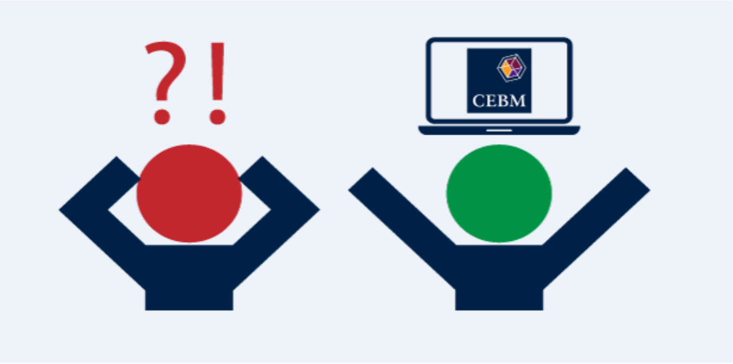
Tip for data extraction in meta-analysis – D8



Can I pool relative risk data that are given for different definitions of high vs low ?

Kathy Taylor

I'd like to thank my colleague, Dr Thomas Fanshawe, for his help in writing this blog post.

In this blog post I'm going to look at another problem that can arise in systematic reviews of prognostic studies. In post D1, I showed how to rescale hazard ratios (HRs), relative risks (RRs) and odds ratios (ORs) that express the change in risk associated with a specific change in a predictor variable on a continuous scale. I gave two examples of studies which reported HRs for cardiovascular mortality associated with increases in 24-hour systolic blood pressure variability and I rescaled the HR from the first study from an increase of 15.6mmHg (1-SD) of blood pressure variability to an increase of 5mmHg (HR 1.01, 95% CI 0.94 to 1.03), so that the data could be pooled with the HR that was reported for a 5mmHg increase by the second study (HR 1.17, 95% CI 0.64 to 2.13). Then in post D4, I showed a log-linear dose response trend estimation method that can be employed to deal with the situation of data reported for multiple categories of the exposure variable, instead of for a continuous variable. The trend estimation method uses relative risk data for all the categories.

This blog post is also about rescaling in the context of varied categorical risk data, but in this case, only two categories are considered, for 'high vs low' values of the predictor, and this many not necessarily encompass data for all categories. These relative risks could be reported for the top third vs the bottom third or the top quartile vs the bottom quartile. This also involves assuming a log-linear association between the predictor and the relative risk of the disease and a method to pooling data was used in a meta-analysis by Danesh et al.

The authors set the logRR of the risk of disease among individuals in the top third vs those in the bottom third as the desired common logRR. I will refer to this logRR as logRR3. I will also refer to the logRR for quartiles as logRR4 and so on. The authors estimated:

$$logRR3 = \frac{2.18}{2.54} \times logRR4$$

$$logRR3 = \frac{2.18}{2.80} \times logRR5$$

The authors also pooled data expressed on a continuous scale and estimated:

$$logRR3 = 2.18 \ \times logRR \ \text{ for a 1-SD difference in the predictor}$$

The standard error of logRR3 may be calculated from its confidence interval (CI) as I have shown previously in post D3 as

$$SE \ = \frac{(upper\ CI \ - \ lower\ CI)}{3.92}$$

for a 95% confidence interval where upperCI is the upper limit and lowerCI is the lower limit of the CI of loggRR3. Bear in mind that the scaling factors also apply to the CI. For example, if a study reports logRR4 with a CI, then

$$SE \ = \frac{2.18}{2.54} \times \frac{(upper\ CI \ - \ lower\ CI)}{3.92}$$

where upperCI is the upper limit and lowerCI is the lower limit of the CI of loggRR4.

Here’s a tip….

Applying another trend estimation approach can help you pool relative risks for high vs low categories

Where did the equations come from?

(You can skip this if you are only interested in carrying out the calculations)

For analysis the values of the predictor need to be normally distributed and this may require the use of the transform. The authors log-transformed C-relative protein values.

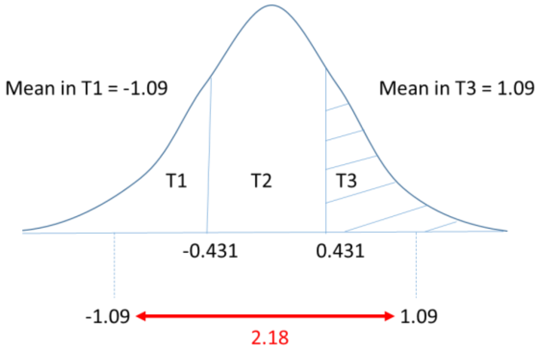
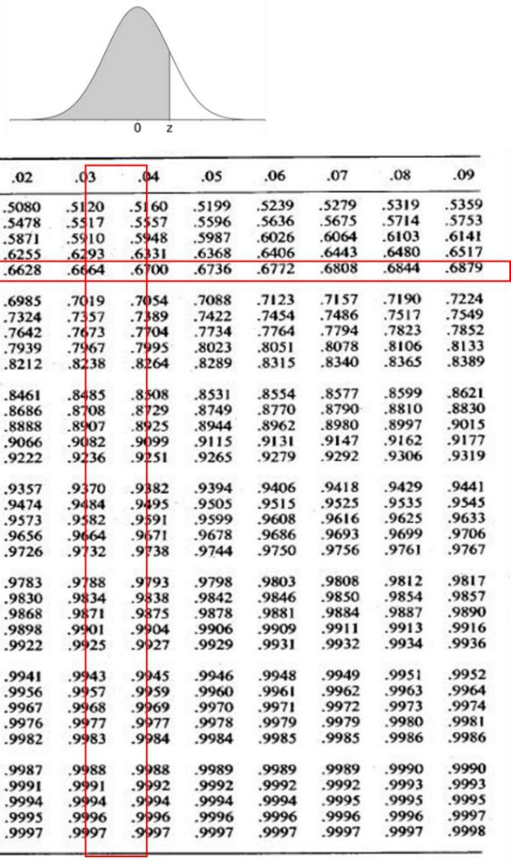


