

Accounting for baseline differences in meta-analysis

Anna Chaimani

Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece; annachaimani@gmail.com



INTRODUCTION

Substantial clinical differences between the studies included in a meta-analysis may compromise the overall applicability of the summary estimate.¹ Several approaches have been suggested to account for heterogeneity, such as conducting a random effects meta-analysis and computing predictive intervals, and include several study-level characteristics as predictors of the observed effect size.²⁻⁴ If individual participant data are available, the investigation of associations between outcome and patient-level characteristics is straightforward by applying regression techniques. However, at the standard pair-wise meta-analysis level, usually only trial-level results are given. This does not always allow researchers to adequately explore the association between predictors and effect sizes, as it is possible that population characteristics do not vary consistently within and/or across trials.

One potentially important source of heterogeneity in a meta-analysis is the baseline severity of study participants with respect to the disease under investigation.⁵ This is evident in mental health trials, for instance, antidepressants appeared to be more effective when administered to severely depressed populations rather than to patients with mild depression.⁶⁻¹²

The possible association between baseline severity and relative effects can be explored via meta-regression, a tool used in meta-analysis to explore the impact of moderator variables or predictors on the study effect size.¹³ This is straightforward when the required data are available (eg, by using the average score of the participants on a rating scale at the beginning of each study as a predictor for the relative effects).

However, particularly for dichotomous outcomes, we usually do not have this type of information and we need to express baseline severity using surrogate predictors. In practice, the underlying or control group risk (defined as the probability for a success in the control arm) is often used as a proxy for a collection of unavailable population and setting characteristics, which might affect the relative treatment effect.¹⁴⁻¹⁵ Several meta-regression models have been presented that control the estimated summary effect for the observed variability in underlying risk.¹⁴⁻¹⁹ These models differ with respect to the assumption they make on how underlying risks are related across studies. A common problem when investigating the association between the underlying risk and the effect size is that the former is involved in the estimation of the latter and therefore they are inherently inter-related. This problem is also known as the 'regression towards the mean' phenomenon.^{5, 20} In addition, the relationship between underlying risk and relative effects may be dependent on the follow-up of the trials or the effect size we use in the meta-analysis.³ Thus, the Cochrane Handbook does not recommend the use of effect sizes that assume a strong association between the outcome and underlying risk.³

In this paper, we offer a brief overview of the available meta-regression models and the possible assumptions that can be employed to account for differences in baseline characteristics between the studies of a meta-analysis, and we use two examples from the mental health literature to present and discuss the findings from alternative models.

METHODS

We employed meta-regression models to investigate how changes in baseline severity or the underlying risk impact on the relative effect,

using aggregated data for the trial effect as well as for the predictor (eg, baseline severity).

Accounting for differences in baseline severity

Mental health trials very often evaluate the effectiveness of drugs using standardised rating scales that measure the severity of symptoms with respect to the clinical condition under consideration. It might be the case that the extent of improvement of symptoms (eg, reduction in the mean endpoint score) is partly dependent on the initial (baseline) score of the participants. If this dependence is not consistent in direction and magnitude between control and experimental group, it would affect the relative effectiveness of the treatments.

Systematic reviews should set narrow enough inclusion criteria for study populations to preserve important discrepancies in baseline severity across the eligible studies, when this characteristic is a potential 'effect modifier'.³ However, researchers sometimes decide to include a wider population in their review to expand the applicability (external validity) of their findings as well as to explore factors that may impact the results.

In that case the average value of the participants' baseline scores from each study may be used as a predictor in a meta-regression model with the relative effect (eg, mean difference) as the dependent variable. The outcome of such a model would be a constant and a slope (regression coefficient) parameter. The regression coefficient would give information on how much the relative effect is on average increased or decreased for a specific change in the mean baseline score of a study. The constant would represent the estimated summary effect when the explanatory variable equals zero; hence this estimate would correspond to studies with zero average baseline score. However, observing zero baseline severity is not a realistic scenario and we should not make inferences for such extreme values. Thus, it is common to subtract an observed value (usually the mean, minimum or maximum score across studies) from the average score of each trial and use these differences as explanatory variable. Then, the constant of the meta-regression model would represent the summary effect that corresponds to studies with that specific observed value (eg, the mean baseline severity from all studies).

