Trials

COMMENTARY Open Access

All I want for Christmas…is a precisely defined research question

Suzie Cro[1†] and Rachel Phillips[1*†]

Each year, Santa Claus and his reindeer have the monumental task of delivering billions of presents across the world. With an ever-increasing population [1] and number of presents to deliver, Santa must continually find new ways to optimise his present delivery. This includes reindeer performance. Over recent years, he had noticed reindeers who ate carrots before starting their 24-h Christmas Eve journey tired more easily than those that ate Mrs Claus' homemade mince pies, and he started to wonder if this was related to their chosen pre-flight meal. Santa decided to put his theory to the test and as part of his delivery improvement research programme conducted a randomised controlled trial to investigate the impact of a reindeer's pre-flight meal on performance. On Christmas Eve 2023, Santa gathered nine reindeers (Dasher, Dancer, Prancer, Vixen, Comet, Cupid, Donner, Blitzen and, of course, Rudolph [2]) from the population of flying reindeers and randomly assigned each one to a pre-flight meal of carrots or a mince pie. On Christmas morning, immediately after successfully delivering all the presents carefully made in his workshop, Santa collected different measures of performance from each reindeer. Afterwards, he shared the data he collected with his friendly statistical elf and asked them to answer the question: *does the reindeer's pre-flight meal impact performance?* But this research question is too vague. To make an informed decision about which pre-flight meal performs best, the statistical elf reminds Santa that he needs to ask a more precise question in line with the International Conference on Harmonisation E9 Addendum on Estimands and Sensitivity Analysis in Clinical Trials (ICH E9 (R1)). Otherwise, the statistical elf might not give Santa the answer to the question he wants answered. This would result in a waste of Santa Claus' research time, costs and efforts in doing the trial. To avoid any confusion and ensure Santa gets the answer to his question of interest, just like in any trial of a healthcare intervention, he needs to specify five key attributes [3–6]:

1) What *treatment conditions* are being compared—a full bunch of five carrots versus a single mince pie.
2) What *population* is the treatment effect being explored for—population of flying reindeers.
3) What *outcome (or endpoint)* is being measured to assess performance on—tiredness score, using a VAS on a scale of 0 = not tired at all to 100 = most tired measured on Christmas morning.
4) What *summary measure* (i.e. statistic) is to be calculated to compare the treatment conditions—the average between intervention group difference is suitable for the numerical tiredness outcome score. In other settings, for different outcome types, different summary measures may be more appropriate. For example, for a binary outcome, a risk ratio or risk difference; for a count outcome, a rate ratio; or for time to event outcome, a hazard ratio may be appropriate; for each of the aforementioned outcome types, other summary measure options exist [7–9], and the most appropriate should be chosen based on the trial objective.

†Suzie Cro and Rachel Phillips contributed equally to this work.

*Correspondence:
Rachel Phillips
r.phillips@imperial.ac.uk
[1] Imperial Clinical Trials Unit, School of Public Health, Imperial College London, Stadium House, 68 Wood Lane, London W12 7RH, UK

