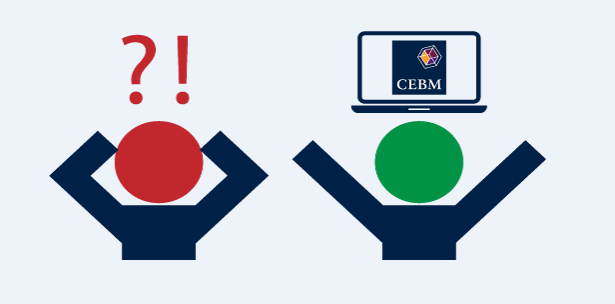
**Tip for data extraction for meta-analysis - 9**



Wanting a particular reference category in categorical risk data

Kathy Taylor

Previously, I showed a [step-by-step guide](https://bit.ly/2YdRVtV) and [worked example](https://bit.ly/2JNN3rL) of a trend estimation method for summarising categorical risk (quantile or dose-response) data, using the trend estimation method of [Greenland and Longnecker](https://www.ncbi.nlm.nih.gov/pubmed/1626547), the STATA [glst command](https://www.stata-journal.com/article.html?article=st0096) and the R [dosresmeta command](https://cran.r-project.org/web/packages/dosresmeta/index.html). In my last post I also showed that you could deal with the problem unbounded limits of categories by imputing values derived from the ranges of other categories. In this post I will look at the problem of wanting a particular reference category which may not be the category that’s reported. I will present three different examples.

**Example 1 – switching the reference category**

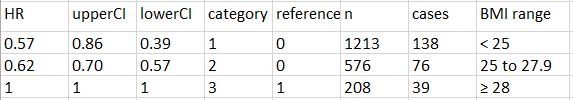
You may want to change the reference category from that with the lowest exposure to the category with the highest exposure. Looking again at the data from [one](https://www.ncbi.nlm.nih.gov/pubmed/22579085) of the studies in the worked example in my previous post (Table 1), the reference category has the lowest exposure (body mass index).

Table 1. Cumulative incidence data on body mass index and risk of atrial fibrillation



To change the reference category to that with the highest exposure we need to divide all the hazard ratios (HRs) by 1.74 (the HR of the category with the highest exposure), divide all the lower confidence interval limits by 1.16 (the lower confidence limit of the highest exposure category) and divide all the upper confidence interval limits by 2.56 (the upper confidence limit of the highest exposure category). Note that you need to swop the upper and lower limits of the confidence intervals (Table 2) because the transformed lower limit become upper limits.

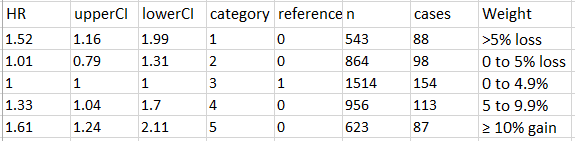
Table 2. With highest exposure as reference category



**Example 2 – separating data and switching the reference category if necessary**

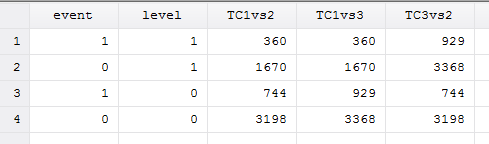
Sometimes an inner category is the reference category, as in Table 3, which shows data from a [study](https://www.ncbi.nlm.nih.gov/pubmed/24907285) of weight change and risk of atrial fibrillation. In this case, the reference category divides the categories into weight gain and weight loss. It would not be appropriate to include weight gain and weight loss data in the same meta-analysis, so these data need to be analysed separately, with the reference category featuring in both analyses. Having separated the data, the reference category may be changed, if necessary, as shown in Example 1

Table 3. Cumulative incidence data on weight change and risk of atrial fibrillation



**Example 3 – setting the reference category when deriving relative risks from event data**

In cases where categorical data are reported with rates, unadjusted estimates of relative risks (RRs) may be estimated, and as part of this process, you can chose the reference category. A [study](https://www.ncbi.nlm.nih.gov/pubmed/12114036) which featured in [Perez et al](https://www.ncbi.nlm.nih.gov/pubmed/29289762) presented rates of the first major vascular event in a trial of simvastatin verses placebo for various baseline categories including those of total cholesterol <5.0, ≥ 5.0 and <6.0, and ≥6.0 mmol/L for categories 1, 2 and 3 respectively. In the intervention group, the event rates for categories 1, 2 and 3 were 360/2030 (18%), 744/3942 (19%) and 929/4297 (22%) respectively. You can estimate RRs from these data by using a generalised linear model function (glm) in STATA and the method of [Chêne and Thompson](https://www.ncbi.nlm.nih.gov/pubmed/8797521). The data are read into STATA as shown below



Looking at the column TC1vs2 (the comparison between category 1 and category 2), the first row gives the number with events (event=1) in category 1. The second row gives the number with no event (event=0) in category 1. The next two rows give the numbers with events and without events for category 2. The reference category is indicated by level=1.

**glm event ib1.level [fweight = TC1vs2], fam(bin) link(log) nolog eform**

estimates the RR of category 2 compared to category 1 (reference) as 1.06 (0.95 to 1.19).

**glm event ib1.level [fweight = TC1vs3], fam(bin) link(log) nolog eform**

estimates the RR of category 3 compared to category 1 (reference) as 1.22 (1.09 to 1.36).

In the above commands **event** is the dependent variable and **level** is the independent variable. Frequency weights are applied using **fweight**. The outcome is binary so the family distribution is binomial, shown as **fam(bin)** and the link function between the covariate and outcome is specified as log in **link(log)**, so a log-binomial function is used. **nolog** reduces the output and **eform** exponentiates the output to produce relative risks. Level is specified as **ib1.level** as a factor variable and setting level=1 as the base or reference level.

The RRs together with the numbers of events and total patients for each category produce cumulative incidence data. Recall that I described different types of categorical data in an earlier [post](https://bit.ly/2YdRVtV).

*Here’s a tip…*

*When dealing with categorical risk data, it may be possible to switch or set the reference category*

My next blog post will focus on situations where categorical risk data are incomplete.

**Dr Kathy Taylor teaches data extraction in Meta-analysis,** [**https://www.conted.ox.ac.uk/courses/meta-analysis</link**](https://www.conted.ox.ac.uk/courses/meta-analysis%3c/link)**> This is a short course that is also available as part of our MSc in Evidence-Based Health Care**

[**https://www.conted.ox.ac.uk/about/msc-in-evidence-based-health-care**](https://www.conted.ox.ac.uk/about/msc-in-evidence-based-health-care)**, MSc in Medical Statistics**

[**https://www.conted.ox.ac.uk/about/msc-in-ebhc-medical-statistics**](https://www.conted.ox.ac.uk/about/msc-in-ebhc-medical-statistics)**, and MSc in Systematic Reviews**

[**https://www.conted.ox.ac.uk/about/msc-in-ebhc-systematic-reviews**](https://www.conted.ox.ac.uk/about/msc-in-ebhc-systematic-reviews)

**Follow updates on this blog and related news on Twitter @dataextips**