Accounting for differences in underlying risk

In the absence of information on the baseline severity of participants, researchers may consider using the underlying risk as a proxy. This characteristic is more likely to be reported in trials with dichotomous outcomes; for example, mental health trials often measure the number of responders and non-responders in each study arm by defining a minimum score reduction in a rating scale as the threshold for response (ie, more than 50% reduction between baseline and end point). In this case the explanatory variable representing underlying risk in the meta-regression model would be the number of responders over the total number of randomised participants in the control arm of each study.

An advantage of using underlying risk as explanatory variable is that it may reflect several population characteristics on the top of baseline severity for which information might not be available. Nevertheless, an important limitation of this approach is the inherent mathematical relationship between underlying risk and relative effects (the regression towards the mean), which may lead to false-positive significant

associations.^{15 19 21} To overcome this problem, transformations of the study's underlying risks have been suggested. A probably better approach is the use of a statistical (hierarchical) model in which the true rather than the observed underlying risk is associated to the true relative effect.¹⁸

Several different assumptions have been suggested in the literature to model the relationship of underlying risks between the different studies:^{14 17–19 22 23}

1. The underlying risks across the different studies of a meta-analysis are unrelated;¹⁴
2. The study-specific underlying risks share a common normal distribution.^{18 19 22} This entails the contention that underlying risks are different yet related in the sense that they have a mean value, which is the most probable value we may observe in a trial, and the probability of observing a larger or smaller value thins out symmetrically as we move away from the mean value;
3. The underlying risks come from two different normal distributions.¹⁷ Under this assumption we allow the underlying risks to be much different than when using the previous scenario (ie, a single common distribution for all studies), although again we impose some level of dependence among them.

An important advantage of the first modelling assumption is that it can be easily fitted in any software. In addition, this approach possibly better 'preserves the benefits of randomisation',²⁴ whereas the dependence of underlying risks between different studies (approaches 2 and 3) is a strong and potentially unrealistic assumption.¹⁶ On the other hand, assuming a common distribution for the underlying risks can occasionally lead to more precise and less biased estimates of the study's underlying risks.¹⁶

Other approaches have been suggested (including the use of asymmetrical distributions or distributions with heavier tails across the underlying risks^{16 23}), however, the appropriateness of each assumption is a matter of debate. Most importantly, the clinical insight on the comparability of underlying risk across trials is crucial to adequately inform the choice of a reasonable model.

RESULTS

We applied the meta-regression models discussed in this paper on two datasets of antidepressant trials published in the scientific literature by Kirsch *et al*¹⁰ and Undurraga and Baldessarini.⁶ The outcome in both meta-analyses was improvement of symptoms for depression measured as a continuous outcome. According to the original publications, in all analyses we assumed that the different antidepressants were equivalent.

The first example consists of 35 trials comparing four active drugs (fluoxetine, venlafaxine, nefazodone and paroxetine) against placebo with available data on the initial (baseline) score of the participants and the change score from baseline at the end of the follow-up. Using random effects meta-analysis the estimated summary standardised mean difference (SMD) was 0.32 (95% CI: 0.24 to 0.41), which suggested a beneficial effect of antidepressants over placebo. The

presence of low-to-moderate heterogeneity was implied by the I^2 (43% with 95% CI: 15% to 62%) and the heterogeneity SD (τ) was 0.16.

The second data set originally involved 124 trials providing information on the number of responders (as dichotomous outcome) in each study arm. For simplicity, in our example we restricted the data to studies that evaluated the relative effectiveness of the same treatments considered by Kirsch and colleagues. Therefore, we included 38 trials that compared three active drugs (fluoxetine, venlafaxine and paroxetine) with placebo. The summary OR estimated from this subset of studies (under the random-effects model) was 2.05 (95% CI: 1.81 to 2.32), in favour of antidepressants over placebo. The estimated heterogeneity was $\tau=0.22$ with an I^2 equal to 36% (95% CI: 4% to 57%).

THE IMPACT OF BASELINE SEVERITY

We explored the possible relationship between the (observed) initial severity score of the participants and the relative effectiveness of antidepressants in the Kirsch data by performing meta-regression. To obtain a meaningful interpretation of the estimated constant value, we used the difference between each study's average baseline score and the mean (baseline) score from all studies (mean=25.54) as predictor. This model gave a statistically significant coefficient for the effect of baseline severity (table 1) suggesting that if two studies have a 10-unit difference in participant's baseline score, then the SMD for the study with a more severe population would be on average 0.5 units larger; it is interesting that this change equals the criterion for clinical significance.²⁵ Also, accounting for differences in baseline severity across studies explained 25% of the estimated heterogeneity.

