Regression in Meta-Analysis

Michael Borenstein
Larry V. Hedges
Julian P.T. Higgins
Hannah Rothstein

Draft – Please do not quote

This draft released January 6, 2015
Please send comments to Biostat100@GMail.com
An updated copy of this manual will be posted at
<table>
<thead>
<tr>
<th>Part 1: Data files and downloads</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 2: Overview of Meta-regression</td>
<td>14</td>
</tr>
<tr>
<td>Part 3: The BCG example</td>
<td>15</td>
</tr>
<tr>
<td>Part 4: Meta-regression is observational</td>
<td>19</td>
</tr>
<tr>
<td>Part 5: Fixed-effect vs. Random-effects</td>
<td>21</td>
</tr>
<tr>
<td>Putting regression in context</td>
<td>29</td>
</tr>
<tr>
<td>Fixed-effect model</td>
<td>30</td>
</tr>
<tr>
<td>Basic analysis (Case A)</td>
<td>30</td>
</tr>
<tr>
<td>The traditional approach</td>
<td>30</td>
</tr>
<tr>
<td>Test of effect size</td>
<td>31</td>
</tr>
<tr>
<td>Test of the statistical model</td>
<td>31</td>
</tr>
<tr>
<td>The regression approach</td>
<td>31</td>
</tr>
<tr>
<td>Test of effect size</td>
<td>32</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>32</td>
</tr>
<tr>
<td>Summary</td>
<td>33</td>
</tr>
<tr>
<td>Subgroups analysis (Case B)</td>
<td>33</td>
</tr>
<tr>
<td>The traditional approach</td>
<td>33</td>
</tr>
<tr>
<td>Is the common effect size zero for each subgroup?</td>
<td>34</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>35</td>
</tr>
<tr>
<td>The regression approach</td>
<td>36</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>37</td>
</tr>
<tr>
<td>Summary</td>
<td>39</td>
</tr>
<tr>
<td>Continuous covariate (Case C)</td>
<td>39</td>
</tr>
<tr>
<td>The traditional approach</td>
<td>39</td>
</tr>
<tr>
<td>The regression approach</td>
<td>39</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>40</td>
</tr>
<tr>
<td>Summary</td>
<td>41</td>
</tr>
<tr>
<td>In context</td>
<td>41</td>
</tr>
<tr>
<td>Random-effects model</td>
<td>42</td>
</tr>
<tr>
<td>Basic analysis (Case A)</td>
<td>42</td>
</tr>
<tr>
<td>The traditional approach</td>
<td>42</td>
</tr>
<tr>
<td>Test of effect size</td>
<td>43</td>
</tr>
</tbody>
</table>
Heterogeneity ............................................................................................................................. 43
The regression approach................................................................................................................. 44
Test of effect size .......................................................................................................................... 45
Test of the model........................................................................................................................... 45
Goodness of fit ............................................................................................................................ 46
Comparison of Model 1 with the null model .............................................................................. 47
Summary ......................................................................................................................................... 48
Subgroups analysis (Case B) ................................................................................................................ 48
The traditional approach ................................................................................................................ 49
Note on computing $T^2$ ................................................................................................................ 50
Is the mean effect size zero for each subgroup? ........................................................................ 51
Test of the model........................................................................................................................... 51
Heterogeneity ............................................................................................................................. 51
The regression approach................................................................................................................. 52
Test of the model........................................................................................................................... 54
Goodness of Fit ............................................................................................................................. 54
Comparison of Model 1 with the null model .............................................................................. 56
Summary ......................................................................................................................................... 58
Continuous covariate (Case C) ............................................................................................................ 58
The traditional approach ................................................................................................................ 58
The regression approach................................................................................................................. 59
Prediction equation ....................................................................................................................... 60
Test of the model........................................................................................................................... 60
Goodness of fit ............................................................................................................................ 60
Comparison of Model 1 with the null model .............................................................................. 62
Summary ......................................................................................................................................... 63
In context ............................................................................................................................................ 64
Part 6: Meta-regression in CMA ........................................................................................................ 65
What’s new in this version of meta-regression? ............................................................................. 66
The covariates and the predictive model ....................................................................................... 67
Quick Start ........................................................................................................................................ 68
Step 1: Enter the data ..................................................................................................................... 69
Insert column for study names ...................................................................................................... 69
Insert columns for effect size data ............................................................................................... 70
Insert columns for moderators (covariates) .............................................................................. 74
Part 19: Some caveats ............................................................................................................................... 293
  Statistical power for meta-regression ........................................................................................................ 294
  Multiple comparisons ................................................................................................................................... 295
Part 20: Technical Appendix ..................................................................................................................... 297
  Appendix 1: The dataset ...................................................................................................................... 298
  Appendix 2: Understanding Q ........................................................................................................... 302
  Appendix 3: Tests of heterogeneity .................................................................................................. 318
  Appendix 4: Computing $\tau^2$ in the presence of subgroups ......................................................... 327
  Appendix 5: Creating variables for interactions .............................................................................. 331
  Appendix 6: Plotting a curvilinear relationship ................................................................................. 334
  Appendix 7: Plotting interactions ......................................................................................................... 338
    Plotting the interaction of two categorical covariates ...................................................................... 339
    Plotting the interaction of a categorical covariate by a continuous covariate ............................. 343
    Plotting the interaction of two continuous covariates .................................................................... 347
  Appendix 8: Interpreting regression coefficients .............................................................................. 352
  Appendix 9: Meta-regression in stata .................................................................................................. 353
References ................................................................................................................................................ 358
<table>
<thead>
<tr>
<th>Figure</th>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td>Main results</td>
<td>Assessing the impact of a set</td>
<td>213</td>
</tr>
<tr>
<td>146</td>
<td>Setup</td>
<td>Defining a set of covariates</td>
<td>214</td>
</tr>
<tr>
<td>147</td>
<td>Setup</td>
<td>Naming a set of covariates</td>
<td>215</td>
</tr>
<tr>
<td>148</td>
<td>Setup</td>
<td>Naming a set of covariates</td>
<td>215</td>
</tr>
<tr>
<td>149</td>
<td>Regression</td>
<td>Main results</td>
<td>216</td>
</tr>
<tr>
<td>150</td>
<td>Main results</td>
<td>Removing a set of covariates</td>
<td>216</td>
</tr>
<tr>
<td>151</td>
<td>Setup</td>
<td>Interaction of two categorical covariates</td>
<td>220</td>
</tr>
<tr>
<td>152</td>
<td>Main results</td>
<td>Interaction of two categorical covariates</td>
<td>221</td>
</tr>
<tr>
<td>153</td>
<td>Plot</td>
<td>Interaction of two categorical covariates</td>
<td>223</td>
</tr>
<tr>
<td>154</td>
<td>Plot</td>
<td>Interaction of two categorical covariates</td>
<td>224</td>
</tr>
<tr>
<td>155</td>
<td>Plot</td>
<td>Interaction of two categorical covariates</td>
<td>225</td>
</tr>
<tr>
<td>156</td>
<td>Setup</td>
<td>Interaction of categorical and continuous covariates</td>
<td>227</td>
</tr>
<tr>
<td>157</td>
<td>Main results</td>
<td>Interaction of categorical and continuous covariates</td>
<td>228</td>
</tr>
<tr>
<td>158</td>
<td>Plot</td>
<td>Interaction of categorical and continuous covariates</td>
<td>230</td>
</tr>
<tr>
<td>159</td>
<td>Plot</td>
<td>Interaction of categorical and continuous covariates</td>
<td>231</td>
</tr>
<tr>
<td>160</td>
<td>Setup</td>
<td>Interaction of categorical and continuous covariates</td>
<td>232</td>
</tr>
<tr>
<td>161</td>
<td>Setup</td>
<td>Interaction of two continuous covariates</td>
<td>234</td>
</tr>
<tr>
<td>162</td>
<td>Main results</td>
<td>Interaction of two continuous covariates</td>
<td>235</td>
</tr>
<tr>
<td>163</td>
<td>Plot</td>
<td>Interaction of two continuous covariates</td>
<td>236</td>
</tr>
<tr>
<td>164</td>
<td>Plot</td>
<td>Interaction of two continuous covariates</td>
<td>237</td>
</tr>
<tr>
<td>165</td>
<td>Plot</td>
<td>Interaction of two continuous covariates</td>
<td>238</td>
</tr>
<tr>
<td>166</td>
<td>Setup</td>
<td>Curvilinear relationship</td>
<td>240</td>
</tr>
<tr>
<td>167</td>
<td>Main results</td>
<td>Curvilinear relationship</td>
<td>241</td>
</tr>
<tr>
<td>168</td>
<td>Plot</td>
<td>Curvilinear relationship</td>
<td>242</td>
</tr>
<tr>
<td>169</td>
<td>Setup</td>
<td></td>
<td>244</td>
</tr>
<tr>
<td>170</td>
<td>Data-entry</td>
<td>Missing data for latitude</td>
<td>245</td>
</tr>
<tr>
<td>171</td>
<td>Basic analysis</td>
<td>Missing data for latitude</td>
<td>246</td>
</tr>
<tr>
<td>172</td>
<td>Regression</td>
<td>Setup</td>
<td>Latitude in list and checked</td>
</tr>
<tr>
<td>173</td>
<td>Regression</td>
<td>Main results</td>
<td>Missing data</td>
</tr>
<tr>
<td>174</td>
<td>Table of missing data</td>
<td></td>
<td>248</td>
</tr>
<tr>
<td>175</td>
<td>Setup</td>
<td>Latitude in list, unchecked</td>
<td>249</td>
</tr>
<tr>
<td>176</td>
<td>Main results</td>
<td>Latitude in list, unchecked</td>
<td>250</td>
</tr>
<tr>
<td>177</td>
<td>Setup</td>
<td>Latitude must be removed from list</td>
<td>251</td>
</tr>
<tr>
<td>178</td>
<td>Setup</td>
<td>Latitude removed from list</td>
<td>252</td>
</tr>
<tr>
<td>179</td>
<td>Main results</td>
<td>Latitude removed from list</td>
<td>252</td>
</tr>
<tr>
<td>180</td>
<td>Data entry</td>
<td></td>
<td>254</td>
</tr>
<tr>
<td>181</td>
<td>Basic analysis</td>
<td></td>
<td>254</td>
</tr>
<tr>
<td>182</td>
<td>Meta-regression</td>
<td></td>
<td>255</td>
</tr>
<tr>
<td>183</td>
<td>Select by study name</td>
<td></td>
<td>255</td>
</tr>
<tr>
<td>184</td>
<td>Select by study name</td>
<td></td>
<td>256</td>
</tr>
<tr>
<td>185</td>
<td>Create a moderator for filtering</td>
<td></td>
<td>257</td>
</tr>
<tr>
<td>186</td>
<td>Filter by moderator</td>
<td></td>
<td>258</td>
</tr>
<tr>
<td>187</td>
<td>Regression using a filter</td>
<td></td>
<td>259</td>
</tr>
<tr>
<td>188</td>
<td>Select by moderator</td>
<td></td>
<td>260</td>
</tr>
<tr>
<td>189</td>
<td>Select by moderator</td>
<td></td>
<td>261</td>
</tr>
<tr>
<td>190</td>
<td>Filter by moderator</td>
<td></td>
<td>261</td>
</tr>
<tr>
<td>191</td>
<td>Defining several models</td>
<td>Setup</td>
<td>262</td>
</tr>
<tr>
<td>192</td>
<td>Defining several models</td>
<td>Setup</td>
<td>262</td>
</tr>
</tbody>
</table>
PART 1: DATA FILES AND DOWNLOADS

This manual http://www.meta-analysis.com/pages/cma_manual.php
CMA program http://www.meta-analysis.com/

BCG data in CMA format

BCG data in Excel™ format http://www.meta-analysis.com/downloads/BCG.xls

Excel™ files for plotting interactions
http://www.meta-analysis.com/downloads/Plot-interaction-of-Hot-x-Year-C.xls
http://www.meta-analysis.com/downloads/Plot-interaction-of-Latitude-x-Year-C.xls
http://www.meta-analysis.com/downloads/Plot-of-curvilinear-relationship.xls
PART 2: OVERVIEW OF META-REGRESSION

INTRODUCTION

In primary studies, multiple-regression is the statistical technique employed to assess the relationship between covariates and a dependent variable. In these studies the unit of analysis is the subject, with covariates and the outcome measured for each subject.

With a few modifications, the same technique can be used in meta-analysis. In this case, the unit of analysis is the study, with covariates and outcomes measured for each study. We sometimes use the term “meta-regression” to refer to the use of regression in meta-analysis.

With these modifications in place, the full arsenal of procedures that fall under the heading of “multiple-regression” in primary studies becomes available to the meta-analyst. For example,

- We can assess the impact of one covariate, or the combined impact of multiple covariates
- We can enter covariates into the analysis using a pre-defined sequence and assess the impact of any covariates, over and above the impact of prior covariates
- We can work with sets of covariates, such as three variables that together define a treatment, or that represent a nonlinear relationship between the predictor variable and the effect size.
- We can incorporate both categorical (for example, dummy-coded) and continuous variables as covariates.

This book is intended as a resource to explain how to run and interpret a meta-regression. It is also intended as a manual to show how to use the program CMA to perform a meta-regression.

In Part 1 we provide links to the data files referenced in this manual
In Part 2 we provide an overview of this manual
In Part 3 we introduce the BCG example
In Part 4 we explain that meta-regression is an observational analysis
In Part 5 we provide an overview of fixed-effect vs. random-effects models
In Part 6 we provide a step-by-step guide for running a meta-regression in CMA
In Part 7 we discuss how to interpret the results of a meta-regression
In Part 8 we discuss the computation and meaning of $R^2$
In Part 9 we discuss how to use and customize the regression plot
In Part 10 we discuss computational options
In Part 11 we explain how to work with categorical variables
In Part 12 we discuss when it makes sense to omit the intercept
In Part 13 we discuss how to work with “sets” of covariates
In Part 14 we explain how to work with interactions
In Part 15 we discuss missing data
In Part 16 we show how to run a regression on subsets of the studies
In Part 17 we show how to define and compare different predictive models
In Part 18 we explain how to work with complex data structures
In Part 19 we explain some caveats about meta-regression
In Part 20 we provide a technical appendix
PART 3: THE BCG EXAMPLE

We will use the “BCG analysis” as the motivating example in this book.

“BCG” refers to the Bacillus Calmette-Guerin (BCG) vaccine, which is intended to prevent tuberculosis (TB). This vaccine had been studied in a series of 13 controlled trials between the years 1933 and 1968, with some trials suggesting that the vaccine was effective in reducing the incidence of TB, and others suggesting that it was not. With the re-emergence of TB in the United States in recent years (including many drug-resistant cases) the question of whether or not BCG was actually effective took on a new urgency. Colditz et al. (1994) conducted a meta-analysis to synthesize the data from these trials.

Figure 1 shows a random-effects meta-analysis based on the BCG studies. The effect size is the risk ratio, as indicated by the labels [A]. In this example a risk ratio of less than 1.0 indicates that the vaccine reduced the risk of TB, a risk ratio of 1.0 indicates no effect, and a risk ratio higher than 1.0 indicates that the vaccine increased the risk of TB.

The summary risk ratio [B] is 0.4896 with a 95% confidence interval of 0.3449 to 0.6950, and a p-value of 0.0001. Thus, there is strong evidence that the vaccine is effective in preventing TB.

Equally important, however, is the variation in the treatment effect, with the risk ratio in individual studies ranging from 0.2049 (approximately an 80% risk reduction) in one study [C] to 1.5619 (approximately a 56% risk increase) in another [D]. While some of the observed variance in effects is probably due to sampling error, a substantial amount of the variance reflects real differences in the treatment effect (more on this later). Obviously, it’s imperative to understand why the vaccine is more effective in some studies than in others.
Among the studies in the meta-analysis, there appears to be a relationship between climate and effectiveness, such that studies performed in colder locations tended to show a stronger effect. If this relationship is real, it could be explained by either of two mechanisms. First, persons in colder climates may be less likely to have a natural immunity to TB. It follows that the population in these climates would be more susceptible to TB, and more likely to benefit from the vaccine. Second, it’s likely that the drug would be more potent in the colder climates. This follows from the fact that in warmer climates the heat could cause the drug to lose potency.

Optimally, a researcher would be able to code each study for the prevalence of natural immunity and for drug potency, and then use these as predictors of effectiveness in a meta-regression. However, these predictors were not available for the analysis. Therefore, the analysts elected to use “Latitude” (actually, the absolute value of latitude) as a surrogate for these covariates, the assumption being that studies more distant from the equator tended to employ populations with less natural immunity and more potent vaccine.

This was the strategy adopted by Berkey et al. (1995), who employed meta-regression to assess the relationship between latitude and treatment effect. Given the post hoc nature of this analysis, a positive finding would not be definitive, but would suggest a direction for additional research.

This regression has been used as an example in many texts on meta-analysis, including Borenstein et al. 2009, Egger et al. 2001, Sutton et al 2000, Hartung et al. (2008)). We will use this example here as well, to allow readers to compare our analysis with the analyses presented in the other texts. Note that each text presents the data in a somewhat different format, and here we follow the format employed by Hartung. In addition to the original variables, we created new variables for the purposes of this text. For example, we classified studies as “Hot” or “Cold” based on the latitude, created variables that are centered versions of the originals, and created variables to represent interactions among the original variables.

Part of the dataset is shown in Figure 2. The full set of variables is described in Table 1. In [Appendix 1: The dataset] we show how to create all of the new variables in Excel™.
Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latitude</td>
<td>Integer</td>
<td>This is absolute latitude, which is simply the latitude ignoring the sign. Low values are closer to the equator, high values are more distant.</td>
</tr>
<tr>
<td>Allocation</td>
<td>Categorical</td>
<td>This is the method employed to assign people to the treated or control conditions. The three possible classifications are randomized, alternate, and systematic.</td>
</tr>
<tr>
<td>Year</td>
<td>Integer</td>
<td>This is the year the study was conducted.</td>
</tr>
<tr>
<td><strong>Variables related to climate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latitude-C</td>
<td>Decimal</td>
<td>Latitude, centered to have a mean of zero</td>
</tr>
<tr>
<td>Latitude-C²</td>
<td>Decimal</td>
<td>The square of Latitude-C</td>
</tr>
<tr>
<td>Climate</td>
<td>Categorical</td>
<td>We dichotomized latitude, and classified each study’s location as Hot (latitude under 34) or Cold (latitude over 41)</td>
</tr>
<tr>
<td>Hot</td>
<td>Integer</td>
<td>This is a numeric version of Climate, coded 1 for Hot and 0 for Cold.</td>
</tr>
<tr>
<td>Cold</td>
<td>Integer</td>
<td>This is a numeric version of Climate, coded 1 for Cold and 0 for Hot.</td>
</tr>
<tr>
<td><strong>Variables related to study year</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year-C</td>
<td>Integer</td>
<td>Year, centered to have a mean of zero</td>
</tr>
<tr>
<td>Time</td>
<td>Categorical</td>
<td>We dichotomized Year, and classified each study’s time as Early (pre-1945) or Recent (post-1945)</td>
</tr>
<tr>
<td>Early</td>
<td>Integer</td>
<td>This is a numeric version of Time, coded 1 for Hot and 0 for Cold.</td>
</tr>
<tr>
<td>Recent</td>
<td>Integer</td>
<td>This is a numeric version of Time, coded 1 for Cold and 0 for Hot.</td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot x Year-C</td>
<td>Decimal</td>
<td>Interaction of Hot and Year-C (Dichotomous and continuous)</td>
</tr>
<tr>
<td>Hot x Recent</td>
<td>Integer</td>
<td>Interaction of Hot and Recent (Dichotomous and Dichotomous)</td>
</tr>
<tr>
<td>Latitude-C x Year-C</td>
<td>Decimal</td>
<td>Interaction of Latitude-C and Year-C (Continuous and Continuous)</td>
</tr>
</tbody>
</table>

The letter “C” added to a variable indicates that the variable has been centered (See appendix)

**Types of variables**

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>Studies belong to discrete groups, such as “Cold” and “Hot”. This cannot be used directly in the analysis, but is used to create numeric variables (dummy-variables).</td>
</tr>
<tr>
<td>Integer</td>
<td>Studies have values on an integer scale, and can take on any whole number.</td>
</tr>
<tr>
<td>Decimal</td>
<td>Studies have values on a continuous scale and can take on any whole or decimal number.</td>
</tr>
</tbody>
</table>
In Figure 1 the treatment effect (or effect size) was displayed as a risk ratio. While the risk ratio has the advantage of being an intuitive index, the analyses are actually performed using the log of the risk ratio, and then converted to risk ratios for display. Since one of our goals in this book is to explain the mechanics of the analyses, we will generally be working with the log units.

For example, where Figure 1 showed the risk ratios, Figure 3 shows the same forest plot using log units, as indicated by the labels [E]. In this case, a log risk ratio of less than 0.0 indicates that the vaccine reduced the risk of TB, a log risk ratio of 0.0 indicates no effect, and a log risk ratio greater than 0.0 indicates that the vaccine increased the risk of TB. The summary effect [F] is a log risk ratio of $-0.7141$. 

![Figure 3](image-url)
PART 4: META-REGRESSION IS OBSERVATIONAL

When we work with primary studies we need to be aware of the difference between a randomized study and an observational study.

In a randomized trial, participants are assigned at random to a condition, such as treatment versus placebo. The randomization is intended to ensure that the subjects in the two conditions are similar in all respects except for the treatment. Therefore, assuming that the randomization works properly, any differences that emerge between groups can be attributed to the treatment.

By contrast, in an observational study we compare pre-existing groups, such as workers with a college education versus those who did not attend college. While we can report on differences between groups, we cannot attribute these differences to the presence or absence of a college education because the groups differ in other ways as well. For example, it is possible (indeed likely) that subjects who had a college education also had other advantages, including better skills in an array of areas.

Consider how this plays out in a meta-analysis.

Assume we start with a set of randomized experiments that assess the impact of an intervention. If the effect in each study can serve to establish causality because of the randomization process, then the summary effect can also serve to establish causality. If the effect in each study is due to the intervention, then the overall effect is due to the intervention.

However, even if the individual studies are randomized trials, once we move beyond the goal of reporting a summary effect and proceed to perform a subgroup analyses or meta-regression, we have moved out of the domain of randomized experiments, and into the domain of observational studies.

In this example, if the effect size is different in hot climates than in cold climates we cannot assume that this is because of climate. While we choose to label one group of studies “Hot climate” and another “Cold climate”, it is possible that the two groups differ from each other in other ways as well, and that it might be these other factors (instead of climate, or in addition to climate) that are responsible for the difference in effects.

For example, it turns out that the “Hot” studies, where the vaccine was less effective, tend to be more recent. We’d like to think that the vaccine was less effective because of storage conditions in the hot climates, and that if we used better storage the efficacy would improve. However, it’s also possible that there were unrelated changes over the decades that caused the vaccine to become less effective.

That said, in primary observational studies, researchers sometimes use regression analysis to try and remove the impact of potential confounders. This is not a perfect solution since there may be other confounders of which we are not aware, but this approach can help to isolate the impact of specific factors and generate hypotheses to be tested in randomized trials. The same holds true for meta-regression.
This approach is potentially useful only when there are enough studies to isolate the unique impact of each factor. Many meta-regressions are based on relatively small numbers of studies, and so it may not be possible to adjust for potential confounds.

Of course, even when there are enough studies to adjust for known confounds, we cannot be certain that we’ve identified all possible confounds. Therefore we can’t use this approach to prove a causal relationship.

There is one exception to the rule that subgroup analysis and regression cannot prove causality. This is the case where not only was assignment to treatment condition randomized, but also assignment to subgroup was randomized. In this case we know that the only systematic difference between subgroups is the one captured by subgroup membership.

The pharmaceutical example discussed later is a case in point. In this hypothetical example we enrolled 1000 patients and assigned some to studies that would test a low dose of the drug vs. placebo, and others to studies that would test a high dose of the drug vs. placebo. Here, the assignment to subgroups is random. The same would apply if the patients were assigned to ten studies where the dose of drug was varied on a continuous scale, and we used meta-regression to test the relationship between dose and effect size.

This set of circumstance (the drug company example) is extremely rare in practice. We present it here primarily to illustrate the conditions that would be needed before we could draw a causal inference from a subgroup analysis or regression. Our main point is that in the absence of these conditions we cannot draw a causal inference.
PART 5: FIXED-EFFECT VS. RANDOM-EFFECTS

Statistical models

Before turning to regression, we briefly review the statistical models typically employed in meta-analysis – the fixed-effect model and the random-effects model.

To understand the difference between these two models it might help to distinguish between the ideas of a population versus a universe.

One study samples subjects from a population, which is defined as people meeting a specific set of criteria. A second study samples subjects from a population, which is defined as people meeting a specific set of criteria. If the two sets of criteria are the same in all relevant respects, then we can say that both studies are sampling from the same population. Equivalently, if the true effect size (the effect size that we would see if there was no sampling error) is identical in both populations, then we can think of them as the same population.

However, if the criteria differ in any material respect, then we would say that the two studies sample from different populations. In this case each study has its criteria that define the study’s population, and we need a second set of criteria that tell us what kinds of studies (what populations) we want to include in the meta-analysis. This second set of criteria defines the universe of populations.

For example, suppose that each study samples persons from the lung-cancer outpatient clinic at the hospital where the study is being conducted.

- If there are ten studies in the analysis and all were conducted at the same hospital (let’s also assume at the same time) then all studies are being sampled from the same population. The true effect size in all the studies is the same.

- If there are ten studies in the analysis and each was conducted at a different hospital, then each study is sampled from a different population. The true effect size probably differs (possibly by a little, or possibly by a lot) from one hospital to the next.

In the second case (ten hospitals) if we decide to include all these hospitals in the meta-analysis, it’s because all the populations are from the same universe. We would probably define the universe as clinics where the patients are similar enough so that the studies are all addressing the same fundamental question.

To this point we’ve focused on the patients in discussing the difference between a population and a universe, but the distinction between the two depends also on other aspects of the study. For example, studies may be from the same population if they all run for exactly two weeks. If some studies run for two weeks and others for three weeks, then the two kinds of studies are based on different populations but are drawn from the same universe. Similarly, studies may be from the same population if they all use the identical measure of outcome. If some studies use one measure while others use a similar measure, then they are based on different populations but are drawn from the same universe.
It should be clear that we are defining a population very narrowly. Under this definition studies are only drawn from the same population if they are essentially replicates of each other in all material respects. This means that not only the subjects but also the methods, the specifics of the intervention, and the outcome measures are, for all intents and purposes, identical across studies. This criterion will rarely be met in practice, and will never be met when we’re working with studies conducted independently of each other.

With this as background we can discuss the difference between the fixed-effect and the random-effects models.

- The fixed-effect model applies if all studies are drawn from a single population (the identical subjects and methods). The studies share a common effect size, and so the effect size is fixed, or constant.

- The random-effects model applies if the studies are drawn from a universe of populations. The true effect size varies from one population to the next, and the studies are sampled at random from this universe.

The selection of a model is critically important for several reasons.

First, it sets up a framework for the analysis, establishing what questions we can ask and how to interpret the results.

- Under the fixed-effect model we assume that all studies share a common true effect size, and our goal is to estimate this common parameter.

- Under the random-effects model we allow that the true effect size might vary from study to study, and our goal is to estimate the mean of these parameters.

Second, the selection of a model affects how weights are assigned to the studies. This affects both the summary estimates themselves and also the precision of the summary estimates.

- Under the fixed-effect model there is only one level of sampling (the subjects in each study are sampled from all subjects in the study’s population) and therefore only one source of sampling error (the observed effect in each study differs from the true effect for that study’s population). The error variance for each study is quantified as $V$, and the weight assigned to each study is the inverse of this variance, or $1/V$.

- Under the random-effects model there are two levels of sampling (the subjects in each study are sampled from all subjects in the study’s population, and the study populations are sampled from the universe of study populations) and therefore two sources of sampling error (the observed effect in each study differs from the true effect for that study’s population, and the mean true effect for the sampled studies differs from the mean for the universe of studies). The first error variance is quantified as $V$ and the second as $T^2$. The total error variance for each study is then $V+T^2$, and the weight assigned to each study is the inverse of this variance, or $1/(V+T^2)$.
We typically encounter the idea of fixed-effect vs. random-effects models in the context of a simple analysis, where we have a single set of studies. In this case, the distinction between the two models is relatively straightforward: the fixed-effect model applies if the studies share a common effect size, and the random-effects model applies otherwise.

However, the same idea can be extended to the case where we have discrete subgroups of studies, and even further, to the case where we have studies that range along a continuum on some dimension(s). Our goal here is to explain these extensions, and for that purpose we will show how the difference between fixed-effect and random-effects models plays out in (A) a simple analysis, (B) a subgroups analysis, and (C) a meta-regression.

**Case A - A simple analysis**

Consider the case where we are working with a single set of studies. The fixed-effect model applies if all the studies share a common effect size. The random-effects model applies if the effect size might vary from study to study, across all studies in the analysis.

**Case B - Subgroups**

Consider the case where we want to compare the effect size for two or more subgroups of studies (for example, studies in cold climates vs. studies in hot climates). The fixed-effect model applies if all studies within a subgroup share the same effect size parameter. The random-effects model applies if the effect size parameters might vary from study to study, for studies within a subgroup.

**Case C – Regression**

Consider the case where we want to look at effect size in relation to a continuous covariate (for example, latitude). The fixed-effect model applies if all studies at the same latitude share the same effect size parameter. The random-effects model applies if the effect size parameters vary from study to study, for studies at the same latitude.

While we have presented these three cases as being distinct from each other, the fact is that they can all be subsumed under the same general principle. Specifically,

- We use the fixed-effect model when all studies with the same predicted value have the same true effect size.
- We use the random-effects model when studies with the same predicted value have different true effect sizes.

