



Comparison of Conditional and Marginal Models in Assessing a Child Nutrition Intervention in Armenia

Arin A. Balalian¹ · Sharon Daniel² · Hambardzum Simonyan³ · Vahe Khachadourian⁴

Accepted: 22 November 2021 / Published online: 8 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Introduction Child malnutrition is a major issue in conflict zones. Evidence-based interventions and their thorough evaluation could help to eliminate malnutrition. We aimed to assess the causal effect of a community-based multidisciplinary nutrition program for children in a chronic conflict zone near the northeastern border of Armenia on two main outcomes: stunting and anemia. We further compared the interpretations and public health relevance of the obtained effect estimates.

Methods In 2016, the study measured hemoglobin and anthropometric measures and collected data from the children's caregivers. We used propensity score matching analyses, inverse probability weighting, and overlap weighting methods to examine the average treatment effects among treated population (ATT), and among population with overlapping weights (ATO).

Results The ATT for stunting among children who participated in the intervention program estimated by propensity score matching analyses (PSM-ATT) was (1.95; 95%CI 1.15–3.28). Nevertheless, children who took part in the program had a lower risk of anemia (0.28; 95%CI 0.19–0.42). The ATT, estimated by inverse probability weighting (IPTW-ATT), was slightly lower for stunting (1.82; 95%CI 1.16–2.86) while similar for anemia (0.33; 95%CI 0.23–0.46) compared to PSM-ATT. Compared to the IPTW-ATT and PSM-ATT the ATO was lower for stunting (1.75; 95%CI 1.14–2.68) and similar for anemia (0.31; 95%CI 0.22–0.43).

Discussion Marginal models could be used in similar quasi-experimental settings to identify the causal effect of interventions in specific populations of interest. Nonetheless, these methods do not eliminate threats to internal validity. Thorough study design and accurate data collection are necessary to improve the efficiency of marginal models.

Keywords Stunting · IPW · Propensity score matching · Intervention effects · Evaluation · Armenia

Significance

Elimination of malnutrition could be achieved by evidence-based interventions which could be identified by rigorous evaluations. We recently assessed the effect of a

multidisciplinary program in a frozen conflict zone. We reported higher odds of stunting and lower odds of anemia among children receiving the program compared to children in the control group. However, the conditional models estimate the effect of interventions at subgroup levels. In community interventions, the causal effects of interventions in specific populations are of interest and could be estimated by marginal models. We compared the findings from marginal models to conditional models and their public health relevance.

Introduction

Child malnutrition remains an unresolved issue, particularly in lower and middle-income countries (de Onis et al., 2013). Although the global prevalence of stunting, an indicator of poor linear growth in children, has generally decreased over

✉ Arin A. Balalian
arin.balalian@columbia.edu

¹ Mailman School of Public Health, Columbia University, 722 West 168th St., New York, NY, USA

² Department of Public Health and Pediatrics, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

³ Fund for Armenian Relief of America, Khorenatsi Street 22, Yerevan, Armenia

⁴ Department of Psychiatry, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY, USA

the past twenty years, this trend is due to disproportionate progress in infant and child feeding and maternal and prenatal health in high-income countries (de Onis et al., 2013). The global prevalence of anemia—also considered an indicator of malnutrition and poor health—has declined slowly, by only 12% between 1995 and 2011 (World Health Organization, 2018). This grim situation warrants massive efforts to reduce malnutrition. With current trends in reduction of prevalence of stunting (de Onis et al., 2013), it is unlikely that goals set by the World Health Assembly will be met (McGuire, 2015). Elimination of malnutrition—one of the sustainable development goals—could be achieved by informed and evidence-based interventions (United Nations Development Programme, 2019).

Numerous interventions have sought to reduce the prevalence of stunting, though not all have been successful. In a recent randomized cluster trial conducted in rural Zimbabwe, infant and young child feeding combined with a water sanitation and hygiene intervention was not superior to infant and young child feeding alone in reducing stunting (Humphrey et al., 2019). In a systematic review of the literature regarding intervention programs to reduce stunting in lower and middle-income countries, several approaches, including education and counseling for nutrition, vitamin supplementation, immunization, water sanitation and hygiene interventions, food safety, security, and social safety net interventions exhibited benefits in a number of settings (Hossain et al., 2017).

Identifying interventions to reduce the prevalence of stunting is challenging and requires robust and rigorous tools for evaluation study design and analysis. Among the several studies assessing the effect of interventions on childhood stunting included in the recent systematic review (Hossain et al., 2017), only one intervention was assessed by a randomized trial design (Arifeen et al., 2009). Although the populations that the effect estimates can be inferred to are implicit in the randomized trials, the interpretability of the observed effect estimates and the populations these effects could be inferred to are debatable in observational studies. Observational evaluation studies in the field of maternal and child health have often used parametric methods such as linear or logistic regressions (conditioning on potential confounding) to assess the intervention effects on outcomes such as stunting (Fenn et al., 2012; Haddad et al., 2014; Lima et al., 2010). In the absence of thorough discussions about the estimated effects, it could be challenging to identify the populations or sub-populations that will benefit from the potential effects of public health interventions.

