Clinical trials

https://doi.org/10.1038/s41591-022-02032-6

Endpoints and estimands: understanding trials of weight-loss drugs

Amanda I. Adler

Challenges that characterize weight-loss drug trials necessitate good trial design — as highlighted by the STEP 5 trial of semaglutide.

Anetlanda / Alamy Stock Photo.

To market a drug for weight loss, its manufacturer must convince the regulators (for example, the US Food and Drug Administration) of the drug's safety, effectiveness and quality. For weight loss, the 'widespread public desire' for drugs makes the regulators' role particularly important[1], so guidance strictly specifies the design of trials[2]. In line with this, recent randomized trials of pharmacological treatments for weight loss included co-primary endpoints and estimands — concepts that may be unfamiliar, and occasionally confusing, to clinicians and consumers. Both of these were incorporated in the STEP 5 trial of semaglutide for weight loss, presented by Garvey et al. in this issue of *Nature Medicine* — which showed (by using two estimands) substantial and sustained weight loss with the drug relative to weight loss with placebo[3].

The STEP 5 trial randomly assigned participants to receive weekly subcutaneous injection of semaglutide (2.4 mg) or placebo, each in addition to interventions of diet and physical activity supported by trial personnel. The trial recruited people whose body-mass index was over 30 mg kg/m^2, or at least 27 kg/m^2 with at least one weight-related comorbidity (specifically, high blood pressure, dyslipidemia, obstructive sleep apnea or cardiovascular disease). Evaluation of the use of subcutaneous semaglutide for weight loss in people with diabetes mellitus, a clear-cut weight-related comorbidity, occurred separately.

In the STEP 5 trial, investigators chose not one but two primary endpoints: percentage change in body weight from baseline to week 104 (a continuous outcome); and whether participants had lost at least 5% of their body weight (a 'yes/no' outcome). By doing this, the investigators did not give themselves 'two chances to win'; in contrast to multiple primary endpoints, co-primary endpoints must both attain statistical significance for declaration of a successful trial[4]. This regulatory requirement[2] increases the chance that positive results reflect the truth. Besides, if a trial is big enough, it may show statistically significant differences with only trivial weight loss. Including an endpoint that reflects clinically meaningful weight loss (in this case, ≥5% of body weight) improves the probability of clinical benefit. The STEP 5 investigators tested endpoints hierarchically, analyzing percent weight change first, and proceeding to the binary endpoint and then secondary endpoints until reaching a nonsignificant test result.

The STEP 5 study also included 'estimands', which are answers to carefully specified clinical questions. The 'estimand framework'[5] therefore requires that investigators explicitly state the questions a trial will answer. The framework facilitates this and expands on the popular PICO (population–intervention–comparison–outcome) framework, which is also used to frame clinical trial questions. The estimand framework includes the PICO components, plus the following: a summary measure of the statistical approach (for example, an odds ratio from logistic regression for losing at least 5% of baseline weight); and intercurrent events. Reflecting the only substantive update in 22 years to the International Council for Harmonisation E9 Statistical Principles for Clinical Trials[5,6], this update on estimands states, logically, "the design of a trial needs to be aligned to the estimands"[5].

The addendum also encourages "a thoughtful envisioning of intercurrent events"; that is, describing unambiguously how analyses will handle them. Intercurrent events are incidents that happen between randomization and the end of a trial. These might include stopping treatment, using 'rescue' treatments, switching between treatment arms, or even dying before the end of the study. Intercurrent events occur in all trials, but plague weight-loss trials in particular[2]. For example, although participants in blinded weight-loss trials are masked to treatment, those with a mirror or scale are not blind to the outcome; if they are not losing weight, they may redouble their dieting efforts or, in frustration, stop treatment. They may turn to other effective treatments, including bariatric surgery. Those who lose weight successfully may exercise more, leading to even greater, albeit non-pharmacological, weight loss[7]. Because many intercurrent events may mean certain data are unavailable, the estimand framework encourages clarity on how analyses will handle this.