Thus,

- In Case A the frame of reference is all studies, and the predicted value for any study is the mean of all studies.
- In Case B the frame of reference is all studies within the subgroup, and the predicted value for any study is the corresponding subgroup mean.
- In Case C the frame of reference is all studies at the same latitude, and the predicted value for any study is the predicted value given by the regression equation.
If the predicted value for all studies within the frame of reference is the same, then the only source of variance is within-study variance ($V$) and so the fixed-effect model applies. Otherwise we need to account also for between-study variance ($\tau^2$) and so the random-effects model applies. Here, we present three examples to show how this idea applies to the three cases outlined above.

Case A - A simple analysis

Case A1. A pharmaceutical company draws a sample of 1000 patients for a randomized controlled trial, but cannot work with all at the patients at one time because of limited space. Therefore, the patients are randomly assigned to one of ten cohorts, and each cohort starts treatment in a different week. If we assume no training effect and no seasonal effect, and ensure that all procedures are identical from one cohort to the next, it follows that the treatment effect should be the same for all cohorts. If we treat each cohort as a separate study and use meta-analysis to synthesize the results, then the fixed-effect model would apply.

Case A2. Ten universities form a consortium to run studies with the identical protocol (one at each university) and then use meta-analysis to synthesize the results. This is similar to Case A1, but the assumption that the effect size will be the same in all studies is more tenuous. It might be possible to ensure that the intervention is identical at all the universities, but it might not. It might be possible to ensure that the subjects are identical in all relevant respects at all the universities, but it might not. Therefore, it’s possible that the fixed-effect model would apply, but it might not. In this case it would probably be a good idea to use the random-effects model. If it turns out empirically that the effect sizes do differ, then the random-effects model will have been the correct choice. If it turns out empirically that the effect sizes do not differ, the random-effects weights will be identical to the fixed-effect weights, and so there is no price to pay for having chosen the random-effects model.

Case A3. The vast majority of simple meta-analyses are not similar to Case A1 nor to Case A2. Rather, they involve studies planned and performed by different teams of researchers without prior coordination. For example, we may locate 10 studies in the literature that seem to address the same fundamental question. We may decide that the studies are similar enough that it makes sense to perform a synthesis (for example, they all tested the same intervention) but there is no reason to assume that the true effect size will be identical in all studies. Rather, it is likely that the impact of the intervention will be affected (at least a little) by details of the sample (age, history), of the intervention itself (dose, duration), and/or the outcome measure (one test or another). If the studies are fundamentally the same (in the sense that they address the same basic question) then meta-analysis may enable us to identify the core impact of the intervention, cutting through the noise created by these differences. However, to identify this impact properly we need to take account of this noise by using random-effects weights. In this case, the random-effects model is a more plausible fit for the data.

Case B - Subgroups

The same idea is easily extended to the case of an analysis where we want to compute the summary effect size for two subgroups of studies, and then compare the two.

Case B1. In Case A1 (above), the pharmaceutical company randomly assigned patients to one of ten identical cohorts to test Drug-A vs. Control. Suppose that the next year the company did exactly the same thing with a new sample of patients, to test Drug-B vs. Control. The Drug-A studies are Subgroup-
A and the Drug-B studies are Subgroup-B. There is no expectation that the effect size will be the same in the two subgroups. Indeed, we may expect that the effect size will be greater in one subgroup than the other. However, we do expect all studies in the first group to have the same effect size as each other, and all studies within the second group to have the same effect size as each other. The fixed-effect model applies here.

Case B2. In Case A2 (above), the consortium planned to run ten identical studies (one at each university). They hoped that studies would be identical in all ways that could impact on the effect size, but could not ensure that this would be the case. Suppose that the next year they continued the arrangement, and ran ten more studies, this time using a variant of the intervention. The same logic applies as in Case A2, the only difference being that the logic now applies within subgroups. While it’s possible that all studies within a subgroup share the same effect size, it’s also possible that the effect size varies. Therefore, it’s probably safer to apply the random-effects model to compute the summary effect size within subgroups.

Case B3. In Case A3 (above), researchers located 10 published studies that were similar, but not identical. Logic dictates that the effect size will vary from study to study, and therefore the random-effects model is appropriate. The same example can be extended to subgroups. Suppose researchers located 10 studies that assessed the impact of Drug-A vs. Control (Subgroup-A), and another 10 that assessed the impact of Drug-B vs. control (Subgroup-B). We may expect that the effect sizes within either subgroup to be similar to each other, but we have no reason to expect that they will be identical to each other. Therefore, the random-effects model is a more plausible fit for the data within subgroups.

Case C – Regression

Finally, the same idea can be extended to the case where we use a continuous covariate (or a set of covariates) to predict effect size.

Case C1. In Case A1 (above), the pharmaceutical company randomly assigned patients to one of ten identical cohorts to test Drug-A vs. Control. We can extend Case A1 to a situation where the company runs ten identical cohorts, and assume that it repeats this process five times, each time with a different dose of the drug. The meta-analysis looks at the relationship of dose with effect size. For all studies at the same dose, the effect size should be the same. The fixed-effect model makes sense here.

Case C2. In Case A2 (above), the consortium planned to run ten identical studies (one at each university). We can extend Case A2 to a situation where the consortium runs ten studies based on the same protocol, and assume that it repeats this process five times, each time with a longer intervention. The meta-analysis looks at the relationship of the intervention’s duration with effect size. For all studies of the same duration, the effect size might be the same, but might not. The random-effects model is probably the better choice.

Case C3. In Case A3 (above), when studies are performed using different protocols, logic dictates that the effect size will vary from study to study. We can extend Case A3 to a situation where we locate all studies that assessed the impact of an intervention, and then code them based on the dosage. While studies with a similar dose may tend to have similar effect sizes, logic dictates that the effect size for any given dose will still vary. The random-effects model is a more plausible fit for the data.
How the model affects the analysis

The selection of a statistical model must be based on the sampling frame (as outlined above) and not the fact that one model will yield a more desirable estimate of the effect size or its precision (as outlined below). That said, it’s helpful to understand how the selection of one model or the other affects the estimates of effect size and precision.

Again, it’s easiest to explain this for the simple case, and then extend the example to the case of subgroups and regression.

Recall that \( V \) represents the within-study variance (the variance of the observed effects about the true effect for that study’s population), while \( T^2 \) represents the between-study variance (the variance of true effects about the mean true effects for all studies in the universe of populations).

The basic idea is that uncertainty (and therefore weights) for the fixed-effect model weights are based on \( V \), whereas uncertainty (and therefore weights) for the random-effects model weights are based on \( V + T^2 \). The only thing that changes as we move from Case A to Case B to Case C is the frame of reference for estimating \( T^2 \), as follows.

- When we are working with a single population, \( T^2 \) reflects the dispersion of true effects across all studies, and is therefore computed for the full set of studies.
- When we are working with subgroups, \( T^2 \) reflects the dispersion of true effects within a subgroup, and is therefore computed within subgroups.
- When we are working with regression, \( T^2 \) reflects the dispersion of true effects for studies with the same predicted value (that is, the same value on the covariates) and is therefore computed for each point on the prediction slope. As a practical matter, of course, most points on the slope have only a single study, and so this computation is less transparent than that for the single population (or subgroups) but the concept is the same.

The practical implications of using a random-effects model rather than a fixed-effects model are the same for all three cases (simple analysis, subgroups, and regression). To wit,

- The random-effects model will lead to more moderate weights being assigned to each study. As compared with a fixed-effect model, the random-effects model will assign more weight to small studies and less weight to large studies.
- Under the random-effects model, the confidence interval about each coefficient (and slope) will be wider than it would be under the fixed-effect model.
- Under the random-effects model, the \( p \)-values corresponding to each coefficient and to the model as a whole are less likely (on average) to meet the criterion for statistical significance.
These points are evident in Figure 4 and Figure 5, which show the regression of log risk ratio on latitude using fixed-effect and random-effects weights, respectively.
Relative weights

Under the fixed-effect model (Figure 4) the study weights tend to be more extreme, with large studies getting substantially more weight than small studies. Under the random-effects model (Figure 5), the study weights tend to be more moderate, with relatively small differences between the studies.

The selection of a model will have a substantial impact on the regression line if studies that fall outside the pattern happen to be especially large. These studies will tend to pull the regression line towards themselves, and this will be more important under fixed effect weights than under random-effect weights. (Conversely, a small study will have more impact on the regression line when we apply random-effects weights). In the BCG example the larger studies tend to fall within the pattern of the others, and so the regression lines are similar in the two plots.

Absolute weights

Under the fixed-effect model there is only one source of sampling variance (within-study). Therefore the weights (which are the inverse of the variance) are relatively large, which yields a relatively narrow confidence interval (Figure 4). Under the random-effects model there is an additional source of sampling variance (between-study). Therefore the weights are smaller, which yields a wider confidence interval (Figure 5).
PUTTING REGRESSION IN CONTEXT

In primary studies, we sometimes think of analyses as belonging to one of three types.

- **Case A.** Simple analysis, where the goal is to estimate the mean
- **Case B.** Analysis of variance, where the goal is to estimate the mean effect in two (or more) subgroups of subjects, and then see if (and how) the mean varies by subgroup
- **Case C.** Regression, where the goal is to estimate the mean effect for subjects that share the same values on one (or more) covariates, and then see if (and how) the mean varies as a function of the covariate values.

In fact, though, regression is a general system that includes all three of these cases. In other words, we can use regression not only in Case C but also in Cases A and B. If we did so, we would get the identical answers using regression that we get using a simple analysis or analysis of variance.

The same holds true for meta-analysis. While we tend to employ meta-regression only for Case C, we also have the option to use it for Case A or B. We will take advantage of this fact to help explain meta-regression. Specifically,

- For Case A we will perform an analysis using the traditional approach and then using regression, to show the correspondence between the two.
- For Case B we will perform a subgroups analysis using the traditional approach and then a regression, to show the correspondence between the two.
- For Case C, there is no simpler approach, and so we will move directly to the regression.

The goals of a fixed-effect analysis are different than the goals of a random-effects analysis, and for that reason we will address each separately. We’ll run through this sequence of cases (A, B, C) using the fixed-effect model, and then for the random-effects model.

In the text that follows we focus on conceptual issues. The statistical formulas that underlie these issues are presented in Appendix 2: Understanding Q and Appendix 3: Tests of heterogeneity.
FIXED-EFFECT MODEL

Basic analysis (Case A)

Here, we present a meta-analysis of the BCG studies. This is a basic meta-analysis in the sense that our goal is simply to estimate the mean effect size for the full set of studies. We will perform this exercise (1) using the traditional approach and then (2) using meta-regression, to show the correspondence between the two.

The traditional approach

Figure 6 and Figure 7 are screen-shots from CMA using the traditional approach to meta-analysis. Note that the effect size is in log units [A]. Each row shows the effect size and confidence interval for one study.

The lines marked “Fixed” in Figure 6 [B] and in Figure 7 [C] show the summary (common) effect size as −0.4303, with a 95% confidence interval of −0.5097 to −0.3509. The Z-value for a test of the null is −10.6247 with a corresponding p-value of < 0.0001.
Test of effect size

*Is the effect size zero?*

The line labeled “Fixed” in Figure 6 [B] and Figure 7 [C] show the effect size is $-0.4303$ with a standard error of $0.0405$. The $Z$-value for a test of the null is $-10.6247$ with a corresponding $p$-value of $< 0.0001$. We would conclude that the common effect size is probably not zero.

![Figure 7](image)

**Figure 7 | Basic analysis | Fixed-effect | Log risk ratio**

Test of the statistical model

*Is the data consistent with the fixed-effect model?*

In Figure 7, the section labeled Heterogeneity [D] presents statistics that address the heterogeneity in effect size. The $Q$-value is $153.2330$ with 12 degrees freedom and a corresponding $p$-value under $0.0001$. This tells us that the true effect size probably varies across studies, which means that the data are not consistent with the assumptions of the fixed-effect model.

The regression approach

We can perform the same analysis using meta-regression. Figure 8 shows the screen in CMA where we define the model. Since our goal here is to estimate the common effect size (that is, the intercept) we have included no covariates except for the intercept [E]. The results are shown in Figure 9.

![Figure 8](image)

**Figure 8 | Regression setup | Intercept only**
Test of effect size

*Is the effect size zero?*

Since there are no covariates, the predicted effect is simply the intercept, and so the question “Is the effect size zero?” is addressed by a test of the intercept.

In Figure 9, the regression equation \( F \) gives the predicted effect size for all studies as \( Y = -0.4303 \) with a standard error of 0.0405, variance of 0.0016, and confidence interval of \(-0.5097\) to \(-0.3509\). The \( Z \)-value for a test of the null is \(-10.6247\) with a corresponding \( p \)-value of < 0.0001. We would conclude that the effect size is probably not zero. Note that these numbers match the numbers from the traditional analysis in Figure 6 [B] and Figure 7 [C].

Analysis of variance

In Figure 9, results for the fixed-effect regression are presented using an analysis of variance framework, where the total [I] weighted sum of squares (WSS, or \( Q \)) is partitioned into the part explained by the predictive model [G] and the residual [H].

Model

The line labeled “Model” [G] displays the WSS explained by the predictive model. Since there are no covariates in this example, this line has no relevance here. The \( Q \)-value and \( df \) are displayed as 0.0, and the \( p \)-value as 1.0.
Residual

The line labeled “Residual” [H] displays the WSS not explained by the model and tests the hypothesis that all studies share a common (true) effect size. Since the $Q$ value is 152.2330 with $df = 12$ and $p < 0.0001$ we conclude that the true effect size probably varies across studies. Thus, the data are not consistent with the assumptions of the fixed-effect model. Note that the numbers are identical to those in Figure 7 [D].

Total

The line labeled “Total” [I] displays the total WSS for the full set of studies (with no predictors). In this case the total WSS is the same as the residual WSS, since both are based on the variance across all 13 studies. The $Q$ value is 152.2330 with $df = 12$ and $p < 0.0001$. Again, the numbers are identical to those in Figure 7 [D].

Summary

Our goal here was to show the correspondence between a traditional analysis and a regression for a simple analysis.

- In a meta-analysis with no covariates we want to estimate (and test) the effect size. This question is addressed by the common effect ($-0.4303$) in the traditional analysis, and by the intercept ($-0.4303$) in the regression. In both cases the standard error is 0.0405 and the $p$-value is $< 0.0001$, which tells us that the true effect size is probably not zero.

- We also want to know if the data are consistent with the fixed-effect model. This question is addressed by the $Q$-test for heterogeneity in the traditional analysis, and by the $Q$-test for the residual in the regression. In both cases the $Q$-value is 152.2330 with $df = 11$ and $p < 0.0001$. This tells us that the true effect size probably varies across studies, and so the assumptions of the fixed-effect model have been violated.

Subgroups analysis (Case B)

Above, we found that the impact of the vaccine varied from study to study. The researchers hypothesized that this variation might be explained by the fact that studies were conducted in various locations, and that the vaccine was more effective in colder climates. To test this hypothesis we can classify each study as being either “Cold” or “Hot” based on its latitude, and then perform an analysis (a) to estimate the effect size in each subgroup of studies, and (b) to compare the effect size for the two subgroups. (This differs from the original analysis, where the researchers used latitude as a continuous covariate rather than creating two groups).

The traditional approach

In Figure 10 the studies have been divided into subgroups as follows.
In Figure 10 the studies have been divided into subgroups.

- The six “Cold” studies are at the top, followed by their common effect, a log risk ratio of $-0.9986$ [A].
- The seven “Hot” studies are at the bottom, followed by their common effect, a log risk ratio of $-0.1115$ [B].

The same statistics are also shown in the “Fixed effect” section of Figure 11. The lines labeled [A] and [B] in Figure 11 correspond to the lines labeled [A] and [B] in Figure 10.

Is the common effect size zero for each subgroup?

For the cold studies this is addressed by Figure 10 [A]. The same numbers are displayed in Figure 11 [A]. The effect size is $-0.9986$ with a standard error of 0.0676, variance of 0.0046, and confidence interval of $-1.1310$ to $-0.8662$. The Z-value for a test of the null is $-14.7808$ with a corresponding $p$-value of $<0.0001$. We would conclude that the common effect size for studies in cold climates is probably not zero.

For the hot studies this is addressed by Figure 10 [B]. The same numbers are displayed in Figure 11 [B]. The effect size is $-0.1115$ with a standard error of 0.0506, variance of 0.0026, and confidence interval of $-0.2107$ to $-0.0124$. The Z-value for a test of the null is $-2.2042$ with a corresponding $p$-value of 0.0275. We would conclude that the common effect size for studies in hot climates is probably not zero.
Analysis of variance

In Figure 11, the section labeled Heterogeneity [C] shows how the total variance is partitioned into its component parts. Where analysis of variance in a primary study is based on sums of squares (SS), the subgroups analysis in a meta-analysis is based on weighted sums of squares (WSS, called $Q$). Still, the basic idea is the same. The total $Q$ can be partitioned into its component parts – the $Q$ explained by the subgroups and the $Q$ within subgroups (and thus unexplained, or residual).

**Total within**

The fixed-effect model requires that all studies within the same subgroup share the same true effect size. This assumption is tested by the $Q$-statistic where $Q$ and its degrees of freedom are computed within subgroups and then summed across subgroups. Here, $Q = 41.7894$, $df = 11$, and $p < 0.0001$. This tells us that effects probably do vary within subgroups, and the fixed-effect model is not valid [D].

**Total between**

The line labeled “Total between” is a test of the predictive model. Here, it addresses the question “Does effect size vary by subgroup?” The $Q$-value of 110.4436 with $df = 1$ and $p < 0.0001$ tells us that it probably does vary by subgroup [E].

**Overall**

The line labeled “Overall” reflects the total dispersion. It addresses the question “Do the effects vary from each other if we ignore the subgroups and compute the variance of all studies about the grand mean?” The $Q$-value of 152.2330 with $df = 12$ and $p < 0.0001$ tells us that they probably do vary [F]. Note that this is the same question we asked in the simple analysis (with no subgroups), and so it follows that the statistics for this section (Overall) are identical to those we saw for the simple analysis in Figure 7 [D].
Note that the variance components are additive. The $Q$-value within the Hot studies plus the $Q$-value within the Cold studies yields the total $Q$-value within subgroups. Then, the $Q$-value within subgroups plus the $Q$-value between-subgroups yields the total $Q$-value.

**The regression approach**

We can perform the same analysis as a meta-regression.

Figure 12 shows the screen where we define the model.

- The first covariate is the intercept.
- The second covariate is the variable called Climate: Hot. This covariate will address the question of whether or not the effect size varies by climate. The sub-designation (Hot) follows the convention that variables are named and coded for the presence of an attribute. Since the variable is called “Hot”, Cold studies will be coded 0 while Hot studies will be coded 1.

![Figure 12 | Regression Cold vs. Hot | Setup](image-url)
Results are displayed in Figure 13.

In Figure 13, the Analysis of variance shows how the total WSS is partitioned into its component parts – the WSS explained by the Model (here, subgroups) and the Residual WSS (here, within subgroups).

**Model**

The line labeled “Model” asks if the predictive model (climate) explains any of the variance in effect size. Put another way, it asks if the dispersion of effects about the regression line is smaller when the regression line is based on climate rather than based solely on the grand mean. The analysis shows that $Q = 110.4436$ with $df = 1$ and $p < 0.0001$, so we conclude that the predictive model probably explains (at least) some of the variance in effect size.

**Residual**

The line labeled “Residual” asks if the data are consistent with the model’s assumption of a common effect size for all studies with the same climate. The $Q$ value is $41.7894$ with $df = 11$ and $p < 0.0001$. We conclude that the data are not consistent with the assumptions of the fixed-effect model.
**Total**

The line labeled “Total” \([L]\) asks if the between-study variance for the full set of studies (with no subgroups) is zero. This analysis is the same whether or not there are subgroups and so, as in the prior analysis, \(Q = 152.2330, \ df = 12, \ p < 0.0001.\)

Note that the variance components are additive. The \(Q\)-value for the residual plus the \(Q\)-value for the model yields the total \(Q\)-value.

**Prediction equation**

The prediction equation \([I]\) is \(-0.9986 + 0.8870 \times \text{Climate}.\)

Since Climate is coded 0 for Cold and 1 for Hot, the predicted value for Cold studies is

\[
Y = -0.9986 + 0 \times 0.8870 = -0.9986, \tag{1.1}
\]

while the predicted value for Hot studies is

\[
Y = -0.9986 + 1 \times 0.8870 = -0.1116. \tag{1.2}
\]

Note that these are the same numbers we saw in the subgroups analysis.

- Figure 10 [A] and Figure 11 [A] showed the mean effect size for the Cold studies as \(-0.9986,\) which is the same number we see in (1.1).
- Figure 10 [B] and Figure 11 [B] showed the mean effect size for the Hot studies as \(-0.1196,\) which is the same number we see in (1.2).
Summary

The total $Q$ of each effect size about the grand mean can be partitioned into its component parts – the $Q$ due to the variation in effect size that can be explained by subgroup membership, and the part that cannot. The traditional approach and the regression approach use somewhat different nomenclatures but are identical mathematically and yield precisely the same answers.

- The $Q$-between (in the traditional model) and the $Q$-model (in the regression) are both 110.4436 with $df = 1$ and $p < 0.0001$. Each tells us that effect size probably differs between subgroups.
- The $Q$-within (in the traditional model) and the $Q$-residual (in the regression) are both 41.7894 with $df = 11$ and $p < 0.0001$. Each tells us that the assumptions of the fixed-effect model have been violated.
- The $Q$-total in each case is 152.2330 with $df = 12$ and $p < 0.0001$. Each tells us that effect sizes vary when we ignore subgroups and work with deviations of all studies from the grand mean.
- The $Q$-values are additive. $Q$-between plus $Q$-within equals $Q$-total.

Continuous covariate (Case C)

Immediately above (Case B) we divided the studies into “Hot” or “Cold” climates, which allowed us to perform a subgroups analysis. In the original paper, the researchers worked with the absolute latitude of each study as a continuous covariate. We turn to that analysis now.

The traditional approach

There is no mechanism to work with a continuous covariate in the traditional framework.

The regression approach

Figure 14 shows the screen where we define the regression. We will use intercept and Latitude to predict the effect size. The covariate will address the question of whether or not effect size is related to latitude. Figure 15 shows the results of this analysis.
In the section labeled Analysis of variance [A] the WSS-total is partitioned into its component parts – the WSS explained by latitude (**Model**) and the WSS-residual.

**Model**

The lines labeled “Model” [B] addresses the hypothesis that the predictive model (latitude) explains any of the variance in effect size. Put another way, it asks if the dispersion of effects about the regression line is smaller when the regression line is based on latitude rather than based solely on the grand mean. Since $Q = 121.4999$ with $df = 1$ and $p < 0.0001$, we conclude that the predictive model probably explains (at least) some of the variance in effect size.

**Residual**

The line labeled “Residual” [C] addresses the hypothesis that the data are consistent with the model’s assumption of a common effect size for all studies at the same latitude. The $Q$ value is $30.7331$ with $df = 11$ and $p = 0.0012$. We conclude that the data are not consistent with the assumptions of the fixed-effect model. Rather, the true effect size does vary from study to study, even for studies at the same latitude.
**Total**

The line labeled “Total” [D] addresses the hypothesis that the variance for the full set of studies (with no predictors) is zero. This analysis is the same whether or not there are subgroups or covariates and so, as in the prior analyses, \( Q = 152.2330, \, df = 12, \, p < 0.0001. \)

**Summary**

The total \( Q \) of each effect size about the grand mean can be partitioned into its component parts – the \( Q \) due to the variation in effect size that can be explained by latitude, and the part that cannot.

- The \( Q \)-value for the model is \( Q = 121.4999 \) with \( df = 1 \) and \( p < 0.0001 \), which tells us that effect size is related to latitude.
- The \( Q \)-value for the residual is \( 30.7331 \) with \( df = 11 \) and \( p = 0.0012 \), which tells us that the assumptions of the fixed-effect model have been violated.
- The \( Q \)-value for the total is \( 152.2330 \) with \( df = 12 \) and \( p < 0.0001 \), which tells us that effect sizes vary when we ignore latitude and work with deviations of all studies from the grand mean.

**In context**

We presented three cases to show the correspondence between regression and a traditional analysis. In Case A there were no covariates; in Case B there was a categorical covariate; and in Case C there was one continuous covariate.

For the traditional analysis the fixed-effect model requires that all studies share the same true effect size (Case A), or that all studies within a subgroup share the same effect size (Case B). This assumption is tested by the \( Q \) value based on deviations from the grand mean (Case A), or by the \( Q \) value based on deviations from each study’s subgroup mean (Case B).

The regression model is a more general model and allows us to cover all cases by saying that the effect size must be identical for all studies with the same predicted value. In Case A the predicted value is the grand mean (as it was for the traditional analysis). In Case B the predicted value is the subgroup mean (as it was for the traditional analysis). In Case C the predicted value is the point on the regression line corresponding to the regression equation.
RANDOM-EFFECTS MODEL

Immediately above, we showed how to interpret the regression under the fixed-effect model. Now, we show how to interpret the same regression under the random-effects model. Many of the statistics are different but, more fundamentally, many of the questions addressed by the analysis are different.

Basic analysis (Case A)

Here, we present a meta-analysis of the BCG studies. This is a basic meta-analysis in the sense that our goal is simply to estimate the mean effect size for the full set of studies. We will perform this exercise (1) using a traditional approach and then (2) using meta-regression, to show the correspondence between the two.

The traditional approach

Figure 16 is a screen-shot from CMA using the traditional approach to meta-analysis. Note that the effect size is in log units [A]. Each row shows the effect size and confidence interval for one study.

The line marked “Random” in Figure 16 [B] shows the summary (mean) effect size as \(-0.7141\), with a 95% confidence interval of \(-1.064\) to \(-0.3638\). The Z-value for a test of the null is \(-3.9952\), and the corresponding \(p\)-value is 0.0001. The same numbers are displayed in Figure 17 [C].
Test of effect size

**Is the mean effect size zero?**

The line labeled “Random” in Figure 16 [B] and Figure 17 [C] shows that the effect size is −0.7141 with a standard error of 0.1787. The Z-value for a test of the null is -3.9952 with a corresponding p-value of less than 0.0001. We would conclude that the mean effect size is probably not zero.

Heterogeneity

**Is there any unexplained variance in the true effect sizes?**

The mean effect size is −0.7141. Is it possible that all the observed variance about this mean reflects sampling error, or is there evidence that some of this variance reflects differences in the true effect size across studies?

This is addressed in Figure 17 by the section labeled “Heterogeneity” [D]. The Q-value is 152.2330 with df = 11 and p < 0.0001. This tells us that it’s unlikely that all of the variance is due to sampling error. We conclude that the true effect size probably does vary from study to study.

Note that the heterogeneity section for the random-effects analysis (Figure 17) is identical to the one for the fixed-effect analysis (Figure 7). In both cases the heterogeneity statistics are based on the weights $1/V$ (that is, where the only sampling error is within-studies). While the numbers are identical for the fixed-effect and the random-effects models, the interpretation differs. Under the fixed-effect model, the presence of heterogeneity in the true effects tells us that the statistical model does not match the data. Under the random-effects model, by contrast, this heterogeneity is employed to estimate $T^2$, which is then incorporated into the weights assigned to each study.

**How much variance is there?**

This is addressed in the section labeled “Tau-squared” in Figure 17 [E].

The between-studies variance ($T^2$) is estimated as 0.3088. The between-studies standard deviation $T$ is simply the square root of $T^2$, or 0.5557.

**What proportion of the observed variance is true variance?**


Some of the observed variance is due to real differences in effect size, while some reflects sampling error. The $I^2$ statistic \([D]\) reflects the proportion of variance that is due to real differences (and thus potentially explainable by covariates). In this case $I^2$ is 92.1173\%, which means that almost all of the observed variance reflects real differences in study effects.

### The regression approach

We can perform the same analysis using meta-regression. Figure 18 shows the screen in CMA where we define the model \([F]\). We’ve used no covariates except for the intercept.

![Figure 18 | Regression | Intercept | Setup](image.png)
Figure 19 shows the results of this analysis.

![Figure 19 | Regression | Intercept | Main results | Random-effects](image)

**Main results for Model 1, Random effects (MM), Z-Distribution**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.7141</td>
<td>0.1787</td>
<td>-1.0644</td>
<td>-0.3638</td>
<td>-3.9952</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**Statistics for Model 1**

- **Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero**
  
  \( Q = 0.0000, \text{df} = 0, p = 1.0000 \)

- **Goodness of fit: Test that unexplained variance is zero**
  
  \( \text{P} = 0.3088, \text{P} = 0.5557, I^2 = 92.12\%, Q = 152.2330, \text{df} = 12, p = 0.0000 \)

**Comparison of Model 1 with the null model**

- **Total between-study variance (intercept only)**
  
  \( \text{P} = 0.3088, \text{P} = 0.5557, I^2 = 92.12\%, Q = 152.2330, \text{df} = 12, p = 0.0000 \)

- **Proportion of total between-study variance explained by Model 1**

  \( R^2 \text{ analog} = 0.00 \)

**Number of studies in the analysis**

13

**Test of effect size**

**Is the mean effect size zero?**

Since there are no covariates, the predicted effect is simply the intercept, and so the question “Is the mean effect size zero?” is addressed by a test of the intercept.

In Figure 19 the regression equation \([G]\) gives the predicted effect size for all studies as \( Y = -0.7141 \) with a standard error of 0.1787 and confidence interval of \(-1.0644 \) to \(-0.3638 \). The \(Z\)-value for a test of the null is \(-3.9952\) with a corresponding \(p\)-value of \(< 0.0001\). We would conclude that the mean effect size is probably not zero. Note that these numbers match the numbers in Figure 16[B] and Figure 17[C].

**Test of the model**

The line labeled Model \([H]\) addresses the hypothesis that the covariates explain any of the variance in effect size. Since there are no covariates in this model, this section is not relevant. The \(Q\)-value is shown as 0.0, the df as 0, and the \(p\)-value as 1.0.
Goodness of fit

Is there any unexplained variance in the true effect sizes?

The predicted effect size for each study is simply the intercept, $-0.7141$. Is it possible that all the observed variance about the mean reflects sampling error, or is there evidence that some of this variance reflects differences in the true effect size across studies? This line is called “Goodness of fit” since the presence of true variance means (by definition) that some variance remains unexplained. That is, the prediction model does not “Fit” (fully explain) the variance in effect sizes.

