Author/s:
Moreno-Betancur, M;Lynch, JW;Pilkington, RM;Schuch, HS;Gialamas, A;Sawyer, MG;Chittleborough, CR;Schurer, S;Gurrin, LC

Title:
Emulating a target trial of intensive nurse home visiting in the policy-relevant population using linked administrative data

Date:
2023-02-01

Citation:
Moreno-Betancur, M., Lynch, J. W., Pilkington, R. M., Schuch, H. S., Gialamas, A., Sawyer, M. G., Chittleborough, C. R., Schurer, S. & Gurrin, L. C. (2023). Emulating a target trial of intensive nurse home visiting in the policy-relevant population using linked administrative data. International Journal of Epidemiology, 52 (1), pp.119-131. https://doi.org/10.1093/ije/dyac092.

Persistent Link:
https://hdl.handle.net/11343/327313

International Epidemiological Association

International Journal of Epidemiology, 2023, 119–131
https://doi.org/10.1093/ije/dyac092
Advance Access Publication Date: 18 May 2022
Original article

OXFORD

More on RCTs

Emulating a target trial of intensive nurse home visiting in the policy-relevant population using linked administrative data

Margarita Moreno-Betancur [1,2,3]* John W Lynch,[4,5,6] Rhiannon M Pilkington,[4,5] Helena S Schuch [4,5,7] Angela Gialamas,[4,5] Michael G Sawyer,[5,8] Catherine R Chittleborough,[4,5] Stefanie Schurer[9] and Lyle C Gurrin[1]

[1]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, VIC, Australia, [2]Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Parkville, VIC, Australia, [3]Clinical Epidemiology and Biostatistics Unit, Department of Paediatrics, University of Melbourne, Parkville, VIC, Australia, [4]School of Public Health, University of Adelaide, Adelaide, SA, Australia, [5]Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia, [6]Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, UK, [7]Postgraduate programme in Dentistry, Federal University of Pelotas, Pelotas, Brazil, [8]School of Medicine, University of Adelaide, Adelaide, SA, Australia and [9]School of Economics, University of Sydney, Sydney, NSW, Australia

*Corresponding author. Clinical Epidemiology and Biostatistics Unit, Department of Paediatrics, University of Melbourne, 50 Flemington Road, Parkville, Victoria 3052, Australia. E-mail: margarita.moreno@unimelb.edu.au

Received 6 July 2021; Editorial decision 1 March 2022; Accepted 21 April 2022

Abstract

Background: Populations willing to participate in randomized trials may not correspond well to policy-relevant target populations. Evidence of effectiveness that is complementary to randomized trials may be obtained by combining the 'target trial' causal inference framework with whole-of-population linked administrative data.

Methods: We demonstrate this approach in an evaluation of the South Australian Family Home Visiting Program, a nurse home visiting programme targeting socially disadvantaged families. Using de-identified data from 2004–10 in the ethics-approved Better Evidence Better Outcomes Linked Data (BEBOLD) platform, we characterized the policy-relevant population and emulated a trial evaluating effects on child developmental vulnerability at 5 years ($n = 4160$) and academic achievement at 9 years ($n = 6370$). Linkage to seven health, welfare and education data sources allowed adjustment for 29 confounders using Targeted Maximum Likelihood Estimation (TMLE) with SuperLearner. Sensitivity analyses assessed robustness to analytical choices.

Results: We demonstrated how the target trial framework may be used with linked administrative data to generate evidence for an intervention as it is delivered in practice in the community in the policy-relevant target population, and considering effects on

outcomes years down the track. The target trial lens also aided in understanding and limiting the increased measurement, confounding and selection bias risks arising with such data. Substantively, we did not find robust evidence of a meaningful beneficial intervention effect.

Conclusions: This approach could be a valuable avenue for generating high-quality, policy-relevant evidence that is complementary to trials, particularly when the target populations are multiply disadvantaged and less likely to participate in trials.

Key words: Causal inference, generalizability, linked data, nurse visiting programme, social disadvantage, target trial, targeted maximum likelihood estimation, transportability

Key Messages

- For complex interventions targeted at multiply disadvantaged populations, there may be uncertainty about the applicability of results from randomized trials to the policy-relevant population in a given setting.
- With interventions that have been implemented in a community, combining the 'target trial' framework with whole-of-population administrative linked data is a possible approach to generate high-quality, policy-relevant evidence that complements trial evidence.
- The target trial framework is a powerful tool for planning analyses of intervention effects, especially with complex data sources such as administrative linked data, aiding development of clear estimand definitions and understanding and limiting bias risks.
- A key strength of whole-of-population administrative linked data is that they enable emulation of the policy-relevant target population that may be missed in trials.
- Administrative linked data may present complexities that increase the risk of measurement, confounding and selection bias relative to data collected for research purposes.

Introduction

Recent methodological work[1–3] describes a framework for evaluating intervention effects with 'real-world' data using the concept of a 'target trial'. This is defined as the hypothetical randomized controlled trial (RCT) that cannot be conducted due to resourcing, feasibility, timeliness or ethical reasons, but may potentially be emulated with relevant observational data. A compelling example of this approach examining effects of COVID vaccination has been recently reported.[4] By requiring specification of each component of the target trial's protocol, this framework aids to develop a refined definition of the target estimand and an appropriate analysis strategy that reduces the risk of bias.[5]

We consider this framework in the context of the South Australian Family Home Visiting programme (FHVP), an intensive postnatal nurse home visiting programme for socially disadvantaged mothers, started in 2004–05.[6] Similar to the USA's Nurse-Family Partnership[7] and the UK's Family-Nurse Partnership,[8] FHVP is based on the 'Family Partnership Model'[9] and aims to enhance maternal and child outcomes through parental support up to 2 years postnatally. There is mixed evidence regarding the effectiveness of such programmes from RCTs. Three US RCTs[10–12] beginning in the 1980s and 90 s found the intervention to be effective on some maternal and child outcomes in the short (2–6 years) and long (15 years) term, whereas a pragmatic RCT testing a similar intervention in the UK, 'Building Blocks', reported in 2016 little evidence of meaningful effects on outcomes at 2 years.[13] In Australia, the Miller Early Childhood Sustained Home visiting (MECSH) RCT in 2011 reported mixed evidence of programme benefits,[14] and the 'right@home' RCT in 2019 reported improvement at 2 years of about 0.2 SD in six of the 10 continuous primary outcomes and odds ratios of about 1.4–1.8 in two of the three binary primary outcomes.[15]

As well as the mixed evidence provided by these RCTs, a further consideration in assessing potential programme implementation in a given setting is whether the results of these studies are applicable to the particular policy-relevant target population. This may be defined as the population that: (Criterion a) needs and can benefit from the programme; and (Criterion b) would actually take up an offer of participating in it.[16] Critically, Ccriterion (a) is a social construct that will change with time, place, politics and social values. US trials of early nurse visiting programmes in the 1980s had young maternal age as an eligibility criterion,

presumably because this was perceived as a 'problem'.[17] Criterion (b) is important when evaluating interventions that rely upon voluntary receipt of service, with individuals who would take up the programme comprising the bulk of the intended social investment.[16] In considering who should be offered supportive programmes, providers routinely consider 'readiness to engage' as a criterion for making an offer of service. Even where the RCT's target population is relevant to a new setting, the actual trial sample is the result of complex selection processes and may end up not being representative of the original target.[18]

So as complementary evidence to the RCTs, it is desirable to evaluate intervention effects among those who actually did take up similar programmes in real service settings. Such evidence may help policy makers better understand the potential yields of their programmatic investments. We use the FHVP as case study to demonstrate how the 'target trial' framework might be applied to emulate a trial using population-wide linked administrative data, showing that a key benefit is the possibility of emulating the results of an RCT in the policy-relevant population.

