Re-analysis of data from a cluster RCT entitled "health literacy and exercise-focused interventions on clinical measurements in Chinese diabetes patients"

Yasaman Jamshidi-Naeini,[a] Lilian Golzarri-Arroyo,[a] Colby J. Vorland,[b] Andrew W. Brown,[b] and David B. Allison [a]*

[a]Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA
[b]Department of Applied Health Science, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA

eClinicalMedicine
2022;53: 101686
Published online xxx
https://doi.org/10.1016/j.eclinm.2022.101686

Wang et al.[1] examined the effects of three health literacy and exercise interventions on HbA1c (the primary outcome) in a four-arm cluster randomized trial (cRCT), but did not account for clustering and nesting. Eight Healthcare Centers (CHCs) were randomly assigned to four conditions (two per condition). Repeated measurements occurred at four timepoints after enrollment. Nominally significant *p*-values were reported for intervention effects on HbA1c. We reanalyzed the data accounting for clustering and nesting, and present the results herein.

An assumption underlying the validity of typical inferential statistical methods is independence of observations: the outcome of each respondent is not related to the outcomes of other respondents. In cRCTs, clusters are randomized, but inferences about the intervention effects are often intended to individuals. In the study, individuals from the same CHC (cluster) are expected to be more similar than those from different clusters, resulting in a pattern of correlated data so 'errors' (model residuals) are not independent across individual participants.[2,3] The inherent correlation among observations from the same cluster typically inflates type I error rates. Not accounting for non-independence within clusters can lead to incorrect estimation of the variance; *p*-values that are smaller than what a *valid analysis* will produce for intervention effects; and thus invalid inferences about intervention effects.[4] By valid analysis we mean an analysis which under the null hypothesis, with continuously distributed data and a continuously distributed test-statistic, produces a sampling distribution of *p*-values that is uniform on the interval [0,1]. An invalid analysis refers to a testing procedure that produces any other sampling distribution of *p*-values.

In the study, generalized estimating equations (GEE) models with random effect for individual subjects were used. This accounts for repeated measurements nested within participants, but not for participants nested within CHCs. The authors of the study collegially shared their data with us. We were able to *reproduce* published results per the original methods[1] with negligible differences: the *p*-value for change of HDL within the control group at 24-month follow-up (our *p*-value = 0.042, *p*-value reported in the study ≥0.05), and differences that may reasonably arise due to rounding. We then reanalyzed data for the intervention effects (each intervention compared to control) on HbA1c and the secondary outcomes from baseline to each follow-up point. Herein, we present a model that accounts for clustering and nesting effects of the design within the context of other methodologic choices of the original authors, which involved null hypothesis significance testing based on *p*-values and a 0.05 alpha level. Thus, debate about the relative value of *p*-values and frequentist testing is not within the scope of current work.

To reanalyze the data, we used linear mixed models (LMM) instead of GEE (used in the study), as it is suggested that GEE has unreliable type I error rates for cRCTs with few clusters and should be avoided.[5] GEE is a population-averaged approach[6] using an asymptotic z test which assumes large sample sizes. Specifically, GEE based methods use empirical-sandwich estimation for standard errors. When the degrees of freedom are limited, empirical-sandwich estimation leads to unreliable type I error rates in hypothesis testing. That is, when the number of clusters per condition is small, the increased variability of the sandwich variance estimator substantially inflates the type I error.[7–10] Thus, it would not be appropriate to account for the clustering effect of CHCs and the repeated measures using GEE. Our LMM models included district as a covariate to account for stratification of CHCs by two city districts, random effect for subjects nested within CHCs to account for clustering effect of CHCs, and random effect for repeated measures nested within individual participants (SAS software, Version 9.4).

We found no evidence supporting efficacy of the interventions in improving HbA1c levels, whereas it

E-mail address: allison@iu.edu (D.B. Allison).