The $Q$-value is 152.2330, with $df = 11$ and $p < 0.0001$. We conclude that the true effect size probably does vary from study to study. These statistics are identical to those in Figure 17 [D] for the traditional analysis.

How much variance is there?

The between-studies variance ($T^2$) is estimated as 0.3088. The between-studies standard deviation ($T$) is then the square root of $T^2$, or 0.5557. These values correspond to the values in Figure 17 [E].

What proportion of the observed variance is true variance?

Some of the observed variance is due to real differences in effect size, while some reflects sampling error. The $I^2$ value reflects the proportion of variance that is due to real differences (and thus potentially explainable by covariates). In this case $I^2$ is 92.1173%, which means that almost all of the observed variance reflects real differences in study effects. This corresponds to the value in Figure 17 [D].

Graphic

In Figure 20 we’ve plotted all 13 studies, and we’ve also plotted the regression line [L]. Since there are no covariates in the predictive model, the regression line is horizontal. That is, the predicted value for every one of the studies is the intercept (which is also the mean) of $-0.7141$. The $Q$-statistic was computed by working with the deviation of every study from this predicted value. Note that this graphic applies to both the traditional analysis (where the predicted value is the mean) and the regression (where the predicted value is the intercept) since these values are identical.
In addition to plotting the mean treatment effect [L] we can also plot the distribution of (true) treatment effects about this mean. Concretely, if we assume that these effects are normally distributed with a mean of \(-0.7141\) and a standard deviation \((T)\) of 0.5557, we would expect some 95% of all true effects to fall in the approximate range of \(-1.8033\) to 0.3751 (that is, the mean +/- 1.96 \(T\)). This range is represented by the normal curve superimposed on the plot, which is centered at \(-0.7141\) and extends from \(-1.8033\) [N] to 0.3751 [M]. The figure is drawn to scale, and we can see that the curve captures almost all of the effect sizes.

**Note**

The curve is intended to capture most of the dispersion in *true* effects, not in observed effects. As it happens, in this example the observed effects and the true effects are very similar (\(I^2\) is 92%, which means that almost all of the observed dispersion is real) and for that reason, most of the effects fall within the curve. However, this will not always be the case. For example, suppose that \(I^2\) had been 25%. In this case the *true* dispersion would have been less, and the curve would have been smaller. While the curve would still capture most of the dispersion in *true* effects, many of observed effects would fall outside the curve.

**Comparison of Model 1 with the null model**

Returning to Figure 19, the goal of section [J] is to estimate \(R^2\), the proportion of between-study variance \((T^2)\) explained by the model. For this purpose we need two estimates of \(T^2\) – with predictors [I] and without predictors [J]. We will then use these two numbers to compute \(R^2\) [K].

**Total between-study variance**

The line labeled [J] displays the unexplained between-study variance \((T^2)\) when there are no predictors in the model, which is 0.3088. This is the same value displayed in the traditional analysis (Figure 17 [E]).
Line [J] gives the original $T^2$ while line [I] gives the residual $T^2$ after we’ve entered all the covariates. In this example, since there are no covariates in the model, the two values are identical. When there are covariates in the model, the estimate of $T^2$ on line [J] may be less than the estimate on line [I], and the difference would be used to compute $R^2$, the proportion of the original variance explained by the covariates.

**Proportion of variance explained**

This section of the screen [K] is used to report what proportion of the total variance can be explained by the predictive model. In this example there are no covariates. Therefore, the estimate of $T^2$ on line [J] is the same as the estimate on line [I] and $R^2$ is shown as 0.00.

**Summary**

Our goal here was to show the correspondence between a traditional analysis and a regression for a simple analysis.

- In a meta-analysis with no covariates we want to estimate (and test) the effect size. This question is addressed by the mean effect (~0.7141) in the traditional analysis, and by the intercept (~0.7141) in the regression. In both cases the standard error is 0.1787 and the $p$-value is 0.0001, which tells us that the true mean effect size is probably not zero.

- We also want to know if there is evidence that the true effect size varies across studies. This question is addressed by the $Q$-test for heterogeneity in the traditional analysis, and by the $Q$-test for the residual in the regression. In both cases the $Q$-value is 152.2330 with $df = 11$ and $p < 0.0001$, which tells us that the true effect sizes probably do vary.

- Finally, we want to estimate the variance in true effect sizes. This estimate, called $T^2$, is 0.3088 in both cases. This variation is incorporated into the model, and affects the weights assigned to each study.

**Subgroups analysis (Case B)**

Immediately above (Case A) we found that the impact of the vaccine varied from study to study. The researchers hypothesized that this variation might be explained by the fact that studies were conducted in various locations, and that the vaccine was more effective in colder climates. To test this hypothesis we can classify each study as being either “Hot” or “Cold” and then perform an analysis (a) to estimate the effect size in each subgroup of studies, and (b) to compare the effect size for the two subgroups. (The researchers used latitude as a continuous covariate rather than creating two groups).
The traditional approach

In Figure 21 the studies have been divided into subgroups.

- The six “Cold” studies are at the top, followed by their mean effect, a log risk ratio of $-1.1987$ [A].
- The seven “Hot” studies are at the bottom, followed by their mean effect, a log risk ratio of $-0.2784$ [B].

![Figure 21 | Subgroups Cold vs. Hot | Random-effects](image)

The same information is displayed in Figure 22. Here, the top section is labeled “Fixed-effect” and reports statistics based on fixed-effect weights. The bottom one is labeled “Mixed-effects” and reports statistics based on random-effects weights within subgroups. (The label “Mixed-effects” refers to the fact that we use the random-effects model within subgroups, but not between subgroups.)

In Figure 22 we are working with the section labeled “Mixed effects” (which means that we’re using random-effects weights within subgroups). The lines marked [A] and [B] correspond to lines [A] and [B] in Figure 21, and show the mean effect size, standard error, variance, and confidence interval for the two subgroups.
Note on computing $T^2$

CMA offers options for computing $T^2$ in the presence of subgroups. Select the option to compute $T2$ within subgroups and then pool the estimates across subgroups, as shown in Figure 23 and Figure 24. For an explanation of these options, see Appendix 4: Computing $\tau^2$ in the presence of subgroups.

To apply this option in CMA, on the analysis screen:

- Select Computational options > Mixed and random effects options
- Select the option to Assume a common among-study variance across subgroups
- Select the option to Combine subgroup using a fixed-effect model
Is the mean effect size zero for each subgroup?

For the Cold studies this is addressed by Figure 21 [A] and Figure 22 [A]. The mean effect size is $-1.1987$ with a standard error of 0.1769, and confidence interval of $-1.5445$ to $-0.8518$. The test that the mean is zero is addressed by the z-value of $-6.7740$ and corresponding p-value of $< 0.0001$. We would conclude that the mean effect size in the universe of Cold studies is probably not zero.

For the Hot studies this is addressed by Figure 21 [B] and Figure 22 [B]. The mean effect size is $-0.2784$ with a standard error of 0.1522, and confidence interval of $-0.5767$ to $+0.0199$. The test that the mean is zero is addressed by the z-value of $-1.9289$ and corresponding p-value of $< 0.0674$. This fails to meet the traditional criterion alpha of 0.05, and so by this criterion we cannot reject the null that the effect size is zero (that the vaccine has no impact).

Test of the model

Is effect size related to subgroup membership?

This question is addressed by the section labeled “Mixed-effects analysis”. The lines marked Cold [A] and Hot [B] give the mean effect size for each group, based on random-effects weights. The line marked “Total between” is a test of the difference between these two values ($-1.1987$ vs. $-0.2784$). The $Q$-value for this difference is 15.5445 with 1 df, with a corresponding p-value of 0.0001 [C]. We conclude that effect size probably does differ by subgroups.

Heterogeneity

Is there any unexplained variance in the true effect sizes?
Immediately above, we saw that we can use information about a study’s subgroup to improve our ability to predict that study’s effect. That is, by using the subgroup mean rather than the grand mean to predict a study’s effect size, we are able to make a more accurate prediction. But does subgroup membership enable us to completely predict that study’s effect – do all studies within a subgroup share a common effect size? Or, is there variance in true effects within subgroups?

To test the assumption that there is no variance in true effect sizes within subgroups we compute $Q$ and $df$ within subgroups and then sum these values across subgroups as shown in Table 2. This table is taken from the section labeled [D] in Figure 22.

<table>
<thead>
<tr>
<th></th>
<th>$Q$</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold</td>
<td>20.3464</td>
<td>5</td>
<td>0.0011</td>
</tr>
<tr>
<td>Hot</td>
<td>21.4431</td>
<td>6</td>
<td>0.0015</td>
</tr>
<tr>
<td>Total</td>
<td>41.7894</td>
<td>11</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

For $Q = 41.7894$ with 11 $df$, the $p$-value is < 0.0001. This tells us that the true effect sizes do vary from study to study, even within subgroups. Put another way, the model is incomplete – knowing whether a study falls into the Cold or Hot subgroup does not allow us to completely predict its effect size.

**How much variance is there?**

The $Q$-values computed immediately above are employed to estimate the variance of true effect sizes ($\tau^2$) within subgroups. For the Cold subgroup $\tau^2$ is 0.1383. For the Hot subgroup $\tau^2$ is 0.0741. The combined estimate (computed within subgroups and combined across subgroups) is not shown on this screen, but is 0.0964 (see Appendix 4: Computing $\tau^2$ in the presence of subgroups). In each case the standard deviation of true effect sizes ($\tau$) is the square root of the variance. The combined estimate of $\tau$ is 0.3105.

**What proportion of the observed variance is true variance?**

Some of the observed variance within subgroups is due to real differences in effect size, while some reflects sampling error. The $I^2$ value reflects the proportion of this variance that is due to real differences (and thus potentially explainable by covariates). For Cold this value is 75.4256%, and for Hot this value is 72.0189%. The combined estimate is not shown on this screen, but is 73.6775%. This means that most of the within-subgroup variance reflects real differences in study effects (see Appendix 4: Computing $\tau^2$ in the presence of subgroups).

**The regression approach**

We can perform the same analysis as a meta-regression. Figure 25 shows the screen where we define the model.

- The first covariate is the intercept.
- The second covariate is the variable called Climate: Hot. This covariate will address the question of whether or not the effect size varies by climate. The sub-designation (Hot) follows the
convention that variables are named and coded for the presence of an attribute. Since the variable is called “Hot”, Cold studies will be coded 0 while Hot studies will be coded 1.

Figure 25 | Regression | Climate | Setup

Figure 26 | Regression | Climate | Main results | Random-effects

Figure 26 shows the results of the analysis.
It is important to note that the results shown on this page are collated from two separate analyses.

- We run one regression with the intercept and climate as predictors. This is the basis for the sections which report on the impact of each covariate [I], the test of the model [J], and the goodness of fit [K].

- We run a second regression with only the intercept. This is the basis for the section that reports the value of $T^2$ with no covariates (that is, the true variance about the grand mean) [L].

- Then we use the estimate of $T^2$ with covariates from [I] and without covariates [L] to compute the proportion of variance explained, or $R^2$ in section [M].

**Test of the model**

*Is effect size related to subgroup membership?*

The question of whether or not effect size is related to subgroup membership is addressed by the section labeled “Test of the model” [J]. The $Q$-value is 15.5445, and with $df = 1$ the $p$-value is 0.0001. These are the same values that we saw in Figure 22 [E]. We conclude that effect size does differ by subgroup.

Equivalently (since there is only one covariate in the model), the $Z$-value for climate is 3.9426 with a $p$-value of 0.0001. (The test of climate is based on $Z$, which is a standardized difference. The test of the model is based on $Q$, which is a squared index. When there is only one covariate, $Z^2$ is equal to $Q$. Here, $3.9426^2$ is equal to 15.5445).

**Goodness of Fit**

*Is there any unexplained variance in the true effect sizes?*

Immediately above, we saw that we can use information about a study’s subgroup to improve our ability to predict that study’s effect. But does this information enable us to completely predict that study’s effect – do all studies within a subgroup share a common effect size? Or is there variance in true effects within subgroups? This is called a Goodness of fit test, since we can say that the model provides a good for the effects if there is no evidence of unexplained heterogeneity.

To address this question we compute $Q$ working with the deviation of each study from its predicted effect, which is $-1.1987$ for the Cold studies and $-0.2784$ for the Hot studies. Computed in this way, $Q$ is 41.7894 with 11 $df$ and the corresponding $p$-value is less than 0.0001 [K]. This tells us that the true effect size varies from study to study, even within subgroups. Put another way, the model is incomplete – knowing whether a study falls into the Cold or Hot subgroup does not allow us to completely predict its effect size. This is the same value we saw in the traditional analysis Figure 22 [G].

*How much variance is there?*
In this same section [K] the program shows that $T^2$, the variance of true effect sizes about the subgroup mean, is 0.0964. It follows that $T$, the standard deviation of true effect sizes about the subgroup mean is 0.3105. These values refer to the dispersion of true effects within each of the subgroups, and are assumed to be the same for all subgroups.

**What proportion of the observed variance is true variance?**

The $I^2$ statistic [K] is 73.6775%, which means that nearly three-fourths of the observed variance that remains (that is, within subgroups) reflects real differences in study effects.

Note that the program reports $I^2$ for two separate analyses. The one on line [L] is the proportion of the total variance that represents between-study (true) variance, and can potentially be explained by study-level covariates. The one on line [K] is the proportion of the within-subgroups variance that represents between-study (true) variance, and can potentially be explained by study-level covariates.

**Graphic**

In Figure 27 we’ve plotted all 13 studies, and the regression line. (While the regression line is actually a line that intersects the Cold and Hot columns at specific points, we’ve taken the liberty of drawing a horizontal line at the points of intersection.)

The predicted value for the Hot studies is $-0.2784$ with a standard deviation of 0.3105. If we assume that the true effects are normally distributed about each predicted value we would expect the true effects for Hot studies to fall in the range of $-0.2784$ plus/minus 1.96 times 0.3105, or $-0.8870$ to 0.3302. In Figure 27 we have superimposed a normal curve on the study points to reflect this span of true effects [N].

The predicted value for the Cold studies is $-1.1199$ with a standard deviation of 0.3105. If we assume that the true effects are normally distributed about each predicted value we would expect the true effects for Cold studies to fall in the range of $-1.1199$ plus/minus 1.96 times 0.3105, or $-1.7285$ to $-0.5113$. In Figure 27 we have superimposed a normal curve on the study points to reflect this span of true effects [O].

Note that the plot shows the observed effects. By contrast, the curves are intended to capture some 95% of the true effects, which are assumed to fall closer to the predicted values.
Comparison of Model 1 with the null model

Returning to Figure 26, the intent of the section labeled “Comparison of Model 1 with the null model” is to report the proportion of variance explained by the model, an index analogous to $R^2$ in primary regression.

The index is a ratio of the explained variance to the total variance, and to get both numbers we need to run two separate regressions.

To get the initial amount of variance we run a regression with no covariates and compute $T^2$. Here, $T^2$ is 0.3088, which is the variance of all studies about the grand mean \[L\]

To get the variance that remains with the covariates, we run a regression with the covariates and compute $T^2$. Here, $T^2$ is 0.0964, which is the variance of all studies about the regression line \[K\]

**Proportion of variance explained \[M\]**

If the initial (total) $T^2$ is 0.3088 and the unexplained (residual) $T^2$ is 0.0964, then the ratio

$$\frac{T^2_{\text{Residual}}}{T^2_{\text{Total}}} = \frac{0.0964}{0.3088} = 0.6878 \quad (1.3)$$

gives us the proportion of variance that is *not* explained by the covariates. $R^2$, the proportion of variance that *is* explained by the covariates is then

$$R^2 = 1 - \left( \frac{T^2_{\text{Residual}}}{T^2_{\text{Total}}} \right) = 1 - \left( \frac{0.0964}{0.3088} \right) = 0.6878 \quad (1.4)$$

We show this graphically in Figure 28, which juxtaposes Figure 20 with Figure 27. At left, the normal curve \[P\] reflects the unexplained variance in effects when the predicted value for each study is the grand mean. At right, the normal curves \[Q,R\] represent the variance in effects when the predicted
value for each study is the corresponding subgroup mean. This is the variance not explained by subgroup membership. The variance at the right (0.0964) is less than the variance at the left (0.3088), which tells us that by using climate as a covariate we can reduce the unexplained variance – or (equivalently) explain some of the variance.

Equation (1.3) gives the ratio of the variance at right to the variance at left (the ratio of not explained to total). Then in equation (1.4) we subtract this value from 1.0 to get the value of $R^2$.

An equivalent approach to computing $R^2$ is as follows. If the initial (total) $T^2$ is 0.3088 and the unexplained (residual) $T^2$ is 0.0964, it follows that the difference (0.2124) is the $T^2$ explained by the model. Then we can compute $R^2$, the proportion explained by the model, as

$$R^2 = \frac{T^2_{\text{Explained}}}{T^2_{\text{Total}}} = \frac{0.2124}{0.3088} = 0.6878.$$ (1.5)

**Prediction equation**

The prediction equation [I] is $-1.1987 + 0.9203 \times \text{Climate}$.

Since Climate is coded 0 for Cold and 1 for Hot, the predicted value for Cold studies is

$$Y = -1.1987 + 0 \times 0.9203 = -1.1987,$$ (1.6)

while the predicted value for Hot studies is

$$Y = -1.1987 + 1 \times 0.9203 = -0.2784.$$ (1.7)

These are the same values that we saw for the subgroup means in Figure 22 [C] and [D].
Summary

The Q statistic

- The Q-Between (in the traditional model) and the Q-Model (in the regression) are both 15.5445, with $df = 1$ and $p = 0.0001$. Each tells us that effect size differs between subgroups.

- The Q-within (in the traditional model) and the Q for goodness of fit (in the regression) are both 41.7894, with $df = 11$ and $p < 0.0001$. Each tells us that the true effect size varies, even within subgroups.

- The Q-total in each case is 152.2330, with $df = 12$ and $p < 0.0001$. Each tells us that effect sizes vary when we ignore subgroups and work with deviations of all studies from the grand mean.

The $I^2$ statistic

- The $I^2$ statistic tells us what proportion of the variation in observed effects reflects variance in true effects rather than sampling error.

- When there are no covariates [L] the $I^2$ value is 92.1173%, which tells us that 92% of the observed variance is real, and may potentially be explained by covariates.

- When we use climate as a covariate [K] the $I^2$ value is 73.68%, which tells us that some 74% of the observed variance about the subgroup means is real, and may potentially be explained by additional covariates.

The $R^2$ Index

- The between-study variance is estimated at 0.0964 within subgroups, as compared to 0.3088 for the population as a whole. It follows that the variance (in log units) explained by subgroups is 0.2124. The ratio of explained/total corresponds to an $R^2$ of 0.6878, meaning that 68.7824% of the variance in true effects can be explained by climate. This is reflected in Figure 28, where the range of true effects about subgroup means is smaller than the range of true effects about the grand mean.

Continuous covariate (Case C)

In the prior analysis we classified each study as either “Cold” or “Hot” based on its absolute distance from the equator. In the original paper, the researchers did not classify the studies as Cold or Hot. Rather, they worked with the absolute latitude of each study as a continuous covariate. We turn to that analysis now.

The traditional approach

There is no mechanism to work with a continuous covariate in the traditional framework.
The regression approach

Figure 29 shows the screen where we define the regression using the intercept and latitude to predict the effect size. Figure 30 shows the results of this analysis.
As before, it is important to note that the results shown on this page are collated from two separate analyses.

- We run one regression with the intercept and climate as predictors. This is the basis for the sections which report on the impact of each covariate [A], the test of the model [B], and the goodness of fit [C].

- We run a second regression with only the intercept. This is the basis for the section that reports the value of $T^2$ with no covariates (that is, the true variance about the grand mean) [D].

- Then we use the estimate of $T^2$ with covariates from [C] and without covariates [D] to compute the proportion of variance explained, or $R^2$ in section [E].

**Prediction equation**

At the top [A] the program shows the coefficient for the intercept and for each covariate, along with the standard error, confidence interval, and significance test.

**Test of the model**

*Is effect size related to latitude?*

In Figure 30 the prediction equation [A] is $0.2595 - 0.0292 \times \text{Latitude}$.

This question of whether or not effect size is related to latitude is addressed by the section labeled “Test of the model” [B]. The $Q$-value is 18.8452 with 1 df and $p$-value of < 0.0001. We conclude that effect size probably is related to latitude.

Equivalently (since there is only one covariate in the model), the $Z$-value for latitude is $-4.3411$ with a $p$-value of < 0.0001. (The test of latitude is based on $Z$, which is a standardized difference. The test of the model is based on $Q$, which is a squared index. When there is only one covariate, $Z^2$ is equal to $Q$. Here, $-4.3411^2$ is equal to 18.8452).

**Goodness of fit**

*Is there any unexplained variance in the true effect sizes?*

Immediately above, we saw that we can use information about a study’s latitude to improve our ability to predict that study’s effect. But does this information enable us to completely predict that study’s effect? That is, do all studies at the same latitude share a common effect size? Or is there variance in true effects among studies at the same latitude?

This is addressed by the Goodness of Fit [C]. To compute a measure of dispersion we work with the deviation of each study from that study’s predicted effect size, where the predicted effect size is a function of each study’s latitude. Computed as a deviation from this predicted value, $Q$ is 30.7331 with 11 df, and the corresponding $p$-value is 0.0012. This tells us that the true effect size varies from study to study, even within latitudes. Put another way, the model is incomplete – knowing a study’s latitude
does not allow us to completely predict its effect size. (Unlike subgroups, we may not have multiple
studies at the same latitude, but the idea is the same – for each study we compute the deviation from
the prediction line to the observed effect size).

**How much variance is there?**

The program [C] shows that $T^2$, the variance of true effect sizes at any point on the regression line, is
0.0633. It follows that the $T$, the standard deviation of true effect sizes at any point on the regression
line is 0.2516.

**What proportion of the observed variance is true variance?**

- The $I^2$ statistic tells us what proportion of the variation in observed effects reflects variance in
true effects rather than sampling error.
- When there are no covariates [D] the $I^2$ value is 92.1173%, which tells us that 92% of the
  observed variance is real, and may potentially be explained by covariates.
- When we use latitude as a covariate [C] the $I^2$ value is 64.21%, which tells us that some 64% of
  the observed variance about the regression line is real, and may potentially be explained by
  additional study-level covariates.

**Graphic**

In Figure 31 we’ve plotted all 13 studies and the regression line. The Q-statistic was computed by
working with the deviation of every study from the regression line.

![Figure 31 | Dispersion of effects about regression line for latitude](image)
The estimate of the variance ($\tau^2$) is 0.0633 and of the standard deviation ($\tau$) is 0.2516. If we assume these effects are normally distributed about each predicted value we would expect the true effects for all studies to fall at the predicted value plus/minus 1.96 $\tau$, or within 0.4931 on either side of the predicted value. This holds true for any point on the regression curve, but for illustrative purposes we have superimposed a normal curve at a few arbitrary points on the regression line [F,G,H] to reflect this range.

Note that the plot shows the observed effects. By contrast, the curves are intended to capture some 95% of the true effects, which are assumed to fall closer to the regression line.

**Comparison of Model 1 with the null model**

Returning to Figure 30, the intent of the section labeled “Comparison of Model 1 with the null model” is to report the proportion of variance explained by the model, an index analogous to $R^2$ in primary regression.

The index is a ratio of the variance explained to the total variance, and to get both numbers we need to run two separate regressions.

To get the initial amount of variance we run a regression with no covariates and compute $\tau^2$. Here, $\tau^2$ is 0.3088, which is the variance of all studies about the grand mean [C].

To get the variance that remains with the covariates, we run a regression with the covariates and compute $\tau^2$. Here, $\tau^2$ is 0.0633, which is the variance of all studies about the regression line [D].

**Proportion of variance explained**

To get the final amount of variance we run a regression with the covariates and compute $\tau^2$. This value, reported above [C] as 0.0633, is the variance of studies about their predicted value.

If the initial (total) $\tau^2$ is 0.3088 and the unexplained (residual) $\tau^2$ is 0.0633, then the ratio

$$\frac{\tau^2_{\text{Residual}}}{\tau^2_{\text{Total}}} = \frac{0.0633}{0.3088} = 0.7950$$

(1.8)

gives us the proportion of variance that is not explained by the covariates. $R^2$, the proportion of variance that is explained by the covariates is then

$$R^2 = 1 - \left(\frac{\tau^2_{\text{Residual}}}{\tau^2_{\text{Total}}}\right) = 1 - \left(\frac{0.0633}{0.3088}\right) = 0.7950$$

(1.9)

We show this graphically in Figure 32, which juxtaposes Figure 20 with Figure 31. At left, the normal curve [I] reflects the unexplained variance in true effects when the predicted value for each study is the grand mean. At right, the normal curves [J, K, L] represent the variance in true effects when the predicted value for each study is the corresponding point on the regression line. This is the variance not
explained by latitude. The variance at the right (0.0637) is less than the variance at the left (0.3088), which tells us that by using latitude as a covariate we can reduce the unexplained variance – or (equivalently) explain some of the variance.

Figure 32 | Dispersion about grand mean vs. dispersion about regression line

An equivalent approach to computing $R^2$ is as follows. If the initial (total) $T^2$ is 0.3088 and the unexplained (residual) $T^2$ is 0.0633, it follows that the difference (0.2455) is the $T^2$ explained by the model. Then we can compute $R^2$, the proportion explained by the model, as

$$R^2 = \frac{T^2_{\text{Explained}}}{T^2_{\text{Total}}} = \frac{0.2455}{0.3088} = 0.7950 \quad (1.10)$$

Summary

The $Q$ statistic

- The $Q$-Model is 18.8452 with $df = 1$ and $p < 0.0001$. This tells us that effect size is related to latitude.

- The $Q$-value for goodness of fit is 30.7331 with $df = 12$ and $p = 0.0012$. This tells us that the true effect size varies, even within studies at the same latitude.

- The $Q$-total is 152.2330 with $df = 12$ and $p < 0.0001$. This tells us that effect sizes vary when we ignore latitude and work with deviations of all studies from the grand mean.

The $I^2$ statistic

- The variance in effect sizes that is observed within a given latitude is partly due to real differences (which can potentially be explained by additional study-level covariates) and partly due to within-study sampling error. When there are no covariates $[D]$ the $I^2$ value is 92.1173%, which tells us that 92% of the observed variance is real, and may potentially be explained by
covariates. When we use latitude as a covariate the $I^2$ value is 64.2080%, which tells us that some 64% of the remaining variance (F, G, H in Figure 31) is real, and may potentially be explained by additional covariates.

The $R^2$ statistic

- The between-study variance is estimated at 0.0633 at any given point on the regression line based on latitude, as compared to 0.3088 for the regression line based on the grand mean. This corresponds to an $R^2$ of 0.7950, meaning that 79.50% of the true variance in effects can be explained by latitude.

In context

We presented three cases to show the correspondence between regression and a traditional analysis. In Case A there were no covariates; in Case B there was a categorical covariate; and in Case C there was one continuous covariate.

In Case A we tested the effect size by looking at the mean (for the traditional analysis) or the intercept (for the regression). In Case B we looked at the relationship between effect size and subgroup (for the traditional analysis), or between the effect size and the covariate (for regression). In Case C we looked at the relationship between effect size and the covariate.

For the traditional analysis, to estimate the variance in true effect sizes we computed the $Q$ value based on deviations from the grand mean (Case A); or by the $Q$ value based on deviations from each study's subgroup mean (Case B).

The regression model is a more general model and allows us to cover all cases by saying that we computed the $Q$-value based on deviations from each study's predicted value. In Case A the predicted value is the grand mean (as it was for the traditional analysis). In Case B the predicted value is the subgroup mean (as it was for the traditional analysis). In Case C the predicted value is the point on the regression line corresponding to the regression equation.

In any case, once we had $Q$ we used it to estimate the true variance for the relevant population (all studies, studies within subgroups, or studies at the same latitude). In addition to estimating the variance of effects ($T^2$) and the standard deviation of effects ($T$), we were able to report what proportion of the observed variance was real ($I^2$) and what proportion of the original variance was explained by the predictive model ($R^2$).
PART 6: META-REGRESSION IN CMA
WHAT’S NEW IN THIS VERSION OF META-REGRESSION?

- The prior version of CMA (Version 2) included a module to perform simple regression (one covariate).
- The current version (Version 3) incorporates a full-fledged regression module, which allows for any number of covariates.

Additionally, the new module includes an array of sophisticated options, including the following.

- Define “Sets” of variables, such as a set of covariates that together capture the impact of a categorical variable, or a set of covariates that represent the linear, curvilinear, and cubic relationship of dose with effect size.
- Automatically create and code dummy variables for categorical covariates.
- Select from an array of computational options, including the choice to use the Z distribution or the Knapp-Hartung adjustment, to use method of moments (MM), full maximum likelihood (ML), or restricted maximum likelihood (REML) for estimating $\tau^2$.
- The regression plot incorporates many options, including the ability to display both confidence intervals and prediction intervals for the regression line.
- Export data and residuals to Excel™ for further processing.

Advanced options

- The program allows you to define two or more prediction models. For example, define one model that includes a series of nuisance variables and another that includes these variables plus variables that represent the treatment. The program displays the proportion of variance explained by each model and also a test that compares models.
THE COVARIATES AND THE PREDICTIVE MODEL

In the prior chapter our goal was to show the correspondence between meta-regression and a traditional meta-analysis. For that reason we included no more than one predictor in each analysis.

Of course, a key strength of meta-regression is that it allows us to include two or more predictors in an analysis, as we will be using multiple predictors in most of the examples that follow.

The interpretation of a meta-regression is basically the same as that of regression in a primary study. The analysis will yield a set of statistics for each covariate, as well as set of statistics for the model. The statistics for each covariate reflect the impact of that covariate, with all other covariates held constant. The statistics for the full model reflect the combined impact of all covariates.

- If we have covariates A and B and want to know the impact of each covariate ignoring the other (that is, ignoring the potential confound) we would run two analyses. The first analysis would include only A. The second would include only B.

