DIABRISK-SL trial: further consideration of age and impact of imputations

Authors: Efstathia Gkioni[1,3], Ketevan Glonti[2,3], Susanna Dodd[1] , Carrol Gamble[1]

Affiliations:

[1] Department of Biostatistics, University of Liverpool, United Kingdom, e.gkioni@liverpool.ac.uk, c.gamble@liverpool.ac.uk, S.R.Dodd@liverpool.ac.uk
[2] Faculty of Humanities and Social Sciences, University of Split, Split, Croatia. kglonti@unist.hr
[3] INSERM, U1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), Methods of therapeutic evaluation of chronic diseases Team (METHODS), Paris, F-75014 France; Paris Descartes University, Sorbonne Paris Cité, France

Corresponding Author: M.Sc. Efstathia Gkioni, e.gkioni@liverpool.ac.uk

Abstract

Type 2 diabetes mellitus (T2DM) is a major cause of morbidity and mortality worldwide. Early interventions may help to delay or prevent onset of cardio-metabolic endpoints of clinical importance.

Wijesuriya et al. recently published in BMC Medicine results of a randomised controlled trial (RCT) in Sri Lanka testing the effect of two lifestyle modification programmes of different intensities in participants aged 6-40 with risk factors for T2DM. The intervention measured the impact of the programmes on the primary composite endpoint consisting of various predictors of cardio-metabolic disease. A wide range of diabetes prevention strategies include structured dietary plans and exercise/activity modification. The authors concluded that the intensive programme significantly reduced the incidence of predictors of cardio-metabolic disease.

The authors delivered a large-scale intervention with restricted resources. The widespread acceptance of the intervention was demonstrated with the high uptake rate. However, we believe that further analysis is required to fully understand the potential for benefit, particularly in relation to age, retention and missing data.

Keywords

Randomised controlled trial, Lifestyle modification programme, type 2 diabetes mellitus

Age differentiation

The intervention implemented by Wijesuriya et al. [1] is well described in general; however, given the wide age range of participants included in the study (5-40 years), further details are required regarding the intervention delivered to children. It is unclear whether the study provided a nuanced intervention for children below 18 with different approaches taking into account the different age brackets with their respective developmental stages [2]. Subgroups within the paediatric population may show differential responses to the same intervention because of physiological and educational differences in children of different ages, therefore adequate power to avoid type II errors in age-specific subgroup analyses is key [3]. In their previous paper that looked into the prevalence of cardio-metabolic risk factors in the study population that was screened for the RCT participation, the authors provided demographic and anthropometric characteristics according to different age groups [4]. It is unclear why this information was not provided for the recruited study participants.

Furthermore, given that children within the lower age ranges do not have independence over their food choices and activity options, the engagement of their primary carers is key [5]. Therefore, it would have been meaningful to provide more details on whether carers had been involved in the intervention and how they handled the situation where carers and children reported differently.

Further exploration of age could also provide important information regarding age-specific effects of the interventions on the outcomes assessed. Consideration of age within the statistical analysis is limited to its categorisation above or below the age of 18 years. Given the rising levels of childhood obesity and its long-term consequences [6] it is important to understand whether the intervention is equally effective across all age groups or whether resources should be targeted to particular age groups.

Retention and missing data

In their study protocol [7] the specified follow-up period is five years in order to detect a 25% reduction in the relative risk between the two groups. However, in the published article [1] the authors report a median follow-up of three years, with a range from one to four years. It would be of interest to know whether the reduction in the follow up period was influenced by retention of study participants. More details on the group-specific retention rates would also be useful because a differential retention between groups may indicate non adherence and biased results [8]. Given the different intensity of the interventions being compared in the study, there may be treatment-group specific differences between the participant engagement and consequently retention.

Furthermore, the authors used the Last Observation Carried Forward (LOCF) method to handle missing data for participants whose measurements were missing. This method substitutes a single reasonable value for a missing observation assuming no change since last observed value prior to dropout [9]. This method of imputation relies on the assumption that the probability of missing data occurs completely at random and that the probability of dropout is not related to variables such as disease severity, group assignment or side effects of intervention [10]. However, the assumptions of stability and randomness may not be realistic for the presented study, as the reasons causing the missing data are not known.

Imputation of a single value for the missing data is not recommended since the underlined assumptions often seem to be unrealistic and they are not scientifically justified [11]. In an anti-obesity drug trial, Jorgensen et al. used different imputation methods for the missing values, the baseline carried forward approach (BOCF), where the missing weight measurements were substituted with the baseline weight, the last observation carried forward and the multiple imputation (MI) method where the missing data are replaced by imputed values, sampled from the predictive distribution based on the observed data [12]. The MI and LOCF methods resulted in similar differences between the treatment and placebo, while the MI method introduces greater uncertainty, which is a more realistic scenario.

MI models impute data several times in order to allow estimation of the full uncertainty about the missing data. This method therefore incorporates not only the variability of the outcome but also the uncertainty about the missing observations. Imputation uses the available information to make better allowances for patients with missing data. Since the mechanism behind the missingness is not known in this trial, and it is possible that the missing data are not missing at random [13], the MI approach could provide more reliable results in comparison with the LOCF approach used in this study [14]. The bias introduced by the MI analyses could be avoided if the variables predictive of missing values are included in the imputation model.

