Answer Key to Suggested Activity Questions for Part 2

Reading   Olson R, Wipfli B, Thompson SV, Elliot DL, Anger WK, Bodner T, Hammer LB, Perrin NA. Weight Control Intervention for Truck Drivers: The SHIFT Randomized Controlled Trial, United States. American Journal of Public Health. 2016:e1-e9.

Questions

1. The primary criteria for choosing a research question are 1) whether the trial will address an important public health question, and 2) whether there is preliminary evidence of feasibility and efficacy for the intervention. How would you evaluate this study with respect to those two criteria?

    The trial did address an important question. As noted in the paper, "Obesity is twice as prevalent among US truck drivers compared with the general population (69% vs 31%). Regulated medical conditions associated with obesity, such as uncontrolled hypertension, may disqualify drivers from working. In addition to creating stressful precarious employment, obesity and associated sleep disorders place drivers at personally imperceptible—yet very real—increased risk of crash involvement… Obstructive sleep apnea roughly doubles drivers' crash risk. Large truck crashes, although more rare per vehicle mile traveled than those involving personal vehicles, are 20% to 55% more likely to result in a fatality. Thus, improving the well-being, health, and safety of commercial truck drivers is a public health priority."

    With regard to preliminary evidence of feasibility and efficacy for the intervention, the investigators refer to a pilot study of the SHIFT intervention (Olson R, Anger WK, Elliot DL, Wipfli B, Gray M. A new health promotion model for lone workers: results of the Safety & Health Involvement For Truckers (SHIFT) pilot study. J Occup Environ Med. 2009 Nov;51(11):1233-46. doi: 10.1097/JOM.0b013e3181c1dc7a. PubMed PMID: 19858740). In this study, they piloted the intervention with 29 drivers from 4 companies for six months. They found that the new intervention was "substantially more engaging and effective with truck drivers than previous education based tactics." As such, they did have preliminary evidence of feasibility and efficacy.

2. What was the major design employed in this trial? Was it a single factor design? A cohort or cross-sectional design?

    This was a group-randomized or cluster-randomized trial; those terms are interchangeable. Study condition was crossed with time, so there were two factors in the design. The unit of assignment was a company terminal, "defined as a company-owned facility with driver services or amenities beyond parking (e.g., drivers' lounge, laundry, maintenance)… obtained informed consent before data collection. Ultimately, 452 drivers fully enrolled at baseline (86.9% of planned sample); 275 returned at 6 months (Figure 1)." So this was a cohort design, as the same terminals and truck drivers were involved at baseline and at six months follow-up.

3. Was any effort made to ensure that the units of assignment would be balanced across study conditions with respect to potential confounding variables?   Did the team employ a priori matching or stratification, or constrained randomization?.

Yes. "Five companies participated with driver employment levels ranging from about 500 to more than 2000 drivers… transportation. At each company, we selected an even number of terminals, matched in pairs by size (number of drivers), and then we randomized 1 terminal from each pair to the intervention condition and assigned the other to the control condition." This is an example of a priori matching.

4. The four primary threats to internal validity in a GRT are selection, differential history, differential maturation, and contamination. How would you evaluate this study with regard to those four threats to internal validity?

The investigators randomized 11 terminals to each condition after pair matching on the number of drivers. A priori stratification is preferred, because it avoids certain analytic problems that can occur with matching. Matching or stratification on the primary outcome, measured at baseline, is preferred, because it usually provides the highest correlation with the primary outcome at follow-up. In this case, they did not have access to driver weight data by terminal prior to the baseline data collection and instead matched on a variable likely to be related to ease of implementing the intervention. This approach should have helped with selection. In addition, the investigators adjusted for variables measured at baseline that reflected differences between the two conditions, and that should have reduced the threat of selection.

We have no information about where the terminals were located, so it is difficult to judge whether differential history might be a problem. And we have no information on the heterogeneity of the terminals with respect to weight gain/loss by drivers, so it is difficult to judge whether differential maturation might be a problem. Randomization will help address either threat of course. Randomization does not provide any protection against contamination, but the investigators did take steps to prevent contamination. Drivers were not told their condition assignment was dependent on their terminal, and "intervention feedback and results were not posted at terminals, and were withheld form corporate leadership and control drivers until data collection was completed." These steps would help prevent contamination, as did the fact that the study lasted only 6 months, so there was not much opportunity for drivers to participate in the intervention program then leave that terminal to work at a control terminal. It would have been even better if only one terminal per company had been involved.

5. One of the major threats to the statistical validity in a GRT is low power. The most important factors are the number of groups per condition and the expectations for the intraclass correlation and variance for the primary outcome. Did the investigators describe their power or sample size calculations for this study? Did they address these three factors?

The investigators randomized 11 terminals to each study condition, for a total sample size of 22 terminals. They recruited 452 drivers and retained all of them in the analysis based on intention to treat. They said that…"On the basis of an a priori power analysis, we selected a target sample size of 520 drivers to provide a 0.99 probability of detecting a body weight effect of the magnitude observed in the pilot." There was no mention of an intraclass correlation in the paper, or any other indicator of dependency among observations taken on drivers from the same terminal. The investigators did not recruit the number of drivers they sought, but they did observe significant effects on their primary outcome, so they had adequate power for that outcome.