We recently assessed the effect of a multidisciplinary program implemented in restive (Balalian et al., 2017) Tavush province in Armenia (Simonyan et al., 2019). The border regions of Tavush province have been affected by a chronic frozen conflict with neighboring Azerbaijan for

more than 20 years (A. N. I. Armenian studies Research Center, 2015; The Economist, 2016). In 2013, the Fund for Armenian Relief of America initiated a multidisciplinary community development program to break the cycle of poverty (BCP). The details of this program have been previously described (Simonyan et al., 2019). In brief, BCP included food supplementation and healthcare capacity building at the family, kindergarten, and community level, provision of medical equipment and supplies, as well as other activities aimed at improving women's and children's health (Supplementary Table 1). The communities were selected based on their proximity to the border and available resources. In 2016, BCP was expanded to include nine new communities (Simonyan et al., 2019).

We reported that the odds of stunting were significantly higher (OR 1.92; 95%CI 1.13–3.26) and the odds of anemia significantly lower (OR 0.24; 95%CI 0.16–0.36) among children who lived in communities included in the BCP program compared with children living in communities not included in the program (Simonyan et al., 2019). However, the OR estimates were obtained by logistic regression models after conditioning on potential confounding variables. Furthermore, the conditional models estimate the effect of intervention at the subgroup level, though in community interventions, the effect of intervention at the specific populations (e.g. treated population) is of interest. Marginal models allow the estimation of average treatment effect in the entire population (ATE), average treatment effect among treated(exposed) (ATT), and average treatment effect using overlapping weights among overlapped populations (ATO). These effects are calculated using methods based on propensity scores and have improved interpretability, significant policy, and public health implications (Li et al., 2018; Pang et al., 2016).

The current study aimed to examine the causal effect of the BCP intervention program on anemia and stunting at the population level by applying and comparing three advanced marginal epidemiological methods: propensity score matching, inverse probability weighting (IPTW), and overlap weighting (OW). We further sought to demonstrate the differences in the effect estimates offered by each of these methods in comparison with conditional models, their interpretation, and their relevance to the evaluation of public health interventions.

Methods

Study Setting and Sampling

The current study is a secondary analysis of data collected in 2016 to evaluate FAR BCP program. The study settings and sampling methods have been previously described (Balalian

et al., 2017; Simonyan et al., 2019). Briefly, in 2016 we sampled 983 children living in five intervention (BCP program) communities ($n=347$) and eight control communities ($n=636$). We estimated the required number of children within each age group to detect a 5% difference in population prevalence of stunting and anemia between exposed and unexposed groups using *proc power* procedure in SAS (with $\alpha=0.05$ and $\text{power}=0.80$).

Measurements and Definition of Variables

The primary and secondary outcomes of the study were stunting and anemia. Each child's weight and height (recumbent length for children younger than 24 months) were measured using electronic scales and measuring boards or stadiometers by physicians at each study site. Stunting was defined as a height at least two standard deviations ($<2\text{SD}$) below the median height-for-age of the WHO growth standards (World Health Organization, 1995). Children's hemoglobin (Hgb) levels were measured using the HemoCue® HB 301, a device designed for quick analysis of capillary blood Hgb (Morris et al., 2007; US Food & Drug Administration, 2019). A child was considered anemic if the blood hemoglobin level was less than 110 g/L. The socioeconomic and demographic characteristics of the children's family and diet in the past 24 h were recorded using a self-administered survey and food frequency questionnaire completed by children's caregivers (Simonyan et al., 2019).

Children were considered "exposed" to the BCP program if they were living in one of the five communities where the program was implemented during 2013–2016.

Ethical Considerations

The study protocol was approved by the Institutional Review Board at the Center of Medical Genetics and Primary Health Care, Yerevan, Armenia. Caregivers gave oral consent to participate in the study.

Statistical Methods

Missing Data

We used multiple imputation methods with the fully conditional specification procedure using 200 iterations between each imputation to run 15 imputations (Van Buuren et al., 2006), to impute missing data for the following covariates: (a) *length at birth* ($n=72$); (b) *whether child had ever had diarrhea* ($n=21$); (c) *father's employment* ($n=15$); (d) *mother's education* ($n=12$); (e) *mother's employment* ($n=5$); and *child's body mass index* ($n=3$). Data analyses were performed using SAS 9.4 and R (v. 3.6.0).

