Application of Prediction Intervals
in Meta-Analyses with Random Effects

Joint Statement from IQWiG, GMDS and IBS-DR

Authors: Ralf Bender, Oliver Kuß, Armin Koch, Carsten Schwenke & Dieter Hauschke

Version 4.1 of the General Methods of the Institute for Quality and Efficiency in Health Care (IQWiG) was published on 28 November 2013 (IQWiG, 2013). The Act on the Reform of the Market for Medicinal Products (Arzneimittelmarktneuordungsgesetz, AMNOG) became effective in January 2011 (Deutscher Bundestag, 2010). In addition to further details on the implementation of new requirements specified by AMNOG, the new methods version contains, among other things, explanations on the application and interpretation of prediction intervals in meta-analyses with random effects. As the application of this method is relatively new in the area of systematic reviews, to achieve acceptance of those decisions by the Federal Joint Committee (G-BA) that are based on IQWiG reports, it seems desirable to reach a scientific consensus on how prediction intervals in meta-analyses with random effects can be adequately applied and interpreted in benefit assessments.

The application of prediction intervals in meta-analyses is based on Skipka (2006), as well as on Higgins, Thompson & Spiegelhalter (2009). This method has been gaining importance in practical applications since the publications by Riley, Higgins & Deeks (2011) in the *British Medical Journal* and Guddat et al. (2012) in *Systematic Reviews.*

In meta-analyses with random effects, prediction intervals quantify the extent of existing heterogeneity. In contrast to a confidence interval, which quantifies the precision of an estimated effect, a $(1-\alpha)$ prediction interval covers the true effect of a single (new) study with probability $1-\alpha$. As with confidence intervals, the level of $1-\alpha = 0.95$ is generally used. It should be noticed that the estimated intervention effect can be statistically significant in cases where the prediction interval contains the value of the zero effect. Some of the literature even suggests drawing robust conclusions on the benefit or harm of interventions from meta-analyses only by means of prediction intervals (see, e.g., Graham & Moran (2012)). However, the German Society for Medical Informatics, Biometry and Epidemiology (GMDS), the International Biometric Society – German Region (IBS-DR) and IQWiG agree that the confidence interval and the corresponding *p*-value, and not the predication interval, are the decisive measures for the assessment of the statistical significance of an effect.

As prediction intervals describe the extent of the existing heterogeneity, it is particularly meaningful to calculate and present prediction intervals if, due to marked heterogeneity, no interpretable overall effect can be estimated and displayed. In these heterogeneous situations, inferring benefit conclusions with regard to medical interventions is more difficult than in a homogeneous situation with an available estimated overall effect. In heterogeneous situations, the prediction interval is a valuable aid in the evaluation of whether and, if yes, how clearly, the individual (heterogeneous) study effects point in the same direction. This in turn enables a transparent operationalization for inferring conclusions on benefit regarding medical interventions, with graded certainty of conclusions

in the case of heterogeneous data. For instance, in the case of studies with a corresponding certainty of results, proof of an effect is present, if the prediction interval does not cover the value of the zero effect, even if no overall effect can be meaningfully provided due to marked heterogeneity. IQWiG's precise approach is described in Version 4.1 of the General Methods (IQWiG, 2013).

When applying prediction intervals, the practical problem exists that the standard method for calculating prediction intervals according to Higgins, Thompson & Spiegelhalter (2009) is based on the *t*-distribution, whereas the standard method for interval estimation of the overall effect according to DerSimonian & Laird (1986) is based on the normal distribution. This use of different test distributions can lead to a disproportionately large difference in the width between the confidence and prediction interval in a homogeneous meta-analysis. In future, at least within the Cochrane Collaboration, it is planned to change the standard method for the interval estimation of the overall effect in meta-analyses with random effects, by using the *t*-distribution according to Knapp & Hartung (2003) as a basis for statistical inference. GMDS, IBS-DR and IQWIG agree that the use of a consistent methodology is meaningful for calculating confidence and prediction intervals in meta-analyses with random effects.

GMDS, IBS-DR and IQWiG agree that the application of prediction intervals is a valuable supplement to the present methods for meta-analyses with random effects, in particular in cases where, due to marked heterogeneity, it is not possible to present an interpretable overall effect. GMDS, IBS-DR and IQWiG support international efforts to establish a consistent methodology for calculating confidence and prediction intervals in meta-analyses with random effects.

Cologne, 7 March 2014

References

[1] DerSimonian, R. & Laird, N. (1986): Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177-188.

[2] Deutscher Bundestag (2010): Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung [Act on the Reform of the Market for Medicinal Products in Statutory Health Insurance] (Arzneimittelmarktneuordnungsgesetz – AMNOG). *Bundesgesetzblatt* Teil 1(67), 2262-2277.

[3] Graham, P.L. & Moran, J.L. (2012): Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *J. Clin. Epidemiol.* **65**, 503-510.

[4] Guddat, C., Grouven, U., Bender, R. & Skipka, G. (2012): A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Syst. Rev.* **1**, 34.

[5] Higgins, J.P.T., Thompson, S.G. & Spiegelhalter, D.J. (2009): A re-evaluation of random-effects meta-analysis. *J. R. Stat. Soc. A* **172**, 137-159.

[6] IQWiG (2013): *General Methods, Version 4.1 of 28 November 2013*. Institute for Quality and Efficiency in Health Care (IQWiG), Cologne.

[7] Knapp, G. & Hartung, J. (2003): Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693-2710.

[8] Riley, R.D., Higgins, J.P.T. & Deeks, J.J. (2011): Interpretation of random effects meta-analyses. *BMJ* **342**, d549.

[9] Skipka, G. (2006): The inclusion of the estimated inter-study variation into forest plots for random effects meta-analysis – a suggestion for a graphical representation. *XIV Cochrane Colloquium; 23-26 October 2006, Dublin, Ireland; Program and Abstract Book*, p. 134.

The English version is a translation of the German original document of 7 March 2014.
In cases of doubt the German original document applies:
http://www.gmds.de/pdf/publikationen/stellungnahmen/140307_Praediktionsintervalle.pdf