In addition, in this study, it is unclear how many participant measurements were observed and how many were imputed. It is important for the readers to know the extent of imputation required and whether the analysis accounted for differential retention [15]. Thus, the robustness of the conclusions reached and any differences in retention rates between trial arms could be investigated, in order to aid interpretation of the findings and support future trial designs.

However, deficiencies in the reporting of missing data seems to be a common problem. Only half of the articles in review of Rezvan et al. [16] reported both the proportion of missing data and complete cases for the variables of interest. Sterne and colleagues also identified lack of reporting of the multiple imputation approach with only seven out of fifty-nine articles reporting results from both imputed and complete cases analyses. Thus, guidelines have been suggested to improve reporting of missing data analysis methods [17].

Despite the fact that there is no universal method for handling incomplete data in a clinical trial, there are six principles that should be considered, including the reasons causing the missingness, the primary set of assumptions about the missing data mechanism and clarification of whether the values that are missing are meaningful for analysis [11]. Although it is not possible to determine whether data are missing at random or missing not at random, sensitivity analyses addressing biases caused by data that are missing not at random are recommended to assess the robustness of findings.

Conclusion

The DIABRISK-SL is a large low-cost educational intervention. Therefore, it is important to take advantage of the sample size and evaluate the available information for different age groups. The analysis of participants below 18 years of age without differentiation into smaller age categories could be considered a missed opportunity to help those of an early age to establish a healthy lifestyle and prevent the incidence of T2DM. Provision of additional information regarding attrition and missing data would allow greater reassurance regarding the robustness of the results and conclusions.

Abbreviations

T2DM: type 2 diabetes mellitus

RCT: Randomised Controlled Trial

LOCF: Last Observation Carried Forward

BOCF: Baseline Observation Carried Forward

MI: Multiple Imputation

Declarations

- Ethics approval and consent to participate: “Not applicable”
- Consent for publication: “Not applicable”
- Availability of data and materials: “Not applicable”
- Competing interests: "The authors declare that they have no competing interests"
- Funding: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 676207.
- Authors' contributions: “The commentary was prepared by EG, KG, SD and CG after the careful consideration of the research article for the DIABRISK-SL trial and its protocol. All authors read and approved the final manuscript."
- Acknowledgements: We would like to thank the members of the MiRoR Project and Marie Sklodowska-Curie Actions for their support.

References

1. Wijesuriya, M., et al., *A pragmatic lifestyle modification programme reduces the incidence of predictors of cardio-metabolic disease and dysglycaemia in a young healthy urban South Asian population: a randomised controlled trial.* BMC Med, 2017. **15**.
2. Waters, E., et al., *Interventions for preventing obesity in children.* Cochrane Database of Systematic Reviews, 2011(12).
3. Klassen, T.P., et al., *Children Are Not Just Small Adults: The Urgent Need for High-Quality Trial Evidence in Children.* PLOS Medicine, 2008. **5**(8).
4. Wijesuriya, M., et al., *High Prevalence of Cardio-Metabolic Risk Factors in a Young Urban Sri-Lankan Population.* PLOS ONE, 2012. **7**(2).
5. Madden, L., et al., *Questioning assent: how are children's views included as families make decisions about clinical trials?* Child: Care, Health and Development, 2016. **42**(6).
6. Sahoo, K., et al., *Childhood obesity: causes and consequences.* Journal of Family Medicine and Primary Care, 2015. **4**(2).
7. Wijesuriya, M., et al., *DIABRISK - SL Prevention of cardio-metabolic disease with life style modification in young urban Sri Lankan's - study protocol for a randomized controlled trial.* Trials, 2011. **12**.
8. Dodd, S., I.R. White, and P. Williamson, *A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes.* Trials, 2017. **18**(1).
9. Kenward, M.G. and G. Molenberghs, *Last Observation Carried Forward: A Crystal Ball?* Journal of Biopharmaceutical Statistics, 2009. **19**(5).
10. Molnar, F.J., B. Hutton, and D. Fergusson, *Does analysis using "last observation carried forward" introduce bias in dementia research?* Canadian Medical Association Journal, 2008. **179**(8).
11. Little , R.J., et al., *The Prevention and Treatment of Missing Data in Clinical Trials.* New England Journal of Medicine, 2012. **367**(14).
12. Jørgensen, A.W., et al., *Comparison of Results from Different Imputation Techniques for Missing Data from an Anti-Obesity Drug Trial.* PLOS ONE, 2014. **9**(11).
13. Molenberghs, G., et al., *Analyzing incomplete longitudinal clinical trial data.* Biostatistics, 2004. **5**(3).
14. Schafer, J.L., *Multiple imputation: a primer.* Statistical Methods in Medical Research, 1999. **8**(1).
15. Nich, C. and K.M. Carroll, *'Intention-to-treat' meets 'missing data': implications of alternate strategies for analyzing clinical trials data.* Drug and alcohol dependence, 2002. **68**(2).
16. Hayati Rezvan, P., K.J. Lee, and J.A. Simpson, *The rise of multiple imputation: a review of the reporting and implementation of the method in medical research.* BMC Medical Research Methodology, 2015. **15**(1).
17. Sterne, J.A.C., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* BMJ, 2009. **338**.