CHAPTER 17

Prediction Intervals

INTRODUCTION

When we report the results of a meta-analysis we often focus on the summary effect size and its confidence interval. These give us an estimate of the mean effect size and its precision, *but they say nothing about how the true effects* are distributed about the summary effect.

In a fixed-effect analysis this is appropriate, since we assume that the true effect is the same in all studies. In a random-effects analysis, however, we need to consider not only the mean effect size, but also how the true effects are distributed about this mean. A mean effect size (say, a standardized mean difference) of 0.50 where all true effects are clustered in the range of 0.40 to 0.60 may have very different implications than the same mean effect where the true effects are scattered over the range of 0.00 to 1.00.

Our goal in this chapter is to show how we can use a prediction interval to describe the distribution of true effect sizes. We will review how the prediction interval is used in primary studies, and then show how the same mechanism can be used for meta-analysis.

PREDICTION INTERVALS IN PRIMARY STUDIES

Suppose we are interested in math scores for a population of children. We want to create a prediction interval, defined as the interval within which a new student's score would fall if that student were selected at random from this population. The 80% prediction interval would include that score 80% of the time, the 95% interval would

Introduction to Meta-Analysis, Second Edition.
Michael Borenstein, Larry V. Hedges, Julian P.T. Higgins, and Hannah R. Rothstein.

include that score 95% of the time, and so on. As such, the interval yields an intuitive picture of the distribution of scores.

If we somehow knew the population mean (μ) and standard deviation (σ), and were willing to assume that the scores are normally distributed, we could create a prediction interval, using

$$LL_{pred} = \mu - Z^{\alpha}\sqrt{\sigma^2} \tag{17.1}$$

and

$$UL_{pred} = \mu + Z^{\alpha}\sqrt{\sigma^2}, \tag{17.2}$$

where Z^{α} is the Z-value corresponding to the desired confidence level (for the 95% interval, Z^{α} would be 1.96). For example, if μ is 0.50 and σ is 0.10, then the lower and upper limits of a 95% prediction interval are

$$LL_{pred} = 0.500 - 1.96 \times 0.100 = 0.3040$$

and

$$UL_{pred} = 0.500 + 1.96 \times 0.100 = 0.6960.$$

Formulas (17.1) and (17.2) are intuitive but are not useful in practice because they assume that we know both μ and σ exactly. When these values are estimated from the sample (as they almost always are) we instead use the formulas

$$LL_{pred} = \overline{X} - t^{\alpha}_{df}\sqrt{S^2 + \frac{S^2}{n}} \tag{17.3}$$

and

$$UL_{pred} = \overline{X} + t^{\alpha}_{df}\sqrt{S^2 + \frac{S^2}{n}}, \tag{17.4}$$

where $\overline{X}$ is the sample mean, t^{α}_{df} is the t-value corresponding to (for example if $\alpha = 0.05$) the 95% interval when there are df degrees of freedom, and S is the standard deviation of scores in the sample. These formulas have the same structure as (17.1) and (17.2) but to allow for error in the estimates of μ and σ they incorporate the following changes. First, we multiply by t rather than Z. Second, t is multiplied by a quantity that involves both the variance of the observations (the standard deviation squared, or S^2) and also the variance of the mean (the standard error squared, or S^2/n).

For example, if $X = 0.50$, $S = 0.10$ and $n = 30$, then

$$LL_{pred} = 0.5000 - 2.0452 \times \sqrt{0.1000^2 + \frac{0.1000^2}{30}} = 0.2921$$

and

$$UL_{pred} = 0.5000 + 2.0452 \times \sqrt{0.1000^2 + \frac{0.1000^2}{30}} = 0.7079,$$

where a t-value of 2.045 corresponds to the t-value for alpha of 0.05 with 29 df. In Excel, the function TINV(0.05,29) returns 2.0452.

Note that the prediction intervals based on the statistics (0.292 to 0.708) are wider than those based on the parameters (0.304 to 0.696).