- If we have covariates A and B, and we want to know the impact of each with the other held constant, we would run one analysis that includes both A and B. The statistics for each covariate reflect the impact of that covariate with all other covariates partialled, or held constant. The statistics for the model reflect the contribution of A and B as a set.

- If we have covariates A and B, and we want to know the impact of each with the other held constant, and also assess the interaction A x B, we would run one analysis that includes A, B, and AB. The statistics for AB give the impact of the interaction over and above the main effects. The statistics for the model reflect the impact of the two main effects plus the interaction.
QUICK START

1) On the data-entry screen
   a) Create a column for study name
   b) Create a set of columns for the effect size
   c) Identify one or more columns as “Moderators” and set the subtype to either “Integer”, “Decimal”, or “Categorical
   d) Enter the data

   Or, simply start CMA and then open the BCG file. Be sure to use the file BCGP if your computer uses a period to indicate decimal places, or BCGC if it uses a comma for this purpose.

2) On the main analysis screen
   a) Optionally, select the effect size index
   b) Optionally, select the studies to be included in the regression
   c) Optionally, specify how to work with studies that included multiple subgroups, outcomes, time-point, or comparisons.
   d) Click Analyses > Meta-regression 2

3) On the regression screen – define the regression
   a) Select the covariates to be included in the model
   b) Optionally, define “Sets” of covariates
   c) Optionally, define multiple models
   d) Optionally, select [Computational options]
   e) Run the analysis

4) On the regression screen – navigate the results
   a) Click the [Fixed] or [Random] tab at the bottom of the screen to select the model
   b) Click the model name (when several models have been created)
   c) Use the toolbar to move between the main analysis screen, the scatterplot, diagnostics, increments, model comparisons, and other screens

5) On the regression screen – save the analysis (Optionally)

6) On the regression screen – export the results (Optionally)

In this manual we use the BCG data as the motivating example.

See Part 1: Data files and downloads for location of files
**STEP 1: ENTER THE DATA**

**Insert column for study names**

In Figure 33 [A], click Insert > Column for > Study names.

Figure 33 | Data-entry | Step 01

Figure 34 [B], the program creates a column labeled “Study name”.

Figure 34 | Data-entry | Step 02
Insert columns for effect size data

In Figure 35, click Insert > Column for > Effect size data.

The program opens a wizard (Figure 36) that allows you to specify the kind of summary data you will enter:

- Select <Show all 100 formats> [D]
- Click [Next] [E]
In the wizard (Figure 37)

- Select the top option button [F]
- On this screen, Click [Next] [G]
In Figure 38, drill down to

- Dichotomous (number of events)
- Unmatched groups, prospective (e.g., controlled trials, cohort studies)
- Events and sample size in each group [H]

Then, click <Finish>

Note that we will be entering events and sample size (N) for each group. Some of the texts that use the BCG example report events and non-events rather than events and N.
The program creates columns as shown in Figure 39. It also opens a wizard that allows you to label the columns.

- Enter Vaccine/Control as names for the two groups [I]
- Enter TB/Ok as names for the two outcomes [J]

Then, click [Ok]

The program applies the labels as shown in Figure 40 [K].
Insert columns for moderators (covariates)

Next, we need to create columns for the moderator variables. As shown in Figure 41,

- Click Insert > Column for > Moderator variable [L]

![Figure 41 | Data-entry | Step 09]
The program opens a wizard (Figure 42)

- Set the variable name to “Latitude” [M]
- Set the column function to Moderator [N]
- Set the data type to Integer [O]

Then, click [Ok]
As shown in Figure 43, Click Insert > Column for > Moderator variable

- Set the variable name to “Year”
- Set the column function to Moderator
- Set the data type to Integer

(This is the year the study was performed, not the year of publication)

Then, click [Ok]
As shown in Figure 44, Click Insert > Column for > Moderator variable

- Set the variable name to “Allocation”
- Set the column function to [Moderator]
- Set the data type to [Categorical]

This moderator tracks the mechanism utilized to assign people to be vaccinated (or not). The possibilities are random, alternate, and systematic.

Then, click [Ok]
As shown in Figure 45, Click Insert > Column for > Moderator variable

- Set the variable name to “Climate”
- Set the column function to [Moderator]
- Set the data type to [Categorical]

This moderator tracks the climate. The possibilities are Cold and Hot.

Then, click [OK]
Customize the screen

The program initially displays the odds ratio (Figure 46).

- We want to work with the risk ratio rather than the odds ratio.
- Additionally, we want to display the risk ratio in log units.

Therefore, we need to customize the display as follows.

- Right-click in any yellow column
- Click <Customize computed effect size display> [A]

![Figure 46 | Data-entry | Step 14](image-url)
The program displays this wizard (Figure 47)

- Tick the box for Risk ratio [B]
- Tick the box for Log risk ratio [B]
As shown in Figure 48, we can set Log risk ratio as the default effect size, and also hide the odds ratio:

- In the drop-down box, select Log risk ratio as the primary index [C]
- Un-check the box for odds ratio [D]
- Un-check the box for log odds ratio [D]

Then click [Ok]

The screen now looks like Figure 49.

---

**Figure 48 | Data-entry | Step 16**

**Figure 49 | Data-entry | Step 17**
Enter the data

You can enter the data manually, or copy and paste from Excel™ or another source (see Appendix 1: The dataset)

In Figure 50 you enter effect-size data into the white columns [E]. The program automatically computes the values in the yellow columns [F].

- You may continue to add the other moderators, as enumerated in the appendix
- Or, open the file BCG.cma. There are two versions of this file, one using a period to indicate decimal places and one using a comma. Use the one that corresponds to your computer’s settings.
STEP 2: RUN THE BASIC META-ANALYSIS

To run the analysis, click [Analysis > Run Analysis] as shown in Figure 51.
The main analysis screen

The program displays the main analysis screen (Figure 52).

The current effect size [A] is “Log risk ratio”. If you want to switch to another effect size, click [Effect measure: Log risk ratio] on the toolbar and select an alternate index.

The next few pages outline the main analysis in CMA using the traditional approach. However, this is optional. As soon as you arrive at the main analysis screen (Figure 52) you can click [Analysis > Meta-regression 2] to proceed immediately to the regression module.

The initial meta-analysis

In Figure 52, the <Fixed> tab [B] is selected, so the program is displaying the results for a fixed-effect analysis [C]. The effect size is in log units [A].
Click the tab [D] for <Random>. The program [E] displays results for a random-effects analysis.

Figure 53 | Basic analysis | Random-effects | Log risk ratio
Display moderator variables

Next, we want to display the moderator variables on the plot.

Note that this is optional, and has no effect on the regression.

In Figure 54, click View > Columns > Moderators [F]

The program displays a list of all variables that had been defined as moderators on the data-entry screen. Drag and drop each of the following onto the main screen, to the right of the “p-value” column (Figure 55 [G]). Latitude, Year, Allocation, and Climate.

Figure 54 | Basic analysis | Display moderators

Figure 55 | Basic analysis | Display moderators
The screen should now look like Figure 56.

You can right-click on any column and sort by that column. Here, the studies are sorted by latitude.

Since the data had been sorted by latitude on the data-entry screen, the program initially displays the studies in that sequence. It appears that the effect size is minor (near 0 in log units) for studies in warmer climates (toward the top) [H] and larger (as extreme as −1.58) for studies in colder climates (toward the bottom) [I].
Display statistics

Click <Next table> to display the statistics shown in Figure 57.

Using random-effects weights \([J]\), the summary log risk ratio is \(-0.7141\). The \(Z\)-value is \(-3.995\) with a corresponding \(p\)-value of 0.0001. Thus, we can reject the null hypothesis that \(\log\) risk ratio is 0.0 (or equivalently, that the risk ratio is 1.0). If we assume that the studies are valid we can conclude that, the vaccine (on average) probably does prevent TB.

At the same time, there is also a substantial amount of dispersion in the effect size. Tau-squared \([K]\) is 0.3088 and Tau is 0.5557. To get a general sense of the true dispersion we can assume that the true effects are balanced about the random-effects estimate of the mean effect, and that some 95\% of all true effects fall within 1.96 \(T\) of this mean. Then (in log units) most true effects fall in the range of \(-1.8032\) to \(+0.3750\). This corresponds to risk ratios of approximately 0.16 (a strongly protective effect) to 1.46 (a harmful effect).

It would be very important to understand the reason for this dispersion, and for this purpose we turn to meta-regression.

Figure 57 | Basic analysis | Display statistics for heterogeneity
**STEP 3: RUN THE META-REGRESSION**

At this point we proceed to the meta-regression.

On the analysis screen (Figure 58), select Analysis > Meta-regression 2.

Note.

If you don’t see any regression option you may have a lite or standard version of the program, rather than the professional version.

If you see an option for Meta-regression but not Meta-regression 2, you have Version 2 of CMA rather than Version 3.
The Interactive Wizard

The program displays the screen shown Figure 59.

The interactive wizard will walk you through all the steps in running the regression. To display or hide the wizard, use the Help menu.

Figure 59 | Run regression | Step 02
Add covariates to the model

When you initially open the regression module the program displays the following

- The main screen [A]
- A list of available covariates [B]

We need to move the covariates from the wizard [B] onto the main screen [A].

Add variables in the sequence shown here (allocation, year, and latitude) to recreate the example that we use in this text.

- Click “Allocation” on the wizard
- Click [Edit reference group] and select [Random].
- Click [Add to main screen] [C]
- Click “Year” on the wizard and then click [Add to main screen] [C]
- Click “Latitude” on the wizard and then click [Add to main screen] [C]
The model is shown in Figure 61.

Note that “Allocation” is displayed as two lines [D], which are linked by a bracket. Since allocation is a categorical variable the program automatically creates dummy variables to represent allocation. See next chapter for a full discussion.

Tick the check-boxes for all covariates [E]

The covariates are controlled by the “Covariates” toolbar [F]. On this toolbar,

- [Show covariates] shows or hides the wizard
- [Remove covariates] allows you to remove a covariate from the main screen
- [Move up] and [Move down] allow you to edit the sequence of covariates
- The blue and red checks allow you to add (or remove) checks from a series of check-boxes
Set computational options

The program allows you to specify various options for the computations.

Click Computational options to display the menu in Figure 62.

Each of these options is discussed in Part 10: Computational options.

To follow the example in this text, set the options as follows.

- Method for estimating $T^2$ (Method of moments)
- One-tailed or two-tailed test for $p$-values (Two-tailed)
- Confidence level (95%)
- Display the variance inflation factor (Off)
- $Z$ distribution or the Knapp-Hartung adjustment for $p$-values and confidence intervals ($Z$)
Run the regression

To run the regression, simply click “Run regression” on the toolbar [A] in Figure 62
STEP 4: NAVIGATE THE RESULTS

Main results screen (Fixed effect)

After you run the regression

- Click [Main Results] [A]
- Click [Fixed] at the bottom to select the statistical model [B]

For a full discussion of how to interpret the output for a fixed-effect analysis, see Part 7: Understanding the results.
Main results screen (Random effects)

After you run the regression

- Click [Main Results] [C]
- Click [Random] at the bottom to select the statistical model [D]

For a full discussion of how to interpret the output for a random-effects analysis, see Part 7: Understanding the results.
**Difference between the fixed-effect and random-effects displays**

Under the fixed-effect model we assume that there is one source of sampling error (within-study variance), whereas under the random-effects model we allow that there may be two sources of sampling error (within-study variance and between-study variance). Since the weight assigned to each study is the inverse of the variance, the weight assigned to each study depends on the model. In the following pages, this will be evident in the fact that for the statistics that are presented under both models (such as the effect size and its standard error), the value depends on the statistical model.

Perhaps more fundamentally, the statistical model determines what statistics we choose to present. The results screen is quite different for the fixed-effect vs. the random-effects model, reflecting the fact that the model determines what questions we can ask of the data.

When we work with the fixed-effect model we assume that all studies share a common effect size. We don’t need to estimate $\tau^2$ (the between-study variance), since this is assumed to be zero. If $\tau^2$ is assumed to be zero, we don’t estimate $I^2$ (the ratio of between-study variance to total variance) since this must be zero. Nor do we estimate $R^2$, the proportion of between-study variance explained by the predictors, since this must also be zero. For example, see Figure 71.

By contrast, when we work with the random-effects model we allow that the true effect size may vary from one study to the next, and therefore these statistics help us to understand this variation. We can estimate the variation of effects sizes (a) about the grand mean and (b) about the regression line. By comparing the two we can compute $R^2$, the proportion of variance explained by the predictors. We can also compute $I^2$ for each case (with and without covariates), and this tells us what proportion of the observed variance reflects variation in true effect sizes rather than random error. For example, see Figure 76.
Plot

To display the plot

- Click [Scatterplot] on the menu bar to navigate to the plot [A]
- Select [Fixed] or [Random] on the tab at the bottom of the screen [B]
- To specify the variable for the X-axis,
  - Right-click on the X-axis label [C] or
  - Click on the drop-down tool [D]

For a full discussion of the plot see page 157.
Other screens

To navigate to other tables of results, click “More results” [A] in Figure 67 and then select any of the following.

About the predictive model

- Covariance (see page 132)
- Correlation (see page 133)
- Diagnostics (see page 125)
- R-squared graphic (see page 145)

About the data included in (or excluded from) the analysis

- All studies (see page 244)
- Valid studies (see page 244)

Statistics that compare different predictive models

- Increments (see page 134)
- Models summary (see page 262)
- Compare models (detailed) (see page 262)
- Compare models (p-value) (see page 262)
**STEP 4: SAVE THE ANALYSIS**

Once you’ve run a meta-regression you can save the predictive model as shown in Figure 68.

- Click File > Save regression file as ...
- This will save the regression template with an extension of .cmr.

The .cmr file saves the instructions for the analysis, NOT the data. By analogy, programs such as SPSS™, SAS™, and stata™ allow you to save a set of commands in one file and the data in another file. The commands can then be applied to any data file that has the same variables.

- The .cmr file, saved here, is analogous to the command file in the other programs.
- The .cma file, saved from the data-entry screen, is analogous to the data file in the other programs.

The .cmr file saves the following

- The list of covariates
- The list of models
- The check-boxes for each model
- The sets
- The model names

In another session you can open a data file on the main data-entry screen. Then, return to the regression module and click File > Open file to open the .cmr file and re-run the analysis.

The .cmr file can be used with the same dataset that was used to create it, or with another dataset that includes the same variables. For example,
• You may return to the data-entry screen and add new studies
• You may return to the main analysis screen and edit the study filters
• You may be working with an entirely different data set that has the same variables as the first one.

In any of these cases, navigate to the regression module and click File > Open to open the .cmr file.

When you open a .cmr file the program simply recreates the main MR screen as though you had entered it manually. The .cmr file does not save the statistical settings that were in place when the file was created. These include the method employed to estimate $T^2$, the use of $Z$ or Knapp-Hartung, the confidence level, the choice of a one-sided or two-sided test.
STEP 5: EXPORT THE RESULTS

The program offers two options for exporting the results of any analysis.

- Export the results to Excel™. Then, you can perform additional computations within Excel™, and/or format the results and copy them as a table to other programs.

- Copy the results to the clipboard as a picture. Then, paste this picture into Word™ or any other program.

Figure 69 shows an example for the main analysis screen.

- Click [File > Save results as Excel™ file and open] [A]
- Provide a name for the Excel™ file

![Figure 69 | Export results]
The program creates the Excel™ file shown in Figure 70.

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|
| Main results for Model 1, Random effects (MM), Z-Distribution |       |
|       |       |       |       |       |       |       |
| 5     | 6     | 7     | 8     | 9     | 10    | 11    |
|       |       |       |       |       |       |       |
| Set   | Covariate | Coefficient | Standard Error | 95% Lower | 95% Upper | Z-value | 2-sided P-value |
| 7     | Intercept | -29.2359 | 45.7929 | -119.0007 | 60.5289 | -0.6383 | 0.5232 |
| 8     | Allocation: Alternate | 0.4855 | 0.4809 | -0.4570 | 1.4279 | 1.0096 | 0.3127 |
| 9     | Allocation: Systematic | 0.4374 | 0.3778 | -0.2831 | 1.1978 | 1.2106 | 0.2290 |
| 10    | Year | 0.0145 | 0.0232 | -0.0305 | 0.0603 | 0.0394 | 0.3225 |
| 11    | Latitude | -0.0190 | 0.0159 | -0.0503 | 0.0122 | -1.1924 | 0.2331 |

Figure 70 | Export results

The same idea applies to any screen that displays results.
PART 7: UNDERSTANDING THE RESULTS
MAIN RESULTS

- To run the regression, click Run Regression.
- The program will display the Main Results screen
- Click [Fixed] or [Random] to select the statistical model

The following pages show the results screen for each statistical model.

The top of the screen is similar for the two models (Figure 72 and Figure 77). For either model it shows the impact of each covariate with other covariates partialled. The difference between the two screens is that one is based on fixed-effect weights while the other is based on random-effects weights.

After that, however, the screens differ in some fundamental ways which reflect the difference in the two statistical models. While this holds true for any predictive model, we’ll use the case of one categorical covariate (Climate) as an example.

For the fixed-effect model (Figure 72) the program displays a table similar to the analysis of variance table we see for primary studies. This table includes a row for

a) The model: The WSS for the deviation of the subgroup means from the grand mean
b) The residual: The WSS for the deviation of all studies from their subgroup means
c) The total: The WSS for the deviation of all studies from the grand mean

The program does not present any statistics for between-study variance ($T^2$) nor for proportion of between-study variance explained by the model ($R^2$). For studies that share the same predicted value, $T^2$ is assumed to be zero, and so there is no reason to report it. If $T^2$ with the predictive model in place is assumed to be zero, then $R^2$ also has no real meaning (by definition, once we apply the model, no between-study variance remains).

By contrast, for the random-effects model (Figure 77) the program does not display an analysis of variance table. The idea of partitioning the WSS only works if the weight for each study is constant. This condition is met under the fixed-effect model since the weights are always the same (based on within-study variance), but not under the random-effects model since the weights (which are based also on $T^2$) change when we introduce covariates (and thus the frame of reference for computing $T^2$).

For the random-effects analysis we do want to present statistics for $T^2$, and for $R^2$. To do this, we need to run a series of distinct analyses and then collate the results. Specifically, we run one analysis to get $T^2$ with the covariates, and another to get $T^2$ without the covariates. The change in $T^2$ gives us the amount of variance explained by the predictive model, and this value over the original $T^2$ gives us the proportion of variance explained ($R^2$). To highlight the fact that these statistics are coming from separate analyses, the statistics are not presented in a table, but rather in separate sections on the screen.
MAIN RESULTS, FIXED-EFFECT ANALYSIS

To navigate to this screen

Click [Run regression] [A]

Figure 71 | Setup
The toolbar changes as shown in Figure 72.

- Click “Main results” [B]
- Click [Fixed] [C]
Test of the model

Analysis of variance

In section [D] the total WSS (Weighted sum of squares, $Q$) is partitioned into the following.

Model

The model [E] is the test that the predictive model explains any of the variance in effect size. Put another way, it asks if the dispersion of effects about the regression line is smaller when the regression line is based on the covariates rather than based solely on the grand mean. Here, $Q = 128.2186$ with $df = 4$ and $p < 0.0001$, so we conclude that the predictive model explains (at least) some of the variance in effect size.

Residual

The residual [F] is the test that the data are consistent with the model’s assumption of a common effect size for all studies with the same predicted value. The $Q$ value is $24.0144$ with $df = 8$ and $p = 0.0023$. We conclude that the data are not consistent with the assumptions of the fixed-effect model.

Total

The total [G] is the test that the variance for the full set of studies (with no predictors) is zero. The $Q$-value is $152.2330$ with $df = 12$ and $p < 0.0001$.

In a primary study, the total sum of squares (SS) is the sum of the SS explained by the model and the SS residual. Similarly, in a meta-analysis (with fixed-effect weights) the total weighted sum of squares (WSS) is the sum of the WSS explained by the model and the WSS residual. As shown in Figure 72 [D],

$$WSS_T = WSS_M + WSS_{RES} = 152.2330 = 128.2186 + 24.0144$$

Similarly, the total $df$ is the sum of the model and the residual df ($12 = 4 + 8$).

$$df_T = df_M + df_{RES} = 12 = 4 + 8$$

Impact of individual covariates [H]

In Figure 72, the test of the model [E] is an omnibus test for the full set of covariates. It tells us that the set as a whole is related to effect size. By contrast, the rows at the top [H] address the unique impact of each covariate – that is, the impact of each covariate when all of the other covariates are held constant.

Since the effect size is the risk ratio, all analyses are carried out in the log metric and all coefficients are in the log metric. In this example, virtually all predicted effects are less than zero, so 0 is no effect, −1 is a large effect, and −2 is a very large effect (see Figure 73). In this example, therefore, a negative coefficients means that as the covariate gets larger the vaccine is more effective (see Appendix 8: Interpreting regression coefficients).
To understand the direction of the effect size as a function of covariates, it’s helpful to work with the scatterplot as discussed immediately below.

Allocation

Allocation type is a categorical covariate with three groups (Randomized, Alternate, and Systematic), and therefore is represented by a set of two dummy variables. In Figure 72, the test of this set yields $Q = 6.3651$ with $df = 2$ and $p = 0.0412$. Thus, there is evidence that effect size is related to allocation type. The relationship between allocation type and effect size with other covariates partialled in displayed in Figure 73.

For a more specific analysis we can look at each line within the set. Alternate allocation has a coefficient of 0.6320 (the vaccine is less effective in studies that employed alternate allocation as opposed to randomized allocation) and a $p$-value of 0.0366. Systematic allocation has a coefficient of 0.3062 (the vaccine is less effective in studies that employed systematic allocation as opposed to randomized allocation). However, as will be discussed in the chapter on caveats (page 293), these findings may be due to a confound with other factors.

Click on Scatterplot and select Allocation to produce the plot shown in Figure 73.

There is a column for each allocation type (Random, Alternate, and Systematic). In each column the program displays the observed effects sizes as well as the summary effect size and the confidence interval for the summary effect size.

In this example the “Fixed” tab is selected at the bottom of the screen, so all statistics are based on the fixed-effect model. The reader will note that the summary effect size for the alternate allocation studies falls outside the range of the actual effect sizes for the two studies in this group. This reflects the fact...
that these means are adjusted for other covariates (and serves as a caution against performing these kinds of adjustments with a small number of studies).

Year

The coefficient for Year is +0.0235, which means that for every increase of one year the log risk ratio will increase by 0.0235 (vaccine was less effective in later trials). The coefficient plus/minus 1.96 times the standard error (0.0159) yields the 95% confidence interval for the coefficient, which is −0.0076 to +0.0545. The coefficient divided by its standard error yields a Z value of 1.4795, and the corresponding p-value of 0.1390. Thus, when latitude and allocation method are held constant, the relationship between year and effect size is not statistically significant.

![Regression of Log risk ratio on Year](image)

Click on Scatterplot and select Year to produce the plot shown in Figure 74.

The regression line shows that as the Year increases, the effect size moves closer to zero. Since the effect size is a log risk ratio, this means that the treatment effect (when adjusted for the other covariates) declined over the years.

The confidence interval shows the range of regression lines that are consistent with the data – in other words, we could pivot the regression line on its axis either counter-clockwise or clockwise (as suggested by the arrows) until it encountered the confidence interval. The uncertainty is such that the true regression line could be either in an upward or downward direction. This corresponds to the p-value of 0.1390 for Year in Figure 72, and the fact that the confidence interval for the coefficient included both negative and positive values (−0.0076 to +0.0545)
**Latitude**

The coefficient for latitude is $-0.0213$, which means that for every increase of one unit (degree) in latitude the log risk ratio will decrease by 0.0213 (vaccine is more effective at greater latitudes). The coefficient plus/minus 1.96 times the standard error (0.0084) yields the 95% confidence interval for the coefficient, which is $-0.0378$ to $-0.0048$. The coefficient divided by its standard error yields a $Z$ value of $-2.526$, and the corresponding $p$-value of 0.0115. Thus, even when year and allocation method are held constant, the relationship between latitude and effect size is statistically significant.

Figure 75 | Plot | Latitude | Fixed-effect

Click on [Scatterplot] and select Latitude to produce the plot shown in Figure 75.

The regression line shows that as the absolute Latitude increases, the effect size moves further from zero. Since the effect size is a log risk ratio, this means that the treatment effect (when adjusted for the other covariates) increases as we move further from the equator.

The confidence interval shows the range of regression lines that are consistent with the data – in other words, we could pivot the regression line on its axis either counter-clockwise or clockwise (as suggested by the arrows) until it encountered the confidence interval. While there is substantial uncertainty, all likely regression lines are in the same (downward) direction. This corresponds to the $p$-value of 0.0115 for Latitude in Figure 72, and the fact that the confidence interval for the coefficient includes only negative values ($-0.0378$ to $+0.0048$)
Summary

The model

The total $Q$ of each effect size about the grand mean can be partitioned into its component parts – the $Q$ due to the variation in effect size that can be explained by the covariates, and the part that cannot.

- Model. The $Q$-value for the model is 128.2186 with $df = 4$ and $p < 0.0001$, which tells us that effect size is related to at least one of the covariates.
- Residual. The $Q$-value for the residual is 24.0144 with $df = 8$ and $p = 0.0023$, which tells us that the assumptions of the fixed-effect model have been violated.
- Total. The $Q$-value for the total is 152.23 with $df = 12$ and $p < 0.0001$, which tells us that effect sizes vary when we ignore subgroups and work with deviations of all studies from the grand mean.

Individual covariates

Where the test of the model is an omnibus test for the full set of covariates, the table at the top addresses the impact of each covariate with all other covariates held constant.

- Alternate allocation is associated with a smaller effect size (but see the chapter on caveats). The $p$-value for allocation is 0.0415.
- Studies that fall further from the equator showing more impact of the vaccine. The $p$-value for year is 0.1390.
- Studies that fall further from the equator showing more impact of the vaccine. The $p$-value for latitude is 0.0115.
MAIN RESULTS, RANDOM-EFFECTS ANALYSIS

To navigate to this screen

Click [Run regression] [A]
The toolbar changes as shown here

- Click “Main results” [B]
- Click [Random] [C]

![Figure 77 | Main results | Random-effects](image)

### Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Set</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-25.2359</td>
<td>45.7992</td>
<td>-119.0007</td>
<td>60.5289</td>
<td>-0.3383</td>
<td>0.7322</td>
</tr>
<tr>
<td>Allocation: Alternate</td>
<td>0.4865</td>
<td>0.4869</td>
<td>-0.4570</td>
<td>1.4379</td>
<td>1.0096</td>
<td>0.3127</td>
</tr>
<tr>
<td>Allocation: Systematic</td>
<td>0.4574</td>
<td>0.3778</td>
<td>-0.2831</td>
<td>1.1978</td>
<td>0.1206</td>
<td>0.2269</td>
</tr>
<tr>
<td>Year</td>
<td>0.0148</td>
<td>0.0232</td>
<td>-0.0308</td>
<td>0.0680</td>
<td>0.0122</td>
<td>0.9124</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0150</td>
<td>0.0159</td>
<td>-0.0500</td>
<td>0.0122</td>
<td>-1.1924</td>
<td>0.2331</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero

\[ Q = 13.1742, \text{df} = 4, \ p = 0.0105 \]

Goodness of fit: Test that unexplained variance is zero

\[ \tau^2 = 0.1194, \tau_a = 0.3456, I^2 = 66.69\%, \ Q = 24.0144, \text{df} = 8, \ p = 0.0023 \]

Comparison of Model 1 with the null model

Total between-study variance (intercept only)

\[ \tau^2 = 0.3008, \tau_a = 0.5057, I^2 = 92.12\%, \ Q = 152.2330, \text{df} = 12, \ p = 0.0000 \]

Proportion of total between-study variance explained by Model 1

\[ R^2 \text{analog} = 0.61 \]

Number of studies in the analysis: 13
The results presented in Figure 78 are based on three separate analyses. Each of these analyses yields specific items of information, which are pulled together on this screen.

The differences among the three analyses are shown in Table 3.

### Table 3

<table>
<thead>
<tr>
<th>Section</th>
<th>Function</th>
<th>Covariates</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>Random-effects estimates</td>
<td>Yes</td>
<td>V+T²</td>
</tr>
<tr>
<td>E</td>
<td>Variance not explained by model</td>
<td>Yes</td>
<td>V</td>
</tr>
<tr>
<td>F</td>
<td>Original variance</td>
<td>No</td>
<td>V</td>
</tr>
</tbody>
</table>

Section D reports statistics for an analysis that employs random-effects weights and includes the covariates. This provides a test of the model, and is also the analysis used in the table at the top of the screen.
Section E reports statistics for an analysis that includes the covariates but assigned weights based on V. This provides a goodness-of-fit test. Specifically, we use this analysis to estimate the residual \( T^2 \), the variance not explained by the covariates.

Section F reports statistics for an analysis that does not include covariates and assigns weights based on V. This allows us to estimate the original \( T^2 \), the total amount of variance.

Section [G] is based on the analyses in sections [E] and [F]. Section [E] gives us the variance that cannot be explained by the covariates, while section [F] gives us the total variance. We can use these to compute the ratio of explained to total, which is presented in section [G] (See Part 8: The \( R^2 \) index).
Test of the model [D]

*Is effect size related to the covariates?*

The test of the model is a simultaneous test that all covariates (except the intercept) are zero. The $Q$-value is 13.1742 with $df = 4$ and $p = 0.0105$. We reject the null and conclude that at least one of the covariates is probably related to effect size.

Goodness of fit [E]

*Is there any unexplained variance in the true effect sizes?*

Immediately above, we saw that the covariates improve our ability to predict that study’s effect. But does this information enable us to completely predict that study’s effect – do all studies with the same values on all covariates share a common effect size? Or is there variance in true effects among studies with the same predicted value?

The $Q$ statistic, based on the deviation of each study from its predicted value, is 24.0144, with 8 $df$ and a corresponding $p$-value of 0.0023. This tells us that the true effect size probably varies from study to study, even for studies that are identical on all covariates. Put another way, the model is incomplete – knowing a study’s allocation type, year, and latitude does not allow us to completely predict its effect size.