Methods

Target trial specification

The protocol for the target trial is outlined in Table 1 (first column).

The eligibility criteria of the target trial are a means for defining the policy-relevant target population in the South Australian setting. To define the population who needs and can benefit from the programme according to that policy context [Criterion (a) mentioned in the Introduction], a first criterion mirrors how eligibility for the FHVP is determined in South Australia, which is through a case review by a multidisciplinary team (see Supplementary Material, available as Supplementary data at *IJE* online for further details). A second criterion is that individuals would accept the offer if they received it, i.e. they are 'ready to engage' [Criterion (b)].

Participants are randomized to two parallel arms: usual care in South Australia plus at least one FHVP nurse visit, versus usual care only. The maximum FHVP consists of 34 nurse visits to the family home over the first 2 years of the child's life. The suggested visiting frequency is weekly until the child is aged 8 weeks, fortnightly until they are 9 months old, and then monthly. The programme is divided into six modules mapping to key developmental stages, and including age-appropriate material and core activities to be covered during the visits. Full details have been described elsewhere[6] (see also Supplementary Material, available as Supplementary data at *IJE* online).

Assessment of two sets of outcomes is blinded and independent of intervention administration, as objective measures are obtained via data linkage. Child development outcome measures are derived from the Australian Early Development Census (AEDC),[19] a teacher-completed instrument administered triennially since 2009 for Australian children in their first year of full-time school (age around 5 years). It assesses five domains: physical health and wellbeing; social competence; emotional maturity; language and cognitive skills (school-based); communication skills and general knowledge. For each domain, children receive a score that is characterized using cut-offs established using the 2009 AEDC national results. In this study, scores were dichotomized (developmentally vulnerable: yes/no) using cut-offs based on the lowest 10th percentile of the 2009 AEDC national results.[19] Consistent with national AEDC reporting, outcome measures are developmental vulnerability on one or more domains (usually termed 'DV1') and on each domain. Academic achievement outcomes are derived from the National Assessment Program—Literacy And Numeracy (NAPLAN) in school year 3 (age around 8–9 years).[20] This test is administered annually to all Australian students and covers four domains: reading, writing, language conventions (spelling, grammar and punctuation) and numeracy. Outcome measures used are domain-specific binary indicators of whether or not the child is at or below the national minimum standard for that domain.[21]

The causal effect measure is the difference in expected outcomes in the intervention arm versus the control arm in the target population.

Target trial emulation

We next describe the emulation of the target trial using the Better Evidence Better Outcomes Linked Data (BEBOLD) platform, which includes state-wide, de-identified linked administrative data on all birth cohorts from 1991 onwards.[22] This study used data within BEBOLD on the FHVP, which was initiated by the South Australian Child and Family Health Service ('CaFHS') in 2004,[6] as well as birth registration, perinatal (including maternal data), public housing, child protection, education, school enrolment and child development data. The probabilistic matching routines used by Australian data linkage systems typically estimate a 0.1–0.5% false linkage rate.[23,24] The reporting of this study has been in accordance with the checklist in the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement. See Supplementary Material for details on investigator access to the database population and data cleaning methods.

Table 1 Protocol of a target trial to estimate the effect of the FHVP on developmental and learning outcomes, and how each aspect was emulated in BEBOLD

	Target trial	Emulation with BEBOLD
Eligibility criteria	Individuals who are (a) deemed eligible for the FHVP in a case review by a multidisciplinary team and (b) 'ready to engage', i.e. would accept the FHVP if they received an offer	Individuals in corresponding analytical sample (AEDC or NAPLAN) who were deemed eligible for the FHVP in the case review process conducted in South Australia, were offered the FHVP and received at least 1 FHVP nurse visit, the latter being a proxy for 'readiness to engage' (this is the 'target sample')
Trial arms	Intervention arm: Individuals are provided usual care plus at least one FHVP nurse visit	Intervention arm: Individuals who were deemed FHVP-eligible by case review, were offered the FHVP and received at least one FHVP nurse visit (this is equal to the target sample)
	Control arm: Individuals are provided usual care	Control arm: Individuals who were deemed FHVP-eligible by case review but did not receive an offer
Assignment procedures	Random assignment to either arm at recruitment, unblinded to participants and program-administering personnel (e.g. nurses)	Adjustment for 29 baseline confounders listed in Table 2 using a targeted maximum likelihood estimation (TMLE) method that adjusts for both confounding and missing outcome data, with incomplete confounders dealt with via the missing covariate indicator method (MCIM)
Follow-up period	From birth to ages 5 and 9 years	From birth to ages 5 and 9 years
Outcomes and outcome measures	Child development at 5 years as measured by AEDC • Indicator of developmental vulnerability in any of the five AEDC domains • Indicators of developmental vulnerability in each of the five AEDC domains	The same outcome measures as in the target trial, with the difference being that linkage consent is not sought individually. In particular, NAPLAN linkage was not available for private schools in BEBOLD, which is an important reason for missing outcome data
	Academic achievement at 9 years as measured by the Year 3 NAPLAN • Indicators of being at or below the national minimum standard for each of the four NAPLAN domains	
	Outcome measures obtained through data linkage, for which participant consent would be sought at recruitment. Outcome assessment is blinded and systematic	
Causal effect measure	Difference in expected outcomes in intervention arm versus control arm in the target population	Under a set of assumptions, the causal effect measure is equal to the 'average treatment effect in the treated' in the difference scale (see main text) and is identifiable from the data (see Supplementary Material, available as Supplementary data at *IJE* online)

FHVP, Family Home Visiting Program; BEBOLD, Better Evidence Better Outcomes Linked Data platform; AEDC, Australian Early Development Census; NAPLAN, National Assessment Program—Literacy And Numeracy.

Analytical samples

We created two analytical samples from BEBOLD by restricting records to children born within specific calendar periods occurring after the FHVP had been rolled out, and which ensured sufficient follow-up to capture either AEDC or NAPLAN. For AEDC, we restricted the sample to children born: 1 January 2004 to 30 April 2004; 1 May 2006 to 30 April 2007; and 1 May 2009 to 30 April 2010, capturing AEDC assessments in 2009, 2012 and 2015, respectively (see Supplementary Material for details). For NAPLAN outcomes, we restricted the sample to children born between 1 January 2004 and 30 April 2007.

Table 1 (second column) summarizes how each aspect of the target trial was emulated for each of the outcomes using the corresponding analytical sample.