PREDICTION INTERVALS IN META-ANALYSIS

We can follow a similar approach in meta-analysis. If we somehow knew the mean effect size (μ) and the standard deviation of true effect sizes (τ), and were willing to assume that the effect sizes are normally distributed, we could create a prediction interval using

$$LL_{pred} = \mu - Z^{\alpha}\sqrt{\tau^2} \tag{17.5}$$

and

$$UL_{pred} = \mu + Z^{\alpha}\sqrt{\tau^2}. \tag{17.6}$$

This is similar to (17.1) and (17.2) except that τ^2, the variance of the true effects in a meta-analysis, has replaced σ^2, the variance of the scores in a primary study.

Actually, we introduced this idea in Chapter 16 when we discussed the interpretation of T as an estimate of the standard deviation of true effect sizes. For example, if μ is 0.358 and τ^2 is 0.0373, then the 95% prediction interval is

$$LL_{pred} = 0.358 - 1.96 \times \sqrt{0.0373} = -0.0205$$

and

$$UL_{pred} = 0.358 + 1.96 \times \sqrt{0.0373} = 0.7365.$$

In a forest plot we would typically use a simple line (from –0.020 to 0.737) to represent the prediction interval, but in Figure 17.1 we use a bell curve to convey the idea that the true effect sizes are expected to be normally distributed within this range. Note that the bell curve has been truncated at either end (–0.020 and 0.737) so that it covers 95% of expected true effects.

Formulas (17.5) and (17.6) assume that we actually know the values of μ and τ, and make no allowance for error in these estimates. Higgins *et al*. propose the following

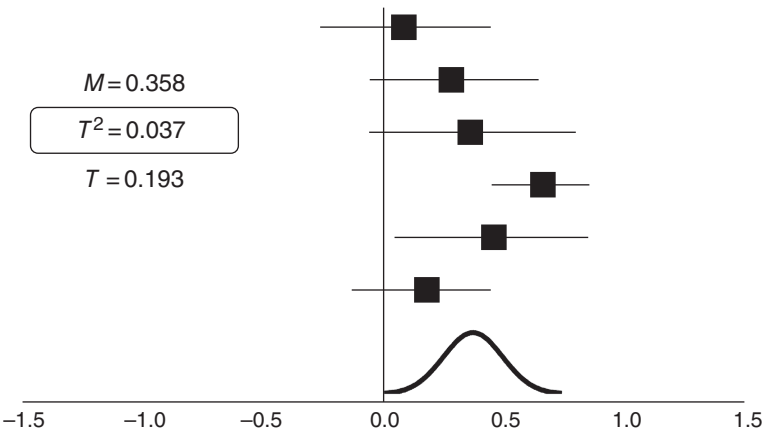

Figure 17.1 Prediction interval based on population parameters μ and τ^2.

formulas for computing a prediction interval when these values are estimated from the sample. The formulas are

$$LL_{pred} = M^* - t^{\alpha}_{df}\sqrt{T^2 + V_{M^*}} \tag{17.7}$$

and

$$UL_{pred} = M^* + t^{\alpha}_{df}\sqrt{T^2 + V_{M^*}}, \tag{17.8}$$

where M^* is the mean effect size in the sample, T^2 is the sample estimate of the variance of true effect sizes, and V_{M^*} is the variance of M^*. The factor t is the t-value corresponding to (for example if $\alpha = 0.05$) the 95% interval when there are df degrees of freedom.

These formulas have the same structure as (17.5) and (17.6) but we multiply by the t-value (rather than the Z-value) and apply this factor to a quantity that involves both the variance of the *true effects* (T^2) and the variance of the *mean effect* (V_{M^*}). The degrees of freedom *(df)* is often taken as the number of studies minus 2 (that is, $k - 2$).

For example, if $k = 6$, $M^* = 0.3582$, $T^2 = 0.0373$, and $V_{M^*} = 0.0111$, then

$$LL_{pred} = 0.3582 - 2.7764 \times \sqrt{0.0373 + 0.0111} = -0.2525$$

and

$$UL_{pred} = 0.3582 + 2.7764 \times \sqrt{0.0373 + 0.0111} = 0.9690.$$

The value 2.7764 is the t-value corresponding to alpha of 0.05 with 4 *df*. In Excel, the function = TINV(0.05,4) returns 2.7764.