We estimated the causal effects of the BCP program on stunting and anemia by matching the observations using propensity scores and weighting the observations using the inverse of the propensity scores and by overlap weights.

Propensity Score Calculation

We calculated the propensity scores for each child by fitting a logistic regression model where exposure to BCP program was set as the outcome. The models included: *kindergarten attendance*, *child age by month*, *weight of the child at birth*, *length of the child at birth*, *minimum dietary diversity*, *any history of previous diagnosis of intestinal parasite infection*, *any history of diarrhea*, *mother's and father's height*, *mother's education*, *mother's employment status*, *father's employment status*, *monthly household expenditure*, *presence of sewage and water system in residence*, *access to electricity and natural gas at home*, *having participated in community education events* and *having received any printed materials related to maternal and child health and safety*, all of which were used as covariates in the multivariable models of the original study (Simonyan et al., 2019). Since 15 imputed datasets were created, the mean probability of being exposed from the 15 imputed databases was used as the propensity score for each observation to conduct the matched and IPTW and OW analysis, as suggested by Mitra and Reiter (2012).

Propensity Score Matching (PSM)

Children who lived in communities that participated in the BCP program were matched with children living in the unexposed communities. Matching was performed by the propensity to reside in areas that engaged in the BCP program.

We matched the exposed and unexposed children (ratio of 1:1) using the greedy matching method. The caliper was set to 0.2 standard deviations of the logit of the propensity score (Austin, 2011; Rosenbaum & Rubin, 1985). Finally, we compared the matched exposed and unexposed groups and calculated the ATT of stunting and anemia and corresponding 95% confidence intervals (95%CI) by fitting negative-binomial models using the *proc genmod* procedure in SAS 9.4.

Inverse Probability Weighting (IPTW)

We calculated the ATT by the stabilized IPTW (IPTW-ATT) using the *proc genmod* procedure in SAS 9.4 (Hernan & Robins, 2010). Inverse population weighting was used in order to standardize the sample in both the exposed and unexposed groups. The inverse probability weights were calculated based on the propensity for exposure to the BCP program. Each observation was assigned a weight, such that

the contributions of under- and over-represented observations in the unexposed group were leveled with the distribution of covariates in the exposed group. “Exposed” observations were given a weight of $P_{(\text{Exposed})}/(\text{propensity score to be included in the FAR program})$, and “unexposed” observations were assigned a weight of $P_{(\text{Unexposed})}/(1 - \text{propensity score to be included in the FAR program})$ (Williamson et al., 2012). We further stabilized the weights to preserve the original sample size (Hernan & Robins, 2010). The IPTW ensures that the distribution of measured covariates is independent of exposure assignment (Austin, 2011; Rosenbaum & Rubin, 1985).

Overlap Weights (OW)

The overlap weighting seeks to weight the study population to reduce the contribution of the outlier observations that have the highest or lowest propensity for receiving the intervention. Overlap weights increase the contribution of the observations with a similar propensity to receive the intervention by assigning higher weights to them. This weighting method does not inflate the sample size and ensures a balance in the measured baseline characteristics between the groups, mimicking the conditions achieved by randomized controlled trials (Li et al., 2018). The observations exposed to the FAR BCP intervention are weighted based on the propensity of not being exposed to the intervention ($1 - \text{propensity score to be included in the FAR program}$), while the unexposed observations are weighted based on the propensity score of receiving the intervention ($\text{propensity score to be included in the FAR program}$). The ATO corresponds to the effect of BCP intervention among the participants that share the most overlap in characteristics and are assigned a higher weight. The ATO was calculated (Li et al., 2018) using *proc genmod* procedure in SAS 9.4.

Results

The sociodemographic and dietary characteristics of study participants are presented in Table 1. In 2016, 1300 children were invited to participate in the evaluation study. The response rate was 85% (Supplementary Fig. 1). The final sample included 347 children from exposed and 636 from unexposed communities. The propensity for BCP intervention ranged from 0.06 to 0.87 (Table 1). We matched 308 children exposed to the FAR program to 308 unexposed based on the propensity scores. A total of 367 observations were excluded from the propensity score matching analysis by the matching process. Most of the excluded unexposed observations had higher propensities to participate in the FAR program. In general, propensity score matching achieved its goal of obtaining a balanced propensity of

exposure among those exposed and unexposed to the FAR program (Fig. 1). The average treatment effects among the children exposed to the program calculated by propensity score matching (PSM-ATT) was similar to the odds ratio estimates calculated in the conditional models by Simonyan et al. (2019): (PSM-ATT 1.93; 95%CI 1.15–3.28) for stunting and 0.28 (95%CI 0.19–0.42) for anemia, respectively (Fig. 2; Supplementary Table 2).