			All participants								Participants with suboptimal clinical measurements[d]							
			Control		Health Literacy		Exercise		Comprehensive		Control		Health Literacy		Exercise		Comprehensive	
		Comparison vs. baseline/ control	GEE[a]	LMM[b]	GEE[a]	LMM[b]	GEE[a]	LMM[b]	GEE[a]	LMM[b]	GEE[a]	LMM[b]	GEE[a]	LMM[b]	GEE[a]	LMM[b]	GEE[a]	LMM[b]
A1C	3 months	Baseline	0.276	0.758	<0.001	0.314	0.011	0.338	<0.001	0.306	0.847	0.897	<0.001	0.270	0.004	0.230	<0.001	0.265
		Control [c]	Ref.	Ref.	<0.001	0.353	0.007	0.371	<0.001	0.347	Ref.	Ref.	0.008	0.384	0.031	0.345	0.015	0.378
	6 months	Baseline	0.337	0.770	<0.001	0.224	<0.001	0.051	<0.001	0.202	0.757	0.988	<0.001	0.164	<0.001	0.029	<0.001	0.170
		Control [c]	Ref.	Ref.	<0.001	0.286	<0.001	0.108	<0.001	0.266	Ref.	Ref.	0.005	0.316	<0.001	0.109	0.008	0.323
	12 months	Baseline	0.016	0.425	<0.001	0.110	<0.001	0.141	0.011	0.461	0.287	0.510	<0.001	0.084	<0.001	0.099	0.009	0.427
		Control [c]	Ref.	Ref.	<0.001	0.095	<0.001	0.114	<0.001	0.282	Ref.	Ref.	<0.001	0.095	<0.001	0.106	0.014	0.308
	24 months	Baseline	0.610	0.866	<0.001	0.086	<0.001	0.057	0.838	0.991	0.543	0.882	<0.001	0.060	<0.001	0.031	0.631	0.922
		Control [c]	Ref.	Ref.	<0.001	0.176	<0.001	0.135	0.635	0.912	Ref.	Ref.	<0.001	0.209	<0.001	0.139	0.977	0.970
SBP	3 months	Baseline	0.349	0.623	0.903	0.939	0.228	0.543	0.605	0.706	<0.001	0.038	<0.001	0.016	<0.001	0.049	0.170	0.259
		Control [c]	Ref.	Ref.	0.564	0.768	0.868	0.931	0.333	0.540	Ref.	Ref.	0.678	0.814	0.652	0.860	0.286	0.450
	6 months	Baseline	0.156	0.452	0.089	0.320	<0.001	0.129	0.463	0.662	<0.001	0.002	<0.001	0.005	<0.001	0.010	<0.001	0.019
		Control [c]	Ref.	Ref.	0.772	0.859	0.237	0.560	0.666	0.823	Ref.	Ref.	0.490	0.713	0.137	0.515	0.185	0.453
	12 months	Baseline	0.897	0.951	0.564	0.725	0.054	0.310	0.872	0.909	<0.001	0.008	<0.001	0.037	<0.001	0.047	<0.001	0.036
		Control [c]	Ref.	Ref.	0.623	0.770	0.158	0.441	0.983	0.970	Ref.	Ref.	0.269	0.547	0.175	0.479	0.387	0.605
	24 months	Baseline	0.995	0.995	0.066	0.217	0.804	0.855	0.699	0.829	<0.001	0.004	<0.001	0.005	<0.001	0.048	<0.001	0.016
		Control [c]	Ref.	Ref.	0.188	0.373	0.858	0.893	0.792	0.875	Ref.	Ref.	0.793	0.910	0.097	0.376	0.439	0.662
DBP	3 months	Baseline	0.002	0.038	0.231	0.418	0.060	0.145	0.721	0.740	<0.001	<0.001	<0.001	0.013	<0.001	<0.001	<0.001	0.012
		Control [c]	Ref.	Ref.	0.002	0.044	0.478	0.629	0.110	0.195	Ref.	Ref.	0.165	0.333	0.722	0.774	0.279	0.373
	6 months	Baseline	0.540	0.635	0.185	0.338	0.184	0.276	0.484	0.579	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	<0.001	0.001
		Control [c]	Ref.	Ref.	0.175	0.314	0.585	0.648	0.942	0.951	Ref.	Ref.	0.165	0.336	0.924	0.948	0.291	0.466
	12 months	Baseline	0.389	0.474	0.200	0.290	0.060	0.122	0.527	0.601	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	<0.001	<0.001
		Control [c]	Ref.	Ref.	0.128	0.214	0.429	0.523	0.883	0.894	Ref.	Ref.	0.284	0.404	0.771	0.838	0.730	0.795
	24 months	Baseline	0.622	0.704	0.849	0.872	0.993	0.982	0.082	0.199	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	<0.001	0.001
		Control [c]	Ref.	Ref.	0.618	0.703	0.720	0.804	0.440	0.506	Ref.	Ref.	0.239	0.446	0.559	0.705	0.364	0.495
LDL	12 months	Baseline	0.970	0.973	<0.001	0.042	0.051	0.521	0.399	0.820	0.081	0.592	0.009	0.108	0.160	0.714	0.051	0.624
		Control [c]	Ref.	Ref.	<0.001	0.120	0.164	0.631	0.565	0.853	Ref.	Ref.	0.002	0.131	0.857	0.907	0.977	0.978
	24 months	Baseline	<0.001	0.192	0.011	0.236	0.559	0.864	0.118	0.590	<0.001	0.130	0.419	0.906	<0.001	0.199	<0.001	0.221
		Control [c]	Ref.	Ref.	<0.001	0.092	0.021	0.406	0.117	0.561	Ref.	Ref.	0.006	0.240	0.753	0.860	0.564	0.817
HDL	12 months	Baseline	0.088	0.248	0.014	0.122	0.075	0.249	0.433	0.727	0.001	0.003	0.009	0.118	0.001	0.005	<0.001	0.016
		Control [c]	Ref.	Ref.	0.674	0.731	0.014	0.116	0.319	0.558	Ref.	Ref.	0.200	0.246	0.789	0.741	0.329	0.336
	24 months	Baseline	0.042	0.334	0.099	0.040	0.136	0.302	0.011	0.162	<0.001	0.012	0.003	0.030	<0.001	0.003	<0.001	0.001
		Control [c]	Ref.	Ref.	0.368	0.338	0.018	0.170	0.482	0.707	Ref.	Ref.	0.945	0.924	0.141	0.204	0.194	0.353