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

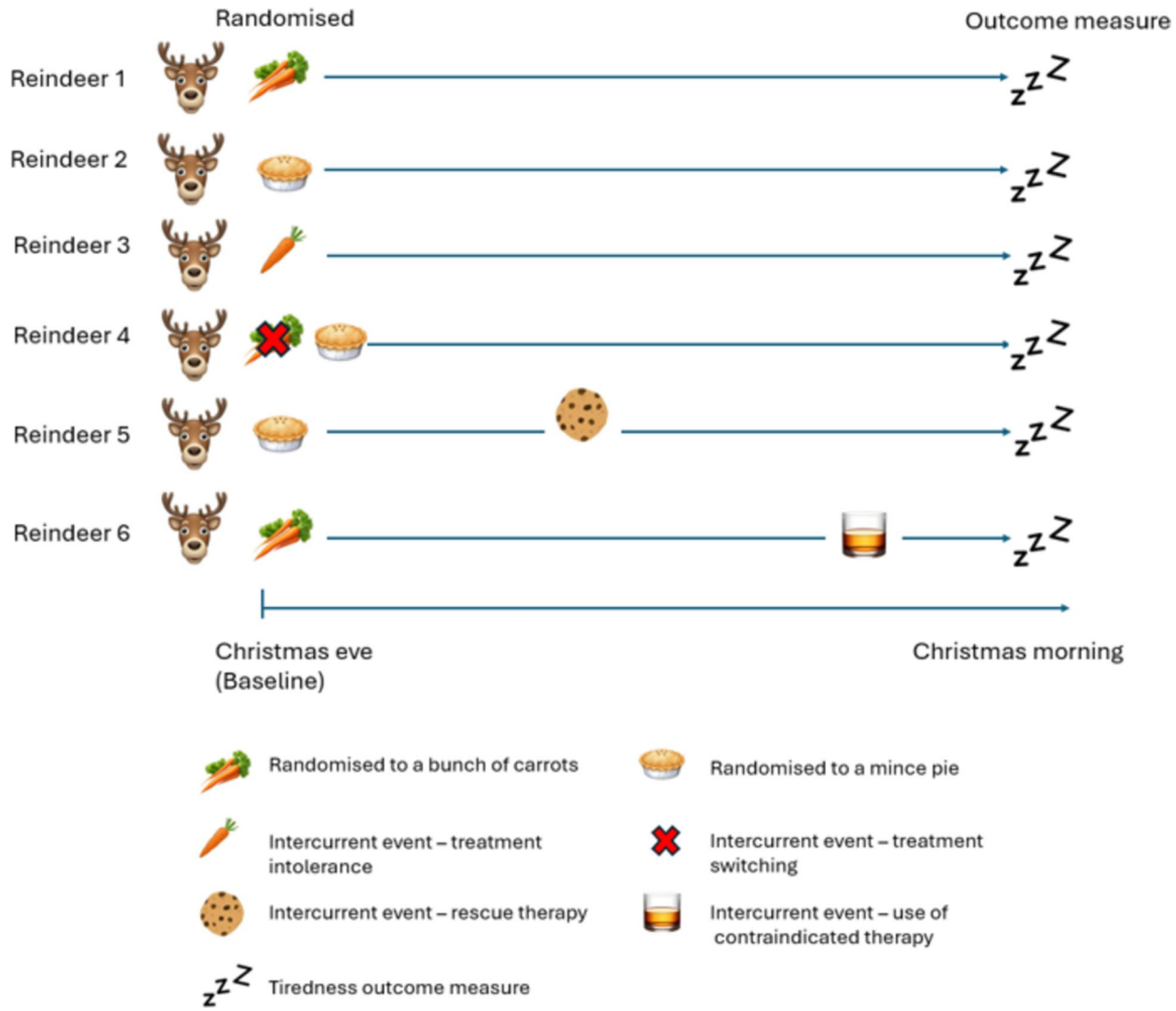


Fig. 1 Potential reindeer experiences over Christmas Eve

Reindeers 1 and 2 adhere to their assigned pre-flight meal. Reindeers 3 to 6 experience different intercurrent events

5) How to handle *intercurrent events*, which are important events that can happen to the reindeers in the trial after randomisation and effect either the interpretation or existence of their outcome. Such events require careful consideration, but first Santa needs to identify these events. So what events does Santa need to consider? Figure 1 shows different reindeer experiences and the types of events that happen to the reindeers after randomisation, including:

   i. Reindeer not stomaching the full bunch of carrots they were assigned as their pre-flight meal and only finishing one carrot.
   ii. Cheeky reindeer switching their allocated pre-flight meal by stealing one of Mrs Claus' freshly made mince pies and leaving the bunch of carrots they were assigned to uneaten.
   iii. Sneaky reindeer, feeling a little peckish on the journey eating snacks left out for Santa Claus.
   iv. Bold reindeer drinking sherry left out for Santa Claus.

Considering how these different intercurrent events are handled is important, since different ways of handling them can lead to quite different conclusions about the most beneficial pre-flight meal. See Table 1 for different strategies for handling intercurrent events. We note that if a reindeer were not to come back and provide

Table 1 Strategies for handling intercurrent events

Strategy	Description
Treatment policy	Use tiredness score regardless of reindeer shenanigans (intercurrent events).
Composite	The reindeer shenanigans (intercurrent events) are absorbed into the tiredness score outcome by assigning a particular tiredness score to the reindeers participating in the shenanigans (intercurrent events). This essentially re-defines the outcome.
While on treatment	Use the tiredness score before* the reindeer shenanigans (intercurrent events).
Hypothetical	Consider what the tiredness score would have been if the shenanigans had not occurred.
Principal stratum	Restrict the population of interest to the reindeers who would not participate in the shenanigans.