The ATT estimate of stunting calculated by IPTW was reduced (IPTW-ATT:1.82; 95%CI 1.16–2.86) compared to the estimates calculated by logistic regression models conditional on confounders (OR 1.92; 95%CI 1.13–3.26). Nevertheless, the ATT estimate for anemia was similar to those calculated by propensity score matching (IPTW-ATT 0.33; 95%CI 0.23–0.46). The effect estimates calculated by overlapping weights were reduced slightly (ATO 1.75; 95%CI 1.14, 2.68) for stunting and (ATO:0.31; 95%CI 0.22, 0.43) for anemia compared to IPTW-ATT and PSM-ATT (Fig. 2; Supplementary Table 2).

Discussion

The current study found a slightly attenuated association between participation in the FAR program and stunting in a propensity score matching analysis, compared with the effect estimate from the models constructed by Simonyan et al. (Simonyan et al., 2019). The ATT obtained from IPTW analysis and ATO calculated by overlapping weights also showed a slightly lower risk for stunting following participation in the intervention program. Despite the identical covariate adjustments in the conditional and marginal models assessing the risk of stunting, there were notable differences between the estimates obtained from these models. However, the estimates of the effect of exposure to the BCP program on the risk of anemia from the conditional and marginal models were very similar.

We posit several reasons for the observed findings of our study. It is implausible that the BCP intervention has caused stunting, we rather believe the observed estimated effects of intervention could reflect the lingering effects of pre-existing (unobserved) community-specific stressors that led them to be selected as target communities for FAR intervention sites. Second, as Simonyan et al. (2019) described, the higher risk of stunting in the intervention communities could be due to surveillance bias among physicians trained by the FAR BCP program or due to a higher proportion of stunting among children in the intervention communities from 2013 compared to the control communities that could not be alleviated by a three-year program. Due to the lack of baseline measures in the control communities, we are unable to determine if intervention and control communities were similar at baseline in terms of outcomes and covariates.

Table 1 Comparison of study participants in communities receiving and not receiving intervention (children 6 month-6 years old)

Variable	Total N = 983 Mean/No (SD/percent)	Intervention N = 347 Mean/No.0 (SD/percent)	No Intervention Mean/No N = 636 (SD/percent)	p value ^a
Hgb	118.4 (± 13.34)	123.0 (± 11.3)	115.9 (± 13.7)	<0.01
Weight at birth (cm)	3124.3 (± 474.6)	3141.6 (± 473.7)	3115.0 (± 475.1)	0.41
Length at birth (cm)	49.6 (± 2.11)	49.7 (± 2.01)	49.6 (± 2.2)	0.46
Mother's height	160.0 (± 6.4)	160.3 (± 6.5)	159.8 (± 6.3)	0.18
Father's height	169.9 (± 7.0)	170.7 (± 6.6)	169.6 (± 7.2)	0.02
Child's current age (month)	39.5 (± 19.1)	39.98 (± 18.77)	39.2 (± 19.3)	0.52
Child's current BMI	15.9 (± 2.21)	16.2 (± 2.2)	15.7 (± 2.2)	0.01
Sex (n,%)				
Female	515 (52.4)	177 (51.0)	338 (53.1)	0.52
Male	468 (47.6)	170 (49.0)	298 (46.9)	
Kindergarten attendance (n,%)				
No	10 (1.0)	4 (1.2)	6 (0.9)	0.96
Yes	469 (47.7)	165(47.5)	304 (47.8)	
Yes	504 (51.3)	178 (51.3)	326 (51.3)	
Child's history of sleeping hungry reported by caregiver (n,%)	3 (0.3)	1 (0.3)	2 (0.3)	0.76
Always	10 (1.0)	4 (1.2)	6 (0.9)	
Never	877 (89.2)	313(90.2)	564 (88.7)	
Often	5 (0.5)	1 (0.3)	4 (0.6)	
Sometimes	88 (8.9)	28 (8.1)	60 (9.4)	
Mother's education (n,%)	12 (1.2)	3 (0.9)	9 (1.4)	<0.01
Higher education	142 (14.4)	62 (17.9)	80(12.6)	
Incomplete secondary school	109 (11.1)	25 (7.2)	84 (13.2)	
Secondary school	464 (47.2)	131 (37.8)	333 (52.4)	
Vocational secondary school	256 (26.0)	126 (36.3)	130 (20.4)	
Mother's employment (n,%)				
Missing	5 (0.51)	3 (0.9)	2 (0.31)	0.16
Maternity leave	56 (5.7)	25 (7.2)	31 (4.9)	
No	758 (77.1)	256 (73.8)	502 (78.9)	
Yes	164 (16.7)	63 (18.2)	101(15.9)	
Father's employment (n,%)				
Missing	15 (1.5)	11 (3.2)	4 (0.6)	<0.01
No	349 (35.5)	83 (23.9)	266 (41.8)	
Other	16 (1.6)	5 (1.44)	11 (1.7)	
Yes	603 (61.3)	248 (71.5)	355 (55.8)	
Monthly expenditures (n,%)	18 (1.8)	6 (1.7)	12 (1.9)	-
< 50,000 (AMD)	265 (27.0)	74 (21.3)	191 (30.0)	0.01
51,000–100,000	392(39.9)	161 (46.4)	231 (36.3)	
101,000–200,000	251 (25.5)	85 (24.0)	166 (26.1)	
201,000–300,000	45(4.6)	18 (5.2)	27 (4.3)	
> 301,000	12 (1.2)	3 (0.9)	9 (1.4)	
Presence of sewage system at residence (n,%)	15 (1.5)	9 (2.6)	6 (0.9)	
No	390 (39.7)	123 (35.4)	267 (42.0)	0.07
Yes	578 (58.8)	215 (62.0)	363 (57.1)	
Participating in community training	4 (0.4)	3 (0.9)	1 (0.2)	
No	397(40.4)	120 (34.6)	277 (43.5)	<0.01
Yes	582(59.2)	224 (64.6)	358 (56.3)	
Receiving printed materials (n,%)	6 (0.6)	3 (0.9)	3 (0.5)	0.83