Figure 17.2 is identical to Figure 17.1 except that this time the prediction interval is based on the sample values M^* and T^2 rather than the population parameters μ, and τ^2. Note that the bell curve is wider in Figure 17.2 (the 95% interval is –0.25 to + 0.97) than in Figure 17.1 (where the interval was –0.02 to + 0.74) which reflects the uncertainty in the estimates.

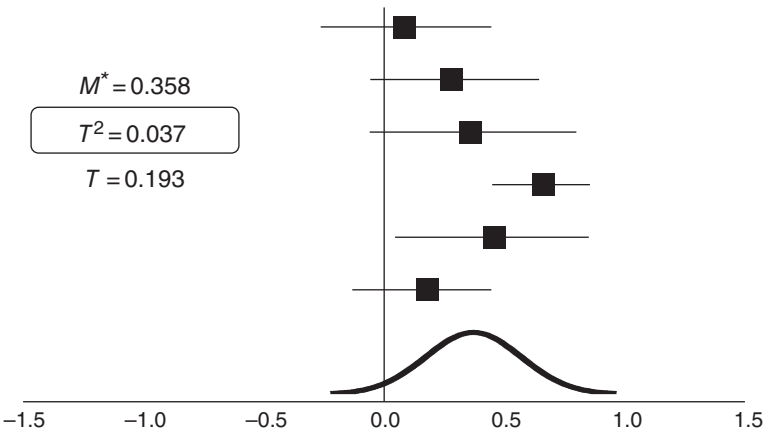

Figure 17.2 Prediction interval based on sample estimates M^* and T^2.

CONFIDENCE INTERVALS AND PREDICTION INTERVALS

Traditionally, the summary line in a forest plot uses a diamond to depict the mean effect size (the center of the diamond) and its confidence interval (the width of the diamond). Now, we want to add a visual indicator of the prediction interval, and we do so by adding a horizontal line to either end of the diamond, as in Figure 17.3.

The meta-analysis line in the plot now shows *two distinct items of information*. First, in 95% of cases the mean effect size falls inside the diamond. Second, in 95% of cases the true effect in a new study will fall inside the horizontal lines. It is important to understand that these two items address two distinct issues. The confidence interval quantifies *the accuracy of the mean*, while the prediction interval addresses the actual *dispersion of effect sizes*, and the two measures are not interchangeable.

As always, how we choose to interpret the effects depends on our goals. We may want to focus on the null effect. If the full diamond exceeds zero then we are 95% certain that the mean effect size exceeds zero. If the full prediction interval exceeds zero then the true effect in 95% of new studies will exceed zero.

Or, we may want to focus on a clinically important effect (say, a standardized mean difference of 0.20). If the full diamond exceeds 0.20 then we are 95% certain that the mean effect size exceeds 0.20. If the full prediction interval exceeds 0.20 then the true effect in 95% of new studies will exceed 0.20.

COMPARING THE CONFIDENCE INTERVAL WITH THE PREDICTION INTERVAL

Earlier, we showed the computation of the prediction interval for a meta-analysis with six studies. Suppose that the meta-analysis included more studies (24, 60, or 1002) with the same pattern as in the first six. In other words, we have the same within-study

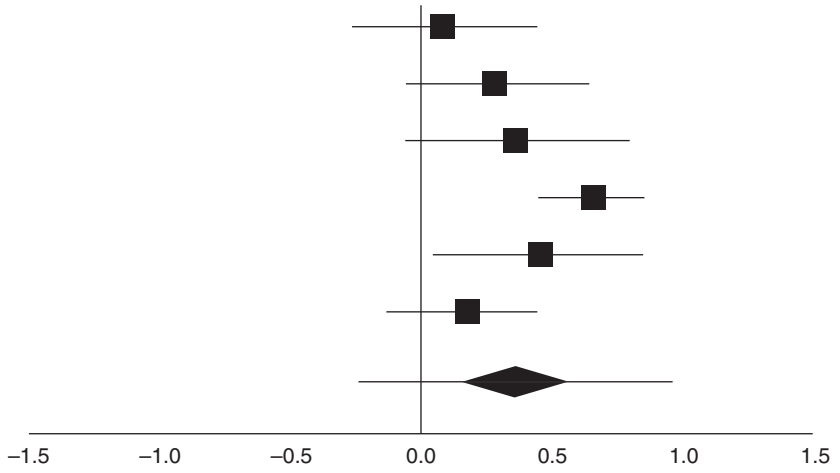

Figure 17.3 Simultaneous display of confidence interval and prediction interval.