*How much variance is there?*

In section [E] the program shows that $T^2$, the variance of true effect sizes at any point on the regression line, is 0.1194. It follows that the $T$, the standard deviation of true effect sizes at any point on the regression line is 0.3455. We can use this to get a sense of how closely the true effects at any point on the regression line are (or are not) clustered together.

In Figure 31 we’ve plotted all 13 studies, the regression line, and a series of normal curves about the regression line. Each normal curve is centered at some point on the regression line, and extends 1.96 $T$ on either side of that line. If the true effects are normally distributed with standard deviation $T$, then 95% of studies with that predicted value will have a true effect size within the range of the normal curve.

For example, consider the normal curve labeled [N]. The regression line crosses the Y axis at −1.5. If we were to run many studies at this latitude, the mean effect in these studies would be −1.5. However, the true effect size in any single study would typically fall somewhere above or below this value. The normal curve tells us that 95% of these studies would have true effects in the range indicated by the curve, approximately from −1.0 to −2.0.

The decision to display normal curves at three specific points is arbitrary. These curves could have been placed at any points on the regression line.
What proportion of the observed variance is true variance?

The variance of the observed effects about the regression line incorporates both within-study variance (error) and between-study variance (that can be potentially explained by additional study-level covariates). The $I^2$ statistic is 66.69%, which tells us that some 67% of the variance of observed effects about the regression line falls into the latter group.

A useful way of using $I^2$ is to help us understand what the distribution of effects would look like if we could plot the true effects rather than the observed effects. An $I^2$ of 67% tells us that the variance of the distribution would shrink by about one-third.

The problem with this number is that it’s in square units, and it’s not intuitive what it means that the variance will shrink by a third. It might be more intuitive to work with the square root of $I^2$, $(I)$, which is 0.8166. If we were looking at the true scores rather than the observed scores, the dispersion of effects about the regression line (in linear units) would shrink by about 18%.

Impact of individual covariates

The test of the model is an omnibus test for the full set of covariates. It tells us that at least one of the covariates is probably related to effect size. By contrast, the table at the top addresses the impact of each covariate with all of the other covariates partialled (or held constant).

Since the effect size is the risk ratio, all analyses are carried out in the log metric, and all coefficients are in the log metric. In this example, virtually all predicted effects are less than zero, so 0 is no effect, −1 is a large effect, and −2 is a very large effect. In this example, therefore, a negative coefficient means that as the covariate gets larger the vaccine is more effective. (The reverse would be true if the predicted values were all positive).
To understand the direction of the effect size as a function of covariates, it’s helpful to work with the scatterplot as discussed immediately below.

**Allocation**

Allocation type is defined as a set of two covariates. The test of the set yields $Q = 1.5402$ with $df = 2$ and $p = 0.46$, and so there is no evidence that effect size is related to allocation type.

For a more specific analysis we can look at each line within the set.

- Alternate allocation has a coefficient of 0.4855 (the vaccine is less effective in studies that employed alternate allocation as opposed to randomized allocation) and a $p$-value of 0.3127.
- Systematic allocation has a coefficient of 0.4574 (the vaccine is less effective in studies that employed systematic allocation as opposed to randomized allocation) and a $p$-value of 0.2260.

None of these $p$-values is statistically significant.

Click on Scatterplot and select Allocation to produce the plot shown in Figure 80.

There is a column for each allocation type (Random, Alternate, and Systematic). In each column the program displays the observed effects sizes as well as the summary effect size and the confidence interval for the summary effect size.

In this example the “Random” tab is selected at the bottom of the screen, so all statistics are based on the random-effects model.
The predicted effect size for studies that employed alternate allocation (center column) or systematic allocation (right-hand column) is closer to zero than the predicted effect size for studies that employed randomized allocation (left-hand column). As above, none of these differences is statistically significant.

**Year**

The coefficient for Year [H] is 0.0148, which means that for every increase of one year the log risk ratio will increase by 0.0148 (the vaccine became less effective over time). The corresponding p-value is 0.5225.

![Figure 81 | Plot | Year | Random-effects](image_url)

Click on [Scatterplot] and select Year to produce the plot shown in Figure 81.

The regression line shows that as the Year increases, the effect size moves closer to zero. Since the effect size is a log risk ratio, this means that the treatment effect (when adjusted for the other covariates) declined over the years.

The confidence interval shows the range of regression lines that are consistent with the data – in other words, we could pivot the regression line on its axis either counter-clockwise or clockwise (as suggested by the arrows) until it encountered the confidence interval. The uncertainty is such that the true regression line could be either in an upward or downward direction. This corresponds to the p-value of 0.5225 for Year in Figure 78, and the fact that the confidence interval for the coefficient included both negative and positive values (−0.0306 to +0.0603).
**Latitude**

The coefficient for latitude [I] is \(-0.0190\), which means that for every increase of one unit (degree) in latitude the log risk ratio will decrease by 0.0190 (vaccine is more effective at greater latitudes). The coefficient plus/minus 1.96 times the standard error (0.0159) yields the 95% confidence interval for the coefficient, which is \(-0.0503\) to 0.0122. The coefficient divided by its standard error yields a Z value of \(-1.1924\), and the corresponding p-value of 0.23. Thus, when year and allocation method are held constant, the relationship between latitude and effect size is not statistically significant.

![Regression of Log risk ratio on Latitude](image)

Click on [Scatterplot] and select Latitude to produce the plot shown in Figure 82.

The regression line shows that as the absolute Latitude increases, the effect size moves further from zero. Since the effect size is a log risk ratio, this means that the treatment effect (when adjusted for the other covariates) increases as we move further from the equator.

The confidence interval shows the range of regression lines that are consistent with the data – in other words, we could pivot the regression line on its axis either counter-clockwise or clockwise (as suggested by the arrows) until it encountered the confidence interval. There is substantial uncertainty in the coefficient. This corresponds to the p-value of 0.2331 for Lattitude in Figure 78, and the fact that the confidence interval for the coefficient (\(-0.0503\) to +0.0122) includes the null value, zero.
In this example none of the individual covariates has a \( p \)-value less than 0.05. Since the model as a whole is statistically significant, the fact that no covariate is statistically significant probably reflects the fact that some of the covariates are correlated with each other. For example, latitude or year might be statistically significant if entered into the equation alone. However, if the two are correlated with each other and compete to explain the same variance, neither has a unique impact that meets the threshold for statistical significance.

**Comparison of Model 1 with the null model**

We want to report what proportion of variance is explained by the predictive model, and for this purpose we need to know how much variance there was initially (with no covariates). For this reason we run a regression with no covariates (the null model) and compute \( T^2 \). Here, \( T^2 \) is 0.3088, which is the variance of all studies about the grand mean.

**Proportion of variance explained**

To get the final amount of variance we run a regression with the covariates and compute \( T^2 \). This value, reported above as 0.1194, is the variance of studies about their predicted value.

- With no covariates in the model \([F]\) the unexplained variance \((T^2)\) is 0.3088
- With covariates in the model \([E]\) the unexplained variance \((T^2)\) is 0.1194
- The difference between these values is the variance explained by the model, or 0.1894

If the initial (total) \( T^2 \) is 0.3088 and the unexplained (residual) \( T^2 \) is 0.1194, then the ratio

\[
\frac{T^2_{\text{Residual}}}{T^2_{\text{Total}}} = \frac{0.1194}{0.3088} \quad (1.13)
\]

gives us the proportion of variance that is not explained by the covariates. \( R^2 \), the proportion of variance that is explained by the covariates is then

\[
R^2 = 1 - \left( \frac{T^2_{\text{Residual}}}{T^2_{\text{Total}}} \right) = 1 - \left( \frac{0.1194}{0.3088} \right) = 0.6133. \quad (1.14)
\]

In Figure 78 this is on the line labeled [G].

We show this graphically in Figure 83. At left, the normal curve [O] reflects the unexplained variance in true effects when the predicted value for each study is the grand mean. At right, the normal curves [P, Q, R] represent the variance in true effects when the predicted value for each study is the corresponding point on the regression line. This is the variance not explained by latitude. The variance at the right is less than the variance at the left, which tells us that by using latitude as a covariate we can reduce the unexplained variance – or (equivalently) explain some of the variance.
An equivalent approach to computing $R^2$ is as follows. If the initial (total) $\tau^2$ is 0.3088 and the unexplained (residual) $\tau^2$ is 0.1194, it follows that the difference (0.1894) is the $\tau^2$ explained by the model. Then we can compute $R^2$, the proportion explained by the model, as

$$R^2 = \frac{\tau_{\text{Explained}}^2}{\tau_{\text{Total}}^2} = \frac{0.1894}{0.3088} = 0.6133 .$$

(1.15)

Summary

The full model

- The Q-Model is 18.85 with $df = 1$ and $p < 0.0001$. This tells us that effect size is related to (at least some of) the covariates.

- The Q-value for goodness of fit is 30.73 with $df = 12$ and $p = 0.0012$. This tells us that the effect size varies, even within studies that share the same value on all covariates.

- The Q-total is 152.23 with $df = 12$ and $p < 0.0001$. This tells us that effect sizes vary when we ignore the covariates and work with deviations of all studies from the grand mean.

The $I^2$ statistic

- The observed variance in effect sizes is partly due to real differences and partly due to within-study sampling error. When there are no covariates [F] the $I^2$ value is 92%, which tells us that
92% of the observed variance is real, and may potentially be explained by covariates. When we use these covariates \( E \) the \( I^2 \) value is 66.69%, which tells us that 66.69% of the remaining variance is real, and may potentially be explained by additional covariates.

**The \( R^2 \) statistic**

- The between-study variance is estimated at 0.1194 at any given point on the regression line based on these covariates, as compared to 0.3088 for the regression line based on the grand mean. This corresponds to an \( R^2 \) of .6133, meaning that 61.33% of the variance in true effects can be explained by the covariates.
DIAGNOSTICS

To navigate to the diagnostics screen

Run the analysis [A]

The toolbar changes as shown here

- Click More > Diagnostics [B]
- Select the statistical model (Fixed or random) from the tabs at the bottom

Figure 85 | Diagnostics
Observed value

This is simply the observed effect size

The Predicted Value

The predicted (fitted) value, $\hat{T}_i$, for the $i^{th}$ study is the value obtained from using the estimated regression coefficients $b_0, b_1, \ldots, b_p$ and the covariate values for the $i^{th}$ study $x_{i1}, \ldots, x_{ip}$ to compute the value of the effect size predicted for that study by the regression model

$$\hat{T}_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip}.$$ (1.16)

The Residual

The (unstandardized) residual value, $e_i$, for the $i^{th}$ study is the difference between the observed value and the fitted value

$$e_i = T_i - \hat{T}_i.$$ (1.17)

If $e_i = 0$, the fitted value and the observed value are identical (the fitted value is exactly on the regression line or plane), but if $e_i$ is far from 0, the predicted value is not close to the observed value.

In meta-analysis, effect sizes and their fitted values from different studies can have very different sampling uncertainties (standard errors). This makes it difficult to interpret differences in the magnitude of residuals from different studies. Standardized or jackknifed residuals attempt to address this problem of comparability by dividing the residual by its standard error.

Studentized Residual

The studentized residual value, $es_i$, for the $i^{th}$ study is the residual divided by its standard error

$$es_i = \frac{e_i}{SE(e_i)}.$$ (1.18)

The standard error of $es_i$ is given by

$$SE(e_i) = \sqrt{\frac{1-h_i}{w_i}},$$ (1.19)

where $w_i$ is the weight given to the $i^{th}$ effect size in the analysis, $s^2$ is the weighted residual mean square, and $h_i$ is the leverage of the $i^{th}$ effect size. Therefore the residual divided by its standard error (the $i^{th}$ studentized residual) is
It is important to note that the standard error of the residual depends on both the residual variance which is determined by the conditional variance of the estimate (and the random effects variance component in random effects models) and the configuration of predictors (including the values for the $i^{th}$ study). Studentized residual $e_{si}$ is on a standard scale, so the values from different studies are more comparable than those of the unstandardized residuals (the $e_i$). If the regression model is correctly specified, the $e_{si}$ have approximately a normal distribution with unit standard deviation, so that $e_{si}$ values greater than 2 in absolute value occur only about 5% of the time by chance and values greater than 2.5 are quite unusual. The actual sampling distribution of $e_{si}$ will often be closer to Student’s $t$-distribution with $k – Q$ degrees of freedom, so slightly larger reference values (than 2 and 2.5) may be appropriate for judging extremeness of residuals when $k – Q$ is small.

**Jackknifed Residual**

The jackknifed residual $e_{ji}$ is similar to the studentized residual in that it is standardized. However the jackknifed residual is the difference between the observed effect size in the $i^{th}$ study and the fitted value of the $i^{th}$ study computed with the $i^{th}$ study deleted from the dataset. That is,

$$e_j = \frac{T_i - \hat{T}_{(i)i}}{SE\left(T_i - \hat{T}_{(i)i}\right)},$$

where $\hat{T}_{(i)i}$ is the fitted value of the $i^{th}$ study computed from all other studies except the $i^{th}$ study. To be precise

$$\hat{T}_{(i)i} = b_{(i)0} + b_{(i)1}x_{i1} + \cdots + b_{(i)p}x_{ip},$$

where $b_{(i)0}, ..., b_{(i)p}$ are the regression coefficients estimated with the $i^{th}$ study removed from the dataset.

The jackknifed residual is designed to better reveal cases where the $i^{th}$ study does not fit the same model as the other studies. By removing the (potentially distorting impact of the $i^{th}$ study from the computation of the regression coefficients, the jackknifed residual sometimes makes it easier to see how different an observed effect size is from what is expected if that study fit the meta-regression model that is appropriate for all of the other studies.

Let the weight of the $i^{th}$ study computed using the variance component estimate with $i^{th}$ study removed be denoted $w_{(i)i}$, then the $i^{th}$ jackknifed residual is equivalent to

$$e_{j_i} = e_i \sqrt{\frac{w_{(i)i}}{1 - h_i}}.$$ 

(1.20)
The sampling distribution of the jackknifed residual is similar to that of the studentized residual (approximately normal) and similar reference values for judging extremeness are appropriate. The actual sampling distribution of $e_j$ will often be closer to Student’s $t$-distribution with $k - Q - 1$ degrees of freedom, so slightly larger reference values (than 2 and 2.5) may be appropriate for judging extremeness of residuals when $k - Q - 1$ is small.

**Leverage**

Leverage is a diagnostic that reveals how much potential influence a particular study can have on the result of the meta-regression. Let $h_i$ be the leverage of the $i^{th}$ study. The values of the leverage are always between zero and one inclusive, that is, $0 \leq h_i \leq 1$. The sum of the leverages $h_1 + \ldots + h_k = Q$, where $Q$ is the total number of predictors including the intercept (that is $Q = p + 1$ when there is an intercept in the model and $Q = p$ if there is no intercept). Thus the average value of the leverage is $Q/k$, and estimates of regression coefficients are most efficient when all the leverage values are close to $Q/k$.

If $h_i = 0$, this implies that the fitted (predicted) value of the effect size for the $i^{th}$ study would be the same even if that study were not part of the data used to estimate the regression coefficients. In one sense, this implies minimal influence. If $h_i = 1$, this implies that the fitted value of the $i^{th}$ study could not be estimated without the data from that study, in other words, the fitted value of that study depends entirely on data from that study. This latter situation is equivalent to saying that there is a regression coefficient (or linear combination of regression coefficients) whose estimate is determined entirely by the data from $i^{th}$ study. In other regression contexts, reference value of $2q/k$ as been suggested as indicating a study of high leverage.

The term leverage arises from a mechanical analogy. Imagine a scatterplot of the effect size versus one predictor. In this one predictor situation, the studies that have $x$ (predictor) values that are far from the center of the data will have high leverage because moving them up or down would have large influence of the regression slope. When there is more than one predictor, there may be studies whose combination of predictor values is far from the center in a multivariate sense. The leverage diagnostic may reveal such multivariate outliers that are not obvious from looking at predictors one at a time.

**Cook’s Distance**

Cook’s distance, $D_i$, for the $i^{th}$ study is a measure of how much the estimated regression coefficients change (on the average) when the $i^{th}$ study is deleted from the dataset. Like the studentized and jackknifed residuals, $D_i$ is standardized, but unlike them it is in a squared (distance-squared) metric.

One can think of $D_i$ as the squared difference between $\mathbf{b}$ the vector of regression coefficient estimates estimated from all studies and the vector $\mathbf{b}_{(i)}$ of regression coefficients estimated from all studies except the $i^{th}$ study, divided by the variance of $\mathbf{b}$ that is

$$D_i = \frac{\left(\mathbf{b} - \mathbf{b}_{(i)}\right)^\prime V^{-1}\left(\mathbf{b} - \mathbf{b}_{(i)}\right)}{p + 1} = \frac{w_i h_i e_i^2}{q \left(1-h_i\right)^2}, \quad (1.24)$$

128
where $V$ is the covariance matrix of $b$. In other regression contexts, the value $4/(k - Q)$ been suggested to help identify studies that have large influence.

**DFITTS**

DFITTS is a diagnostic that describes the change in the fitted (predicted) value of the $i^{th}$ study that would arise as a consequence of deleting the $i^{th}$ study from the data to compute the regression coefficients used to compute the fitted value. DFITTS is defined as

$$DFFITS_i = \hat{T}_i - \hat{T}_{(i)i} = e_i \sqrt{\frac{w_i h_i}{(1 - h_i)^2}}, \quad (1.25)$$

where

$$\hat{T}_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} \quad (1.26)$$

and

$$\hat{T}_{(i)i} = b_{(i)0} + b_{(i)1} x_{i1} + \cdots + b_{(i)p} x_{ip} \quad (1.27)$$

where $b_{(i)}, \ldots, b_{(i)p}$ are the regression coefficients estimated with the $i^{th}$ study removed from the dataset.

Like the jackknifed residual, DFITTS is designed to better reveal cases where the $i^{th}$ study does not fit the same model as the other studies. By removing the (potentially distorting impact of the $i^{th}$ study from the computation of the regression coefficients, the jackknifed residual sometimes makes it easier to see how different an observed effect size is from what is expected if that study fit the meta-regression model that is appropriate for all of the other studies. In other regression contexts, the reference value $2\sqrt{q/n}$ has been suggested for identifying studies with potentially large impact on fitted values.

**Variance**

The variance $v_i$ of the $i^{th}$ study is the conditional (estimation error) variance of the effect size in the $i^{th}$ study. Because $v_i$ depends on the sample size in each study, $v_i$ can vary substantially across studies.

**Tau Squared**

Tau-squared, $\tau^2$, is the estimate of between-studies variance among effect size parameters at any point on the prediction line. An assumption of the meta-regression is that the true variance of effect sizes is the same for all values of the covariate.

**Sum**
Sum is the total variance of the \( i \)th effect size, which is \( v_i \) in fixed effect meta-regression or \( \tau^2 + v_i \) in random effects meta-regression.

**Weight**

The weight of the \( i \)th study, \( w_i \), is the actual (raw) weight assigned to this study in the analysis, namely the reciprocal of the total variance, namely \( w_i = 1/v_i \) in fixed effects meta-regression and \( w_i = 1/(\tau^2 + v_i) \) in random effects meta-regression.

**Percent Weight**

Percent weight for the \( i \)th study is the percentage of the total weight accorded to study, that is \( w_i \) divided by the sum of all study weights.

**How to use the diagnostics**

Regression diagnostics are designed to be simple checks that reveal important features of the data and the regression model fitted to that data. However, the multivariate situation is complex, and diagnostics are typically imperfect.

For example, consider the important feature of colinearity (correlations) among predictors. It is well known that colinearity can degrade the quality of estimates of regression coefficients by increasing their sampling uncertainty. This can occur when two predictors are highly correlated but independent of the others, or when there is a high multiple correlation among predictors (when one predictor is almost a linear combination of several others). These two situations have different implications for the quality of regression estimates. In the former case, only the two coefficients corresponding to the correlated predictors may be poorly estimated. In the latter case, the impact of colinearity may affect more coefficients. A diagnostic designed to reveal colinearity, in general, may be not be able to distinguish between the two types of colinearity. In the former case, only the two coefficients corresponding to the correlated predictors may be poorly estimated. In the latter case, the impact of colinearity may affect more coefficients. A diagnostic designed to reveal colinearity, in general, may be not be able to distinguish between the two types of colinearity. On the other hand, producing diagnostics tailored to a myriad of possible special cases increases the complexity of the suite of diagnostics, defeating the purpose of simple checks on data and the regression model.

We have implemented a set of diagnostics that have proven most useful in regression problems generally and adapted them to meta-regression. All of these diagnostics are related, in that they are different ways of looking at the extent to which the data associated with a study is inconsistent with the meta-regression model that fits the other studies. They approach the problem in different ways however.

- The leverage and Cook’s distance focus on the impact of a study on the estimated regression coefficients.
- The residual and studentized residual focus on the difference between fitted (predicted) effect sizes based on all the data and the observed effect size in each study.
- \( DFITTS \) and the jackknifed residual focus on the difference between fitted values of each effect size when the regression coefficients are estimated with and without a particular study.

It is important to recognize that, because these diagnostics are closely related, it is not surprising that the same studies may be flagged by several of the diagnostics as having high impact or influence. In fact, it more surprising (but not impossible) when a study is flagged by only one of them.
The diagnostics should not be used by themselves to exclude studies from inclusion in a meta-analysis. The diagnostics are intended to help identify studies that have substantial impact on the estimated regression coefficients. The reference values we have given are not intended to be used like critical values in a significance test, but as criteria for further evaluation. Just because a study has high impact on the analysis does not make it incorrect. However, it is useful to know that a certain study has (or a few studies have) substantial impact on the results. In such cases it is crucial to be sure of the integrity of the studies with high impact.

It is also important to know that the impact of a study may change when the set of covariates in the meta-regression is changed or the set of studies is changed (e.g., when a subset of studies is examined). A study that has high impact may have much less impact when a certain covariate is removed from the covariate set or a certain study is removed from the dataset.

**Variance Inflation Factor**

The variance inflation factor $VIF_j$ for the $j^{th}$ covariate is a diagnostic designed to provide information about the colinearity of the covariate set. One of the consequences of colinearity is that it increases the variance (the square of the standard error of) the regression coefficient estimates. If the standard error of a regression coefficient estimate is too large, it may be difficult to meaningfully interpret that estimate. For example, suppose that a particular coefficient expresses the difference between the average of two groups of standardized mean difference effect sizes, that the coefficient estimate is 1.0 and the standard error of that coefficient is 4. In such a case, it is difficult to draw an informative conclusion because the results imply a 95 percent confidence interval for the mean difference between group mean effects of -3 to +5, a range which is consistent with substantially different substantive conclusions.

The $VIF_j$ indicates how much greater the variance of the regression coefficient estimate $b_j$ for the $j^{th}$ covariate is than it would have been if the covariates were totally uncorrelated. A $VIF$ value of 4 for a particular covariate indicates that the standard error of that regression coefficient is twice as large as it would have been if the covariate were uncorrelated with all the other covariates.

A high $VIF_j$ value for the $j^{th}$ covariate does not necessarily mean that the standard error of the coefficient of that covariate is too large for the estimate to be meaningful. For example, suppose that a particular coefficient expresses the difference between the average of two groups of standardized mean difference effect sizes, that the coefficient estimate is 1.0, and the standard error of that coefficient is 0.1. In such a case, one can still draw an informative conclusion even if $VIF = 4$ because the results imply a 95 percent confidence interval for the mean difference between group mean effects of 0.8 to 1.2, a range which is consistent with substantially the same substantive conclusion of a very large difference between the group mean effect sizes.

Note that $VIF_j$ is not a property of the $j^{th}$ covariate alone, but depends on all of the other covariates as well. Therefore removing one covariate from the covariate set may change (sometimes drastically) the $VIF$ values of several other covariates.
COVARIANCE

This page gives us the covariance matrix of the $B$ values.

To understand what these covariances represent, imagine that we draw a sample of studies, run the regression, and get an estimate of $B_{Year}$ and $B_{Latitude}$. We repeat this process $j$ times, and each time get an estimate of $B_{Year}$ and $B_{Latitude}$. Then, we compute the covariance of $B_{Year}$ with $B_{Latitude}$ over the $j$ samples. This covariance would be 0.0003 [$B$]

The same idea applies to all cells in the matrix.
CORRELATIONS

To navigate to this page click More results > Correlation [A]

This page gives us the correlation matrix of the $B$ values.

To understand what these correlations represent, imagine that we draw a sample of studies, run the regression, and get an estimate of $B_{Year}$ and $B_{Latitude}$. We repeat this process $j$ times, and each time get an estimate of $B_{Year}$ and $B_{Latitude}$. Then, we compute the correlation of $B_{Year}$ with $B_{Latitude}$ over the $j$ samples. This correlation would be 0.8444 [B]

The same idea applies to all cells in the matrix.

When the correlation between two covariates is high (close to 1.0 or close to $-1.0$), this tells us that the two are highly confounded, and it is therefore difficult to isolate the unique impact of each. In this example, this is probably why (using random-effect weights) latitude was statistically significant when used by itself, but not when used in conjunction with year.
**INCREMENTS**

We use the term “increments” to refer to the process of adding one covariate at a time to the model, and studying the change in the variance explained. This approach provides some kinds of information that are not available in a single analysis that includes all covariates. For example, consider the [Main results] screen shown in Figure 88.

The test of the model, goodness of fit, estimates of $T^2$ and $R^2$, apply to the full model (Allocation, Year, and Latitude). Suppose we want to know these statistics if (a) we include only allocation, (b) we include allocation and year, and (c) we include allocation, year, and latitude. One way to get this information is to actually run a series of analyses, adding one covariate at each iteration.

While the idea of running a series of analyses will work, it can be a tedious process, and also requires the researcher to collate the results of all the analyses. To address this problem the program has automated the process. When you define a model with covariates X, Y, Z, the program will run an analysis with X, another with X and Y, and another with X, Y, and Z. It will then collate the results, showing the key statistics at each iteration. Additionally, it shows the change in $T^2$ and in $R^2$, as well as a statistical test for the change, at each iteration.
To make it clear how the increments work, we’re going to run a series of analyses and present the results for each. Then we’ll use these to understand the information on the increments screen. In practice, of course, you would need to run only one model (the one with all the covariates) and then jump directly to the increments screen.

First, we include only the intercept.

In Figure 89 we have added all the covariates to the main screen, but only the box for the intercept is ticked, so this will be the only covariate included in the analysis.
With only the intercept in the model (Figure 90),

- For variance explained by the model $Q = 0.0000, df = 0, p = 1.0000$ [A]
- For variance unexplained by the model $Q = 152.2330, df = 12, p = 0.0000$ [B]
- $R^2$ for the model is 0.00 [C]
In Figure 91 we add tick-marks for the two dummy variables that represent Allocation.

Figure 91 | Setup | Intercept + Allocation

With intercept + allocation in the model (Figure 92)

Figure 92 | Main results | Intercept + Allocation
• For variance explained by the model $Q = 1.4349$, $df = 2$, $p = 0.4880$ [A]
• For variance unexplained by the model $Q = 132.3676$, $df = 10$, $p = 0.0000$ [B]
• $R^2$ for the model is 0.00 [C]
In Figure 93 we add a tick-mark for Year and re-run the analysis

**Figure 93 | Setup | Intercept + Allocation + Year**

**Figure 94 | Main results | Intercept + Allocation + Year**

With intercept + allocation + year in the model (Figure 94)
• For variance explained by the model $Q = 10.7159$, $df = 3$, $p=0.0134$ [A]
• For variance unexplained by the model $Q = 30.3951$, $df = 9$, $p = 0.0004$ [B]
• $R^2$ for the model is .56 [C]
Finally, in Figure 95, we add a tick mark for latitude and re-run the analysis.

**Figure 95 | Setup | Intercept + Allocation + Year + Latitude**

With intercept + allocation + year + latitude in the model (Figure 96)
• For variance explained by the model $Q = 13.1752$, $df = 4$, $p = 0.0105$ [A]  
• For variance unexplained by the model $Q = 24.0144$, $df = 8$, $p = 0.0023$ [B]  
• $R^2$ for the model is $0.6133$ [C]
Alternatively, we could have jumped directly to the full model, run the analysis, and gone to the increments page.

To navigate to this page —

- Run the analysis that includes all the covariates (Figure 95)
- Click More results > Increments
- Select the statistical model tab (Fixed or random)

![Comprehensive meta-analysis](image)

**Figure 97 | Main results | Intercept + Allocation + Year + Latitude**

Every row in this table copies information from a separate analysis.

- The row labeled Intercept copies information from Figure 90.
- The row labeled Allocation (the second row labeled Allocation, since this is a set) copies information from Figure 92.
- The row labeled Year copies information from Figure 94.
- The row labeled Latitude copies information from Figure 96.

Each column in this table corresponds to a section in the prior figures

- [A] columns copy information such as $T^2$ from [A] in the prior screens
- [B] columns copy information about goodness of fit from [B] in the prior screens
- [C] columns copy information about $R^2$ from [C] in the prior screens

Additionally, this table presents information about change from one row to the next
• [AA] columns show the change in $T^2$ and in the test of significance
• [CC] column shows the change in $R^2$

Suppose we want information about the model that includes allocation and year. On the line labeled Year, we see that $T^2$ is 0.1349, $R^2$ is .5631, the model is statistically significant ($Q = 10.72$ with $df = 3$ and $p = 0.0134$) but fails to explain all the variance ($Q = 30.40$, $df = 9$, $p = 0.0004$). These statistics were copied from the analysis in Figure 94. If we return to that figure, we’ll see the same numbers.

The columns labelled [AA] and [CC] are unique to this screen, and address the change as we move from one model to the next. The column labeled “Change from prior” gives the change in $T^2$ and in $R^2$. The column labeled “Test of change” is the corresponding test of statistical significance.