Table 1 (continued)

Variable	Total N=983 Mean/No (SD/percent)	Intervention N=347 Mean/No.0 (SD/percent)	No Intervention Mean/No N=636 (SD/percent)	p value ^a
No	249 (25.3)	89 (25.6)	160 (25.2)	
Yes	728 (74.1)	255 (73.5)	473 (74.4)	
Minimum dietary diversity (n,%)				
No	154 (15.7)	73 (21.0)	81 (12.7)	<0.01
Yes	829 (84.3)	274 (79.0)	555(87.3)	
Anemia status(n,%)				
Anemic	237 (24.1)	37 (10.7)	200 (31.4)	<0.01
Non anemic	746 (75.9)	310 (89.3)	436 (68.6)	
Stunting status (n,%)				
Not stunted	894 (91.0)	307 (88.5)	587 (92.3)	0.04
Stunted	89 (9.1)	40 (11.5)	49 (7.7)	
Propensity scores mean (SD); (min–max)	0.35 (0.17); (0.06–0.87)	0.43 (0.17); (0.12–0.87)	0.31(0.15); (0.06–0.80)	N/A

AMD Armenian Drams (currency), N/A not applicable, min minimum, max maximum, SD standard deviation

^ap values are derived from the chi-square test of equality of proportions for categorical variables and t-test for equality of means for numeric variables between the groups of exposed and not-exposed to intervention during 2013–2016

Table adapted with permission from Simonyan et al (2019)