Figure 1 shows that this apparent strong association might be largely influenced by one trial²⁶ in which participants had a considerably smaller baseline score compared to the other studies. However, after excluding the Dunlop study and again running the meta-regression model the results did not materially change (table 1).

The impact of underlying risk

To account for differences in underlying risk across trials in the Undurraga dataset, we first performed the meta-regression analysis using the observed risk for response in placebo arm as predictor (after subtracting the average risk from all studies). This analysis yielded a marginally statistically non-significant coefficient, which implied that a 10% increase in the underlying risk of the participants reduces the logarithm of the OR (lnOR) on average by 0.12 (table 2). In addition, the variability in underlying risk across studies explained 27% of the heterogeneity (τ) that was estimated from the standard random effects meta-analysis (without any predictor).

Similarly to the previous example, a graphical depiction of the studies (figure 2) shows that the trial by Cohn and Wilcox²⁷ is a potential outlier observation. After excluding the Cohn study from the meta-regression model the results were not modified substantially in terms of the coefficient but there was an important decrease of the heterogeneity (table 2).

To illustrate how the different assumptions for underlying risk across trials may affect the results, we additionally fit the meta-regression

Table 1 Results from the meta-regression model that accounted for differences in the average baseline severity score between the studies of the Kirsch dataset with dependent variable as the standardised mean difference (SMD) of antidepressants versus placebo

	All studies		Without the Dunlop study	
	Estimate	95% CI	Estimate	95% CI
Coefficient for baseline severity	0.05	(0.02 to 0.08)	0.06	(0.01 to 0.11)
SMD at mean baseline score (constant)	0.33 (mean score=25.54)	(0.25 to 0.41)	0.33 (mean score=25.78)	(0.25 to 0.41)
Heterogeneity	$\tau=0.12$, $I^2=31\%$ (95% CI: 0% to 55%)		$\tau=0.12$, $I^2=32\%$ (95% CI: 0% to 56%)	

Coefficients correspond to a 1-unit increase in baseline severity score.

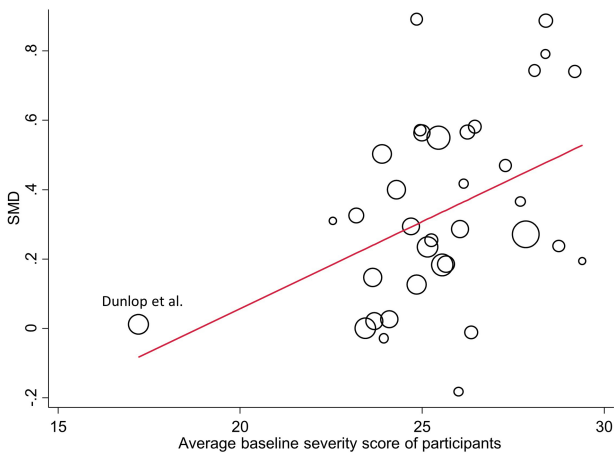


Figure 1 Graphical illustration of the relationship between the average baseline severity score of study participants (ie, the average score from all participants on a continuous symptom scale at the beginning of each study) and the relative effect of antidepressants for the Kirsch data.¹⁰ The size of the circles is proportional to the precision (inverse of the variance) of the studies. The line represents the meta-regression line (SMD, standardised mean difference).

model (without the Cohn study) using the true underlying risk (in a Bayesian environment) as a predictor and assuming both independent and dependent (ie, from a normal distribution) study-specific underlying risks. The results from these approaches are given in table 3. More specifically, when we assumed no relationship between the underlying risks from different studies, the estimated coefficient was smaller in magnitude compared to that derived from the initial meta-regression (which used the observed values as predictor), but marginally statistically significant. The assumption that all underlying risks share a common normal distribution yielded a (non-significant) coefficient close to zero, implying that this characteristic does not seem to affect the relative effect of antidepressants when compared with placebo. The same findings (lack of association) were obtained when we used the ‘weaker’ assumption that the trial’s underlying risks come from two different normal distributions (with common variance; table 3).