* Not applicable when the outcome is only measured at trial end

their tiredness score resulting in missing data (i.e. loss to follow-up), this is not an intercurrent event. Missing data often occur alongside intercurrent events but is not by itself an intercurrent event [10]. It is the intercurrent event that affects the interpretation of the outcome and not the missing data. Missing data would be an issue that the statistical elf needs to handle as part of their statistical analysis. By specifying the five attributes (1–5 above), Santa Claus will clarify precisely what treatment effect he is interested in, referred to as an 'estimand' [10]. This ensures he will get the answer to the question of interest and correctly interpret the trial results to make an informed decision.

For example, Santa could be interested in:

Estimand 1: The average difference (*summary measure*) in tiredness score (*outcome*) between eating a mince pie and a bunch of five carrots (*treatment conditions*) for flying reindeers (*population*) regardless of whether all the carrots or mince pies are eaten or cheeky reindeers switching allocations or sneaky reindeers eating other food on route or bold reindeers indulging in Santa's sherry (*intercurrent events—treatment policy strategy*).

Or he could be interested in:

Estimand 2: The average difference *(summary measure*) in tiredness score (*outcome*) between eating a mince pie and a bunch of five carrots (*treatment conditions*) for flying reindeers (*population*) if all flying reindeers adhere to the allocated pre-flight meal and do not receive any additional foods or sherry (*intercurrent events—hypothetical strategy*).

If Santa had asked his friendly statistical elf to investigate *estimand 1*, then his statistical elf could have proceeded to include all reindeer outcomes in the analysis regardless of the occurrence of any of the reindeer shenanigans (intercurrent events) and use a suitable analysis model that estimates the mean difference in outcome by allocated food allocation, for example, a linear regression model including the tiredness VAS score as outcome and the food allocation as the exposure. This aligns with a conventional intention-to-treat trial analysis including everyone randomised and all outcomes regardless of subsequent events. Santa's elf would have reported that the mean difference in tiredness score for mince pies compared to carrots was -15, demonstrating a small beneficial effect of the mince pies. But this has not accounted for reindeers not tolerating all of their carrot meal, reindeers switching their pre-flight meal from carrots to a mince pie and some reindeers consuming contraband (cookies or sherry) on route.

If Santa had instead asked his friendly statistical elf to investigate *estimand 2*, his statistical elf could have proceeded to set the tiredness VAS score outcomes to missing for all reindeer who experienced an intercurrent event and then used a missing data method to attempt to recreate what the outcomes would have looked like, had the intercurrent events not occurred, for example, a method such as multiple imputation or inverse probability weighting or likelihood-based analyses to recreate outcomes in the hypothetical setting [11, 12]. Alternatively, causal inference methods could be used [13]. For this hypothetical estimand, Santa's elf would have reported that the mean difference in tiredness score was -55. This demonstrates a much bigger effect of the mince pies when accounting for all of the reindeers' shenanigans.

In reality, the statistical elf would have found the sample size was too small (nine reindeers) to estimate these estimands reliably using the discussed methods. We provide fictitious results for the purpose of illustration and would not advise employing the discussed hypothetical estimation approaches in practice with such a small sample size.

Like optimising performance for flying reindeer, in trials of healthcare interventions multiple different questions can be investigated and the answers to these questions may lead to different conclusions on treatment benefit [5]. In such trials, participants may similarly experience different types of intercurrent events which when handled in different ways can result in different impressions of treatment benefit. In a clinical trial setting, the

Table 2 Intercurrent events for Santa's consideration and how they align to clinical trials

Reindeer's intercurrent events*	Clinical trial analogy
Reindeer not stomaching the full bunch of carrots they were assigned as their pre-flight meal and only finishing one carrot	Treatment intolerance
Cheeky reindeer switching their allocated pre-flight meal by stealing one of Mrs Claus' freshly made mince pies and leaving the bunch of carrots they were assigned to uneaten	Treatment switching
Sneaky reindeer, feeling a little peckish on the journey eating snacks left out for Santa Claus	Rescue therapy
Bold reindeer drinking sherry left out for Santa Claus	Use of contraindicated therapy

* Each of these intercurrent events would be considered treatment modifying events as per the Kahan et al. definitions [3]. Truncating events are a different type of intercurrent event that preclude the existence of the outcome, the most common being death, which we have not included in this fictitious example for reasons of merriment

above intercurrent events (i)–(iv) would be analogous to (i) treatment intolerance, (ii) treatment switching, (iii) using rescue therapy, and (iv) taking a contraindicated medication (Table 2). Many more are possible. It is therefore important to have a clear understanding of exactly what treatment effect a trial intends to investigate, i.e. what 'estimand' is of interest. Trial design, conduct and analysis can then be aligned to address this.