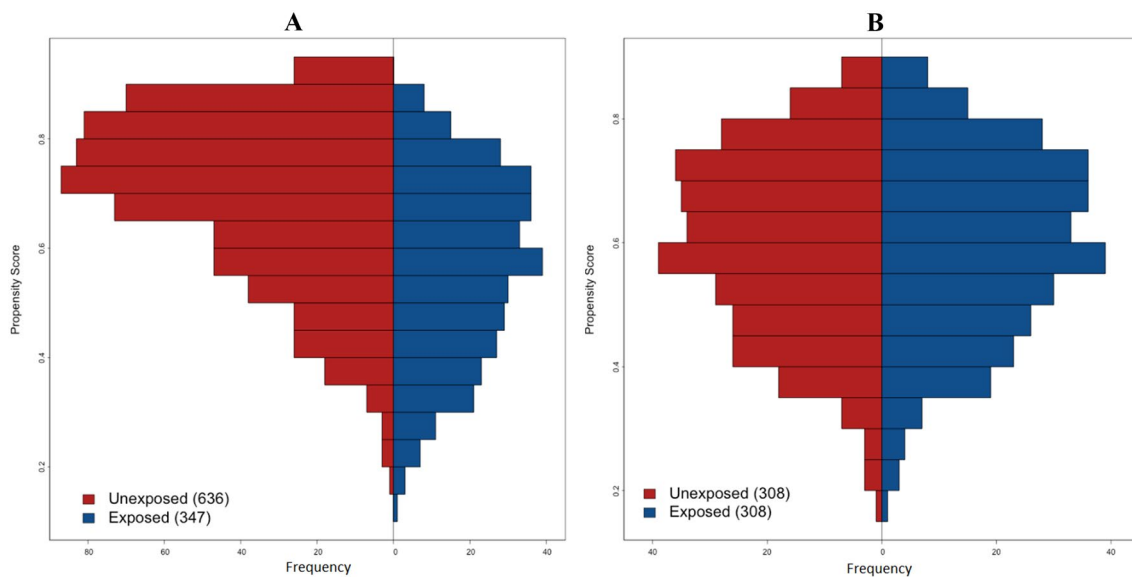


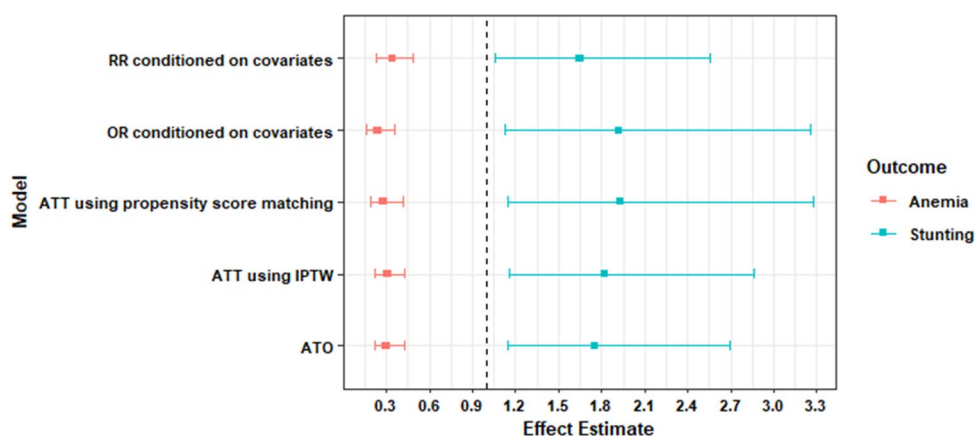
Fig. 1 Top. Mean propensity among exposed and unexposed groups. **A** (Left) Before (n=983) and **B**(right) after 1:1 ratio matching (n=616). Blue bars indicate participants exposed and the red bars

indicate the participants who were not exposed to the program. The number of participants are on the x axis and the propensity score to be exposed to intervention is on the y axis

We observed slight differences in effect estimates across the marginal models and in comparison with the conditional models. The characteristics of the underlying models may explain these observed differences Conditional methods estimate the interventions’ weighted average of stratum-specific effect on health outcomes; they do not necessarily

estimate the ATE and ATT (Pang et al., 2016). Both IPTW and propensity score matching can be used to estimate the ATT. The propensity-score matched sample is a subset of the study population in which the covariates are balanced by matching the exposed and unexposed observations based on propensity score (Austin, 2011; Hernan & Robins, 2010).

Fig. 2 Top. Effect estimates from propensity score matched, IPTW and overlap weighting analysis compared with the results of Simonyan et al. (2019). *OR* odds ratio, *RR* risk ratio, *CI* confidence interval, *ATT* average treatment effect among exposed, *IPTW* inverse probability weighting, *ATO* average treatment effect in the overlapping population using overlapping weights



Propensity score matching also results in a reduction in sample size due to excluding unmatched observations. Since the caliper matching method utilizes more stringent criteria to match exposed with unexposed participants, it did not find unexposed matches for all exposed children, and not all exposed participants were matched with an unexposed participant. Hence, the analytical sample did not coincide with the initially intended population. Thus, the ATT from propensity score matching could be interpreted as effect estimates among the exposed who were matched to an unexposed subpopulation and might vary across the studies (Lunt et al., 2009).

In contrast, the IPTW achieves the covariate balance by creating a pseudo-population in which each individual is weighted by the inverse of the conditional probability of attending the observed intervention. The IPTW can be used to calculate all ATT or ATE depending on the weight used to create the pseudo-population (Austin, 2011). Nonetheless, the extreme probabilities of receiving or not receiving the intervention can inflate variance and reduce the balance between intervention groups.

Overlap weights attempt to address some of these challenges by assigning higher weights to the observations with a similar propensity of receiving or not receiving the intervention (observations with propensity scores closer to 50%) to achieve exact balance in baseline characteristics (Li et al., 2018). The average treatment effect among the overlapped population could be interpreted as the intervention effect estimates among the populations with the most overlap in background characteristics between the exposed and unexposed groups. In our study the ATO refers to the children exposed and unexposed to BCP program who had the most overlap in their background characteristics.

The non-collapsibility of odds ratios between conditional and marginal models (Austin, 2011; Greenland, 1987; Pang et al., 2016) in observational studies may be another source of the difference we found between our

conventional analysis and marginal models. An effect measure is considered collapsible if the conditional and marginal estimates coincide. For example, in randomized clinical trials, the risk differences and difference in means coincide. Nonetheless, odds ratios and hazard ratios do not have the same collapsibility properties, even in the absence of confounding. Therefore, the conditional and marginal models might not result in the same effect estimates (Austin, 2011; Greenland & Morgenstern, 2001). In the absence of confounding, both conditional and marginal models are valid and can be interpreted based on the question the researcher seeks to answer (Pang et al., 2016).

