Three causal lessons from our Simulation Learner

Els Goetghebeur, Gent (B) and Saskia le Cessie, Leiden (NL)

on behalf of STRATOS TG 7

Virtual Biometrics Conference 2020

with Bianca De Stavola; Erica Moodie; Ingeborg Waernbaum

STRATOS www.stratos-initiative.org
The TG7 website: www.ofcaus.org

Three causal lessons from our Simulation Learner

- SUTVA is fiction
- Randomisation: no IV for post randomisation exposures
- Averaging causal effects over observed (experimental) IV may be irrelevant.

Several reasons to perform a simulation

- Show properties of a new method
  (e.g small sample behaviour)
- Compare the performance of different methods under different conditions
- Verify calculations /analysis
  (i.e check (power) calculations, check an R-function)
- Deepen understanding of data and methods to analyse them

Several reasons to perform a simulation

- Show properties of a new method
  (e.g small sample behaviour)
- Compare the performance of different methods under different conditions
- Verify calculations /analysis
  (i.e check (power) calculations, check an R-function)
- Deepen understanding of data and methods to analyse them
  $->$ incl. the target estimand

All the more when you draw inference on potential outcomes

Morris et al.'Using simulation studies to evaluate statistical methods', SIM 2019

The Simulation Learner

TG 7 wrote tutorial on causal questions and principled answers

- Overview of causal concepts
- Overview of analysis methods for time-fixed exposures

The Simulation Learner

TG 7 wrote tutorial on causal questions and principled answers

- Overview of causal concepts
- Overview of analysis methods for time-fixed exposures

Simulation Learner

- Simulated data inspired by existing trial
- Illustrates concepts and methods on data $->$ augments data with alternative exposures and potential outcomes

Franklin et al. 'Plasmode Simulation for the Evaluation of Pharmacoepidemiologic Methods in Complex Healthcare Databases', Comp. Stat. & Data Analysis, 2014

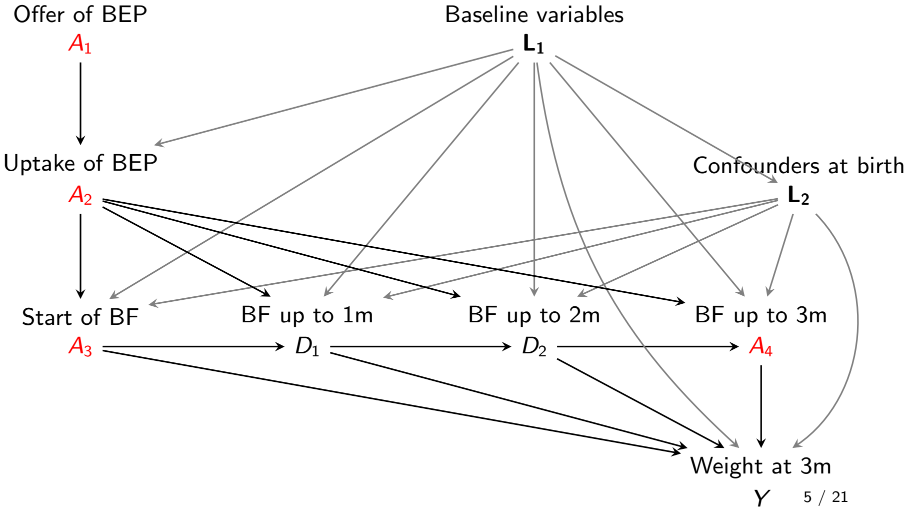
Promotion of Breastfeeding Intervention Trial - PROBIT

(Kramer et al, 2001) our inspiration

- Women living in a low income area of Belarus who gave birth to a full-term singleton baby between June '96- Dec '97
- were (cluster) **randomised to BF encouraging educational program** or not, during their last term of pregnancy.
- All babies were weighed at age 3 months.
- A simulated version of individually *randomised* women **Probitsim** included: 17,044 women with singleton births (8,667 in the active arm and 8,377 in the control arm).

Simulated on www.ofcaus.org

Sketch of data generating model



Possible exposures $\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \mathfrak{a}_4$ relevant for whom?