For example, consider the line labeled “Year”. The table shows that $T^2$ changed by $-0.4247$ (which is the difference between 0.5596 on the prior line and 0.1349 for the current line). It shows that $R^2$ changed by 56.31% (which is the difference between 0.00% on the prior line and 56.31% on the current line). It shows the statistical test for the change yields $Q = 8.43$, $df = 1$, $p = 0.0037$.

Note that the test for change corresponds to the impact of each covariate at the point that it is entered into the model. Thus,

• The change for Allocation corresponds to the impact of allocation with no covariates held constant. The $p$-value of 0.4880 in Figure 97 corresponds to the $p$-value of 0.4880 in Figure 92.

• The change for Year corresponds to the impact of Year with allocation held constant. The $p$-value of 0.0037 in Figure 97 corresponds to the $p$-value of 0.0037 in Figure 94.

• The change for Latitude corresponds to the impact of Latitude with Year and allocation held constant. The $p$-value of 0.2331 in Figure 97 corresponds to the $p$-value of 0.2331 in Figure 96.

Note. The tests in the earlier screens report $Z$ rather than $Q$. In each case you could square the $Z$-value on the earlier figure to get the $Q$-value in Figure 97.
PART 8: THE $R^2$ INDEX

$R^2$ is the proportion of between-studies variance explained by the model. It is analogous to the $R^2$ index commonly reported for the proportion of variance explained by covariates in primary studies.

Consider the example shown in Figure 98. We’ve used latitude as the sole covariate.

![Figure 98 | Setup]
### Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.2595</td>
<td>0.2323</td>
<td>-0.1958</td>
<td>0.7149</td>
<td>1.1172</td>
<td>0.2639</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0292</td>
<td>0.0067</td>
<td>-0.0424</td>
<td>-0.0160</td>
<td>-4.3411</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

### Statistics for Model 1

**Test of the model**: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 18.8452, df = 1, p = 0.0000

**Goodness of fit**: Test that unexplained variance is zero
\( \tau^2 = 0.0633, \tau = 0.2516, I^2 = 54.21\%, Q = 30.7331, df = 11, p = 0.0012 \)

### Comparison of Model 1 with the null model

**Total between-study variance (intercept only)**
\( \tau^2 = 0.3088, \tau = 0.5557, I^2 = 92.12\%, Q = 152.2330, df = 12, p = 0.0000 \)

**Proportion of total between-study variance explained by Model 1**
\( R^2 \) analog = 0.79

**Number of studies in the analysis**: 13

---

Figure 99 | Main results | Latitude | Random-effects
In Figure 99 [A] the program shows that the proportion of variance explained by latitude is 0.79. Before turning to the computation, let’s take a moment and get an intuitive sense of what this means.

![Figure 100 | Dispersion of effects about grand mean vs. dispersion of effects about regression line]

Figure 100 includes two plots, as follows.

The left-hand plot shows the dispersion of all studies about a regression line where there are no covariates. The predicted effect size for each study is the grand mean, and the variance of true effect sizes about the mean ($\tau^2$) is 0.3088. We’ve superimposed a normal curve based on $\tau$, and it covers about 95% of all true effects in this population.

The right-hand plot shows the dispersion of all studies about a regression line based on latitude. The predicted effect size for each study is the regression line, and the variance of true effect sizes about the regression line ($\tau^2$) is 0.0633. We’ve superimposed a series of normal curves based on $\tau$. At any point on the regression line, the curve covers about 95% of all true effects for studies at that latitude.

The normal curves in the right-hand plot are smaller than the normal curve in the left-hand plot. This reflects the fact that by using the regression line at the right to predict effect sizes we are able to make better predictions. In fact, we are able to reduce the unexplained variance by 79.5%. The $R^2$ value describes this reduction in unexplained variance.

At this point we can return to the screen in Figure 99.

- To get the initial amount of variance we run a regression with no covariates and compute $\tau^2$. Here, $\tau^2$ is 0.3088 [B], which is the variance of all studies about the grand mean.

- To get the final amount of variance we run a regression with the covariates and compute $\tau^2$. This value, reported above as 0.0633 [C], is the variance of studies about their predicted value.

- If the initial $\tau^2$ is 0.3088 and the remaining $\tau^2$ is 0.0633, the difference (0.2455) is the $\tau^2$ explained by the model.
• The program [A] then uses these values to compute \( R^2 \), the proportion explained by the model, using either

\[
R^2 = \frac{T^2_{\text{Explained}}}{T^2_{\text{Total}}} = \frac{0.2455}{0.3088} = 0.7950, \tag{1.28}
\]

or equivalently,

\[
R^2 = 1 - \left( \frac{T^2_{\text{Residual}}}{T^2_{\text{Total}}} \right) = 1 - \left( \frac{0.0633}{0.3088} \right) = 0.7950
\]

The test that \( R^2 \) is zero in the population is the same as the test of the model. For the model, \( Q = 18.845, df = 1, p < 0.0001 \) [E], so we can conclude that \( R^2 \) in the population is probably not zero.

Notes

While the logic of \( R^2 \) is the same for primary studies and for meta-regression, the actual computation is different. (In primary studies the computation is based on the observed variance while in regression it is based on the true variance. In primary studies all observations are given the same weight, while in regression each study is given a different weight.) For this reason, when used with meta-analysis, the index is sometimes called the \( R^2 \) analog rather than \( R^2 \).

In Figure 100, the normal curves are drawn to capture 95% of the dispersion in true effects, not in observed effects. The observed variance is assumed to include both within-study variance and between-study variance (\( T^2 \)). We are concerned only with the latter.

The normal curves correspond to two standard deviations on either side of the regression line. By contrast, \( R^2 \) is based the ratio of variances. Therefore, while the variance on the right is 79.5% smaller than the variance on the left, the normal curve is not 79.5% smaller (the ratio of the variances is not the same as the ratio of the standard deviations). Nevertheless, the correspondence is close enough for the purposes of this illustration.
**THE SCHEMATIC FOR $R^2$**

The program features a schematic illustration of $R^2$ (Figure 101).

To navigate to this screen click More > R-squared graphic

---

In this figure, the bar represents the total variance in effects, $T^2$, which we saw earlier Figure 99 [D] is 0.3088. Note that this is not the observed variance (which includes within-study variance and between-study variance) but rather our estimate of the between-study (true) variance.

The full bar reflects the true variance of all effects about the regression line when there are no covariates in the equation (that is, the true variance of all effects about the grand mean). The light-blue and dark-blue parts of the bar reflect the fact that the total variance can be decomposed into parts that can and cannot be explained by the model.
When we include the covariates in the model $T^2$ is 0.0633. This is the variance of true effects about the regression line, or variance that cannot be explained by the model. This is represented by the light blue portion of the bar. If we look back at Figure 100, this is the ratio of the curves at right (squared) to the curves at left (squared).

The dark blue portion of the bar represents the between-study variance that can be explained by the model, which is shown as 0.2455. We get this value by subtraction. If the unexplained variance with no covariates is 0.3088 and the unexplained variance with covariates is 0.0633, then the variance explained by the covariates is the difference, 0.2455.

We define $R^2$ as the proportion of variance that is explained by the covariates. This is the proportion of the bar that is colored dark blue, in this case 0.79.

We can get $R^2$ by taking the ratio of the dark blue (explained) to the total, using

$$R^2 = \frac{T^2_{\text{explained}}}{T^2_{\text{total}}} = \frac{0.2455}{0.3088} = 0.7950,$$

(1.29)

Which is the formula displayed on the screen.

Alternatively, we can take the ratio of the light blue (not explained) to the total and subtract this from 1.0, using

$$R^2 = 1 - \left( \frac{T^2_{\text{residual}}}{T^2_{\text{total}}} \right) = 1 - \left( \frac{0.0633}{0.3088} \right) = 0.7950.$$  

(1.30)

A SEEMING ANOMALY

In both primary studies and meta-analyses $R^2$ is based on two estimates of the variance, i.e., with and without covariates.

In primary studies both estimates are based on the same data, and so are linked. If one estimate is too low the other will be too low also, and the computation of $R^2$ will be largely unaffected.

For meta-analysis, by contrast, the situation is different. In a meta-analysis the two estimates of $T^2$ are based on separate analyses, and it’s possible for one estimate to be low while the other is too high. A simplified version of the possible outcomes is shown in Table 4.

<table>
<thead>
<tr>
<th>$T^2$ with no covariates</th>
<th>$T^2$ with covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underestimate $T^2$</td>
<td>$R^2$ could be accurate (A)</td>
</tr>
<tr>
<td>Overestimate $T^2$</td>
<td>$R^2$ will be too high (C)</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>$T^2$ with no covariates</th>
<th>$T^2$ with covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underestimate $T^2$</td>
<td>$R^2$ could be accurate (A)</td>
</tr>
<tr>
<td>Overestimate $T^2$</td>
<td>$R^2$ will be too low (B)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T^2$ with no covariates</th>
<th>$T^2$ with covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underestimate $T^2$</td>
<td>$R^2$ could be accurate (D)</td>
</tr>
<tr>
<td>Overestimate $T^2$</td>
<td>$R^2$ will be too high (C)</td>
</tr>
</tbody>
</table>

150
Consider the situation in Cell (C), where we overestimate the initial variance and then underestimate the final variance. It seems we have explained more variance than we actually did.

Conversely, consider the situation in Cell (B), where we underestimate the initial variance and then overestimate the final variance. It seems we have explained less variance than we actually did.

While the error is easiest to see in cells B and C, there will also be error in cells A and D. Even if we underestimate both variances (Cell A) or underestimate both variances (Cell D), the magnitude of the errors will invariably differ, which will affect the estimate of $R^2$.

When the true value of $R^2$ is large, these errors are not obvious. If the true value of $R^2$ is .40 and we underestimate that value by .10, we simply assume that the correct value is 0.30. However, if the true value of $R^2$ is near zero and we underestimate the value, then our estimate could fall below zero.

For example, suppose the initial and final values of $T^2$ are 0.20 and 0.19. Suppose further that the initial estimate is low and the second is high, so the observed values are 0.19 and 0.20. In this case it will appear that the unexplained variance has increased, which would mean that $R^2$ is negative. It’s easiest to see this in Cell B, but this can also happen in cells A or D if one estimate has more error than the other, even if both are in the same direction.

Since this must be due to sampling error (a proportion of variance cannot be negative) we simply set the value of $R^2$ to zero.
ASSESSING CHANGE IN THE MODEL

Suppose we run a model with only year as the covariate. Then we run a model with year plus latitude. Just as we can report statistics for either model as compared with the null model (the intercept only) we can also report statistics for the second model as compared with the first.

Specifically,

- We can test the $Q$-value for change in the model, to see if latitude adds any improvement in prediction over and above year.
- We can report the change in $R^2$ with the change in the model.
- And, we can report the change in $R^2$ with the change in the model.

Here, we show how to assess change as we start with only the intercept, add Year, and then add Latitude to the prediction model.

In Figure 103 we define a prediction model with the Intercept, Year, and Latitude.
Run the analysis

- Click More results > Increments
- Select the statistical model tab (Random)
- Click More results > Increments

The table of increments (Figure 105) collates the results of three analyses, as follows

- Intercept only [D]
- Intercept plus year [E]
- Intercept, Year, and Latitude [F]

At each iteration the program also displays the change in $T^2$, the change in $R^2$, and the statistical significance of the change.

Figure 104 | Display increments
To assess the statistical significance of the model vs. the null model (or the change in $R^2$ vs. 0.0) we use the columns at the left (G)

- $T^2$ with no covariates is 0.3088, which serves as our baseline for computing $R^2$.

- When we use year as a covariate $T^2$ drops to 0.2377, and $R^2$ is computed as .2303. The test that $R^2$ is zero is given by $Q = 2.21$, $df = 1$, $p = 0.1368$.

- When we use year and latitude as covariates $T^2$ drops to 0.0921 and $R^2$ is computed as .7018. The test that $R^2$ is zero is given by $Q = 14.30$, $df = 2$, $p = 0.0008$.

To assess the statistical significance of the model vs. the prior model (or the change in $R^2$ vs. the prior $R^2$) we use the columns at the right (H)

- $T^2$ with no covariates is 0.3088, which serves as our baseline for computing $R^2$.

- When we use year as a covariate, $T^2$ drops to 0.2377 and $R^2$ is computed as .2303. The change in $R^2$ for the second line vs. the first (that is, adding year to the null model) is .2303. The test that the change is zero given by $Q = 2.21$, $df = 1$, $p = 0.1368$. For this row, statistics for change versus
the prior model are identical to statistics for change versus the null model (since the prior model is the null model).

- When we use year and latitude as covariates $T^2$ drops to 0.0921 and $R^2$ is computed as .7018. The change in $R^2$ for the third line vs. the second (that is, adding latitude to year) is .4715. The test that the change is zero given by $Q = 9.17, df = 1, p = 0.0025$. 


UNDERSTANDING $I^2$

In Figure 106 the program shows results for a regression that included latitude as a covariate. The screen displays two estimates of $I^2$, as follows

- For a regression with no covariates [A] covariates, $I^2$ is 92.12%
- For a regression with covariates [B], $I^2$ is 64.21%

In any meta-analysis, the dispersion of observed effects can be partitioned into two parts. One is the dispersion of the true effects, and the other is dispersion due to sampling error.

The $I^2$ statistic gives us the ratio of true to total variance. Since we see the variance of the observed effects, but we care about the variance of the true effects, we can use $I^2$ to serve as a link between the two. If we start with the variance of the observed effects and multiply this by $I^2$, we get the variance of the true effects. Put simply, we get a sense of what the dispersion would look like if each study had a really large sample size (and therefore minimal error).

If $I^2$ is near 100%, then a plot of the true effects would look similar to the plot of the observed effects. As $I^2$ moves toward 0%, more and more of the observed variance is simply sampling error, and would disappear if the studies were large enough (and we thus eliminated the sampling error).

In a regression we report estimates for two distinct types of $I^2$.

On row A we report statistics for a regression with no covariates. On this line $T^2$ is the variance of true effects about the regression line (which here is simply the grand mean). $I^2$ tells us that proportion of the variance in observed effects about the regression line would remain if all studies had an extremely large sample size, so that essentially all error was removed.

On row B we report statistics for a regression with covariates. On this line $T^2$ is the variance of true effects about the regression line (which here is based on latitude). $I^2$ tells us that proportion of the variance in observed effects about the regression line would remain if all studies had an extremely large sample size, so that essentially all error was removed.

In both cases, the interpretation of $I^2$ is the same. If we are presented with the variance of observed effects and we want to know the variance of the true effects, we multiply the former by $I^2$ to get the latter. Note that if we multiply the observed variance by $I^2$ the value we get is $I^2$, which is the value presented on rows A and B. While $T^2$ is the number we will actually employ in our computations (for example, to assign weights or to compute a prediction interval), $I^2$ offers a way to get a visual sense of how the plot would change if we could somehow eliminate the error.
Summary

If we start with a plot of the effects (either about the grand mean or about the regression line), some of the observed dispersion reflects differences in the true effects, while some reflects sampling error. If we’re interested in the actual utility of the intervention, then we care about the former, not the latter.

$I^2$ tells us what proportion of the variance reflects the former ($\tau^2$) and what proportion reflects the latter ($V$). It also provides a mechanism that allows us to get a sense of what the plot would look like if it was based on the true effects rather than the observed effects. Specifically, if we construct a normal curve that captures most of the observed effects, we could multiple that curve’s height by a factor of $I$. This gives us the approximate distribution of the true effects.

Critically, $I^2$ is a proportion of variance, not an absolute variance. An $I^2$ near 100% tells us that most of the observed variance is due to variation in true effect sizes, but it does not tell us that this variance is substantial. Conversely, a low value of $I^2$ tells us that only a small proportion of the observed variance is due to variation in the true effect sizes, but does not tell us that this variance is trivial.
PART 9: WORKING WITH THE PLOT

You can create a regression plot with one click and then modify it extensively. We will use the BCG analysis for this illustration.

- Create the model shown in Figure 107.
- Then, click Run regression
At this point the program displays a [Scatterplot] button on the toolbar

- Click [Scatterplot] [A]

Figure 108 | Main results | Random-effects
The program displays the screen shown in Figure 109.

The regression line is based on the regression equation

- The variable on the X-axis varies
- Continuous variables are plotted at their means
- Categorical variables are plotted based on the proportion of studies in each category. For example, suppose that we’ve included “Hot” as a covariate, which is coded 0 for studies in cold climates and 1 for studies in Hot climates. Suppose further than 7/13 studies are coded 1. The mean score for Hot would be 0.54, and the regression would be plotted for studies where Hot is 0.54.

The confidence interval and prediction interval are based on the uncertainty of the coefficient of the variable on the X-axis and the intercept—it does not depend on the uncertainty of the coefficients for the other variables.
To set the variable for the X-axis

In Figure 110,

- Click [Graph by] on the tool bar [A]
- Or right-click on the variable name on the X-axis [B]
The main screen includes four elements – the studies, regression line, confidence interval, and prediction interval. Each can be set to show or hide independently of the others.

For purposes of this tutorial, un-check the buttons as shown in Figure 111 [A].

- Studies <Off>
- Regression line <Off>
- Confidence interval <Off>
- Prediction interval <Off>
Studies

Click [Studies] to display/hide the individual studies as in Figure 112.

The program displays each study as a circle

- To set the circles to be proportionate to the study weight (or not) Click Format > Studies
- To edit the appearance of the circles Click Format > Studies
- To modify the color of the circles Click Color > Edit colors > Studies

Figure 112 | Plot of log risk ratio on Latitude | Studies
Regression line

Click [Regression line] to display/hide the regression line as in Figure 113.

The program displays the regression line. This reflects the predicted effect size (on the Y-axis) for any given value (on the X-axis).

- To edit the appearance of the regression line Click Format > Regression line
- To modify the color of the regression line Click Color > Edit colors > Regression line
CONFIDENCE INTERVAL AND PREDICTION INTERVAL

The confidence interval and prediction interval are two very different indices. The confidence interval reflects the precision with which we estimate the mean value, while the prediction interval reflects the actual dispersion of effects about the mean value. The former is based on the standard error, and the latter on the standard deviation.

For example, consider a simple meta-analysis (no covariates) where we report the mean effects size.

The confidence interval is a measure of precision. If the mean effect is 0.5 with a confidence interval of 0.4 to 0.6, this tells us that the mean effect in this population (the population of studies from which the sample was drawn) probably falls in the range of 0.4 to 0.6. The estimate will become more precise as the number of studies increases, and with an infinite number of studies the confidence interval will have a width that approaches zero.

By contrast, the prediction interval is a measure of dispersion. It does not tell us about the mean effect but rather about the dispersion of effects about that mean. Suppose that the mean effect was 0.5 but the true effects ranged from 0.3 to 0.7. Suppose further that we had an infinite number of studies, and therefore knew the mean effect precisely. The confidence interval would be 0.5 to 0.5, but the prediction interval would be 0.3 to 0.7.

In real life, of course, we don’t have an infinite number of studies and therefore we don’t know the mean precisely. If we estimate the mean as 0.5 and we estimate that 95% of studies will fall within 0.2 on either side of the mean, then the prediction interval will take into account the plus/minus 0.2 and add to that the uncertainty in the mean. This is not as simple as adding the width of the confidence interval to the width of the prediction interval (we actually work with the variances) but that’s the general idea.

These same ideas apply to meta-regression as well. For any covariate we can estimate the coefficient, and then the confidence interval and the prediction interval for any value of the corresponding covariate. We can then display these on the graph.

For example, suppose the regression line shows the relationship between latitude and effect size. We pick a latitude of X and find that the predicted effect is 0.50 with a confidence interval of 0.40 to 0.60 and a prediction interval of 0.30 to 0.70. This means that in the universe of studies from which we sampled

- The mean effect for a study at latitude X is probably in the range of 0.40 to 0.60.
- The effect size for any single study usually falls in the range of 0.30 to 0.70.

Note that the prediction interval only makes sense when we apply random-effect weights. When we apply fixed-effect weights we assume that all studies at any given latitude have the same true effect size. By definition, the deviation of true effects about the predicted effect is zero. It follows that if the CI is 0.40 to 0.60, the PI will also be 0.40 to 0.60.
When working with the confidence interval or the prediction interval we need to base the intervals on one-point or simultaneous computations.

Click [Computational options] > [One point] or [Simultaneous]

- One-point – In 95% of analyses, the confidence interval at any single latitude will include the true mean effect for that latitude.
- Simultaneous – In 95% of analyses, the confidence interval at all latitudes will include the true mean effect for those latitudes.

- One-point – In 95% of analyses, the prediction interval at any single latitude will include the true effect for a study selected at random at that latitude.
- Simultaneous – In 95% of analyses, the prediction interval at all latitudes will include the true effect for a study selected at random at that latitude.

These examples assume that the confidence level has been set to 95%.

Finally, note that the CI and PI are relatively narrow at the mean of X and get wider as we depart from the mean. This is because any error in the coefficient gets multiplied as we depart from the mean of X in either direction.
Confidence interval

Click [Confidence interval] to display/hide the confidence interval as in Figure 114.

The Confidence interval is a measure of actual dispersion. It addresses the mean effect for any given latitude.

In Figure 114,

- In our sample of studies the mean effect size for a study at any given latitude is indicated by the regression line [C].
- In the universe from which we sampled, the mean effect size for a study at any given latitude probably falls in the confidence interval [B] to [D].

Click [Computational options] > [One point] or [Simultaneous]

- One-point – In 95% of analyses, the confidence interval at any single latitude will include the true mean effect for that latitude.
- Simultaneous – In 95% of analyses, the confidence interval at all latitudes will include the true mean effect for those latitudes.
- These examples assume that the confidence level has been set to 95%.

- To set the confidence level (e.g., 90% or 95%) click [Computational options]
- To set the confidence interval to be based on Z or Knapp-Hartung click [Computational options]
- To edit the appearance of the confidence line Click Format > Confidence interval
- To modify the color of the confidence line Click Color > Edit colors > Confidence interval
**Prediction interval**

Click [Prediction interval] to show/hide the prediction interval as in Figure 115.

![Figure 115](image)

**Where the confidence interval is an index of precision, the prediction interval is an index of dispersion.** In Figure 115,

- In our sample of studies the mean effect size for a study at any given latitude is indicated by the regression line [C].
- In the universe from which we sampled, the mean effect size for a study at any given latitude probably falls in the confidence interval [B] to [D].
- In the universe from which we sampled, the true effect size for a single study at any given latitude probably falls in the prediction interval [A] to [E].

Click [Computational options] > [One point] or [Simultaneous]

- One-point – In 95% of analyses, the prediction interval at any single latitude will include the true effect for a study selected at random at that latitude.
- Simultaneous – In 95% of analyses, the prediction interval at all latitudes will include the true effect for a study selected at random at that latitude.
- These examples assume that the confidence level has been set to 95%.

- To set the confidence level (e.g., 90% or 95%) click [Computational options]
- To set the prediction interval to be based on Z or Knapp-Hartung click [Computational options]
- To edit the appearance of the confidence line Click Format > Prediction interval
- To modify the color of the confidence line Click Color > Edit colors > Prediction interval
To identify specific studies

The program allows you to identify any study in the plot.

In Figure 116,

- Click [Identify study] [A]
- Click on any study [B]
- The program displays the study name [C]
Other options for customizing the graph are as follows

Appearance
- Line width: Format > Line width
- Font: Font
- Font size: Format > Font size

Title and labels
- Title: Labels > Title
- X-Axis: Labels > X-axis
- Y-Axis: Labels > Y-axis

Study circles
- Proportionate: Format > Studies
- Line width: Format > Studies

Axes
- Scale for X-axis: Format > X-axis
- Scale for Y-axis: Format > Y-axis
- Decimals: Format > Decimals

Statistical Model
- Fixed: Select Fixed tab at bottom of screen
- Random: Select Random tab at bottom of screen

Predictive model
- Model 1: Select desired model at bottom of screen

Export
- To Word: Files > Export to Word
- To PowerPoint: Files > Export to PowerPoint
- To File: Files > Export to File
- To Clipboard: Files > Copy to clipboard

Comments
- Equation: Show / Hide / Edit (The prediction equation)
- Annotation: Show / Hide / Edit (For the confidence interval and prediction interval)
- Comment 1: Show / Hide / Edit (User’s optional comment)
- Comment 2: Show / Hide / Edit (User’s optional comment)

Decimals
- Equation in plot: Select number
- X-Axis: Select number
- Y-Axis: Select number
Modify the colors

The program maintains two color schemes. These are called Printing and PowerPoint but can actually be used for any purpose. To switch between the schemes click

- Color > Use colors for printing
- Color > Use PowerPoint

After you’ve selected one scheme or the other, you can edit the color for any element on the screen. To modify colors for the current color scheme click

- Color > Edit colors (for current scheme)
Categorical variables

The plot for categorical variables works the same way as for continuous variables.

In Figure 117 we will plot by Allocation. This is a categorical variable that reflects the mechanism by which patients were assigned to either vaccine or placebo. The group names, corresponding to the type of allocation, are “Randomized”, “Systematic”, and “Alternate”.

In Figure 117 we define the prediction model, then click [Run regression].
Figure 118 shows the main results.

![Image of a table and a graph showing statistical results for Model 1, Random effects (MM), Z-Distribution.](image)

**Main results for Model 1, Random effects (MM), Z-Distribution**

<table>
<thead>
<tr>
<th>Set</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>Z-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-29.2359</td>
<td>45.7992</td>
<td>-119.0007</td>
<td>60.5289</td>
<td>-0.6383</td>
<td>0.5232</td>
<td></td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0190</td>
<td>0.0159</td>
<td>-0.0503</td>
<td>0.0122</td>
<td>-1.1524</td>
<td>0.2331</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>0.0148</td>
<td>0.0232</td>
<td>-0.0306</td>
<td>0.0683</td>
<td>0.6594</td>
<td>0.5225</td>
<td></td>
</tr>
<tr>
<td>Allocation</td>
<td>0.4805</td>
<td>0.4809</td>
<td>-0.4570</td>
<td>1.4279</td>
<td>1.0596</td>
<td>0.3127</td>
<td></td>
</tr>
<tr>
<td>Allocation: A</td>
<td>0.4574</td>
<td>0.3778</td>
<td>-0.2831</td>
<td>1.1570</td>
<td>1.1106</td>
<td>0.2260</td>
<td></td>
</tr>
</tbody>
</table>

**Statistics for Model 1**

- **Test of the model**: Simultaneous test that all coefficients (excluding intercept) are zero
  - $Q = 13.1742$, df = 4, $p = 0.0165$

- **Goodness of fit**: Test that unexplained variance is zero
  - $\tau^2 = 0.1194$, $\hat{\tau}^2 = 0.3456$, $I^2 = 66.69\%$, $Q = 24.0144$, df = 8, $p = 0.0023$

- **Comparison of Model 1 with the null model**
  - **Total between-study variance (intercept only)**
    - $\tau^2 = 0.3088$, $\hat{\tau}^2 = 0.5557$, $I^2 = 92.12\%$, $Q = 152.2330$, df = 12, $p = 0.0000$
  - **Proportion of total between-study variance explained by Model 1**
    - $R^2$ analog = 0.61

- **Number of studies in the analysis**: 13

**Figure 118 | Regression | Main results | Random-effects**
In Figure 119,

- Click [Scatterplot] [A]
- Click Graph by > Allocation [B]

The plot shows one column for every category (random, alternate, and systematic). The options to show or hide the studies, regression line, confidence interval, and prediction interval are the same as they were for continuous covariates.

While there are only two dummy variables (Alternate and Systematic) the program automatically adds a column for the reference category (Randomized).
Setting the scales

The program will automatically set the scale for the X-axis and Y-axis. This works well for any single plot. However, but if you want to create a series of plots and ensure that these all employ with the same scale you’ll need to set the Y-axis. Otherwise, the Y-axis may differ from one plot to the next, making it difficult to compare plots.

To set the Y-axis manually click Format > Y-axis.

Figure 120 | Regression | Plot | Setting the scale anchors
PART 10: COMPUTATIONAL OPTIONS

The program allows you to set various computational options.

On the regression screen click Computational Options on the menu.

Figure 121 | Regression | Set statistical options
In primary studies when we perform a significance test we have the option to use either the Z-test or the t-test. We use the Z-test when the population variance is known, and we use the t-test when we are using the sample variance to estimate the population variance.

The choice of a test (t vs. Z) affects the p-value in two ways.

- First, the estimate of the standard error is greater with t than with z, and so the test statistic is smaller.
- Second, when we use the t-test the value required for statistical significance is larger than it is with Z.

The difference between the two tests is most pronounced when the sample size is small. Once the sample size passes thirty the difference between t and Z is minor, and at one-hundred the difference is trivial.

While the choice between t and Z applies to cases where we compare two groups, the same idea applies to cases where we compare more than two groups. Here, the choice is between the F statistic (when the variance is estimated) and chi-squared (when the variance is known). The choices are shown in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Variance estimated</th>
<th>Variance known</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two groups</td>
<td>t</td>
<td>Z</td>
</tr>
<tr>
<td>More than two groups</td>
<td>F</td>
<td>(\chi^2)</td>
</tr>
</tbody>
</table>

We are faced with a similar situation in meta-analysis. Since the variances are often being estimated from the observed data, it would make sense to use the t distribution to test the null hypothesis and to construct confidence intervals. In fact though, researchers have traditionally used the Z distribution for these purposes.

In the case of a fixed-effect model this distinction turns out to have little practical impact. The only source of error is the variance within studies, and since the \(n\) within studies (accumulated across studies) is typically well over thirty, the difference between t and Z is negligible. Therefore, the practice of using z has not been challenged.

However, in the case of a random-effects model, the situation is more complicated. Recall that the error component incorporates two distinct elements – the within-study error and the between-study error. We can justify using Z for the within-study error for the same reason that we justify that approach for the fixed-effect model. However, the between-study variance is based on the number of studies, which is typically small, and the difference between t and Z for this component of the variance is typically substantial.
The solution proposed by Knapp and Hartung is to address each component of the variance separately. Specifically, we would use the $Z$ (or chi-squared) distribution for the within-study variance and the $t$ (or $F$) distribution for the between-study variance.

