Best Practices for Extracting Data

Overview

Several studies have documented the fact that extraction errors in systematic reviews are very common, with extraction error rates ranging from 8% to 63% (Mathes et al., 2017). Unfortunately, no universal recommendations exist regarding how to best extract data. For instance, recommendations vary as to whether data extraction should be conducted by at least two different people (Buchter et al., 2020).

Configuring Data Elements

Before extracting any data, the variables that will be collected must be defined as either continuous, dichotomous, or categorical.

- *Generally, gender is classified as a dichotomous variable (male or female) as well as patient characteristics such as hypertension or diabetes (yes or no). Age is classified as a continuous variable.*

A note about Continuous Variables

The way in which continuous variables will be extracted needs to be determined. This can be done by choosing a central tendency measure (mean or median) and a dispersion measure (standard deviation, median, or range).

There are four options that are typically reported: mean (SD), mean (range), median (range), median (IQR).

- The mean/median or the SD/range/IQR may not be present in all studies included for extraction. If this is the case, leave blank/empty. Do not enter the mean as the median or vice versa.
- Standard Error Versus Standard Deviation - a common mistake in the extraction of continuous outcome data is the use of the standard error instead of the standard deviation (Harris et al., 2019). The standard deviation can be obtained by multiplying the standard error of the mean by the square root of the sample size.

Check Reported Outcomes Carefully

- Check units of data elements to ensure that they are reported in a consistent way. Make sure that elements are being extracted at the same time point (for instance, 7-day versus 28-day mortality) and that the wording of the study is consistent with what is being collected.

- Papers will sometimes report outcomes in ways that are close but not exactly the same. For instance, while the majority of studies may report cardiovascular disease, other studies may report coronary artery disease. Because coronary artery disease is a type of coronary heart disease, it would typically be fine to extract coronary artery disease into a data element called cardiovascular disease.

Best Practices for Data Extraction: NK Software

How to perform data extraction in Nested Knowledge.

References

Büchter RB, Weise A and Pieper D. Development, testing and use of data extraction forms in systematic reviews: a review of methodological guidance. BMC Med Res Methodol 2020;20:259. https://pubmed.ncbi.nlm.nih.gov/33076832/.

Harris RG, Neale EP and Ferreira I. When poorly conducted systematic reviews and meta-analyses can mislead: a critical appraisal and update of systematic reviews and meta-analyses examining the effects of probiotics in the treatment of functional constipation in children. Am J Clin Nutr 2019;110:177-195. https://pubmed.ncbi.nlm.nih.gov/31127810/.

Mathes T, Klaßen P and Pieper D. Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. BMC Med Res Methodol 2017;17:152. https://pubmed.ncbi.nlm.nih.gov/29179685/.

— *Tiffany Yesavage 2022/01/27 00:18*

From:
https://wiki.nested-knowledge.com/ - **Nested Knowledge**

Permanent link:
https://wiki.nested-knowledge.com/doku.php?id=wiki:guide:research:extract

Last update: **2023/07/21 18:16**