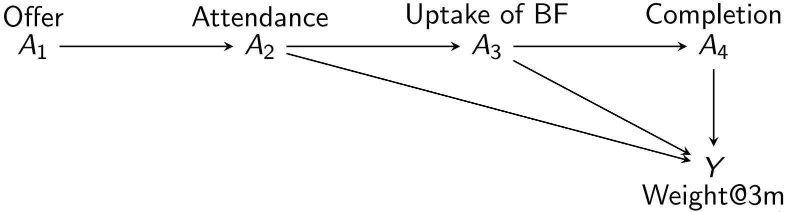
When aiming to change

- the nature of the invitation to the breastfeeding program
- the content of the breastfeeding program to improve uptake of breastfeeding
- one's decision to start breastfeeding
- the supporting measures to improve maintaining breastfeeding for the full 3 months

Potential consequences of setting $\mathfrak{a}_1$

- $A_{2\mathfrak{a}_1(a)}$: the potential value of A_2 if A_1 is set to the value a.
- $A_{3\mathfrak{a}_1(a)} = 1$ would start BF if the programme were offered ($a = 1$) or not ($a = 0$).
- $A_{3\mathfrak{a}_1(a_1),\mathfrak{a}_2(a_2)} = 1$ would start BF if the programme were offered and followed ($a_1 = 1, a_2 = 1$), or offered but not followed ($a_1 = 1, a_2 = 0$) or not offered ($a_1 = 0, a_2 = 0$).

$$Y_{\mathfrak{a}_1(a)} = Y_{\mathfrak{a}_1(a)}(\mathfrak{a}_1 = a, A_{2\mathfrak{a}_1(a)}, A_{3\mathfrak{a}_1(a)}, A_{4\mathfrak{a}_1(a)})$$



Stable Unit Treatment Value Assumption

has an aim to uniquely define the potential outcome. That involves

- No interference: treatment applied to one unit does not affect the outcome for another unit
- There is only a single version of each treatment level
- Consistency: observed outcome to observed exposure identical to the potential outcome to that set exposure

The binary indicator A_1 : being offered the breastfeeding education program or not

- is uniquely well defined, but
- its consequence depends on level of uptake - both of the program and of the actual breastfeeding (duration)
- in the study population

VanderWeele, T. J., Hernan, M. A. (2013). Causal inference under multiple versions of treatment. Journal of causal inference, 1(1), 1-20.

The binary indicator A_3 : starting breastfeeding or not

- is uniquely well defined, but
- its consequence depends on level of uptake - how much milk, for how long the actual breastfeeding ... interactions inevitable

So...

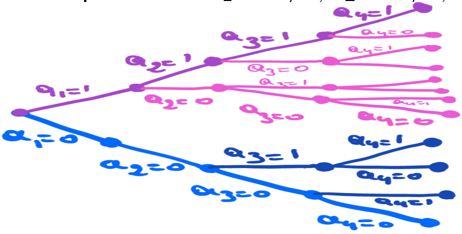
- For transportability/interpretability: description of the nature of exposure in context is needed
- We draw from a distribution of potential outcomes: variation may incorporate variation in treatment (delivery)
- Drug treatment intake daily - uniquely defined? Resolution of your view: taken when, how (with water or whisky) ...

Simulation Learner

We simulate the parallel worlds for the PROBIT-like trial

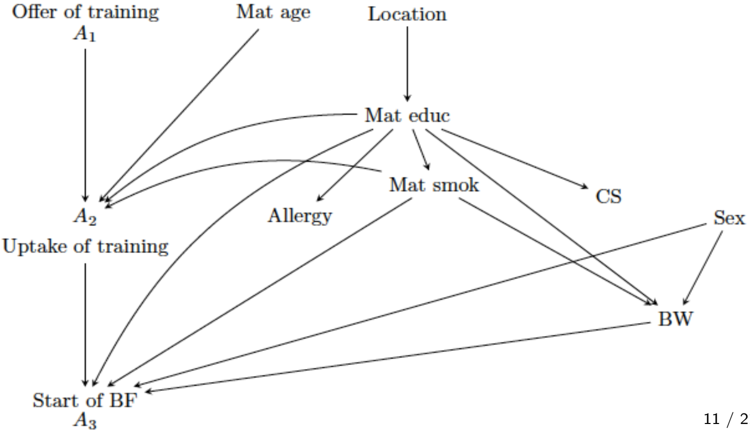
- for the composition of the study population, i.e. mimic the baseline covariate distribution L in which we
- 'let' every person experience each of the possible exposure sequences :

  part of 16 options for $\mathfrak{a}_1 = 0/1, \mathfrak{a}_2 = 0/1, \mathfrak{a}_3 = 0/1, \mathfrak{a}_4 = 0/1$