We note that marginal models provide clarity over conditional models by offering effect estimates comparable in well-defined (sub)populations of interest. However, they do not eliminate threats to internal validity. Calculation of propensity scores depends on accurate measurement of relevant covariates, thorough sampling strategies, and selection of source and target populations to reduce the possibility of bias due to unmeasured confounding, differential measurement errors, and selection bias.

Marginal models based on propensity scores are alternative epidemiological methods that can be applied to estimate the causal effects of interventions on specific well-defined (sub)populations that can benefit from those interventions. These methods may be used in similar quasi-experimental settings, given that sufficient information is obtained to calculate the probability of participating in the intervention program and have significant public health and policy implications.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10995-021-03308-y>.

Acknowledgements AB was supported by the Luys Foundation, Inc., the Armenian General Benevolent Union, the Armenian Relief Society, and the Adrina Casparian Kayaian Memorial Scholarship Fund. The funding for the FAR-Breaking Cycle of Poverty program was provided by the Mardigian Family foundation. The authors would like to

express their gratitude to primary healthcare physicians in Berd region for facilitating data collection.

Author Contributions AB conceptualized and designed the study, conducted data analysis, and wrote the manuscript. SD conducted data analysis, critically reviewed the manuscript for intellectual content and manuscript readiness. HS, supervised design of data collection instruments and data collection and critically reviewed the manuscript for intellectual content and manuscript readiness. VK conceptualized and designed the study reviewed the analysis, and critically reviewed and supervised the manuscript for intellectual content and manuscript readiness.

Funding The authors received no funding to conduct this study.

Availability of Data and Material Not Applicable.

Code Availability The code to reproduce the findings in this study is included in supplementary material.

Declarations

Conflict of interest HS is the director of healthcare programs at Fund for Armenian Relief of America. AB was previously employed at FAR. VK and SD declare no conflict of interest.

Ethics Approval The study protocol was approved by the Center of Medical Genetics and Primary Health Care, Yerevan, Armenia.

Consent to Participate Caregivers gave oral consent to participate in the study, and data were collected without identifiable information. Only the corresponding author had access to the data. Data safety was assured using password protection, and data will be retained for a minimum of three years following publication of the study findings.

Consent for Publication Not Applicable.