- When we are estimating the mean effect size in one set of studies this approach would apply to the test of the null.
- In a subgroups analysis this would apply to the test that compares the subgroup means.
- In a meta-regression it would apply to the test of each covariate and to the test of the model.

Note.

The program allows you to select either option from the statistics menu.

When you select [Z-Distribution] the program uses $Z$ and $Q$ (Figure 123 and Figure 124). When you select [Knapp-Hartung] the program uses $t$ and $F$ (Figure 125 and Figure 126).
Figure 122 shows a prediction model using Allocation, Year, and Latitude as covariates.
Click Computation options > Z-distribution (Figure 123)
Figure 124 shows the main-results screen with this option in effect.

The screen’s title shows that the Z-distribution is being employed:

- [A] The standard errors are based on Z
- [B] The confidence intervals are based on Z
- [C] The test statistics and p-value for individual covariates are based on Z
- [D] The test statistic and p-value for the set is based on Q
- [E] The test statistic and p-value for the model are based on Q
- [F] The test statistic and p-value for Goodness of fit are based on Q
- [G] The test statistic and p-value for the model with only the intercept are based on Q
To select Knapp-Hartung click Computation options > Knapp-Hartung (Figure 125).
Figure 126 shows the main results with this option in effect.

The screen’s title shows that Knapp-Hartung (KH) is being employed

- [A] The standard errors are based on $t$
- [B] The confidence intervals are based on $t$
- [C] The test statistics and p-values for individual covariates are based on $t$
- [D] The test statistic and p-value for the set are based on $F$
- [E] The test statistic and p-value for the model are based on $F$
- [F] The test statistic and p-value for Goodness of fit are based on $Q$, and not $F$. This is because this tests the null that $T^2$ is zero. Since the KH adjustment only affects the $T^2$ part of the variance, when $T^2$ is zero the adjustment is not applied.
- [G] Statistics for the model with only the intercept are based on $Q$, and not $F$. This is because this tests the null that $T^2$ is zero. Since the KH adjustment only affects the $T^2$ part of the variance, when $T^2$ is zero the adjustment is not applied.

In addition to using the $t$-distribution or $F$-distribution for the critical values, the standard error is adjusted for all of these.
Compare Figure 124 which is based on \( Z \), with Figure 126 which is based on Knapp-Hartung.

**Table of coefficients**

When we move from a \( Z \)-score to Knapp Hartung

- The coefficients do not change
- The standard error increases [A]
- The confidence interval width increases [B]
- The \( Z \)-score is replaced by a (smaller) \( t \)-score [C]
- The \( p \)-value becomes larger (less significant) [C]
- The \( Q \)-value for a set is replaced by a (smaller) \( F \)-score [D]
- The \( p \)-value for a set becomes less significant [D]

**Test of the model**

When we move from a \( Z \)-score to Knapp Hartung

- The \( Q \)-value is replaced by a (smaller) \( F \)-value
- The \( p \)-value becomes less significant

**Goodness of fit**

When we move from a \( Z \)-score to Knapp Hartung [F]

- The numbers do not change. This is because the Knapp-Hartung adjustment only applies to the \( T^2 \) part of the variance, but the goodness of fit test is computed assuming \( T^2 \) is zero.

**Comparison of Model 1 with the null model**

When we move from a \( Z \)-score to Knapp Hartung [G]

- The numbers do not change. This is because this comparison employs weights based on within-study variance (\( V \)), and the Knapp Hartung adjustment only affects between-study variance (\( T^2 \)).

**Notes**

While it is always true that the \( p \)-value will be the same or higher (further from zero) for Knapp-Hartung (KH), the extent of the difference depends on the amount of between-study variance and the number of studies. To the extent that the between-study population variance is small and/or the number of studies is large, the between-study error variance will be small, and the difference between the \( Z \) option and the KH option will tend to be relatively small. Conversely, to the extent that the between-study population variance is large and/or the number of studies is low, the difference between the two options will tend to be relatively large.
You do not need to return to the [Modify models] screen to switch between Z and Knapp-Hartung. Rather, if you’re already looking at the results you can simply change the setting and the results will change.

Figure 124 and Figure 126 showed the impact of this option for the main screen, but the impact actually affects all screens that display confidence intervals and/or tests of significance. The option also affects the plots, since the confidence interval and prediction interval depend on the standard error and the statistical distribution.

Since the Knapp-Hartung option is intended to address uncertainty in between-studies variance, most people who use it do so only for the random-effects model. In CMA, the Knapp-Hartung option is only available for random-effects models.

While these adjustments can be applied to any use of the random-effects model (that is, for a single group of studies, for a subgroup analysis, and for meta-regression), to date we have only implemented them for the meta-regression. We plan to update the other modules in the future.

The intent of the Knapp-Hartung adjustment is to improve the accuracy of p-values, confidence intervals, and prediction intervals. Higgins and Thompson (2004) proposed an approach that bypasses the sampling distributions and instead employs a permutation test to yield a p-value. Using this approach we would compute the Z-score corresponding to the observed covariate. Then, we would randomly redistribute the covariates among studies and see what proportion of these re-distributions yield a Z-score exceeding the one that we had obtained. This proportion may be viewed as an exact p-value. This option is not implemented in CMA.
**ONE-POINT OR SIMULTANEOUS CONFIDENCE INTERVALS FOR GRAPH**

When you plot the regression line of effect size on a covariate in a meta-regression, the program allows you to plot the confidence interval. The confidence interval reflects the uncertainty in the predicted value (the height of the regression line) being plotted. For example, if we plot the regression line for effect size on latitude, the confidence interval reflects the uncertainty in the predicted value of effect size for each value of latitude, but it treats all other covariates as fixed at their mean.

There are two useful options for plotting the interval. We can plot an interval that is accurate for any single point on the graph, or an interval that is accurate for all points on the graph simultaneously. To select either option:

Click Computation options > Simultaneous/One-point

- **Accurate for one point** means that if we were to select any *one point* on the regression line at random, in 95% of all possible regressions, the true predicted value for that point would fall within the confidence interval displayed at that point.

- **Accurate for all points** means that if we were to look at *all points* on the regression line, in 95% of all possible regressions, the true predicted value for all the points would fall within the confidence interval displayed at that point. Note that this includes predictions for all possible values of latitude, not only those that happen to appear in the data set.

Obviously the second criterion is stricter (we want to make an inference about the predicted value for all values of latitude rather than one) and therefore, the confidence interval will need to be wider. We do this by using a multiplier based on the Sheffé adjustment.

The difference between the two can be seen by comparing Figure 127 (one-point) with Figure 128 (simultaneous). The regression line is identical in the two, but the confidence interval is wider in the second.

This is true for all points on the plot, but is most evident toward either end of the regression line, since uncertainty in the coefficient becomes more evident as we depart from the mean of the predictor. For example, compare the width of the one-point confidence interval in Figure 127 [A] versus the simultaneous confidence interval in Figure 128 [B].

To facilitate this comparison we used Format > Y-Axis to set the same scale for both plots.

Click Comments > Show annotation to include the details on the plot, as shown in the bottom right-hand corner of each figure.
Figure 127 | Set statistical options | One-point confidence intervals

Figure 128 | Set statistical options | Simultaneous confidence intervals
OPTIONS FOR ESTIMATING $\tau^2$ (MM, ML, REML)

When we select random-effects the program needs to estimate the value of tau-squared ($\tau^2$), the true between-studies variance. (We use the Greek symbol $\tau^2$ to represent the true value, and $T^2$ to represent the sample estimate of that value).

To select a method click [Computational options] as shown in Figure 129.

There are three approaches commonly used to partition the variance and estimate $\tau^2$. These are

- Method of moments (MM), also known as the DerSimonian and Laird method
- Unrestricted maximum likelihood (ML) also known as maximum likelihood
- Restricted maximum likelihood (REML)

Each of these methods has advantages and disadvantages.

If we are not willing to assume that the effect sizes are normally distributed, MM is often the method of choice. The method of moments does not depend on any assumptions about the distribution of the random effects, so it has a robustness characteristic that the two other methods (which involve the assumption that the random effects have a normal distribution) do not have. If we are willing to assume a normal distribution of effects, then statisticians tend to prefer ML or REML, which are more efficient than MM (the estimates have smaller variance).

Between MM and REML, ML tends to yield a more precise estimate of $T^2$ (but with a bias) while REML tends to yield a less biased estimate (but with less precision). With small numbers of studies imprecision can be more important than bias, and so some prefer ML. With more studies, the balance may shift in favor of REML.
Note.

If you have already run the analysis and want to modify the statistical option, you do not need to return to [Modify models] and re-run the analysis. Simply click [Computational options] and make a selection.

All three options for estimating $\tau^2$ can be used for a basic analysis as well as for regression. However, the basic analysis module in CMA offers the MM option only. (We plan to add other options in the future).
**ONE-SIDED VS. TWO-SIDED TESTS**

Many statistical tests allow the option of one-sided or two-sided.

- Two-sided tests are appropriate when an effect in either direction would be meaningful.
- One-sided tests are appropriate when we only need to identify an effect in one direction, and an effect in the other direction would have the same implications as zero effect.

In the overwhelming majority of social-science and medical research, while we may (indeed almost always do) expect the effect to fall in a specific direction, an effect that was statistically significant in the other direction would still be important. For example, if we expected the treatment to improve survival but it turned out to hurt survival, this would be critically important information. However, if the test had been performed as one-tailed then an effect in the reverse direction (that the treatment is harmful) cannot be statistically significant by definition, even if the computed p-value is < 0.0001. Therefore, except in rare instances, the two tailed test is appropriate.

In the event that you select a one-tailed test, note that this applies only to the p-values for individual covariates on the main results screen.

- It does not affect the confidence interval since this is displayed for lower and upper limits.
- It does not affect the p-value for the test of the model. Since this is based on Q (or F) no direction can be specified and it must be two-tailed.
- It does not affect the p-value for a set of covariates. Since this is based on Q (or F) no direction can be specified and it must be two-tailed. (For consistency, this applies even if the set incudes only one covariate.)
- It does not affect the p-value for a test of the increment. Since this is based on Q (or F) no direction can be specified and it must be two-tailed.
- It does not affect the confidence interval nor the prediction interval on the plot. Since these are shown for both the lower and upper limit, they are displayed using multipliers for a two-tailed test.
PART 11: CATEGORICAL COVARIATES

Categorical covariates are covariates that represent a category or group, rather than a numerical score. For example, the covariate “Allocation” reflects the mechanism employed to assign patients to either vaccine or placebo. Each study is coded as “Randomized”, “Systematic”, or “Alternate”.

When we perform a subgroups analysis (as with an analysis of variance in a primary study) we can work directly with categorical covariate and classify each study by its allocation method (e.g., “Systematic”). However, this not possible when we perform a regression, since regression requires that we work with numbers, not labels. Therefore, rather than working with the original variable we create so-called “dummy variables”, numeric variables that stand for a group or category.

In this chapter we will discuss how to create and interpret these dummy variables. As always, we assume that the reader is familiar with the use of dummy variables in primary regression, and our intent is to show how the same rules apply for meta-regression.

Note. The mechanism for working with dummy variables in a regression depends on whether or not we include the intercept in the regression equation. In this chapter we assume that we will include the intercept. The alternative is discussed in [Part 12: When does it make sense to omit the intercept].

Overview

For a categorical variable with $m$ groups, we need to create $m - 1$ dummy variables. Since Allocation has three groups (Randomized, Systematic, Alternate) we need to create two dummy variables.

We need to select any one of the three groups to serve as the “Reference” group. Then, we create a dummy variable for each of the other two groups (but not for the reference group). With three groups, we have three options –

A. We can select Randomized to serve as the reference group. The dummy variables will be Systematic and Alternate.
B. We can select Systematic to serve as the reference group. The dummy variables will be Randomized and Alternate.
C. We can select Alternate to serve as the reference group. The dummy variables will be Systematic and Randomized.

In this chapter we discuss –

- How to create the dummy variables
- How to use these in the regression
- How to select the reference group
Dummy variables

CMA is able to create the dummy variables automatically. We can use the Allocation variable as an example. In Figure 130,

- Click on Show Covariates [A]
- Click on Allocation [B]
- Click on Edit reference group and select [Random] [C]
- Click [Add to main screen] [D]
Since we’ve set “Random” as the reference group, the Dummy variables are “Alternate” and “Systematic”. In Figure 131 [E] the program creates these and adds them to the variable list.

![Figure 131 | Creating dummy variables](image)

As always, tick the boxes [F] to include these variables in the current predictive model. Tick *either* of the two boxes, and the other will be ticked automatically. This is because the two represent allocation, and not because they belong to the same set.

The two dummy variables are Alternate and Systematic. Following the conventions proposed by Cohen, studies are coded “1” if they belong to the dummy-variable’s group name. Thus,

- A study is coded 1 for “Alternate” if it employed alternate allocation, or 0 otherwise.
- A study is coded 1 for “Systematic” if it employed systematic allocation, or 0 otherwise.

Therefore, the first three studies, shown as Figure 132 [G], should be coded as follows.

- Frimodt-Moller et al (Alternate) should be coded 1 for alternate and 0 for systematic
- TB Prevention trial (Random) should be coded 0 for alternate and 0 for systematic
- Comstock et al 1974 (Systematic) should be coded 0 for alternate and 1 for systematic
If you’d like to see the actual codes assigned by the program, proceed as follows.

- Click Run Regression
- Click More results > All studies [H]

Figure 133 [I] shows that the codes have been assigned as expected.

- Frimodt-Moller et al (Alternate) is coded 1 for alternate and 0 for systematic
- TB Prevention trial (Random) is coded 0 for alternate and 0 for systematic
- Comstock et al 1974 (Systematic) is coded 0 for alternate and 1 for systematic
The selection of a reference group has no impact on the model

To this point we’ve introduced the idea of a reference group, and shown that the selection of a reference group determines which dummy variables will be created.

Critically, while the selection of a reference group can be important for some aspects of the analysis (as discussed below), it has no impact on the statistics for the model.

To emphasize this point, we present three versions of the regression.

- In Figure 134 the reference group is “Random”, the dummy variables [A] are Alternate and Systematic
- In Figure 135 the reference group is “Systematic”, the dummy variables [A] are Alternate and Random
- In Figure 136 the reference group is “Alternate”, the dummy variables [A] are Random and Systematic
Main results for Model 1, Random effects (MM), Z-Distribution, Log risk ratio

<table>
<thead>
<tr>
<th>Set</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Intercept</td>
<td>-0.9905</td>
<td>0.3181</td>
<td>-1.6140</td>
<td>-0.3671</td>
<td>-3.1139</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>Allocation: Alternate</td>
<td>0.4780</td>
<td>0.6206</td>
<td>-0.7940</td>
<td>1.7099</td>
<td>0.7603</td>
<td>0.4470</td>
</tr>
<tr>
<td></td>
<td>Allocation: Systematic</td>
<td>0.5822</td>
<td>0.5222</td>
<td>-0.4413</td>
<td>1.6058</td>
<td>1.1149</td>
<td>0.2649</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding Intercept) are zero
Q = 1.4349, df = 2, p = 0.4880

Goodness of fit: Test that unexplained variance is zero
\( \tau^2 = 0.5396, \tau = 0.7480, I^2 = 92.45\% \), Q = 132.2675, df = 10, p = 0.0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
\( \tau^2 = 0.3086, \tau = 0.5557, I^2 = 92.12\% \), Q = 152.2330, df = 12, p = 0.0000

Proportion of total between-study variance explained by Model 1
\( R^2 \) analog = 0.00 (computed value is 0.81)

Number of studies in the analysis 13

Figure 134 | Dummy variables | Allocation with “Randomized” as the reference group
Figure 135 | Dummy variables | Allocation with “Systematic” as the reference group

### Main results for Model 1, Random effects (MM), Z-Distribution, Log risk ratio

<table>
<thead>
<tr>
<th>Set</th>
<th>Allocation</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>Intercept</td>
<td>-0.4083</td>
<td>0.4142</td>
<td>-1.2201</td>
<td>0.4035</td>
<td>0.2242</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Allocation: Alternate</td>
<td>-0.1042</td>
<td>0.6822</td>
<td>-1.4414</td>
<td>1.2329</td>
<td>-0.2528</td>
<td>0.8766</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Allocation: Random</td>
<td>-0.5822</td>
<td>0.5322</td>
<td>-1.0058</td>
<td>0.4413</td>
<td>-1.1149</td>
<td>0.2649</td>
</tr>
</tbody>
</table>

#### Statistics for Model 1

- **Test of the model:** Simultaneous test that all coefficients (excluding intercept) are zero
  \[ Q = 1.4349, \text{df} = 2, p = 0.4880 \]
- **Goodness of fit:** Test that unexplained variance is zero
  \[ \tau^2 = 0.0596, \tau = 0.7480, I^2 = 92.45\%, Q = 132.8676, \text{df} = 10, p = 0.0000 \]

#### Comparison of Model 1 with the null model

- **Total between-study variance (intercept only)**
  \[ \tau^2 = 0.3008, \tau = 0.5557, I^2 = 92.12\%, Q = 152.2330, \text{df} = 12, p = 0.0000 \]
- **Proportion of total between-study variance explained by Model 1**
  \[ R^2 \text{ analog} = 0.00 \text{ (computed value is -0.81)} \]

#### Number of studies in the analysis
13
### Main results for Model 1, Random effects (MM), Z-Distribution, Log risk ratio

<table>
<thead>
<tr>
<th>Set</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Intercept</td>
<td>-0.5125</td>
<td>0.5421</td>
<td>-1.5751</td>
<td>0.5500</td>
<td>-0.9454</td>
<td>0.3444</td>
</tr>
<tr>
<td></td>
<td>Allocation: Random</td>
<td>-0.4780</td>
<td>0.6286</td>
<td>-1.7099</td>
<td>0.7540</td>
<td>-0.7605</td>
<td>0.4470</td>
</tr>
<tr>
<td></td>
<td>Allocation: Systematic</td>
<td>0.1042</td>
<td>0.6822</td>
<td>-1.2329</td>
<td>1.4414</td>
<td>0.1328</td>
<td>0.8786</td>
</tr>
</tbody>
</table>

Statistics for Model 1

- Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
  - $Q = 1.4349, df = 2, p = 0.4880$
- Goodness of fit: Test that unexplained variance is zero
  - $\tau^2 = 0.5595, \tau = 0.7480, I^2 = 92.45\%, Q = 132.3675, df = 10, p = 0.0000$

Comparison of Model 1 with the null model

- Total between-study variance (intercept only)
  - $\tau^2 = 0.3088, \tau = 0.5557, I^2 = 92.12\%, Q = 152.2530, df = 12, p = 0.0000$
- Proportion of total between-study variance explained by Model 1
  - $R^2$ analog = 0.00 (computed value is -0.83)

Number of studies in the analysis: 13

Figure 136 | Dummy variables | Allocation with “Alternate” as the reference group
In all three versions of the analysis, statistics for the model are exactly the same

- The test of the model [B] yields a Q-value of 1.4349 with 2 df and \( p = 0.4880 \).
- The goodness of fit test [B] yields a Q-value of 132.3676 with 10 df and \( p < 0.0001 \).
- The estimates of \( T^2 \) and \( \tau \) are 0.5596 and 0.7480 [B], respectively.
- The estimate of \( I^2 \) with covariates in the model is 92.45% [B]
- The estimate of \( R^2 \) is 0.00% [C]

The reason that the statistics for the model are the same regardless of which group serves as the reference group (and which two dummy variables are included in the regression) is that all three versions incorporate precisely the same information. Concretely, once we know a study’s code on any two dummy variables, we know precisely which allocation method was employed for that study. It follows that (at least for purposes of the model) it doesn’t matter which two dummy variables we used.

Indeed, it must be this way. The set of dummy variables addresses the question “Is allocation related to effect size” and it must be true that we will get the same answer regardless of which mechanism we employ to represent allocation in the regression.

Working with the “Set”

When the program creates a series of dummy variables to represent a categorical covariate, it automatically defines dummy variables these as a “Set”. In Figure 134, Figure 135, and Figure 136 the program has added a column labeled “Set”.

- In this column we see the label “Allocation”, which refers to the categorical variable.
- Brackets indicate the two dummy variables that represent allocation.
- Each of these variables has a two-part name, with the first part reflecting the core variable (Allocation) and the second part indicating the dummy-variable (Alternate).

In our example (where there are three groups) the set includes two covariates [A].

- The line for “Alternate” (if it exists) addresses the impact of Alternate allocation vs other types of allocation.
- The line for “Systematic” (if it exists) addresses the impact of Systematic allocation vs other types of allocation.
- The line for “Random” (if it exists) addresses the impact of Randomized allocation vs other types of allocation.

Thus, each of these lines addresses the impact of a specific allocation type.

By contrast, the “Set” addresses the impact of Allocation in general. This is an omnibus test that asks if there are any differences in effect size among allocation types. These statistics are shown at the right-hand side of the display [D].
In our example Allocation (in the form of dummy variables) is the only covariate in the equation, and so the test of the set is identical to the test of the model.

- The test of the set \([D]\) yields \(Q\)-value of 1.4349 with 2 \(df\) and \(p=0.4880\).
- Similarly, the test of the model \([B]\) yields a \(Q\)-value of 1.4349 with 2 \(df\) and \(p=0.4880\).

Therefore, in this example we really didn’t need to present statistics for the set. We could have simply relied on the statistics for the model.

However, this equality only exists when the variables in the set are the only variables in the model. Typically, this is not the case. Rather, there will often be additional covariates in the model, and when that’s true, the test of the set is quite different from the test of the model.

For example, suppose the model includes Allocation (dummy coded) and also latitude. The Set would test the impact of allocation with latitude partialled. By contrast, the model would test the combined (and simultaneous) impact of allocation and latitude. These are two entirely different issues.

**How to select a reference group**

To this point we’ve shown that the selecting one group rather than another to serve as the reference group has no impact on the statistics for the full model nor the statistics for the set.

Nevertheless, the selection of a reference group does have implications what lines are displayed within the set, and how these lines are interpreted.

Consider the case where we use Random allocation as the reference group (Figure 134). In this case the dummy variables are “Alternate” and “Systematic”.

The predicted effect size for a study is going to be

\[
Y = B_0 + B_1(Alternate) + B_2(Systematic) \tag{1.31}
\]

Studies that employed Random allocation will have a code of 0 on both covariates, and so the predicted value is

\[
Y = B_0 + B_1(0) + B_2(0), \tag{1.32}
\]

or simply

\[
Y = B_0 \tag{1.33}
\]

In other words, for studies in the reference group, the predicted value is simply the intercept.

For a study in either of the other groups, the predicted value is the intercept plus the coefficient for that group. Thus, the coefficient for Alternate gives us the difference between the predicted effect size in that group and the predicted effect size in the reference group.
If we work with Figure 134, where the reference group is Random allocation,

- Random allocation – predicted effect size is the intercept (−0.9905), or −0.9905
- Alternate allocation – predicted effect size is 0.4780 units above the intercept, or −0.5125
- Systematic allocation – predicted effect size is 0.5822 units above the intercept, or −0.4083

If we work with Figure 135, where the reference group is systematic allocation

- Systematic allocation – predicted effect size is the intercept (−0.4083), or −0.4083
- Alternate allocation – predicted effect size is 0.1042 units below the intercept, or −0.5125
- Random allocation – predicted effect size is 0.5822 units below the intercept, or −0.9905

If we work with Figure 136, where the reference group is alternate allocation

- Alternate allocation – predicted effect size is the intercept (−0.5125), or −0.5125
- Random allocation – predicted effect size is 0.4780 units below the intercept, or −0.9905
- Systematic allocation – predicted effect size is 0.1042 units above the intercept, or −0.4083

Thus, the predicted value for each group is the same regardless of which group serves as the reference group. The difference is that each version presents a different set of comparisons.

Note.

The standard error for the reference group is the standard error of that group’s mean effect. By contrast, the standard error for the other groups is the standard error of the difference between that group and the reference group.

Similarly, the p-value for the reference group tests the null that the mean effect size in the reference group is zero. By contrast, the p-value for the other groups tests the null that the difference between that group’s mean effect and the reference group’s mean effect is zero.
Finally, to show the correspondence between this analysis and a subgroups analysis, we show a subgroups analysis (Figure 137) where we’ve grouped by allocation type.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Fixed effect analysis</th>
<th>Mixed effects analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number Studies</td>
<td>Point estimate</td>
</tr>
<tr>
<td>Alternate</td>
<td>2</td>
<td>0.2114</td>
</tr>
<tr>
<td>Random</td>
<td>7</td>
<td>-0.3827</td>
</tr>
<tr>
<td>Systematic</td>
<td>4</td>
<td>-4.3294</td>
</tr>
<tr>
<td>Total-between</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total-between</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>12</td>
<td>0.4280</td>
</tr>
</tbody>
</table>

Note. For the subgroups analysis, Computational options > Random and mixed-effect options must be set to pool estimates of $T^2$.

Creating dummy variables manually

Above, we showed how to create dummy variables automatically. You also have the option of creating dummy variables manually. On the data-entry sheet create a column for a moderator and then (critically) define the moderator type as integer or decimal. Then, enter a value for each study.

While the option to create dummy variables automatically works well in most cases, there are several cases where you’ll need to use the manual option.

Interactions

- The mean effect size for each subgroup [A] is the same as the numbers from the regression
- The Q-value for the model [B] is the same as the numbers from the regression

In this example, Allocation (or rather the dummy variables that represent allocation) is the only covariate in the regression. Therefore, the intercept is simply the mean effect for the reference group and the coefficients represent the difference in mean effects. If the regression model included other covariates then all the statistics would be adjusted for the other covariates.
Suppose you have a categorical variable that is represented by a dummy variable A, and you want to assess the impact of that variable and also its interaction with another variable B. You’ll need to work with the variables A, B, and AB. In this case it will be easier to create A (and then AB) manually.

Alternate coding schemes

Dummy-coding is only one of the options possible for categorical covariates. Texts on multiple regression discuss other options, such as effects-coding and contrast-coding. You can use any of these coding schemes, but you’ll need to create the dummy variables manually.

Regressions with no intercept

The automatic coding scheme is only available when you include the intercept in the equation. When you omit the intercept (see next chapter) the coding scheme changes (in that case, for \( m \) groups you need \( m \) rather than \( m - 1 \) dummy variables) and the automatic function is not available.

If you create the dummy variables manually, you’ll also need to define these as a set manually. This is discussed in Part 13: Working with “Sets” of covariates.
PART 12: WHEN DOES IT MAKE SENSE TO OMIT THE INTERCEPT

In any primary regression or meta-regression we have the option to either include or exclude the intercept from the prediction equation. The decision to omit the intercept usually arises when we are working with a categorical variable, and we will discuss the issue in this context. This decision fundamentally affects the issues addressed by the analysis, as follows.

When we include the intercept in a regression with categorical covariates

- Coefficients reflect differences in effect size across categories.
- Tests of a covariate address the question “Is the covariate related to effect size?”
- The model tests the null hypothesis that no covariate is related to effect size.

When we omit the intercept

- Coefficients reflect the absolute effect sizes within categories.
- Tests of a covariate address the question “Is the effect size zero?” in this category.
- The model tests the null hypothesis that all groups have a mean of zero.

In all chapters up to this point we have assumed that we are including the intercept, which is typically the case. However, there are cases when we may want to omit the intercept.

- First, we will address a technical issue about coding categorical variables. As explained below, the decision to include or omit the intercept impacts the way that we code categorical variables for the analysis.
- Second, we will show how to interpret an analysis where the intercept is omitted, and how this differs from an analysis where the intercept is included.

As always, we caution the reader that this chapter is not a comprehensive treatment of the topic. We assume that the reader is familiar with these issues from multiple regression in primary studies. Our goal here is to review the key concepts, and show how they can be applied in meta-analysis.

The example

The initial data set includes a covariate called latitude. For purposes of this example we need a categorical covariate, so we will create a new variable called Climate, which is coded Hot or Cold (latitude 33 and under, vs. 34 and higher).

Since Climate is categorical, it cannot be inserted directly into the analysis. Rather, we need to create a numerical covariate corresponding to climate and use this in the analysis. We may use any of several schemes for this purpose (dummy coding, effects coding, contrast coding) but what all of these schemes have in common in the usual approach (including an intercept) is that we need \( m - 1 \) covariates to represent a covariate with \( m \) groups. In the current example there are two groups (Cold and Hot) so we need one covariate.
If we elect to use dummy coding, we can create a covariate called Hot and code it 0 for Cold studies and 1 for Hot studies. Alternatively we can create a covariate called Cold and code it 1 for Cold studies and 0 for Hot studies. We could use either of these covariates in the analysis but we could not use both since one of them will be redundant.

This rule, that we need \( m - 1 \) covariates for a categorical variable with \( m \) groups, only applies when we include the intercept in the regression. By contrast, when we omit the intercept we actually need \( m \) covariates. In the current example we would include two covariates (Hot and Cold) in the analysis.

CMA includes a mechanism to create dummy variables automatically, but this mechanism (since it creates \( m - 1 \) covariates) is intended only for cases where we include the intercept. For simplicity in this chapter we will create all dummy variables manually, whether or not we include the intercept in any given analysis.

Figure 138 shows the two dummy variables

- “Hot” is coded 1 if the study was located in a hot climate and 0 otherwise [A]
- “Cold” is coded 1 if the study was located in a cold climate, and 0 otherwise [B]

Thus, a “1” indicates the presence of the attribute (Hot or Cold) while a “0” indicates the absence of that attribute.

Figure 138 | Data-entry | Dummy variables for Hot and Cold

If we include the intercept we will include either Hot or Cold in the prediction equation. If we omit the intercept we will include both Hot and Cold in the prediction equation.

Note. This becomes more complicated when we have two or more categorical variables, such as Climate and Allocation. This is beyond the scope of this manual.
Interpreting the results

A meta-regression without the intercept will give us the mean effect for each group (Hot and Cold). Before proceeding to the regression, we can use a subgroups analysis to see what these means actually are.

On the main analysis screen click Computational options > Mixed and random effects options and then select the option to “pool within-group estimates of tau-squared” as shown in Figure 139. Then click Computational options > Group by and group by Climate. Select the “Random” tab at the bottom of the screen.