- followed by the corresponding potential outcome(s)
- besides the potential exposures and outcomes, we also simulate 'the observed' exposures and outcomes

Data generating model for ‘observed’ A2 and A3 given A1, L1, and L2



Overall		
	$A_1 = 0$	$A_1 = 1$
	N (%)	N (%)
$A_2 = 0$	8377 (100)	3083 (35.6)
$A_2 = 1$	0 (0)	5584 (64.4)
$A_3 = 0$	4226 (50.5)	2782 (32.1)
$A_3 = 1$	4151 (49.5)	5885 (67.9)
All	8377 (100)	8667 (100)

Among those with $A_1 = 1$		
	$A_2 = 0$	$A_2 = 1$
	N (%)	N (%)
$A_3 = 0$	1745 (56.6)	1037 (18.6)
$A_3 = 1$	1338 (43.4)	4547 (81.4)
All	3083 (100)	5584 (100)

'Compliance' with program offer

Consider A_1, the intervention of offering the program
2 types of compliers with the program offer:

- Program offer accepters: $\{A_{2\mathfrak{a}_1(1)} = 1 \text{ and } A_{2\mathfrak{a}_1(0)} = 0\}$
- BF compliers: $\{A_{3\mathfrak{a}_1(1)} = 1 \text{ and } A_{3\mathfrak{a}_1(0)} = 0\} = C$

Representing groups of women

- not directly observed
- most directly impacted by the intervention

'Feasible'? estimands for different purposes

The potential mean weight at three months in the study population under different possible conditions

outcome	interventions	pop	Education low	int	high
$Y_{\mathfrak{a}_1(0)}$	BEP not offered	6017	5914	6057	6141
$Y_{\mathfrak{a}_1(1)}$	BEP offered	6115	6024	6155	6207
$Y_{\mathfrak{a}_2(1)}$	BEP followed	6182	6128	6208	6226
$Y_{\mathfrak{a}_3(0)}$	no BF	5827	5730	5854	5981
$Y_{\mathfrak{a}_1(0),\mathfrak{a}_3(1)}$	BEP not offered, BF started	6214	6154	6248	6246
$Y_{\mathfrak{a}_1(1),\mathfrak{a}_3(1)}$	BEP offered, BF started	6249	6207	6276	6262
$Y_{\mathfrak{a}_2(1),\mathfrak{a}_3(1)}$	BEP followed, BF started	6277	6261	6292	6266
$Y_{\mathfrak{a}_4(1)}$	duration BF $=$ 3 months	6351	6393	6339	6286

- L-specific ITT: $E(Y_{\mathfrak{a}_1(1)} - Y_{\mathfrak{a}_1(0)}|L = \text{low/int/high}) = 110/98/66gr.$
- ATE of attending BEP $E(Y_{\mathfrak{a}_2(1)} - Y_{\mathfrak{a}_2(0)}) = 164gr.$
- ATE of starting BF with or without prior program attendance $E(Y_{\mathfrak{a}_2(1),\mathfrak{a}_3(1)} - Y_{\mathfrak{a}_2(0),\mathfrak{a}_3(1)}) = 63gr.$

The potential mean weight at three months in the study population under different possible conditions