DISCUSSION

Ignoring important variability in one or more population characteristics across studies in a meta-analysis may lead to misleading conclusions. When information on such characteristics is available, meta-regression or subgroup analysis should be employed to assess their impact on the results. Baseline severity has been widely considered as an important effect modifier in meta-analyses of mental health trials. In the Kirsch data, we found an important linear association between baseline severity and relative effects, while the differences in the initial severity of participants across trials explained the amount of heterogeneity. Meta-regression models can also be used to disentangle the effect of disease severity and placebo response on the outcome.

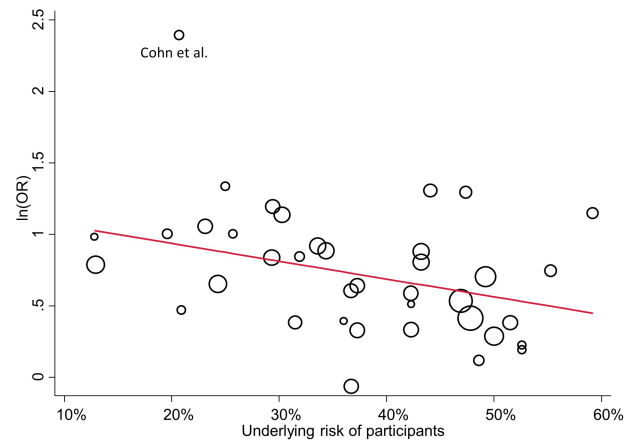


Figure 2 Graphical illustration of the relationship between the underlying risk (ie, the number of successes over the total participants) of the studies and the relative effect of antidepressants for the Undrurraga example.⁶ The size of the circles is proportional to the precision (inverse of the variance) of the studies. The line represents the meta-regression line (ln(OR), logarithm of the OR).

In the absence of available data for important baseline characteristics, underlying risk has been suggested as a surrogate variable for baseline severity, and a variety of additional and possibly unmeasured or unknown factors that may act as effect modifiers. An empirical study that investigated the relationship between underlying risk and treatment effects in 115 meta-analyses with dichotomous outcomes found statistically significant associations in 14% of these meta-analyses.¹⁸ In our example,⁶ the effect of underlying risk became marginally statistically non-significant by excluding one study that seemed to be an outlier. Although this should be verified using appropriate statistical methods for detecting outliers, a graphical depiction is always helpful in identifying observations that possibly do not fit to the rest of the data and might exaggerate associations via meta-regression. Also, in this collection of studies the different assumptions for the underlying risks across trials affected the estimation of the regression coefficients. Where no prior belief exists for the plausibility of each assumption the different suggested models may be employed as a sensitivity analysis.

Finally, despite the importance of baseline characteristics in synthesising the studies of a systematic review, it is unclear whether investigators regularly incorporate them in their analyses. Protocols for meta-analyses should carefully consider the targeted population and include meta-regression with baseline severity or underlying risk among their additional analyses when baseline differences are expected between the eligible studies.

Funding AC received funding from Greek national funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF)—Research Funding Program: ARISTEIA. Investing in knowledge society through the European Social Fund.

Table 2 Results from the meta-regression model that accounted for differences in underlying risk between the studies from the Undrurraga example with dependent variable as the logarithm of the OR (lnOR) of antidepressants versus placebo

	All studies		Without the Cohn study	
	Estimate	95% CI	Estimate	95% CI
Coefficient for underlying risk	-0.12	(-0.23 to 0.02)	-0.10	(-0.19 to 0.00)
lnOR at the average underlying risk (constant)	0.72 (average risk=37.1%)	(0.60 to 0.85)	0.70 (average risk=37.5%)	(0.59 to 0.80)
Heterogeneity	$\tau=0.16, I^2=27%$ (95% CI: 0% to 52%)		$\tau=0.09, I^2=8%$ (95% CI: 0% to 38%)	

Coefficients correspond to a 10% increase in underlying risk.

Table 3 Results from the three meta-regression models that used the true study underlying risks from the Undurraga example with dependent variable as the logarithm of the OR (lnOR) of antidepressants versus placebo

	Independent underlying risks		A common normal distribution		Two normal distributions	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Coefficient for underlying risk	-0.10	(-0.20 to 0.00)	-0.01	(-0.13 to 0.13)	-0.01	(-0.12 to 0.13)
lnOR at the average underlying risk (37.5%) (constant)	0.70	(0.59 to 0.83)	0.69	(0.57 to 0.80)	0.68	(0.57 to 0.80)
Heterogeneity(τ)	0.16	(0.02 to 0.33)	0.15	(0.01 to 0.31)	0.13	(0.01 to 0.31)

Coefficients correspond to a 10% increase in underlying risk.