Emulation of the target population

The challenge in this step is to identify a sample that is representative of the target population as defined by the eligibility criteria, in the sense that it shares the same distribution for a rich vector of measured baseline covariates C_0. The covariates we considered under C_0 are defined in the section 'Confounders' below. To identify such a sample, we made use as explained below of the following binary indicators available for each record: FHVP eligibility as determined by the multidisciplinary case review (denoted E, with $E = 1$ if eligible by case review, $E = 0$ if not); FHVP offer (O, with $O = 1$ if offered FHVP, $O = 0$ if not); and receipt of one or more nurse visits at an age in agreement with the programme design and after the initial FHVP roll-out phase (V, with $V = 1$ if received one or more FHVP visit after 1 April 2004 and within the first 25 weeks of life; $V = 0$ if not).

We used the first indicator, E, to identify all individuals in the sample deemed eligible by case review, enabling us to emulate eligibility Criterion (a). However, we did not have a binary indicator D identifying those who would in addition accept an offer, which would be needed to emulate eligibility Criterion (b). This lack was, first, because not all individuals deemed eligible by case review received an offer (i.e. for some records, $E = 1$ but $O = 0$) which was due to system-related reasons: lack of CaFHS service resources or geographical area, for which we have measured proxies (CaFHS service area and remoteness, denoted P). For individuals who did receive an offer ($O = 1$), their acceptance D was not captured in the data. To circumvent this we made some assumptions, depicted in the causal diagram in Figure 1, in particular that within the eligible population receipt of an offer is independent of the individual's characteristics C_0 given P, $(O \perp C_0 \mid P, E = 1)$. We further assume that V is a reasonable proxy of readiness to engage, considering that $V = D$ when $O = 1$. Under these assumptions it follows that, given P, the sample of individuals who were eligible by case review, received an offer and at least one visit, is representative of the target population in terms of the joint distribution f of C_0. Indeed, $f(c_0|P, E = 1, D = 1) = f(c_0|P, E = 1, O = 1, D = 1) = f(c_0|P, E = 1, O = 1, V = 1)$. We therefore used this sample to emulate the target population and refer to it as the 'target sample', and further included P in the confounder set.

Emulation of trial arms (intervention and control)

We emulated the intervention arm by the sample of eligible individuals who were offered FHVP and received at least one nurse visit, that is with $E = 1$, $O = 1$, and $V = 1$. Therefore, the intervention arm was equal to the target sample. We emulated the control arm using the sample of eligible individuals who were not offered the programme ($E = 1, O = 0$) for what we know were system-related reasons P. It is not known whether these individuals are representative of the target population with respect to C_0, even given P. Therefore, below we use the distribution of C_0 in the target sample/intervention arm to emulate the causal effect measure in the target population.

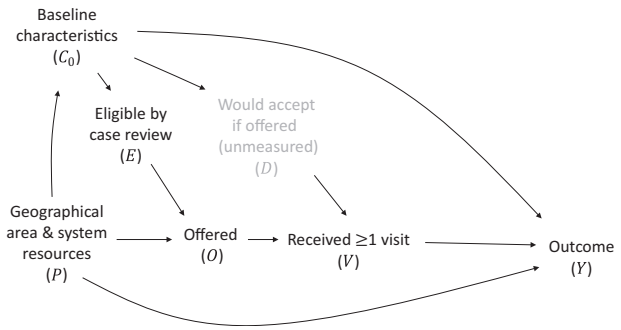

Figure 1 Causal diagram depicting assumptions regarding the observed data on the Family Home Visiting programme (FHVP)

Confounders

Following the causal diagram in Figure 1, we adjusted all analyses for P as well as 27 additional baseline covariates (C_0) identified as potential confounders a priori, based on substantive knowledge. The overall set, denoted by C, is listed in Table 2 and consists of sociodemographic, geographical, maternal health and pregnancy characteristics, including year of birth, birth outcomes, domestic violence indicators, housing type and other data captured specifically as part of the FHVP process.

Emulation of follow-up period and outcomes

Data linkage enables emulation of the follow-up from birth (time zero/baseline) to when AEDC and NAPLAN take place.

Emulation of causal effect measure

Denote by A the arm indicator ($A = 1$ for intervention, i.e. if $E = 1, O = 1, V = 1$, and $A = 0$ for control, i.e. $E = 1, O = 0$); Y the observed outcome; and Y_a the potential outcome of an individual had they been assigned to trial arm $A = a$. It can be deduced from the definitions of A and C and Figure 1 that $Y_a \perp A \mid C$. Given this, and that the intervention arm is representative of the target population (proof above), the causal effect measure of interest is mathematically equivalent to the 'average treatment effect in the treated' (ATT)[25] in the difference scale:

$$\begin{aligned}&\sum\nolimits_c \{P(Y_1 = 1|C = c) - P(Y_0 = 1|C = c)\} f(c|A = 1)\\ &= \sum\nolimits_c \{P(Y_1 = 1|A = 1, C = c)\\ &- P(Y_0 = 1|A = 1, C = c)\} f(c|A = 1)\\ &= P(Y_1|A = 1) - P(Y_0|A = 1) = \text{ ATT}.\end{aligned}$$

In the context of incomplete outcome and confounder data, where the latter are handled via the missing covariate

indicator method (MCIM—see below), the ATT is identifiable under a set of assumptions provided and discussed in the Supplementary Material.[25–28] In our study there were missing data in some confounders and outcomes, with an important source being NAPLAN scores from private schools who do not make their data available for linkage (see Supplementary Material).

Estimation and handling of missing data

We considered confounders with less than 0.1% values missing as complete, excluding the corresponding incomplete records (three and seven records for AEDC and NAPLAN samples, respectively). Missingness in remaining covariates was handled using MCIM, which includes the covariate, its missingness indicator and their interaction in the models in targeted maximum likelihood estimation (TMLE) as described below. MCIM has been commonly used with TMLE[26,29] because it enables incorporation of all available data while avoiding the parametric assumptions of default multiple imputation (MI) strategies using main-effects regression models that could be incompatible with machine learning algorithms and thus lead to bias.[30]

Primary estimation used a version of TMLE for the ATT with the SuperLearner[31,32] that adjusts for both confounding and missing outcome data by relying on three models: one for the outcome risk, given (A, C, M_C), $P(Y = 1|A, C, M_C)$; one for the probability of being in the intervention arm, given (C, M_C), $P(A = 1|C, M_C)$; and one for the probability of having missing outcome, given (A, C, M_C), $P(M_Y = 1|A,\ C, M_C)$.[27,28,32] Here M_Y is a binary indicator of missing outcome and M_C denotes a vector of binary missingness indicators for each incomplete covariate.

Robustness checks

We conducted additional analyses to check robustness of results to the selection of the analytical sample, including an examination of effect modification by period, as well as to the chosen estimation and missing data handling methods (see Supplementary Material).

Results

Supplementary Figures S2 and S3 (available as Supplementary data at *IJE* online) show the flow of participants through the FHVP process, for the AEDC ($n = 45\ 150$) and NAPLAN ($n = 61\ 185$) analytical samples. In each case, 12% of newborns were eligible for FHVP. In the AEDC sample, 38% ($n = 2025$) of eligible newborns ($n = 5300$) went on to receive both an offer and at least one FHVP nurse visit, which constituted the target sample and the intervention arm. The control arm consisted of the 40% ($n = 2135$) who were eligible but did not receive either an offer or a visit. In the NAPLAN sample, 33% ($n = 2381$) of eligible newborns ($n = 7302$) received both an offer and at least one visit, with the control arm consisting of the 55% ($n = 3989$) who did not receive either an offer or a visit.