***Table 1*: Time and intervention effect *p*-values from generalized estimating equation (GEE) models that do not account for clustering and nesting and linear mixed models (LMM) that do account for clustering and nesting.**

[a] Generalized Estimating Equations with repeated measure nested in subject ID as random effect and Gaussian family; the statistical model used by Wang et al., which ignores the clustering effect (i.e., is invalid).

[b] Our reanalyzed results using Linear Mixed Models with repeated measure nested in subject ID and subject ID nested in CHC as random effects, and district as covariate, accounting for the clustering effect of CHCs.

[c] *p*-values between each intervention group compared to the control arm in changes from baseline to follow-up time.

[d] Wang et al. defined suboptimal clinical measurements as HbA1c≥7.0%, SBP>130 mmHg, DBP> 80 mmHg, LDL>2.6 mmol/L, HDL<1.04 (men) or <1.30 mmol/L (women).

was concluded in the study that interventions resulted in statistically significant improvements in HbA1c. Among the 12 between-group HbA1c comparisons, 11 were originally reported as nominally significant (p-value<0.05). None of those were statistically significant when clustering and nesting were accounted for during statistical analyses. In Table 1, we present the p-values derived from GEE models not accounting for clustering and nesting (used in the study), and analyses using LMM (accounting for clustering and nesting). Similarly, none of the between-group p-values originally reported as nominally significant for the secondary outcomes and for the subgroup analyses were statistically significant in our analyses.

One may argue that because in the study interventions were implemented for individual participants (e.g., one-on-one sessions) and not at group level (e.g., participants receiving the intervention simultaneously) clustering has been independent from treatment implementation, and thus, accounting for clustering and nesting is not necessary. We do not consider this to be correct: even if clustering was truly independent from treatment implementation, this does not eliminate the effect of clustering on variance components.

We also need to emphasize that post-hoc calculation of sample's ICC is not an appropriate gatekeeper test to determine whether one should account for clustering.[11] Not accounting for clustering, regardless of the sample ICC's magnitude, potentially leads to miscalculation of variance components and type I error rates above the nominal significance level[4,12] (Table 1).

Statistical analyses of cRCTs that ignore clustering and nesting are invalid, and any drawn conclusions are unsubstantiated. To ensure the integrity of the scientific record, errors in the methods, results, and conclusions of the study should be publicly acknowledged and corrected results should be used instead.

Contributors

Conceptualization: All authors; Data analysis: LGA, YJ-N; Investigation: LGA, YJ-N; Writing original draft: YJ-N; Review and editing: all authors; Funding acquisition: DBA, AWB. All authors approved the final draft.

Data sharing statement

Statistical code is available upon request from the corresponding author. The dataset used for the analyses was shared with us by the corresponding author of https://doi.org/10.1016/j.eclinm.2019.11.004. We did not have any special access privileges that others would not have.