Figure 1. Tertiles

The authors considered tertiles on a standardised normal distribution (with mean=0 and SD=1, Figure 1). The Z value taken from the standard normal distribution table for the area in the upper tail equal to 1/3 is between 0.43 and 0.44 (Figure 2, shaded area is 1-1/3=2/3=0.6667).



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Figure 2. Table of proportions of the standard Normal distribution

A more precise value of the Z value is 0.431 and this can be obtained in R using `qnorm(2/3)`.

The average values of the upper and lower tertiles is obtained by simulating a normal distribution and calculating basic description statistics, including the average, of the area beyond 0.431. The relevant R code is

```
x<-rnorm(1000000)
summary(x[x>qnorm(2/3)])
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.4307  0.6736  0.9663  1.0902  1.3836  4.9406
```

The difference between the average predictor in the lower and upper tertiles is 2 x 1.0902 = 2.18 (Figure 1). This represents the standardised predictor change for tertiles.

A linear relationship between X and Y can be expressed as $Y = A + BX$ where A and B are constants; A is the intercept and B is the slope. Assuming a log linear association between the risk of disease and the baseline value of a predictor variable over the midrange of baseline values (we assume a zero intercept), we can say that

$$logrisk = A + \ B \times predictor$$
$$logrisk\ in\ T3 = A + \ B \times mean\ predictor\ value\ of\ T3$$
$$logrisk\ in\ T1 = A + \ B \times mean\ predictor\ value\ of\ T1$$

$logrisk\ in\ T3 - \log risk\ in\ T1 \ = B \times predictor\ change$ (since logA-logB=logA/B) i.e.

$logRR3 = \ B \times predictor\ change$ **(equation 1)**

$logRR3 \ = \ B \ \times \ 2.18 \Longrightarrow B = \frac{logRR3}{2.18}$ **(equation 2 and Figure 1)**

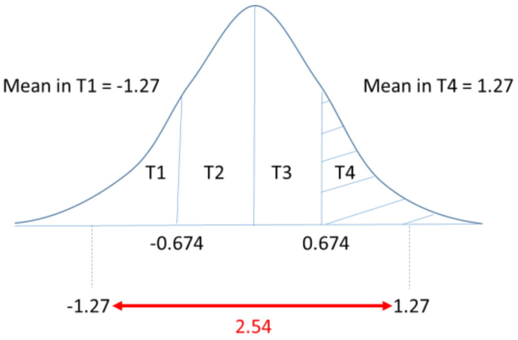


Figure 3. Quartiles

Considering quartiles (Figure 3), the Z score for the area in the upper tail equal to 1/4 is 0.674 and the average value of the upper tertile is 1.27. The difference between the average predictor in the lower and upper tertiles is therefore 2.54. These values are calculated in R using

```
qnorm(3/4)
summary(x[x>qnorm(3/4)])
```

Using equation 1 for quartiles and substituting for B using equation 2

$$logRR4 \ = \ B \ \times 2.54 = \frac{2.54}{2.18} \times logRR3$$

Rearranging

$$logRR3 \ = \frac{2.18}{2.54} \times logRR4$$

Similarly, considering quintiles, the Z score for the area in the upper tail equal to 1/5 is 0.841 and difference between the average value in the upper quintile is 1.40 x 2 = 2.80. These are calculated in R by

```
qnorm(4/5)
summary(x[x>qnorm(4/5)])
```

$$logRR5 \ = \ B \ \times 2.80 = \frac{2.80}{2.18} \times logRR3$$

Rearranging

$$logRR3 \ = \frac{2.18}{2.80} \times logRR3$$

Recall that the Z score is calculated

$$Z = \frac{x - \mu}{SD}$$

and so the Z score is measured in terms of SDs from the mean. For the standard normal distribution the SD is equal to 1. Therefore, the change in 2.18 units is in fact 2.18 SDs and a change in 2.80 units is 2.80 SDs. Using equation 1 for a change in 1-SD quartiles and substituting for B using equation 2

$$logRR \ for \ a \ 1SD \ difference = \ B \times 1 = B = \frac{logRR3}{2.18}$$

Rearranging

$$logRR3 = 2.18 \ \times logRR \ \text{ for a 1-SD difference}$$

Dr Kathy Taylor teaches data extraction in Meta-analysis, https://www.conted.ox.ac.uk/courses/meta-analysis</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, MSc in 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

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