Descriptive statistics by arm are shown in Table 2. Those for the intervention arm, which is also the target sample (columns 2 and 5, for AEDC and NAPLAN, respectively) provide a characterization of the target population (see Supplementary Material for comparison with general population). Compared with the intervention arm, in the control arm (columns 1 and 4, for AEDC and NAPLAN, respectively) mothers were older at birth (median 28–29 years versus 22–23 years in intervention arm), there were lower proportions of Indigenous children (10.9–14.8% versus 23.1%) and there were higher proportions of children from remote areas (5.5–6.2% versus 0.1–2.2% in intervention arm). Over 40% of records in each arm in each sample had at least one incomplete covariate.

Proportions of missing outcome data were high and unadjusted risk estimates based on available data were higher in the intervention arm than control arm (Table 3). In both arms, outcome risks were higher than estimates in the general population: risk of scoring at or below national minimum standard across NAPLAN domains was 21–38% in our sample versus 10–20%,[33] and risk of developmental vulnerability on at least one AEDC domain was 32–37% in our sample versus $\leq 25\%$.[34]

Figure 2 shows estimates of the causal effect of FHVP for AEDC outcomes in the main analytical sample, as well as the common sample (defined as the overlap of the AEDC and NAPLAN samples), for the eight combinations of estimation [TMLE, g-computation ('gcomp'), inverse probability weighting (IPW)] and missing data [MCIM, complete covariates ('CCov'), complete cases (CC), MI] methods.

We would expect that the intervention, if it has any effect, would be beneficial, resulting in a reduced outcome risk (negative risk difference). Contrary to expectations, with the primary approach ('TMLE-MCIM', in bold) in the main sample, we estimated a small adverse effect of FHVP on risk of development vulnerability on one or more domains, with the intervention leading to an absolute increase in risk of 7.4% [95% confidence interval (CI): 4.8 to 10.0%] compared with receiving no intervention. Adverse but smaller effects were also estimated for some domain-specific scores (social competence, emotional maturity, and communication skills and general knowledge), but not for the other two domains (physical health and wellbeing, and school-based language and cognitive skills). In the common sample or with

Table 2 Characteristics of the intervention arm (equal to the target sample) and the control arm in each analytical sample within BEBOLD, South Australia, 2004–10

	AEDC analytical sample			NAPLAN analytical sample		
	Control	Intervention	Missing (%)[c]	Control	Intervention	Missing (%)[c]
n[a]	2135	2025		3989	2381	
Year of birth (%)			0.0			0.0
2004	46 (2.2)	45 (2.2)		873 (21.9)	394 (16.5)	
2005	–	–		1024 (25.7)	733 (30.8)	
2006	1152 (54.0)	621 (30.7)		1492 (37.4)	959 (40.3)	
2007	600 (28.1)	295 (14.6)		600 (15.0)	295 (12.4)	
2008 or later	337 (15.8)	1064 (52.5)		–	–	
Sex of baby: Male (%)	1082 (50.7)	1028 (50.8)	0.0	1989 (49.9)	1200 (50.4)	0.0
Aboriginal and/or Torres Strait Islander: Yes (%)	233 (10.9)	467 (23.1)	0.0	592 (14.8)	549 (23.1)	0.0
Maternal country of birth: Not Australia (%)	241 (11.3)	352 (17.4)	0.0	487 (12.2)	334 (14.0)	0.0
Maternal age at birth: median [IQR]	29 [23, 33]	23 [19, 30]	0.0	28 [22, 33]	22 [19, 29]	0.0
Paternal age at birth: median [IQR]	31 [27, 36]	27 [22, 34]	10.0	31 [25, 36]	27 [22, 34]	10.7
Maternal marital status at birth: Not partnered (%)	489 (22.9)	693 (34.2)	0.0	1109 (27.8)	873 (36.7)	0.0
Previous births (%)			0.0			0.0
First child	832 (39.0)	1359 (67.1)		1726 (43.3)	1416 (59.5)	
Second child	677 (31.7)	440 (21.7)		1187 (29.8)	547 (23.0)	
Third or higher child	625 (29.3)	226 (11.2)		1074 (26.9)	417 (17.5)	
Remoteness (%)			0.0			0.0
Major Cities of Australia	1299 (60.8)	1582 (78.1)		2271 (56.9)	2023 (85.0)	
Inner Regional Australia	271 (12.7)	159 (7.9)		509 (12.8)	80 (3.4)	
Outer Regional Australia	448 (21.0)	239 (11.8)		960 (24.1)	275 (11.5)	
Remote Australia	117 (5.5)	45 (2.2)		249 (6.2)	3 (0.1)	
CaFHS service area (%)			0.0			0.0
Country North and Other	313 (14.7)	167 (8.2)		705 (17.7)	137 (5.8)	
Country South	461 (21.6)	249 (12.3)		906 (22.7)	211 (8.9)	
Metro Central	323 (15.1)	535 (26.4)		928 (23.3)	296 (12.4)	
Metro North	358 (16.8)	615 (30.4)		508 (12.7)	1110 (46.6)	
Metro South	680 (31.9)	459 (22.7)		942 (23.6)	627 (26.3)	
IRSAD decile: median [IQR]	4 [2, 6]	3 [1, 6]	0.1	4 [2, 6]	4 [1, 5]	0.1
Maximum education of parent 1 (%)			32.6			29.1
Year 9 or equivalent or below	84 (5.8)	69 (5.1)		142 (5.1)	93 (5.4)	
Year 10 or equivalent	191 (13.2)	201 (14.8)		370 (13.3)	270 (15.5)	
Year 11 or equivalent	175 (12.1)	185 (13.6)		370 (13.3)	270 (15.5)	
Year 12 or equivalent	164 (11.4)	107 (7.9)		318 (11.5)	180 (10.4)	
Certificate I to IV (including trade certificate)	529 (36.7)	491 (36.1)		978 (35.2)	605 (34.8)	
Advanced diploma/Diploma	118 (8.2)	134 (9.9)		258 (9.3)	160 (9.2)	
Bachelor degree or above	182 (12.6)	173 (12.7)		341 (12.3)	160 (9.2)	
Family employment at birth (%)			0.9			1.1
Mother and father not in labour force	304 (14.4)	387 (19.3)		629 (16.0)	532 (22.5)	
No partner, mother not in labour force	223 (10.5)	312 (15.6)		531 (13.5)	414 (17.5)	
Other	1590 (75.1)	1307 (65.2)		2781 (70.6)	1416 (59.9)	
Housing SA contact (%)			0.0			0.0
No contact	1755 (82.2)	1432 (70.7)		3102 (77.8)	1613 (67.7)	
In public housing	149 (7.0)	194 (9.6)		338 (8.5)	260 (10.9)	
Waiting list, not in public housing	99 (4.6)	178 (8.8)		281 (7.0)	270 (11.3)	
Rental assistance, not in public housing/waiting list	132 (6.2)	221 (10.9)		268 (6.7)	238 (10.0)	
Maternal smoking 2nd half of pregnancy: Yes (%)	617 (29.8)	570 (28.8)	2.8	1215 (31.6)	834 (36.2)	3.4
Pregnancy hypertension—all types: Yes (%)	182 (8.5)	189 (9.3)	0.0	345 (8.7)	227 (9.5)	0.0
Gestational diabetes: Yes (%)	109 (5.1)	82 (4.0)	0.0	187 (4.7)	95 (4.0)	0.0
Antenatal care: Less than 7 antenatal visits (%)	255 (12.7)	260 (13.8)	6.5	541 (14.7)	352 (15.6)	6.8

