Average effects

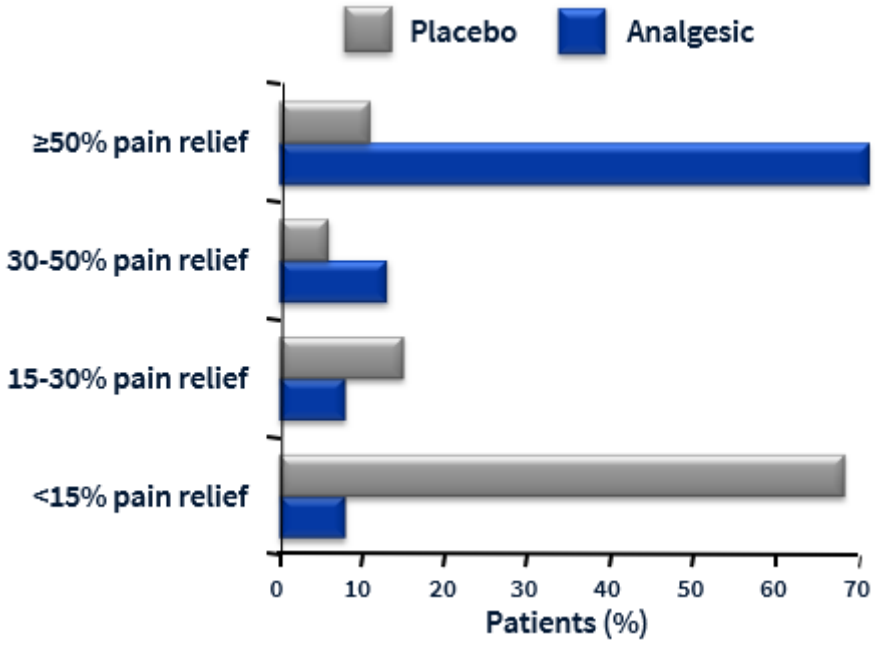
From: Key Concepts for assessing claims about treatment effects and making well-informed treatment choices (Version 2022)

2.3c Be cautious of average differences between treatments.

Explanation

Average effects do not apply to everyone. For outcomes that are assessed using scales (for example, to measure weight, or pain) the difference between the average among people in one treatment group and the average among those in a comparison group may not make it clear how many people experienced a big enough change (for example, in weight or pain) for them to notice it, or that they would regard as important. In addition, many scales are difficult to interpret and are reported in ways that make them meaningless. This includes not reporting the lower and upper 'anchor', for example, whether a scale goes from 1 to 10 or 1 to 100; whether higher numbers are good or bad; and whether someone experiencing an improvement of, say, 5 on the scale would barely notice the difference, would consider it a meaningful improvement, or would consider it a large improvement.

For example, the average difference in pain relief is not only hard to interpret, but misleading. When asked what they would consider treatment success, patients with chronic pain specify a large reduction in pain intensity, by 50% or more *[Moore 2013 (OR)]*. Most people tend to respond to painkillers (or a placebo) in two ways. Some people experience a very good pain relief (50% or more), whereas others experience very little (less than 15%). So, the average pain relief does not reflect what most people experienced in randomized trials of painkillers (analgesics) compared to placebos *[Moore 2013 (OR)]*. In the illustration below, the average difference in pain relief is about 28%. A less misleading and easier to understand way of reporting those results would be the difference between the proportion of participants in the analgesic group and the placebo group who were treated successfully (with ≥50% pain relief). In the illustration below, about 60% more participants were treated successfully with the analgesic compared to placebo.



Basis for this concept

Even if the average difference between a treatment and "no treatment" or a comparison treatment is appreciably less than the smallest change that is important to people, treatment may have an important impact on many people *[Guyatt 1998]*. For example, for some quality-of-life questionnaires, it has been shown that the smallest change that is important to people on a seven-point scale is 0.5. Even if the mean difference between a treatment and a comparison treatment is much less than 0.5, the treatment may have important impacts (change greater than 0.5) on many patients.

Outcomes assessed using scales ("continuous outcomes"), such as pain or quality of life, are easily misinterpreted and it is often difficult to make sense of them, especially when different scales are used in different studies *[Guyatt 2013b , Mayer 2019 (OR)]*.

It is possible to convert continuous outcomes to yes/no outcomes (dichotomous outcomes). This makes it easier to interpret the results, and several methods for doing this have been validated by comparing the results of these conversions and dichotomous outcomes measured in the same trials *[da Costa 2012 (SR), Meister 2015 (SR)]*. However, these methods have several limitations *[Guyatt 2013b]*. They can sometimes be misleading when different studies have used different scales, and they may underestimate or overestimate effects when the comparison group's chance of achieving an important change was ≤20% or >60%, respectively *[da Costa 2012 (SR)]*. There are several other ways of presenting the effects of treatments that have been measured using a scale, all of which have limitations *[Guyatt 2013b]*. Therefore, using more than one presentation is likely to be both informative and, if the message is similar, reassuring. It can also reduce the risk of biased selection of which presentation to use when the messages are different. If the messages are different, and it is not clear which to believe, the treatment effect is less certain.

Implications

When outcomes are assessed using scales, it cannot be assumed that every individual in the treatment comparison groups experienced the average effect. Be wary of differences on scales that are not explained or easily understood.

References

Systematic reviews

da Costa BR, Rutjes AW, Johnston BC, Reichenbach S, Nüesch E, Tonia T, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. Int J Epidemiol. 2012;41(5):1445-59. https://doi.org/10.1093/ije/dys124

Meister R, von Wolff A, Kriston L. Odds ratios of treatment response were well approximated from continuous rating scale scores for meta-analysis. J Clin Epidemiol. 2015;68(7):740-51. https://doi.org/10.1016/j.jclinepi.2015.02.006

Other reviews

Mayer M. Continuous outcome measures: conundrums and conversions contributing to clinical application. BMJ Evid Based Med. 2019;24(4):133-6. https://doi.org/10.1136/bmjebm-2018-111136

Moore RA, Straube S, Aldington D. Pain measures and cut-offs - 'no worse than mild pain' as a simple, universal outcome. Anaesthesia. 2013;68(4):400-12. https://doi.org/10.1111/anae.12148

Other references

Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. BMJ. 1998;316(7132):690-3. https://doi.org/10.1136/bmj.316.7132.690

Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol. 2013b;66(2):173-83. https://doi.org/10.1016/j.jclinepi.2012.08.001