Two estimands were specified in the STEP 5 study. The first, a 'treatment policy estimand', described an average effect of treatment, regardless of adherence and rescue treatments. Analyzing outcomes taking into account only the treatment assigned during randomization (akin to an 'intention to treat' analysis) reflects the experiences of participants. It is, arguably, the estimand most likely to be of use to patients and clinicians. Using available information from people

who showed up for the final weigh-in, investigators imputed what the body weight might have been at the trial's end for people who did not show up. They did this separately by treatment arm and according to whether or not participants were still on treatment; then, they used body-mass index and weight at randomization, sex, and the timing and value of the last available weight (supplementary data 1 of ref. [3]) This 'multiple imputation' approach is probably an improvement on the 'last observation carried forward' approach, which relies on the (unrealistic) assumption that after a participant stops treatment, weight does not change[8].

The second, a 'trial product estimand', addressed the average treatment effect in all randomly assigned participants, assuming they had taken the drug or placebo as intended. This 'hypothetical' estimand (supplementary data 1 of ref. [3]) would be of interest to anyone wanting to know how well semaglutide works when taken properly. For treatments that cause adverse effects and low adherence, this hypothetical estimand may not reflect the 'average' person. Fortunately, semaglutide was well tolerated, with fewer than 6% of participants discontinuing treatment.

The STEP 5 trial showed clear effectiveness by either estimand. However, the trial raises several questions. It showed that treatment works over a longer period than the 68-week duration of the similarly designed STEP 1 trial[9], but how long treatment can and should be extended is still unknown; obesity is a chronic condition, and many people are likely to require treatment for longer than 2 years. The manufacturer proposed to the National Institute of Health and Care Excellence that treatment with semaglutide stop at 2 years regardless of weight loss[10]. Unpopular with stakeholders, this submission did not align with the findings of the STEP 4 withdrawal trial, in which participants treated with semaglutide who swapped to placebo gained weight, while participants who remained on semaglutide lost weight[11]. It is also unknown how well semaglutide might work when accompanied by lifestyle interventions less intense than those offered in STEP 5, which was characterized by regular counseling, personal contact and diary review. Finally, how semaglutide will benefit people with weight lower than that of those included in STEP 5 remains unknown. Future trials to answer these questions will, one hopes, be supported by well-defined estimands.

Amanda I. Adler
Diabetes Trial Unit, Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK.
e-mail: amanda.adler@dtu.ox.ac.uk

Published online: 10 October 2022

References

1. Patel, M. R., Donahue, M., Wilson, P. W. & Califf, R. M. *Am. Heart J.* **151**, 633–642 (2006).
2. US Department of Health and Human Services Food and Drug Administration. https://www.fda.gov/media/71252/download (2007).
3. Garvey, W. et al. *Nat. Med.* https://doi.org/10.1038/s41591-022-02026-4 (2022).
4. US Department of Health and Human Services Food and Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry (2017).
5. European Medicines Agency Committee for Medicinal Products for Human Use. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf (2020).
6. European Medicines Agency. https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials (1998).
7. Wharton, S. et al. *Int. J. Obes. (Lond)* **45**, 923–933 (2021).
8. Simons-Morton, D. G., Obarzanek, E. & Cutler, J. A. *J. Am. Med. Assoc.* **295**, 826–8 (2006).
9. Wilding, J. P. H. et al. *N. Engl. J. Med.* **384**, 989–1002 (2021).
10. National Institute of Health and Care Excellence. https://www.nice.org.uk/guidance/gid-ta10765/documents/final-appraisal-determination-document (2022).
11. Rubino, D. et al. *J. Am. Med. Assoc.* **325**, 1414–1425 (2021).

Competing interests

A.I.A. is an investigator in the ISAP trial, testing oral semaglutide in people at risk for dementia.

A.I.A. is supported by the NIHR Oxford Biomedical Research Centre (BRC).