(Continued)

Table 2 Continued

	AEDC analytical sample			NAPLAN analytical sample		
	Control	Intervention	Missing (%)[c]	Control	Intervention	Missing (%)[c]
Preterm baby (up to 36 weeks): Yes (%)	215 (10.1)	221 (10.9)	0.0	403 (10.1)	268 (11.3)	0.0
Birthweight for GA z-score: mean (SD)	0 (1)	0 (1)	0.0	0 (1)	0 (1)	0.1
Baby in hospital at 28 days: Yes (%)	77 (3.6)	67 (3.3)	0.0	139 (3.5)	80 (3.4)	0.0
Congenital anomalies: Yes (%)	184 (8.6)	160 (7.9)	0.0	309 (7.8)	174 (7.3)	0.0
Congenital anomalies (FHVP[b]): Yes (%)	151 (7.1)	77 (3.8)	0.0	224 (5.6)	94 (3.9)	0.0
Drug- and alcohol-related issues (FHVP[b]): Yes (%)	95 (4.4)	64 (3.2)	0.0	238 (6.0)	131 (5.5)	0.0
Maternal mental health issues (FHVP[b]): Yes (%)	1153 (54.0)	684 (33.8)	0.0	1969 (49.4)	903 (37.9)	0.0
Poor maternal attribution (FHVP[b]): Yes (%)	99 (4.6)	242 (12.0)	0.0	266 (6.7)	270 (11.3)	0.0
Domestic violence: Yes (%)	196 (9.2)	195 (9.6)	0.0	481 (12.1)	415 (17.4)	0.0
Domestic violence (FHVP[b]): Yes (%)	106 (5.0)	69 (3.4)	0.0	245 (6.1)	132 (5.5)	0.0
Child protection services contact in first 30 days: Yes (%)	79 (3.7)	95 (4.7)	0.0	103 (2.6)	59 (2.5)	0.0
Any covariate missing: Yes (%)	927 (43.4)	935 (46.2)	0.0	1744 (43.7)	984 (41.3)	0.0

BEBOLD, Better Evidence Better Outcomes Linked Data platform; AEDC, Australian Early Development Census; NAPLAN, National Assessment Program—Literacy And Numeracy; IQR, interquartile range; CaFHS, South Australian Child and Family Health Service; SA, South Australia; GA, gestational age; FHVP, Family Home Visiting Program; IRSAD, Index of Relative Socio-economic Advantage and Disadvantage.

[a]This gives the total number in each analytical sample, but descriptive statistics for each characteristic are based on the records with available data for that variable in the given sample.

[b]This measure was collected as part of the FHVP eligibility assessment process.

[c]This is the proportion of missing data across both treatment groups.

other non-TMLE methods, such estimates were in general closer to zero and confidence intervals wider. Period-specific analyses (Supplementary Figures S4 and S5, available as Supplementary data at *IJE* online) revealed that the main-sample results were driven by the 2009–10 sample, in which adverse effects were estimated by most approaches for overall risk of developmental vulnerability and three domains. However, effect estimates in the earlier cohort (2006–07) were for the most part small.

For NAPLAN (Figure 3), in the main sample and with TMLE-MCIM, we estimated a large, beneficial effect of FHVP on the risk of reading scores being at or below the national minimum standard, with the intervention leading to an absolute reduction in risk of -20.6% (95% CI -23.5 to -17.7%). There was also a substantial reduced risk when considering writing scores (-15.5%, 95% CI -18.6 to -12.4%), and smaller estimates when considering grammar (-8.2%, 95% CI -11.6 to -4.8%) and spelling (-5.0%, 95% CI -8.0 to -2.0%) scores, while the estimate for numeracy was small (1.4%, 95% CI -1.6 to 4.5%). However, in the common sample or considering other estimation methods, these beneficial effect estimates disappeared. Period-specific analyses (Supplementary Figures S6–S8, available as Supplementary data at *IJE* online) revealed that those beneficial effect estimates, particularly on reading and writing, were driven by the 2004–05 sample, in which TMLE approaches also estimated a large beneficial effect on numeracy. There is a pattern of transition across the three 12-month period-specific cohorts between 1 May 2004 and 30 April 2007, from large beneficial effect estimates through negligible differences to small adverse effect estimates.

Discussion

Our goal was to demonstrate the use of causal inference concepts and methods in whole-of-population linked administrative data to evaluate effects of an intervention targeted to populations experiencing social disadvantage. The methodology followed the steps of the 'causal roadmap',[31] with clear definition of a policy-relevant estimand using the target trial concept, emulation under transparent assumptions that elucidate and help to limit bias risks, estimation using robust statistical methods and sensitivity analyses. A key motivation for this work was to demonstrate the possibility of estimating effects in policy-relevant target populations, which can in principle be characterized in linked databases. It can be difficult for RCTs to enrol the populations that are of policy relevance,[18] particularly for long-term follow-up. RCTs of early interventions rarely are funded for follow-up beyond 2–5 years[35,36] and yet policy-relevant outcomes, such as welfare receipt, may occur decades later. The proposed approach is therefore a potential route to generate high-quality, policy-relevant evidence that is complementary to that produced in RCTs. This approach provides an alternative evidence lens that is particularly significant in the context of target populations experiencing multiple disadvantages.

Table 3 Crude risk of developmental vulnerability on AEDC and scoring at or below national minimum standard on Year 3 NAPLAN in the intervention arm (equal to the target sample) and the control arm in each analytical sample within BEBOLD, South Australia, 2004–10

AEDC analytical sample				NAPLAN analytical sample			
n[a]	Control	Intervention	Missing %[b]	n	Control	Intervention	Missing %[b]
	2135	2025			3989	2381	
AEDC-DV1 (%)	517 (31.8)	577 (36.7)	23.1	NAPLAN-Reading (%)	533 (24.1)	397 (28.9)	43.8
AEDC-Physical health & wellbeing (%)	270 (16.5)	285 (18.1)	22.9	NAPLAN-Numeracy (%)	653 (29.7)	526 (38.2)	43.9
AEDC-Social competence (%)	253 (15.5)	306 (19.4)	22.9	NAPLAN-Writing (%)	469 (21.2)	346 (25.2)	43.8
AEDC-Emotional maturity (%)	223 (13.7)	254 (16.1)	23.0	NAPLAN-Spelling (%)	565 (25.6)	471 (34.2)	43.6
AEDC-Language & cognitive skills (school-based) (%)	177 (10.9)	201 (12.8)	23.1	NAPLAN-Grammar & punctuation (%)	554 (25.1)	412 (29.9)	43.6
AEDC-Communication skills and general knowledge (%)	193 (11.8)	213 (13.5)	22.9				

BEBOLD, Better Evidence Better Outcomes Linked Data platform; AEDC, Australian Early Development Census; NAPLAN, National Assessment Program—Literacy And Numeracy; DV1, Developmental vulnerabilities in at least 1 domain.

[a]This gives the total number in each analytical sample, but crude risks for each outcome are based on the records with available data for that measure in the given sample.