Competing interests None.

doi:10.1136/eb-2014-102035

REFERENCES

1. **Thompson SG.** Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994;**309**:1351–5.
2. **Nikolakopoulou A, Mavridis D, Salanti G.** Demystifying fixed and random effects meta-analysis. *Evid Based Ment Health* 2014;**17**:53–7.
3. **Deeks JJ, Higgins JPT, Altman DG.** Chapter 9: analysing data and undertaking meta-analyses. In: Higgins JPT, Green S. eds. *Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011. <http://www.cochrane-handbook.org>
4. **Riley RD, Higgins JPT, Deeks JJ.** Interpretation of random effects meta-analyses. *BMJ* 2011;**342**:d549.
5. **Sharp SJ, Thompson SG, Altman DG.** The relation between treatment benefit and underlying risk in meta-analysis. *BMJ* 1996;**313**:735–8.
6. **Undurraga J, Baldessarini RJ.** Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. *Neuropsychopharmacology* 2012;**37**:851–64.
7. **Khan A, Leventhal RM, Khan SR, et al.** Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 2002;**22**:40–5.
8. **Barbui C, Furukawa TA, Cipriani A.** Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *Can Med Assoc J* 2008;**178**:296–305.
9. **Fournier JC, DeRubeis RJ, Hollon SD, et al.** Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010;**303**:47–53.
10. **Kirsch I, Deacon BJ, Huedo-Medina TB, et al.** Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;**5**:e45.
11. **Moncrieff J, Wessely S, Hardy R.** Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* 2004;(1):CD003012.
12. **Bech P, Cialdella P, Haugh MC, et al.** Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *Br J Psychiatry J Ment Sci* 2000;**176**:421–8.
13. **Borenstein M, Hedges LV, Higgins JPT, et al.** Meta-regression. In: *Introduction to meta-analysis*. John Wiley & Sons, Ltd, 2009:187–203. <http://onlinelibrary.wiley.com/doi/10.1002/9780470743386.ch20/summary> (accessed 26 Dec 2013).
14. **Thompson SG, Smith TC, Sharp SJ.** Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med* 1997;**16**:2741–58.
15. **Ghidey W, Stijnen T, van Houwelingen HC.** Modelling the effect of baseline risk in meta-analysis: a review from the perspective of errors-in-variables regression. *Stat Methods Med Res* 2013;**22**:307–23.
16. **Sharp SJ, Thompson SG.** Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Stat Med* 2000;**19**:3251–74.
17. **Arends LR, Hoes AW, Lubsen J, et al.** Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Stat Med* 2000;**19**:3497–518.
18. **Schmid CH, Lau J, McIntosh MW, et al.** An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998;**17**:1923–42.
19. **McIntosh MW.** The population risk as an explanatory variable in research synthesis of clinical trials. *Stat Med* 1996;**15**:1713–28.
20. **Bland JM, Altman DG.** Statistic notes: regression towards the mean. *BMJ* 1994;**308**:1499.
21. **Ghidey W, Lesaffre E, Stijnen T.** Semi-parametric modelling of the distribution of the baseline risk in meta-analysis. *Stat Med* 2007;**26**:5434–44.
22. **Schmid CH, Stark PC, Berlin JA, et al.** Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004;**57**:683–97.
23. **Guolo A.** Flexibly modeling the baseline risk in meta-analysis. *Stat Med* 2013;**32**:40–50.
24. **Dewilde S, Hawkins N.** Investigating incoherence gives insight: clopidogrel is equivalent to extended-release dipyridamole plus aspirin in secondary stroke prevention. *J Clin Epidemiol* 2012;**65**:835–45.
25. **National Institute for Clinical Excellence.** *Depression: management of depression in primary and secondary care. Clinical practice guideline No 23*. London: National Institute for Clinical Excellence, 2004:670 p.
26. **Dunlop SR, Dornseif BE, Wernicke JF, et al.** Pattern analysis shows beneficial effect of fluoxetine treatment in mild depression. *Psychopharmacol Bull* 1990;**26**:173–80. <http://psycnet.apa.org/psycinfo/1992-36203-001> (accessed 1 Dec 2014).
27. **Cohn JB, Wilcox C.** A comparison of fluoxetine, imipramine, and placebo in patients with major depressive disorder. *J Clin Psychiatry* 1985;**46**:26–31.