References

- Balalian, A. A., Simonyan, H., Hekimian, K., Deckelbaum, R. J., & Sargsyan, A. (2017). Prevalence and determinants of stunting in a conflict-ridden border region in Armenia—A cross-sectional study. *BMC Nutrition*, 3(1), 85. <https://doi.org/10.1186/s40795-017-0204-9>
- A. N. I. Armenian studies Research Center. (2015). *Armenian Border Tavush*. Retrieved from Yerevan: <http://www.aniarc.am/wp-content/uploads/2015/02/Report-first-draft.pdf>
- Arifeen, S. E., Hoque, D. M. E., Akter, T., Rahman, M., Hoque, M. E., Begum, K., et al. (2009). Effect of the Integrated Management of Childhood Illness strategy on childhood mortality and nutrition in a rural area in Bangladesh: A cluster randomised trial. *The Lancet*, 374(9687), 393–403. [https://doi.org/10.1016/S0140-6736\(09\)60828-X](https://doi.org/10.1016/S0140-6736(09)60828-X)
- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Austin, P. C. (2017). Double propensity-score adjustment: A solution to design bias or bias due to incomplete matching. *Statistical Methods in Medical Research*, 26(1), 201–222. <https://doi.org/10.1177/0962280214543508>
- Bhutta, Z. A., Das, J. K., Rizvi, A., Gaffey, M. F., Walker, N., Horton, S., et al. (2013). Evidence-based interventions for improvement of maternal and child nutrition: What can be done and at what cost? *The Lancet*, 382(9890), 452–477. [https://doi.org/10.1016/S0140-6736\(13\)60996-4](https://doi.org/10.1016/S0140-6736(13)60996-4)
- de Onis, M., Dewey, K. G., Borghi, E., Onyango, A. W., Blossner, M., Daelmans, B., et al. (2013). The World Health Organization's global target for reducing childhood stunting by 2025: Rationale and proposed actions. *Maternal & Child Nutrition*, 9(Suppl 2), 6–26. <https://doi.org/10.1111/mcn.12075>
- Fenn, B., Bulti, A. T., Nduna, T., Duffield, A., & Watson, F. (2012). An evaluation of an operations research project to reduce childhood stunting in a food-insecure area in Ethiopia. *Public Health Nutrition*, 15(9), 1746–1754. <https://doi.org/10.1017/S1368980012001115>
- Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology*, 125(5), 761–768. <https://doi.org/10.1093/oxfordjournals.aje.a114593>
- Greenland, S., & Morgenstern, H. (2001). Confounding in health research. *Annual Review of Public Health*, 22(1), 189–212. <https://doi.org/10.1146/annurev.publhealth.22.1.189>
- Haddad, L., Nisbett, N., Barnett, I., & Valli, E. (2014). Maharashtra's child stunting declines: What is driving them? Findings of a multidisciplinary analysis.
- Hernan, M. A., & Robins, J. M. (2010). *Causal inference*. CRC Boca Raton.
- Hossain, M., Choudhury, N., Adib Binte Abdullah, K., Mondal, P., Jackson, A. A., Walson, J., & Ahmed, T. (2017). Evidence-based approaches to childhood stunting in low and middle income countries: a systematic review. *Archives of Disease in Childhood*, 102(10), 903. <https://doi.org/10.1136/archdischild-2016-311050>
- Humphrey, J. H., Mbuya, M. N., Ntozini, R., Moulton, L. H., Stoltzfus, R. J., Tavengwa, N. V., et al. (2019). Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: A cluster-randomised trial. *The Lancet Global Health*, 7(1), e132–e147. [https://doi.org/10.1016/S2214-109X\(18\)30374-7](https://doi.org/10.1016/S2214-109X(18)30374-7)
- Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521), 390–400. <https://doi.org/10.1080/01621459.2016.1260466>
- Lima, A. L. L. D., Silva, A. C. F. D., Konno, S. C., Conde, W. L., Benicio, M. H. D. A., & Monteiro, C. A. (2010). Causes of the accelerated decline in child undernutrition in Northeastern Brazil (1986–1996–2006). *Revista de Saude Publica*, 44, 17–27. <https://doi.org/10.1590/S0034-89102010000100002>
- Lunt, M., Solomon, D., Rothman, K., Glynn, R., Hyrich, K., Symons, D. P., et al. (2009). Different methods of balancing covariates leading to different effect estimates in the presence of effect modification. *American Journal of Epidemiology*, 169(7), 909–917. <https://doi.org/10.1093/aje/kwn391>
- McGuire, S. (2015). World Health Organization. Comprehensive Implementation Plan on Maternal, Infant, and Young Child Nutrition. Geneva, Switzerland, 2014. *Adv Nutr*, 6(1), 134–135. <https://doi.org/10.3945/an.114.007781>
- Mitra, R., & Reiter, J. P. (2012). A comparison of two methods of estimating propensity scores after multiple imputation. *Statistical Methods in Medical Research*, 25(1), 188–204. <https://doi.org/10.1177/0962280212445945>
- Morris, L. D., Osei-Bimpong, A., McKeown, D., Roper, D., & Lewis, S. M. (2007). Evaluation of the utility of the HemoCue 301

- haemoglobinometer for blood donor screening. *Vox Sanguinis*, 93(1), 64–69. <https://doi.org/10.1111/j.1423-0410.2007.00919.x>
- Pang, M., Kaufman, J. S., & Platt, R. W. (2016). Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Statistical Methods in Medical Research*, 25(5), 1925–1937. <https://doi.org/10.1177/0962280213505804>
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the Propensity Score. *The American Statistician*, 39(1), 33–38. <https://doi.org/10.2307/2683903>
- Simonyan, H., Sargsyan, A., Balalian, A. A., Davtyan, K., & Gupte, H. A. (2019). Short-term nutrition and growth indicators in 6-month-to 6-year-old children are improved following implementation of a multidisciplinary community-based programme in a chronic conflict setting. *Public Health Nutrition*. <https://doi.org/10.1017/S1368980019002969>
- The Economist. (2016). A frozen conflict explodes. *Economist*.
- United Nations Development Programme. (2019). Goal 2: Zero Hunger. Retrieved from <https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-2-zero-hunger.html>
- US Food and Drug Administration. (2019, 11/25/2019). Establishment Registration & Device Listing. Retrieved from <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRL/rl.cfm?lid=128227&lpcd=GKR>
- Van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12), 1049–1064. <https://doi.org/10.1080/10629360600810434>
- Williamson, E., Morley, R., Lucas, A., & Carpenter, J. (2012). Propensity scores: From naive enthusiasm to intuitive understanding. *Statistical Methods in Medical Research*, 21(3), 273–293. <https://doi.org/10.1177/0962280210394483>
- World Health Organization. (1995). Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee.
- World Health Organization. (2018). WHO | Global Nutrition Targets 2025: Anaemia policy brief. *WHO*. Retrieved from https://www.who.int/nutrition/publications/globaltargets2025_policybrief_anaemia/en/#.XB8lvUkexek.mendeley

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.