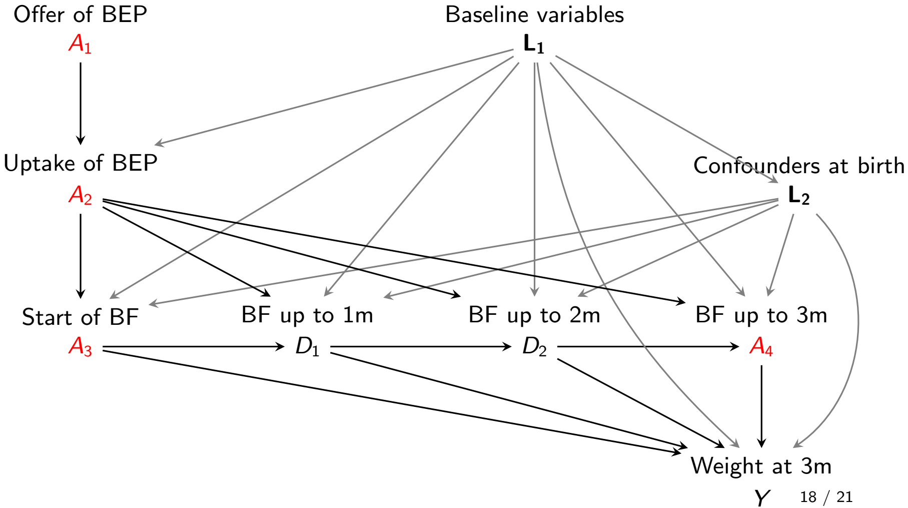
Potential outcome	$A_2 = 1$	$A_1 = 1$ $A_2 = 0$	$A_1 = 1$ $A_3 = 1$	$A_1 = 1$ $A_3 = 0$	$A_1 = 0$ $A_3 = 1$	$A_1 = 0$ $A_3 = 0$
$Y_{\mathfrak{a}_1(0)}$	6047	5964	6149	5733	6274	5761
$Y_{\mathfrak{a}_1(1)}$	6200	5964	6292	5733	6308	5923
$Y_{\mathfrak{a}_2(1)}$	6200	6149	6308	5911	6329	6035
$Y_{\mathfrak{a}_3(0)}$	5849	5788	5871	5733	5893	5761
$Y_{\mathfrak{a}_1(0),\mathfrak{a}_3(1)}$	6226	6193	6251	6133	6274	6153
$Y_{\mathfrak{a}_1(1),\mathfrak{a}_3(1)}$	6282	6193	6292	6157	6308	6191
$Y_{\mathfrak{a}_2(1),\mathfrak{a}_3(1)}$	6282	6270	6308	6212	6329	6225
$Y_{\mathfrak{a}_4(1)}$	6345	6362	6372	6307	6392	6311

Estimated ATE and ATT of A_2 on weight at 3 months (in grams)

Estimand	Estimation method	Estimate	(SE)
ATE			
	True value	165.1	
	Crude regression	196.0	(9.6)
	Regression adjustment (without interactions)	155.4	(9.5)
	Regression adjustment (with interactions)	165.0	(9.7)
	PS stratification† (6 strata)	165.0	(9.4)
	Regression with PS †	156.2	(9.0)
	PS matching (1 match)‡	155.7	(10.1)
	PS matching (3 matches)‡	154.9	(10.1)
	PS IPW†	164.7	(9.3)
	PS DR IPW†	164.7	(9.7)
	IV	146.2	(14.0)
ATT			
	True value	152.8	
	Regression adjustment (with interactions)	148.7	(9.4)
	PS stratification† (6 strata)	148.7	(9.6)
	PS matching (1 match)‡	145.8	(9.8)
	PS matching (3 matches)‡	145.4	(9.7)
	PS IPW†	148.0	(9.6)

* controlled for: maternal age, maternal education, maternal allergy status, smoking status in the first trimester, and area of residence.

Sketch of data generating model



Results for A_3 (Starting breastfeeding)

	$\mathbf{A_1 = 0}$		$\mathbf{A_1 = 1}$	
Estimation method	Estimate	(SE)	Estimate	(SE)
ATE				
True value	386.8		422.3	
Crude regression	503.2	(11.6)	582.0	(12.2)
Regression (simple)	384.3	(2.8)	428.0	(3.3)
Regression (with interactions)	384.7	(3.2)	425.3	(2.7)
Regression with PS *	384.4	(3.2)	425.9	(3.3)
PS stratification* (6 strata)	392.2	(4.1)	442.0	(6.5)
PS matching (1 match)	386.5	(8.1)	429.0	(10.6)
PS matching (3 matches)	380.7	(5.5)	437.2	(7.8)
PS IPW	384.7	(3.8)	426.6	(6.7)
PS DR IPW	384.8	(3.9)	426.7	(7.0)
NO IV	*513.3*	*(44.4)*	–	–

Once an instrument always an instrument?

- $A1$ an instrument for the effect of following BEP, A_2
- $A1$ NO instrument for the effect of starting BF, A_3
- It makes no sense to average causal effect of A_3 over the A_1 distribution in the dataset

Simulation learner is useful because:

- Generates observed data, augmented with potential outcomes under specific assumptions (+/-)
- Gives more insight in data generation process - *and* assumptions
- Actual causal effects are 'known' $->$ examine estimands (gold standard)
- Great help in finding correct ways of analysis (which turned out to be different for A_2 and A_3)
- Enables to compare different analytic methods .
- It is helpful in teaching causal methods, *e.g. on competing risks for in-hospital death, with hydroxychloroquine effect*
- Code of generation and analysis of data is available on www.ofcaus.org