[b]This is the proportion of missing data across both treatment groups.

Substantively, our results did not provide robust evidence of meaningful beneficial or adverse effects of FHVP. Period-specific analyses revealed that the effect, if any, worsened over time, going from beneficial to neutral for NAPLAN, and from neutral to adverse for AEDC. Assuming no measurement error, confounding or selection bias, which might not be reasonable (see below), there are two possible explanations for effect modification by time-period: a change in the intervention itself or a change in the target population over time. For example, it is possible that nurse training might have been better during the early days which, if not sustained, which would entail a change in the intervention over time. A comparison of the measured characteristics of period-specific cohorts did not suggest any substantial differences, except in the CaFHS service area (results available upon request), which points to potential heterogeneity in the delivery of the programme across centres.

The quality of linked administrative data is not comparable to that of data collected for research purposes. Our efforts to mitigate the risk of bias due to measurement error and confounding may not have been sufficient, particularly to correct for the confounding that could explain the higher risks observed in the intervention arm in unadjusted analyses (Table 3), especially in the later periods. If there is residual confounding or measurement error then our causal effect estimates could be biased. Alternative approaches to dealing with the high proportion of missing outcome data, which could also induce bias in causal effect estimates, led to similar results for most outcomes except in a couple of cases with TMLE, notably for some NAPLAN outcomes. Other approaches could be considered,[37] but the lack of linkage to NAPLAN results for private schools would remain a threat to the "recoverability" of the effect.[38] The consequences in terms of bias would be different for each missing data method and are difficult to quantify because of the large number of possible unverifiable and alternative scenarios.[38,39] Nonetheless, it has been found that between-sector differences in a NAPLAN-based measure of student progress are equivalent to about 1 month after school-level disadvantage is taken into account.[40]

The target trial defines the causal effect of interest as the effect of receiving at least one visit, which would correspond to a type of per protocol effect in a real trial. Other possible per protocol effects could examine the effect of receiving a larger number of visits, but estimation of those effects would require data, such as reasons for dropping out of the programme, that were not available for this study. An analysis of the effect of offering the programme would be closer to an intention-to-treat effect,[1] but may be of less policy relevance because it is when individuals

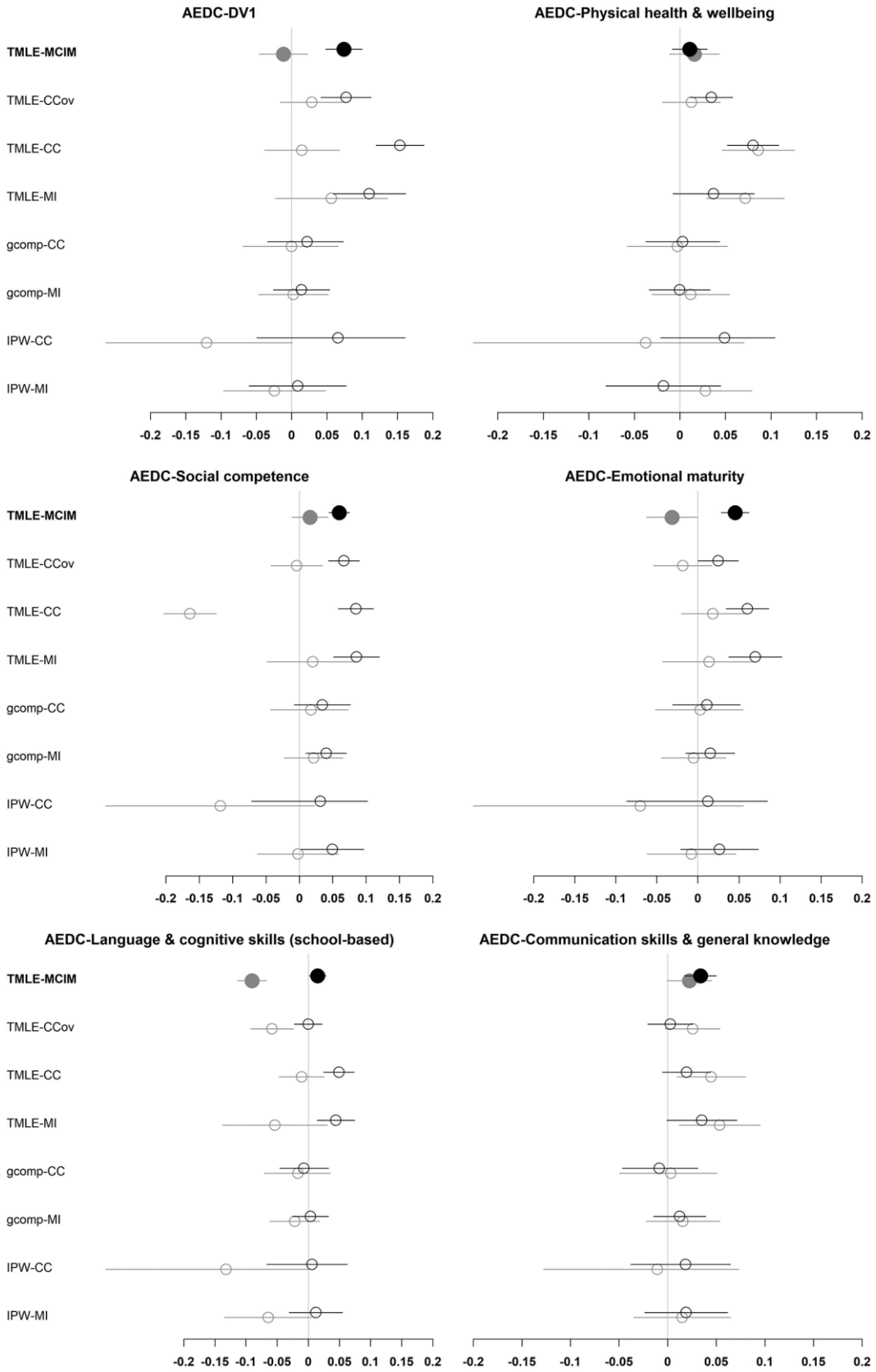

Figure 2 Estimates of the effect of the Family Home Visiting programme (FHVP) on Australian Early Development Census (AEDC) outcomes in BEBOLD, South Australia, 2004–10: adjusted risk differences (intervention—control) estimated with the primary method (in bold) and in sensitivity analyses with other methods, in the main (black) and common (grey) analytical samples, and corresponding 95% confidence intervals. BEBOLD, Better Evidence Better Outcomes Linked Data; CC, Complete Case; CCov, Complete Covariates; DV1, Developmental Vulnerabilities in at least 1 domain; Gcomp, G-computation; IPW, Inverse Probability Weighting; MCIM, Missing Covariate Indicator Method; MI, Multiple Imputation; TMLE, Targeted Maximum Likelihood Estimation