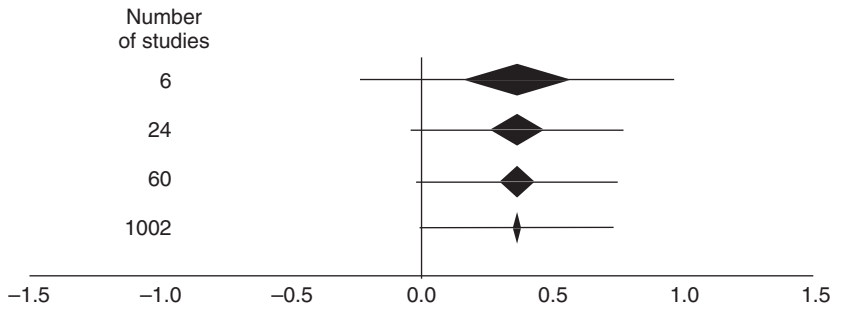

Figure 17.4 Impact of number of studies on confidence interval and prediction interval.

error and the same pattern of dispersion but a more precise estimate of the mean effect size and of the true between-studies dispersion.

In Figure 17.4 we illustrate what the confidence interval and the prediction interval would be for these four hypothetical analyses. While the specific pattern shown here is unique to this analysis, the general trend will apply to any analysis.

With six studies the confidence interval (the diamond) is quite wide, but with 60 studies its width is cut by about half, and with 1002 studies its width is trivial. This follows from the formula for a confidence interval, which is

$$CI_{M^*} = M^* \pm Z\sqrt{V_{M^*}}. \tag{17.9}$$

The confidence interval reflects only error (V_{M^*}), and so we see a consistent decline in the confidence interval width as the number of studies goes from 6 to 1002. With an infinite number of studies the error would approach zero, and so the width of the confidence interval would approach zero.

By contrast, the width of the prediction interval (the line) drops sharply as the number of studies goes from 6 to 24 but shows almost no change beyond that point. This follows from the formula for a prediction interval, which is

$$PI = M^* \pm t\sqrt{T^2 + V_{M^*}}. \tag{17.10}$$

The interval is based on error in estimating the mean (V_{M^*}), which is dependent on the number of studies. The interval is based also on the variance of the studies, T^2, which is not affected by the number of studies. In this example, as the number of studies increases from 6 to 24, V_{M^*} decreases and therefore the interval narrows. Beyond that point the decrease in V_{M^*} is trivial (and T^2 remains constant), so the prediction interval shows little change. With an infinite number of studies, the interval would approach μ plus/minus 1.96 τ.

SUMMARY POINTS

- For a random-effects analysis we want to know both the mean effect size and also how the true effects are distributed about the mean.
- The precision of the mean is addressed by the confidence interval. Since the confidence interval reflects only error of estimation of the mean, with an infinite number of studies its width would approach zero.
- The distribution of true effect sizes is addressed by the prediction interval. Since the prediction interval incorporates true dispersion as well as error, with an infinite number of studies it will approach the actual dispersion of true effect sizes.
- The summary effect in a forest plot has traditionally been represented by a diamond which corresponds to the confidence interval. For random-effects analyses we can modify this to display both the confidence interval and the prediction interval.

Further Reading

Borenstein, M., Higgins J.P.T., Hedges, L.V., and Rothstein, H.R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods* 8 (1): 5–18. https://doi.org/10.1002/jrsm.1230.

Borenstein, M. (2019). *Common Mistakes in Meta-Analysis and How to Avoid Them*. Englewood, NJ: Biostat, Inc.

Borenstein, M. (2020). Research Note: In a meta-analysis, the I^2 index does not tell us how much the effect size varies across studies. *Journal of Physiotherapy* 66 (2): 135–139. https://doi.org/10.1016/j.jphys.2020.02.011.