2$

$\Rightarrow$ in the **actual trial**, for a patient who did not need rescue or discontinue treatment, their actual outcome Y is identical to the outcome they would have in the hypothetical trial where we withhold rescue and prevent discontinuation.

To convincingly argue why consistency would hold, we need to to try and articulate how the ICEs would be prevented.

C.f. Hernán [4] on causal effect of obesity on mortality being ill-defined - effect depends on how you change someone's weight.

Assumptions - no unmeasured confounding

Conditional exchangeability (no unmeasured confounding)

In our case, this means that ICE occurrence at a given visit is independent of final outcome, conditional on measured past.

This holds under the DAG shown previously.

But, we need to measure (and adjust for in the analysis) **all common causes of ICE and outcome** Y.

In a diabetes trial, this means we should adjust for FPG, not just HbA1c, if FPG influences rescue decisions (c.f. Holzhauer *et al* [5])

Assumptions - positivity

Positivity

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.

Positivity is not actually needed for G-formula, but then we are relying on the model to extrapolate beyond the data.

Causal inference estimation methods

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

- G-formula (sometimes called G-computation)
- Inverse probability of treatment weighting (here 'treatment' is A_0, A_1, A_2)
- G-estimation [6]

I will focus on G-formula, and how it relates to MMRM and multiple imputation. See [8] for IPW in this context.

G-formula v1 ('standard version')

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 simulates (imputes!) 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 a 'standard' MMRM analysis, which discards post-ICE data.

If trial did not collect data after ICE, we of course cannot model what happens post-ICE.

G-formula v2 - 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:

- $f(L_1|A_0, L_0)$ (all patients)
- $f(L_2|A_0, A_1 = 0, L_0, L_1)$ (only patients ICE free following visit 1)
- $f(Y|A_0, A_1 = 0, A_2 = 0, L_0, L_1, L_2)$ (only patients ICE free following visit 2)

since these are all we need for step 2.

This (non-standard) version of G-formula is more robust, but less efficient statistically than the first implementation.

G-formula - reflections

After fitting required models, G-formula discards all observed data and simulates new data for all patients.

We then analyse the simulated data.

I anticipate (legitimate) hesitancy to this - can we really base our analysis in the end on a completely simulated dataset?!

Outline

Missing data approaches

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

If based on same data and model assumptions, multiple imputation (MI) and MMRM are (essentially) equivalent [2].

Let's consider MI, where we impute the post-ICE HbA1c values.

How does this compare to the G-formula method?

G-formula and MI

	G-formula v1	G-formula v2	MI
Data used to fit imp. models	Pre and post-ICE	Pre-ICE	Pre-ICE
Data imputed	All times for all patients	All times for all patients	Post-ICE times in patients with ICE

G-formula v1 - 'standard' G-formula
G-formula v2 - modified G-formula where we only fit models using ICE free patients at each visit

G-formula v2 and MI still differ - G-formula replaces all observed data with simulated/imputed values, whereas MI only imputes post ICE data.

G-formula and MI equivalence

In fact, at least for certain (important) model setups, G-formula v2 and MI are the same.

Both methods impute final outcomes for patients who experience ICE, based on same model fits.

For patients with no ICE, the mean of their imputed values in G-formula v2 matches the mean of their observed values (as used by MI).

This is basically because in regression, the mean of the fitted values equals the mean of the dependent variable in the sample.

This can also be used to argue that G-formula v2 and MMRM are the same.

G-formula and MMRM equivalence

Since MMRM $\approx$ MI (when using same data and models),
and MI $\approx$ G-formula v2,
it follows that MMRM $\approx$ G-formula v2 **(when using same data and models)**.

In our paper we make the argument more directly (and carefully) for continuous outcomes, where G-formula does not require simulation at all.

G-formula via multiple imputation

Is it possible to use MI methods but exploit observed post-ICE data when fitting models, like 'standard' G-formula (v1)?

Yes - based on ideas from using MI to create synthetic datasets [10].

This is potentially useful, because we can use MI to handle both missing actual data (assuming MAR) and missing counterfactual data in one approach.

For details of performing G-formula via Bayesian multiple imputation, see [1].

Outline

Conclusions

- Hypothetical estimands require careful specification to be well defined and relevant
- For estimation, need to adjust for all common causes of ICE and final outcome
- Post-ICE data can be exploited for improved power, but this requires more modelling assumptions
- MMRM and MI discarding post-ICE data can be viewed as particular implementations of G-formula method from causal inference
- MI can be adapted to exploit post-ICE data, providing a convenient route to handling missing actual and counterfactual data
- I haven't talked about inverse probability of (treatment/missingness) weight methods - these are identical if we set post-ICE to missing.

Papers

- For more on links between causal inference and missing data methods for hypothetical estimands, see [8]
- For more on performing G-formula using MI methods and software, see [1] and `gFormulaMI` R package on CRAN
- More details on application of all these methods to the diabetes trial mentioned earlier, and the results, are in [9]

References I

[1] J. W. Bartlett, E. Parra, Camila Olarte Granger, R. H. Keogh, E. W. van Zwet, and R. M. Daniel.

G-formula for causal inference via multiple imputation.

arXiv preprint arXiv:2301.12026, 2023.

[2] J. R. Carpenter and M. G. Kenward.

Missing data in randomised controlled trials - a practical guide.

Birmingham, UK. National Health Service Co-ordinating Centre for Research Methodology, 2007.

[3] M. Hernán and J. Robins.

Causal Inference: What If.

Boca Raton: Chapman & Hall/CRC, 2020.

References II

[4] M. A. Hernán and S. L. Taubman.

Does obesity shorten life? the importance of well-defined interventions to answer causal questions.

International journal of obesity, 32(3):S8–S14, 2008.

[5] B. Holzhauer, M. Akacha, and G. Bermann.

Choice of estimand and analysis methods in diabetes trials with rescue medication.

Pharmaceutical statistics, 14(6):433–447, 2015.

[6] F. Lasch, L. Guizzaro, F. Petavy, and C. Gallo.

A simulation study on the estimation of the effect in the hypothetical scenario of no use of symptomatic treatment in trials for disease-modifying agents for alzheimer's disease.

Statistics in Biopharmaceutical Research, 15(2):386–399, 2023.

References III

[7] D. Müller-Wieland, M. Kellerer, K. Cypryk, D. Skripova, K. Rohwedder, E. Johnsson, R. Garcia-Sanchez, R. Kurlyandskaya, C. D. Sjöström, S. Jacob, et al.

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

Diabetes, Obesity and Metabolism, 20(11):2598–2607, 2018.

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

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

Statistics in Biopharmaceutical Research, 15(2):421–432, 2023.

[9] C. O. Parra, R. M. Daniel, D. Wright, and J. W. Bartlett.

Estimating hypothetical estimands with causal inference and missing data estimators in a diabetes trial.

arXiv preprint arXiv:2308.13085, 2023.

References IV

[10] T. E. Raghunathan, J. P. Reiter, and D. B. Rubin.

Multiple imputation for statistical disclosure limitation.

Journal of Official Statistics, 19(1):1, 2003.