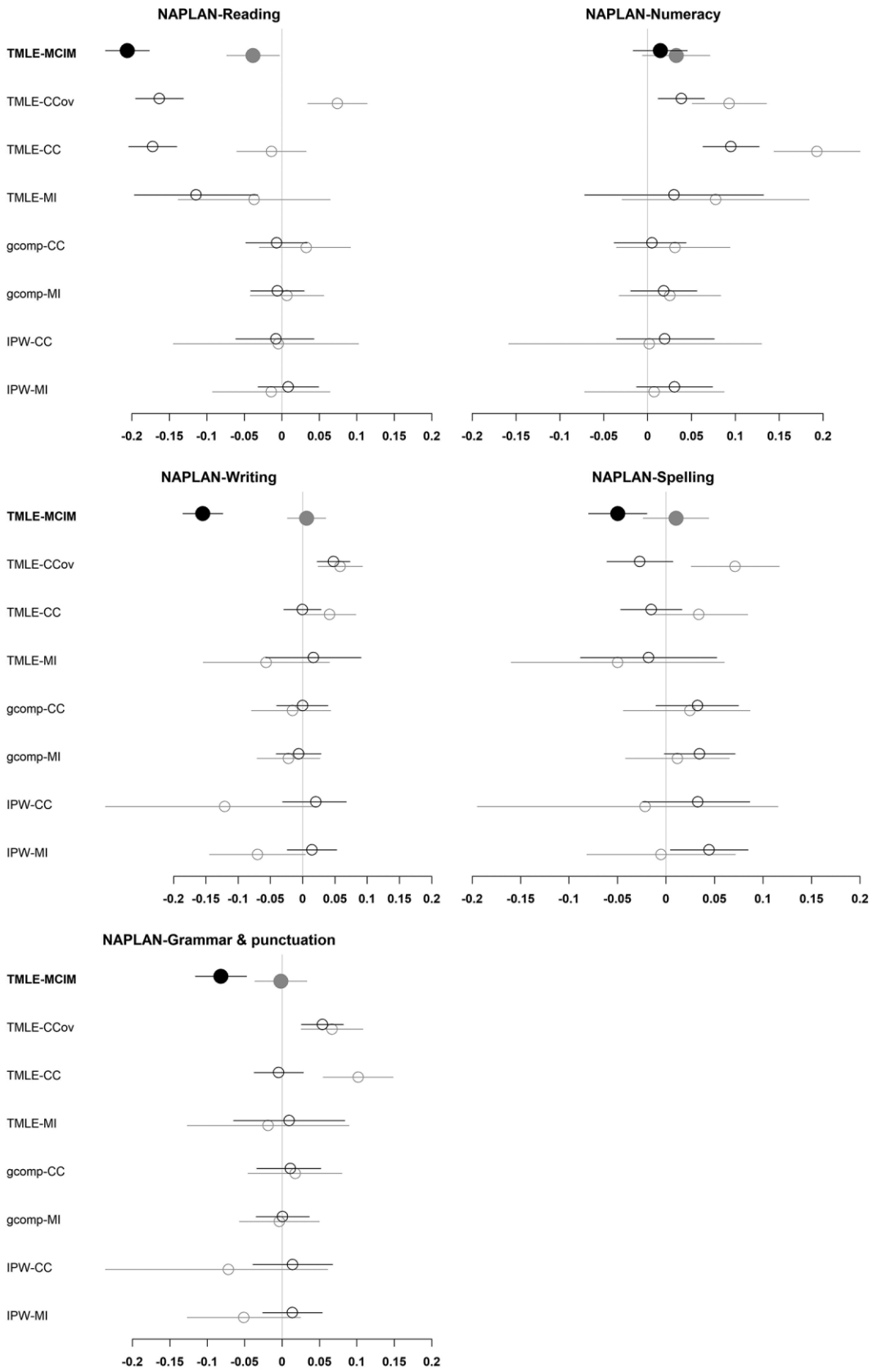

Figure 3 Estimates of the effect of the Family Home Visiting programme (FHVP) on National Assessment Program—Literacy And Numeracy (NAPLAN) outcomes in BEBOLD, South Australia, 2004–10: adjusted risk differences (intervention—control) estimated with the primary method (in bold) and in sensitivity analyses with other methods, in the main (black) and common (grey) analytical samples, and corresponding 95% confidence intervals. BEBOLD, Better Evidence Better Outcomes Linked Data; CC, Complete Case; CCov, Complete Covariates; Gcomp, G-computation; IPW, Inverse Probability Weighting; MCIM, Missing Covariate Indicator Method; MI, Multiple Imputation; TMLE, Targeted Maximum Likelihood Estimation

initiate the programme that the bulk of resourcing is triggered, such as estimating staffing requirements.

Differences in estimates obtained with TMLE versus other methods, in particular the narrower confidence intervals often observed for TMLE, need to be interpreted with caution. Whereas TMLE is robust to mis-specification in terms of parameter estimates, this robustness does not extend to standard error and confidence interval estimation, with bias having been observed in similar sample sizes.[41] This strengthens our rationale for sensitivity analyses and global interpretation of results.

Administrative whole-of-population linked data platforms, coupled with the target trial approach, present an opportunity to complement evidence from RCTs which may not capture the policy population of interest, as well as to produce evidence when RCT evidence is not available or for post-implementation evaluation. We hope that the potential to generate policy-relevant evidence will prompt researchers to collaborate with policy makers in the process of collecting administrative data, to ensure it contains the information needed for high-quality emulation of target trials of real-world interventions. This would help turn purely administrative data systems into 'intelligent information systems' that help underpin continuous quality improvement of what works.[22]

Ethics approval

The project received ethics and site-specific approval from: the South Australian Department of Health and Wellbeing (HREC/13/SAH/106; SSA/13/SAH/146); Women's and Children's Health Network (SSA/14/WCHN/21); and the Aboriginal Health Research Ethics Committee (04–13-538).

Data availability

Data are owned by a third party. The data underlying this article were provided by several Australian State and Commonwealth government agencies under agreements with the researchers, led by author J.W.L, SA NT Datalink as the independent linkage authority and multiple ethics committees. Data are only able to be accessed by researchers who have entered into agreements with the Data Custodians and are approved users by the Human Research Ethics Committee. Data can be accessed through an application and approval process administered by the independent data linkage authority, SA NT Datalink. Example R code for the analyses conducted can be accessed at the first author's GitHub repository [https://github.com/moreno-betancur/EmulateTrial_LinkedData].

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

M.M-B., J.W.L., R.M.P. and L.G.C. conceived the study and designed the analysis plan, which was underpinned by methodological derivations led by M.M-B. All authors approved the analysis plan. R.M.P. and H.S.S. extracted and cleaned the data. M.M-B. conducted the data analyses. M.M-B., J.W.L., R.M.P. and L.G.C. interpreted the findings. M.M-B. led the drafting of the manuscript, with contributions from J.W.L., R.M.P. and .LG.C. All authors critically revised the manuscript and approved the final version.

Funding

This work was supported by the Australian National Health & Medical Research Council (NHMRC) (Centre for Research Excellence grant ID#1099422 to EMPOWER) and the Australian Research Council (Discovery Early Career Researcher Award grant ID# DE190101326 to MMB). The Murdoch Children's Research Institute is supported by the Victorian Government's Operational Infrastructure Support Program.

Acknowledgements

The authors thank the data custodians for contributing de-identified data, and SA NT Datalink for undertaking the data linkage. The findings and views reported are those of the authors and should not be attributed to any particular government department.

Conflict of interest

None declared.