Declaration of interests

In the last thirty-six months prior to the original submission of this manuscript, DBA has received personal payments or promises for same from: American Society for Nutrition; Alkermes, Inc.; American Statistical Association; Big Sky Health, Inc.; Biofortis; California Walnut Commission; Clark Hill PLC; Columbia University; Fish & Richardson, P.C.; Frontiers Publishing; Gelesis; Henry Stewart Talks; IKEA; Indiana University; Arnold Ventures (formerly the Laura and John Arnold Foundation); Johns Hopkins University; Kaleido Biosciences; Law Offices of Ronald Marron; MD Anderson Cancer Center; Medical College of Wisconsin; National Institutes of Health (NIH); Medpace; National Academies of Science; Sage Publishing; The Obesity Society; Sports Research Corp.; The Elements Agency, LLC; Tomasik, Kotin & Kasserman LLC; University of Alabama at Birmingham; University of Miami; Nestle; WW (formerly Weight Watchers International, LLC). Donations to a foundation have been made on his behalf by the Northarvest Bean Growers Association. Dr. Allison was previously an unpaid member of the International Life Sciences Institute North America Board of Trustees. In the last thirty-six months prior to the original submission of this manuscript, AWB has received travel expenses from University of Louisville; speaking fees from Purdue University and University of Arkansas; consulting fees from LA NORC, and Pennington Biomedical Research Center; and award honorarium from the American Society for Nutrition Foundation. His wife is employed by Reckitt Benckiser. The institution of DBA and AWB, Indiana University, has received funds, contracts, or donations to support their research or educational activities from: NIH; USDA; Soleno Therapeutics; National Cattlemen's Beef Association; Eli Lilly and Co.; Reckitt Benckiser Group PLC; Alliance for Potato Research and Education; American Federation for Aging Research; Dairy Management Inc; American Egg Board; California Walnut Commission; Almond Board; Peanut Institute; Mondelez; Hass Avocado Board; Whistle Labs, Inc; Dynamic AQS; Arnold Ventures; the Gordon and Betty Moore Foundation; the Alfred P. Sloan Foundation; Indiana CTSI; Center for Open Science, and numerous other for-profit and non-profit organizations to support the work of the School of Public Health and the university more broadly. Other authors declare no competing interests.

Acknowledgements

Authors supported in part by R25DK099080, R25HL124208, R25GM141507, and the Gordon and Betty Moore Foundation. The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization.

References

1 Wang L, Fang H, Xia Q, et al. Health literacy and exercise-focused interventions on clinical measurements in Chinese diabetes patients: a cluster randomized controlled trial. *EClin Med*. 2019;17:100211. https://doi.org/10.1016/j.eclinm.2019.11.004.

2 Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *Am J Public Health*. 2004;94:423–432.

3 Sainani K. The importance of accounting for correlated observations. *PM&R*. 2010;2:858–861.

4 Brown AW, Li P, Bohan Brown MM, et al. Best (but oft-forgotten) practices: designing, analyzing, and reporting cluster randomized controlled trials. *Am J Clin Nutr*. 2015;102:241–248.

5 Li P, Redden DT. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. *Stat Med*. 2015;34:281–296.

6 Zeger SL, Liang K-Y, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44:1049–1060.

7 Murray DM. *Design and Analysis of Group-Randomized Trials*. USA: Oxford University Press; 1998.

8 Murray DM, Hannan PJ, Wolfinger RD, Baker WL, Dwyer JH. Analysis of data from group-randomized trials with repeat observations on the same groups. *Stat Med*. 1998;17:1581–1600.

9 Hayes RJ, Moulton LH. *Cluster Randomised Trials*. CRC Press; 2017.

10 Murray DM. Design and analysis of group-randomized trials: a review of recent developments. *Ann Epidemiol*. 1997;7:S69–S77.

11 Jamshidi-Naeini Y, Brown AW, Mehta T, et al. A practical decision tree to support editorial adjudication of submitted parallel cluster randomized controlled trials. *Obesity*. 2022;30:565–570.

12 Golzarri-Arroyo L, Dickinson SL, Jamshidi-Naeini Y, et al. Evaluation of the type I error rate when using parametric bootstrap analysis of a cluster randomized controlled trial with binary outcomes and a small number of clusters. *Comput Methods Programs Biomed*. 2022:106654.