The use of estimands in trials was brought to the forefront by the publication of an addendum to the ICH E9 harmonised guideline on statistical principles for clinical trials on estimands and sensitivity analysis in clinical trials in November 2019 (ICH E9 (R1)) [10]. These guidelines are now adopted by regulators worldwide and require the use of estimands. However, research has identified slow uptake of the use of estimands and moreover, that without estimands, it is often unclear what precisely is being estimated in trials, potentially leading to misinterpretation or the use of inappropriate statistical methods [14]. Five years on from the publication of the ICH E9 (R1), we urge trialists to implement the estimand framework into their practice, clearly specifying their precise research question using estimands. For more details on using estimands in trials of healthcare interventions, see a primer by Kahan et al. [3]. Disclaimer: This is an entirely fictitious trial setting. Neither mince pies nor carrots may impact reindeers' performance or indeed be edible by them. But Santa enjoys sherry regardless of the impact on bold reindeers!

Acknowledgements
Not applicable.

Authors' contributions
Both SC and RP conceived of this article and contributed equally to the planning, conduct and reporting over mince pies. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding
No funding was received for this article.

Data availability
Not applicable.

Declarations

Ethics approval and consent to participate
Ethical approval or consent to participate was not required for this fictitious trial.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 18 October 2024 Accepted: 30 October 2024
Published online: 14 December 2024

References

1. World Health Organization. Ageing and health, https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, Accessed 8 Sept 2024.
2. May RL. Rudolph the red-nosed reindeer. Chicago, Illinois, United States of America: Montgomery Ward. Applewood Books, U.S.;1939.
3. Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. BMJ. 2024;384: e076316.
4. Kahan BC, Morris TP, Cro S. We must let the research question drive study methods. BMJ. 2024;384: q173.
5. Cro S, Kahan BC, Rehal S, Chis Ster A, Carpenter JR, White IR, et al. Evaluating how clear the questions being investigated in randomised trials are: systematic review of estimands. BMJ. 2022;378: e070146.
6. Cro S. Time to improve the clarity of clinical trial reports by including estimands. BMJ. 2022;378: o2108.
7. Quartagno M, Morris TP, Gilbert DC, Langley RE, Nankivell MG, Parmar MK, et al. A comparison of different population-level summary measures for randomised trials with time-to-event outcomes, with a focus on non-inferiority trials. Clin Trials. 2023;20(6):594–602.
8. Rombach I, Knight R, Peckham N, Stokes JR, Cook JA. Current practice in analysing and reporting binary outcome data—a review of randomised controlled trial reports. BMC Med. 2020;18(1):147.
9. Kahan BC, Morris TP, White IR, Tweed CD, Cro S, Dahly D, et al. Treatment estimands in clinical trials of patients hospitalised for COVID-19: ensuring trials ask the right questions. BMC Med. 2020;18(1):286.
10. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials 20 November 2019. https://www.ema.europa.eu/en/documents/scientific-guideline/iche9-r1-addendum-estimands-and-sensitivity-analysis-clinical-trials-guideline-statistical-principles-clinical-trials-step-5_en.pdf.
11. Cro S, Morris TP, Kahan BC, Cornelius VR, Carpenter JR. A four-step strategy for handling missing outcome data in randomised trials affected by a pandemic. BMC Med Res Methodol. 2020;20(1):208.
12. Van Lancker K, Tarima S, Bartlett J, Bauer M, Bharani-Dharan B, Bretz F, et al. Estimands and their estimators for clinical trials impacted by the

COVID-19 pandemic: a report from the NISS Ingram Olkin Forum Series on unplanned clinical trial disruptions. Statistics in Biopharmaceutical Research. 2023;15(1):94–111.
13. Olarte Parra C, Daniel RM, Bartlett JW. Hypothetical estimands in clinical trials: a unification of causal inference and missing data methods. Stat Biopharm Res. 2023;15(2):421–32.
14. Kahan BC, Morris TP, White IR, Carpenter J, Cro S. Estimands in published protocols of randomised trials: urgent improvement needed. Trials. 2021;22(1):686.

Publisher's Note

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