References

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016; **183**:758–64.
2. Hernán MA, Alonso A, Logan R *et al.* Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;**19**:766–79.
3. Danaei G, Rodríguez LAG, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: an emulation of randomized trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res* 2013;**22**:70–96.
4. Dagan N, Barda N, Kepten E *et al.* BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N Engl J Med* 2021;**384**:1412–23.
5. Moreno-Betancur M. The target trial: a powerful device beyond well-defined interventions. *Epidemiology* 2021;**32**:291–94.
6. Government of South Australia. Family Home Visiting Service Outline *(Report)*. 2005. https://www.yumpu.com/en/document/read/4748339/family-home visiting-service-outline-child-and-youth-health (1 February 2017, date last accessed).
7. Love JM, Kisker EE, Ross C *et al.* The effectiveness of early head start for 3-year-old children and their parents: lessons for policy and programs. *Dev Psychol* 2005;**41**:885–901.
8. Barnes J, Ball M, Meadows P, McLeish J, Belsky J; FNP Implementation Research Team. Nurse-family partnership program: first year pilot sites implementation in England (Research

Report). 2008. http://www.iscfsi.bbk.ac.uk/projects/files/Year-1-report-Barnes-et-al.pdf (4 June 2017, date last accessed).
9. Davis H, Day C. *Working in Partnership: The Family Partnership Model*. London: Pearson, 2010.
10. Olds DL, Henderson CR, Chamberlin R, Tatelbaum R. Preventing child abuse and neglect: a trial of nurse home visitation. *Pediatrics* 1986;**78**:65–78.
11. Kitzman H, Olds DL, Henderson CR *et al*. Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing. A randomized controlled trial. *JAMA* 1997;**278**:644–52.
12. Olds DL, Robinson J, O'Brien R *et al*. Home visiting by nurses and by paraprofessionals: a randomized controlled trial. *Pediatrics* 2002;**110**:486–96.
13. Robling M, Bekkers MJ, Bell K *et al*. Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomized controlled trial. *Lancet* 2016;**387**:146–55.
14. Kemp L, Harris E, McMahon C *et al*. Child and family outcomes of a long-term nurse home visitation programme: a randomized controlled trial. *Arch Dis Child* 2011;**96**:533–40.
15. Goldfeld S, Price A, Smith C *et al*. Nurse home visiting for families experiencing adversity: a randomized trial. *Pediatrics* 2019;**143**:1–14.
16. Olds DL, Sadler L, Kitzman H. Programs for parents of infants and toddlers: recent evidence from randomized trials. *J Child Psychol Psychiatry* 2007;**48**:355–91.
17. Chittleborough CR, Lawlor DA, Lynch JW. Young maternal age and poor child development: predictive validity from a birth cohort. *Pediatrics* 2011;**127**:e1436–44.
18. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;**210**:2–21.
19. Brinkman SA, Gregory TA, Goldfeld S, Lynch JW, Hardy M. Data Resource Profile: The Australian Early Development Index (AEDI). *Int J Epidemiol* 2014;**43**:1089–96.
20. Australian Curriculum Assessment and Reporting Authority. NAPLAN Achievement in Reading, Writing, Language Conventions and Numeracy. National Report for 2017. 2017. https://nap.edu.au/docs/default-source/default-document-library/naplan-national-report-2017_final_04dec2017.pdf (1 September 2018, date last accessed).
21. Australian Curriculum Assessment and Reporting Authority. *National Assessment Program*, 2016. https://www.nap.edu.au/results-and-reports/how-to-interpret (1 September 2018, date last accessed).
22. Pilkington R, Montgomerie A, Grant J *et al*. An innovative linked data platform to improve the wellbeing of children: the South Australian Early Childhood Data Project. Canberra: Australian Institute of Health and Welfare: Australia's welfare 2019 data insights. Australia's Welfare Series No. 14. Cat. no. AUS 226 2019. 161–79.
23. Holman C, Bass A, Rouse I, Hobbs M. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health* 1999;**23**:453–59.
24. Centre for Health Record Linkage. *Master Linkage Key Quality Assurance (Report)*, 2012. https://www.cherel.org.au/media/24160/qa_report_2012-a.pdf (1 December 2016, date last accessed).
25. Hernan MA, Robins J. *Causal Inference: What If*. Boca Raton, FL: Chapman & Hall/CRC, 2020.
26. Kreif N, Sofrygin O, Schmittdiel J *et al*. Evaluation of adaptive treatment strategies in an observational study where time-varying covariates are not monitored systematically. *arXiv* 1806.11153, June 28, 2018, preprint: not peer reviewed.
27. Lendle SD, Schwab J, Petersen ML, van der Laan MJ. ltmle: an R package implementing targeted minimum loss-based estimation for longitudinal data. *J Stat Softw* 2017;**81**:1–21.
28. Díaz I, van D, Laan MJ. Doubly robust inference for targeted minimum loss–based estimation in randomized trials with missing outcome data. *Stat Med* 2017;**36**:3807–19.
29. van der Laan MJ. *Imputation and missing data in TMLE*. https://vanderlaan-lab.org/2019/05/10/imputation-and-missing-data-in-the-tmle-framework/ (5 May 2019, date last accessed)
30. Bartlett JW, Seaman SR, White IR, Carpenter JR; Alzheimer's Disease Neuroimaging Initiative. Multiple imputation of covariates by fully conditional specification. *Stat Methods Med Res* 2015;**24**:462–87.
31. van der Laan MJ, Rose S, *Targeted Learning. Causal Inference for Observational and Experimental Data*. New York, NY: Springer, 2011.
32. Gruber S, van der Laan M. An R package for targeted maximum likelihood estimation. *J Stat Softw* 2015;**51**:1–35.
33. ACARA. *NAPLAN Results*. 2019. https://reports.acara.edu.au/ (3 August 2020, date last accessed).
34. Commonwealth of Australia. *Findings from the Australian Early Development Census (AEDC)*. 2019. https://www.aedc.gov.au/communities/findings-from-the-aedc (3 August 2020, date last accessed).
35. Lynch JW. What can early interventions really achieve, and how will we know? Australian Institute of Family Studies. *Fam Matters* 2017;**99**:15–22.
36. Smithers LG, Sawyer ACP, Chittleborough CR, Davies NM, Davey Smith G, Lynch JW. A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat Hum Behav* 2018;**2**:867–80.
37. Dashti SG, Lee KJ, Simpson JA, White IR, Carlin JB, Moreno-Betancur M. Handling missing data when estimating causal effects with Targeted Maximum Likelihood Estimation. *arXiv*, doi:10.48550/arXiv.2112.05274.
38. Moreno-Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical causal diagrams to guide the treatment of missing data in epidemiologic studies. *Am J Epidemiol* 2018;**187**:2705–15.
39. Moreno-Betancur M, Chavance M. Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: methodology and application in a clinical trial with drop-outs. *Stat Methods Med Res* 2016;**25**:1471–89.
40. Goss P, Sonnemann J, Emslie O. *Measuring Student Progress. A State-by-state Report Card*. 2018. https://grattan.edu.au/wp-content/uploads/2018/10/Mapping_Student_Progress.pdf (5 July 2021, date last accessed)
41. Naimi AI, Mishler AE, Kennedy EH. Challenges in obtaining valid causal effect estimates with machine learning algorithms. *Am J Epidemiol* 2021; doi:10.1093/aje/kwab201 [Epub 15 July 2021].