![Figure 139 | Basic analysis | Computing $\tau^2$ in the presence of subgroups](image1)

![Figure 140 | Basic analysis | Subgroups Cold vs. Hot](image2)
Figure 140 and Figure 141 show the results of this analysis:

- For the Cold subgroup \([C]\) the mean effect size is \(-1.1987\) with \(SE = 0.1769\). The test addresses the question “Is this effect size zero?” and yields \(Z = -6.7740, p < 0.0001\).

- For the Hot subgroup \([D]\) the mean effect size is \(-0.2784\) with \(SE = 0.1522\). The test addresses the question “Is this effect size zero?” and yields \(Z = -1.8289, p < 0.0674\).

- The test of the between-subgroups variance \([E]\) addresses the question “Does the mean effect size differ by subgroup?” and yields \(Q = 15.5445, df = 1, p = 0.0001\).

To compute the mean for each subgroup in meta-regression we would omit the intercept and include both Cold and Hot as covariates (note that the box for intercept is unchecked).
Because we have omitted the intercept, the statistics reflect the mean effect size for each group.

- For the Cold subgroup [CC] the mean effect size is $-1.1987$ with $SE = 0.1769$. The test addresses the question “Is this effect size zero?” and yields $Z = -6.7740$, $p < 0.0001$.

- For the Hot subgroup [DD] the mean effect size is $-0.2784$ with $SE = 0.1522$. The test addresses the question “Is this effect size zero?” and yields $Z = -1.8289$, $p < 0.0674$.

Note that these numbers are exactly the same as the numbers we saw in the subgroups analysis, Figure 141). Line CC corresponds to line C, and line DD corresponds to line DD.

Finally, in Figure 143, since there is no intercept, the test of the model addresses the null hypothesis that the mean for all groups is zero. The $Q$-value of $49.2323$ with 2 degrees of freedom yields a $p$-value
of < 0.0001. We reject the null and conclude that the mean effect size is probably not zero in at least one of the groups. (The subgroups analysis does not include this test).
PART 13: WORKING WITH “SETS” OF COVARIATES

DEFINING A “SET”

In regression there are times when we use several covariates to capture a concept. For example

- If we have a categorical covariate with \( m \) values, we use \( m - 1 \) covariates to represent this variable in the analysis (as discussed in the preceding chapter).
- If we want to assess the relationship between duration of treatment and effect we might include \( duration, duration^2, \) and \( duration^3 \) as predictors.
- We may have a series of covariates, such as \( income \) and \( education \) that (together) represent the impact of socio-economic status.
- We may have a series of covariates such as \( dose \) and \( duration \) that (together) represent the intensity of a treatment.
- We may have two covariates and also the interaction between, where the three together represent their influence on outcome.

When we define covariates as a Set, the program reports a test of significance for the Set with all other covariates held constant. For example, consider the analysis displayed in Figure 144. The two covariates dummy variables that (as a set) capture the Allocation method.

**Note that in**

Figure 137 there is now a column labeled “Set” [C]. Under this column there is a set called “Allocation” and the program has inserted brackets to show which covariates are included in the set.
The statistics for Allocation: Alternate tells us if the use of alternate allocation is related to effect size (when all other covariates are partialled).

The statistics for Allocation: Systematic tells us if the use of systematic allocation is related to effect size (when all other covariates are partialled).

The statistics for the set \([A]\) tell us if Allocation as a whole (that is, the use of Random, Systematic, or Alternate allocation) is related to effect size.

When the covariates in the set are the only covariates in the prediction equation (as they are in this example) the statistics for the set will be identical to the statistics for the model. Specifically, the \(Q\)-value for the set \([A]\) and the \(Q\)-value for the model \([B]\) are both 1.4349. Therefore, in this example (Figure 144) we could have simply employed the test of the model as the test of Allocation.

However, that is not the case when the model includes additional covariates. Consider Figure 145, where the model includes Latitude and Year as well as the two dummy-variables for Allocation. In this case, if we want to know the impact of Allocation with Latitude and Year partialed, we need to use the statistics for the Allocation set, \(Q = 1.5492, df = 2, p = 0.4609 \) \([A]\). We could not use the statistics for the model \([B]\) to test the impact of allocation, since the test of the model is based on the impact of all four covariates.
## Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Set</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>Z-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>-29.2359</td>
<td>45.7992</td>
<td>-119.0007</td>
<td>60.5289</td>
<td>-0.6383</td>
<td>0.5232</td>
</tr>
<tr>
<td>Latitude</td>
<td></td>
<td>-0.0190</td>
<td>0.0139</td>
<td>0.0050</td>
<td>0.0322</td>
<td>-1.9248</td>
<td>0.2331</td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td>0.0148</td>
<td>0.0232</td>
<td>0.0030</td>
<td>0.0603</td>
<td>0.6354</td>
<td>0.5225</td>
</tr>
<tr>
<td>Allocation</td>
<td>Alternate</td>
<td>0.4855</td>
<td>0.4809</td>
<td>0.4570</td>
<td>1.4279</td>
<td>1.0968</td>
<td>0.3127</td>
</tr>
<tr>
<td></td>
<td>Systematic</td>
<td>0.4574</td>
<td>0.3778</td>
<td>0.2831</td>
<td>1.1578</td>
<td>1.2106</td>
<td>0.2260</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 1.5492, df = 2, p = 0.4609

Goodness of fit: Test that unexplained variance is zero
\( \tau^2 = 0.1194, \tau = 0.3456, I^2 = 66.69\% \), Q = 24.0144, df = 8, p = 0.0023

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
\( \tau^2 = 0.3085, \tau = 0.5557, I^2 = 92.12\% \), Q = 152.2330, df = 12, p = 0.0000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0.61

Number of studies in the analysis 13

---

Figure 145 | Main results | Assessing the impact of a set

---

213
HOW TO CREATE A SET

When we use a Set to represent a categorical variable, CMA creates the dummy variables automatically and links them (with the bracket) automatically.

In all other cases we need to link the covariates manually. For this example we’ll work with Latitude-C and Latitude-C^2 (latitude centered, and its square). We want to create a set that incorporates these two covariates and call the set “Latitude Set”.

In Figure 146

- Move Latitude-C into the model
- Move Latitude-C^2 into the model
- Ensure that the covariates intended for the set are sequential in the list
- Highlight these covariates by pressing (SHIFT) and clicking on the covariate names [B]

Figure 146 | Setup | Defining a set of covariates
In Figure 147

- Click [Link Covariates] [C]
- Enter the name Latitude Set and click [Ok] [D]

The program has now created a set called “Latitude-C Set” [E] which includes the two covariates (Figure 148). When you run the regression the program will display statistics for this set (Figure 149).
HOW TO REMOVE A SET

- Highlight the set’s name [F]
- Click Unlink Covariates [G]
PART 14: INTERACTIONS AND CURVILINEAR RELATIONSHIPS

Suppose we run a regression with two covariates, $X_1$ and $X_2$. Consider what happens in two cases —

- When there is not an interaction
- When there is an interaction

There is no interaction if the impact of $X_1$ on the effect size is constant for all values of $X_2$ (and vice versa). There is an interaction between two variables if the impact of one variable depends in the magnitude of the second variable (and vice versa).

**When there is not an interaction**

When there is no interaction we include two covariates in the model – $X_1$ and $X_2$.

a) $B_1$ gives us the main effect for variable $X_1$, for any value of $X_2$

b) $B_2$ gives us the main effect for variable $X_2$, for any value of $X_1$

When we run a regression with covariates $X_1$ and $X_2$ (but not the interaction), the following is true.

- We assume that the impact of $X_1$ is the same for all values of $X_2$. The $p$-value for $X_1$ is a test of this constant effect.
- We assume that the impact of $X_2$ is the same for all values of $X_1$. The $p$-value for $X_2$ is a test of this constant effect.

**When there is an interaction**

By contrast, consider what happens when there is an interaction of $X_1$ and $X_2$. In this case we create a new variable $X_3$ (defined as $X_1 \times X_2$) and enter all three variables into the prediction equation.

a) $B_1$ gives us the first-order effect for variable $X_1$ when $X_2$ is zero

b) $B_2$ gives us the first-order effect for variable $X_2$ when $X_1$ is zero

c) $B_3$ gives us the impact of the interaction (over and above the first-order effects)

When we run a regression with covariates $X_1$ and $X_2$ and the interaction $X_3$, the following is true:

- We assume that the impact of $X_1$ depends on the value of $X_2$, and so we assess the impact of $X_1$ at a specific value of $X_2$, zero. The $p$-value is a test of the relationship between $X_1$ and the effect size at this specific value of $X_2$.

- We assume that the impact of $X_2$ depends on the value of $X_1$, and so we assess the impact of $X_2$ at a specific value of $X_1$, zero. The $p$-value is a test of the relationship between $X_2$ and the effect size at this specific value of $X_1$. 

217
Centering

To center a variable means to re-scale the variable to have a mean of zero. If the original studies took place in the years 1930, 1935, 1940, 1945, 1950, we could subtract 1940 from each value to yield scores of -10, -5, 0, 5, and 10. If the original covariate is Year, the new one could be called Year-C.

When we include \( X_1 \) and \( X_2 \) (but not the interaction, \( X_3 \)) in the equation, the decision to center (or not) has no impact on the \( p \)-value for the individual covariates.

By contrast, if we include \( X_1, X_2, \) and the interaction \( X_3 \) in the equation, then the decision to center (or not) will have a substantial impact on the \( p \)-values for the individual covariates. This is because we test \( X_1 \) for the case where \( X_2 \) is zero (and vice-versa).

For example, suppose that \( X_1 \) and \( X_2 \) are Latitude and Year.

- Consider the statistics for Year. If we don’t center Latitude, then we test the impact of Year when Latitude is 0. If we do center, then we test the impact of year when Latitude-C is 0, and (it follows) Latitude is around 33.

- Consider the statistics for Latitude. If we don’t center Year, then we test the impact of Latitude when Year is 0. If we do center, then we test the impact of Latitude when Year-C is 0 and the actual year is 1948.

Centering is also important if we want to assess the impact of curvilinear relationships. For example, suppose that we want to see if the relationship between Latitude and effect size has a curvilinear component. For this purpose we need to enter both Latitude and Latitude\(^2\) as covariates.

If we don’t center Latitude, then these two covariates will be highly correlated with each other, and it will be difficult to disentangle the linear from the curvilinear components. By contrast, if we center latitude and square the centered value, the correlation between the two will be low, and we will be able to identify the unique impact of each.

For these reasons, in the examples that include an interaction (or a curvilinear term) we use variables have been centered about their mean (for continuous variables) or dummy coded (as explained below) for categorical variables, so that zero is a meaningful value or category.

The program will not automatically create centered variables, nor variables for the interaction. While it is possible to enter these variables manually, it’s usually easier to copy the original variables to Excel™, create the new variables, and then copy these back into CMA.

Important note.

As always, we assume that the reader who plans to work with interactions has a good understanding of these from primary regression, and focus here on the elements that are specific to meta-analysis and to the use of this program. Similarly, the very brief overview of centering does not fully address the implications of scaling or centering, or other issues which may affect the results. The same rules apply for interactions involving categorical variables, continuous variables, or combinations of the two types. For clarity, we present an example for each of three cases. These are
• The interaction of two categorical covariates
• The interaction of a categorical covariate with a continuous covariate
• The interaction of two continuous covariates
INTERACTION OF TWO CATEGORICAL COVARIATES

The original data set includes the covariates Latitude and Year, both of which are continuous. For purposes of this discussion we need two categorical covariates, and we create them by dichotomizing Latitude and Year (see Appendix 5: Creating variables for interactions).

- Hot is coded 1 if the latitude is 34 or less, and is coded 0 if the latitude is exceeds 34.
- Recent is coded 1 if the Year is 1945 or later, and is coded 0 if the Year is earlier than 1945.
- Hot x Recent is created by multiplying Hot x Recent.

Figure 151 shows the model, Figure 152 shows the main results, which are plotted in Figure 153 and Figure 154. While the program can compute the statistics for interactions it cannot plot these interactions, and therefore these plots were created in Excel™ (see appendix).
### Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.1154</td>
<td>0.2132</td>
<td>-1.5334</td>
<td>-0.6975</td>
<td>-5.2307</td>
<td>0.0000</td>
</tr>
<tr>
<td>Hot</td>
<td>0.8990</td>
<td>0.3814</td>
<td>0.1515</td>
<td>1.6465</td>
<td>2.3573</td>
<td>0.0184</td>
</tr>
<tr>
<td>Recent</td>
<td>-0.3261</td>
<td>0.4142</td>
<td>-1.1380</td>
<td>0.4858</td>
<td>-0.7873</td>
<td>0.4311</td>
</tr>
<tr>
<td>Hot x Recent</td>
<td>0.2391</td>
<td>0.5519</td>
<td>-0.8426</td>
<td>1.3208</td>
<td>0.4332</td>
<td>0.6649</td>
</tr>
</tbody>
</table>

**Statistics for Model 1**

- **Test of the model**: Simultaneous test that all coefficients (excluding intercept) are zero
  - \( Q = 15.2062, \text{df} = 3, p = 0.0016 \)
- **Goodness of fit**: Test that unexplained variance is zero
  - \( Tau^2 = 0.1061, Tau = 0.3257, I^2 = 63.23\%, Q = 24.4785, \text{df} = 9, p = 0.0036 \)

**Comparison of Model 1 with the null model**

- **Total between-study variance (Intercept only)**
  - \( Tau^2 = 0.3088, Tau = 0.5557, I^2 = 92.12\%, Q = 152.2330, \text{df} = 12, p = 0.0000 \)
- **Proportion of total between-study variance explained by Model 1**
  - \( R^2 \text{ analog} = 0.66 \)

**Number of studies in the analysis** 13

---

Figure 152 | Main results | Interaction of two categorical covariates
We can display these results as shown in Table 6.

<table>
<thead>
<tr>
<th>Climate</th>
<th>Time</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early</td>
<td>Recent</td>
<td></td>
</tr>
<tr>
<td>Cold</td>
<td>−1.1154</td>
<td>−1.4416</td>
<td></td>
</tr>
<tr>
<td>Hot</td>
<td>−0.2164</td>
<td>−0.3035</td>
<td></td>
</tr>
</tbody>
</table>

Instructions for creating this table are given in the appendix, but the computation is quite intuitive. Since the covariate values for all cells are a combination of 0s and 1s, the expected effect size in each cell is given by the sum of the relevant coefficients from Figure 152.

- Upper-left cell is the Intercept (−1.1154) = −1.1154
- Upper-right cell is the Intercept plus Recent (−1.1154 −0.3261) = −1.4416
- Lower-left cell is the Intercept plus Hot. (−1.1154 + 0.8990) = −0.2164
- Lower-right cell is the Intercept plus Recent, Hot, and the Interaction. (−1.1154 −0.3261 −0.8990 + 0.2391) = −0.3035

Once we have this table, it’s a simple matter to plot the main effects and their interactions as shown in the following plots.
Does the vaccine’s effect differ as a function of climate?

Since the equation includes the interaction term, the impact of climate is not a main effect. Rather, it is a first-order effect that may differ for Early studies vs. Recent studies.

In Figure 153,

The impact of climate for Early studies (Recent = 0) is indicated by the arrow labeled [B]. The impact of climate for Recent studies (Recent = 1) is indicated by the arrow labeled [BB].

The impact of climate is tested for the Early studies [B] since these are the studies coded 0 for Recent. The $p$-value for this difference is 0.0184 as shown in Figure 152 [B].
Recent (vs. Early)

Did the impact of the vaccine change from the Early studies to the Recent studies?

Since the equation includes the interaction term, the impact of Recent may differ for Cold studies vs. Hot studies. In Figure 154,

The impact of Recent for Cold studies (Hot = 0) is indicated by the arrow labeled [B].
The impact of Recent for Hot studies (Hot = 1) is indicated by the arrow labeled [BB].

The impact of Recent is tested for the Cold studies [B] since these are the studies coded 0 for the covariate (Hot). For these studies the regression line increases slightly as we move from Early to Recent studies. However, the corresponding $p$-value for Recent is 0.4311 as shown in Figure 152 [C]. Thus, there is no evidence that effect size is related to Recent.
Does the relationship between Time and effect size vary by Climate? In Figure 155 the regression line for Time in Hot climates is not strictly parallel to the regression line for Time in Cold climates. However, the differences in slopes are minor.

Does the relationship between Climate and effect size vary by Time? The impact of Climate in the Recent studies is only slightly larger than the impact of Climate in the Early studies (as indicated by the difference in the height of the two arrows in Figure 146).

These two questions are functionally identical, and the same p-value applies to both. Error! Reference source not found. [C] shows the p-value for the interaction is 0.6648.

![Effect size as a function of time and climate](image)

**Figure 155 | Plot | Interaction of two categorical covariates**
The full set

Is there a relationship between Time, Climate, and the interaction (as a set) and the effect size? Since these three are the only covariates in the model, this is addressed by a test of the full model, as shown in Figure 152 [E].

- The $p$-value of 0.0016 allows us to reject the null hypothesis that none of the covariates is related to effect size.

- The $R^2$ analog [F] is 0.66, which tells us that 66% of the initial between-study variance in effect sizes can be explained by this combination of covariates.

- Thus, we conclude that the full model (Time, Climate, and the interaction between them) is able to explain at least some of the variance in effect size.
INTERACTION OF A CATEGORICAL COVARIATE WITH A CONTINUOUS COVARIATE

The original data set includes the covariates Latitude and Year, both of which are continuous. For the purpose of this example we created the following variables based on Latitude and Year (see Appendix 5: Creating variables for interactions).

- Year-C is the study year, centered
- Hot is coded 0 for studies in Cold climates, and 1 for studies in Hot climate
- Year-C x Hot is the interaction

Figure 156 shows the model, Figure 157 shows the main results. While the program can compute the statistics for interactions it cannot plot these interactions, and therefore the plots (Figure 159, Figure 158, and Figure 160) were created in Excel™ (see appendix).

![Figure 156: Setup | Interaction of categorical and continuous covariates](image-url)
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>Z-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.4085</td>
<td>0.2835</td>
<td>-1.9642</td>
<td>-0.8527</td>
<td>-4.9673</td>
<td>0.0000</td>
</tr>
<tr>
<td>YearC</td>
<td>-0.0260</td>
<td>0.0274</td>
<td>-0.0798</td>
<td>0.0278</td>
<td>-0.9482</td>
<td>0.3430</td>
</tr>
<tr>
<td>Hot</td>
<td>1.2243</td>
<td>0.3611</td>
<td>0.5166</td>
<td>1.9321</td>
<td>3.3908</td>
<td>0.0007</td>
</tr>
<tr>
<td>Hot x Year-C</td>
<td>0.0156</td>
<td>0.0327</td>
<td>-0.0486</td>
<td>0.0797</td>
<td>0.4757</td>
<td>0.6343</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 15.5389, df = 3, p = 0.0009

Goodness of fit: Test that unexplained variance is zero
\( \tau^2 = 0.0986, \tau = 0.3140, \, I^2 = 58.76\%, \, Q = 21.8218, \, df = 9, \, p = 0.0095 \)

Comparison of Model 1 with the null model

Total between-study variance (Intercept only)
\( \tau^2 = 0.3088, \tau = 0.5557, \, I^2 = 92.12\%, \, Q = 152.2330, \, df = 12, \, p = 0.0000 \)

Proportion of total between-study variance explained by Model 1
\( R^2 \) analog = 0.68

Number of studies in the analysis 13

Figure 157 | Main results | Interaction of categorical and continuous covariates
In two-way analysis of variance, we sometimes display the results as a 2x2 table, and we can do the same here, as shown in Table 7.

<table>
<thead>
<tr>
<th>Climate</th>
<th>Year-C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-15</td>
</tr>
<tr>
<td>Cold</td>
<td>-1.0184</td>
</tr>
<tr>
<td>Hot</td>
<td>-0.0276</td>
</tr>
</tbody>
</table>

Instructions for creating this table are given in the appendix.

Once we have this table, it’s a simple matter to plot the main effects and their interactions as shown in the following plots.
Hot

Does the vaccine’s effect differ as a function of climate? Since the equation includes the interaction term, the impact of climate is not a main effect, but rather varies depending on the time frame. Here, it will be evaluated for studies where Year-C = 0 (and the actual year is 1948) as indicated by the arrow in Figure 158 [C].

When Year-C is zero (and the actual year is 1948) the effect size for the Cold studies (at the bottom of the arrow) is substantially larger than the effect size for the Hot studies (at the top of the arrow). The p-value for this difference is 0.0007 as shown in [Figure 157 C].
Year-C

Did the impact of the vaccine change over time? Since the equation includes the interaction term, the impact of Year is not a main effect. Rather, it is a first-order effect that may vary with climate.

In Figure 159,

The impact of Year for Cold studies (Hot=0) is indicated by the arrow labeled [B].
The impact of Year for Hot studies (Hot=1) is indicated by the arrow labeled [BB].

The impact of Year is tested for the Cold studies [B] since these are the studies coded 0 for Hot.

The effect size increases (moves away from zero) as we move from 1930 to 1970, but the corresponding \( p \)-value for Year-C is 0.3430 as shown in [Figure 157 B]. Thus, there is no evidence that effect size is related to year.
Does the relationship between Year and effect size vary by Climate? In Figure 158 the regression line for Year in Hot climates is not strictly parallel to the regression line for Year in Cold climates. However, the differences in slopes are minor.

Does the relationship between Climate and effect size vary by Year? The impact of Climate in 1968 is only slightly larger than the impact of Climate in 1933 (as indicated by the difference in the height of the two arrows in Figure 158).

These two questions are functionally identical, and the same p-value applies to both. Figure 157 [D] shows the p-value for the interaction is 0.6343.
The full set

Is there a relationship between Year-C, Hot, and the interaction (as a set) and the effect size? Since these are the only covariates in the model, this is addressed by a test of the full model [Figure 157 E].

- The $p$-value of 0.0008 allows us to reject the null hypothesis that none of the covariates is related to effect size.

- The $R^2$ analog [Figure 157 F] is 0.68, which tells us that 68% of the initial between-study variance in effect sizes can be explained by this combination of covariates.

- Thus, we conclude that the full model (Time, Climate, and the interaction between them) is able to explain at least some of the variance in effect size.
INTERACTION OF TWO CONTINUOUS COVARIATES

In this example we assess the impact of Year-C, Latitude-C, and Year-C x Latitude-C. For instructions on creating the data set used here, see Appendix 5: Creating variables for interactions

- Latitude-C is the latitude, centered
- Year-C is the study year, centered
- Latitude-C x Year-C is the interaction

Figure 161 shows the model, Figure 162 shows the main results, and Error! Reference source not found. shows a plot of these results. While the program can compute the statistics for interactions it cannot plot these interactions, and therefore Error! Reference source not found. was created in Excel™ (see Plotting the interaction of two continuous covariates).

![Figure 161](image_url)
Working with the screen shown in Figure 162 we can create Table 8. Then we use the numbers in this table to create the subsequent plots in Excel™. Details are provided in the appendix.

Table 8

<table>
<thead>
<tr>
<th>Latitude-C</th>
<th>Year-C</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-15</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>-20</td>
<td>0.1319</td>
<td>-0.2796</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>-0.6171</td>
<td>-0.7835</td>
<td></td>
</tr>
<tr>
<td>+21</td>
<td>-1.4035</td>
<td>-1.313</td>
<td></td>
</tr>
</tbody>
</table>
Latitude-C

Does the vaccine’s effect differ as a function of latitude? Since the equation includes the interaction of Latitude and Year, the impact of latitude is not a main effect, but rather varies depending on the Year.

The impact of Latitude on the effect size will be evaluated for studies where Year-C = 0 (and the actual year is 1948) as indicated by the arrow [B] in Figure 163.

When Year-C is zero (and the actual year is 1948) the effect size for the studies where Latitude is 55 (at the bottom of the arrow) is substantially larger than the effect size for the studies where Latitude=13 (at the top of the arrow).

At this Year there is a statistically significant relationship between latitude and effect size, with \( p = 0.0066 \) as shown in Figure 162 [B].

![Figure 163 | Plot | Interaction of two continuous covariates](image-url)
Did the impact of the vaccine change over time? Since the equation includes the interaction of Year by Latitude, the impact of Year is not a main effect. Rather, it is a first-order effect that may vary with Latitude.

In Figure 164,

Since the equation includes the interaction term, the impact of Year-C will be evaluated at the point where Latitude-C = 0 (and the actual latitude is 33). This is represented by Line [C] in Figure 164.

At this latitude the regression line seems to be relatively horizontal. The corresponding $p$-value for Year—C is 0.7594 as shown in Figure 162 [C]. Thus, there is no evidence that effect size is related to year at this latitude.
Hot x Year-C

Does the relationship between Year and effect size vary by latitude? While the lines for the different latitudes in Figure 165 are not strictly parallel, the differences in slopes are minor and not statistically significant.

Does the relationship between latitude and effect size vary by Year? The impact of latitude in 1968 is only slightly larger than the impact of latitude in 1933 (as indicated by the difference in the height of the two arrows in Figure 165).

These two questions are functionally identical, and the same p-value applies to both. Figure 162 [D] shows the p-value for the interaction is 0.7181.
The full set

Is there a relationship between Year-C, Latitude-C, and the interaction (as a set) and the effect size? Since these are the only covariates in the model, this is addressed by a test of the full model, as shown in Figure 162 [E].

- The $p$-value of 0.0097 allows us to reject the null hypothesis that none of the covariates is related to effect size.

- The $R^2$ analog [F] is 0.60, which tells us that 60% of the initial between-study variance in effect sizes can be explained by this combination of covariates.

- Thus, we conclude that the full model (Year, Latitude, and the interaction between them) is able to explain at least some of the variance in effect size.
CURVILINEAR RELATIONSHIPS

Earlier, we established that there is a linear relationship between latitude and effect size. Suppose we want to test the hypothesis that the relationship between latitude and effect size is actually curvilinear – for example, that the vaccine’s impact is relatively constant as we move from a latitude of 13 to 30, but then increases as we move from 30 to 55. Or, that the vaccine’s impact increases sharply as we move from a latitude of 13 to 30, but is relatively unchanged beyond that point.

A curvilinear relationship can be seen as a kind of interaction, and that is the approach we take here. In any interaction we ask of the impact of one covariate depends on the level of another covariate. Typically, the two covariates are distinct (A and B). Here, they are the same (A and A) but the idea is the same. We are asking if the impact of A depends on the level of A.

When working with curvilinear (or higher-order) relationships it’s generally a good idea to center variables, and that’s the practice we follow here. To assess the hypothesis that there is a curvilinear relationship between latitude and effect size we’ll need two covariates

- Latitude-C is simply Latitude centered, to have a mean of zero.
- Latitude-C² is the square of Latitude-C.

For information on how to create these variables see Appendix 5: Creating variables for interactions.

Figure 166 shows the model, Figure 167 shows the main results, and Figure 168 shows a plot of these results. While the program can compute the statistics for curvilinear relationships it cannot plot these relationships, and therefore Figure 168 was created in Excel™ (see appendix).
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5501</td>
<td>0.1862</td>
<td>-0.9151</td>
<td>-0.1850</td>
<td>-2.3535</td>
<td>0.0091</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0316</td>
<td>0.0075</td>
<td>-0.0452</td>
<td>-0.0199</td>
<td>-4.2294</td>
<td>0.0000</td>
</tr>
<tr>
<td>Latitude-C2</td>
<td>-0.0008</td>
<td>0.0007</td>
<td>-0.0022</td>
<td>0.0006</td>
<td>-1.1264</td>
<td>0.2600</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 17.8997, df = 2, p = 0.0001

Goodness of fit: Test that unexplained variance is zero
Tau² = 0.0747, Tau = 0.2733, I² = 64.91%, Q = 28.4961, df = 10, p = 0.0015

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau² = 0.3086, Tau = 0.5557, I² = 92.12%, Q = 152.2330, df = 12, p = 0.0000

Proportion of total between-study variance explained by Model 1
R² analog = 0.76

Number of studies in the analysis  13

Fixed | Random |
Model 1 |
In this example, where the effects are in log units and range downward from zero, an effect size of zero reflects no effect while an effect size of −1.5 reflects a substantial effect. In this example, as the latitude increases the effect size increases (moves away from zero).

We want to evaluate the linear component of this relationship, the curvilinear component, and then the two (as a set).

Latitude-C

The line for Latitude-C in Figure 167 [A] addresses the linear relationship between latitude and effect size. The p-value is < 0.0001, which tells us that the linear relationship is statistically significant.

Note. Since we have included an interaction, the linear component varies as a function of latitude. Therefore, the coefficient for Latitude-C is not a slope but rather the tangent to the curved line where Latitude-C is zero.

Latitude-C²

The line for Latitude-C² in Figure 167 [B] addresses the curvilinear component of the relationship between latitude and effect size. The line in the plot is curvilinear, with a slope that is initially shallow but increases as the altitude increases. Is this line a better fit for the data than a straight line would be? This is addressed by the p-value for this covariate, which is 0.2600. There is no evidence that the relationship is curvilinear.

Test of the model

Since the only covariates in the model are Latitude-C and Latitude-C², the model [C] tests the null hypothesis that both coefficients are zero. In Figure 167 [C] the Q-value is 17.90, and with 2 degrees of
freedom the $p$-value is 0.0001. We can conclude that there is a relationship between these covariates (as a set) and the effect size.
PART 15: MISSING DATA

In regression for meta-analysis, as in regression for primary studies, there are many options for dealing with missing data. The program takes a very simple approach to missing data, as follows. If a study is missing data for the outcome or for any of the covariates in the covariate list that study is excluded from the analysis.

Note that this exclusion is based on all covariates listed on the main screen, and not only on the covariates that are checked.

Figure 169 | Setup
To show how the program handles missing data we need to create some missing data. For this purpose,

- Return to the data-entry screen (Figure 170)
- Highlight the three cells in the latitude column for studies 9, 10, 11 [A]
- Press the Delete key

Figure 170 | Data-entry | Missing data for latitude
Run the analysis.

The main analysis screen (Figure 171) shows that the three studies are missing latitude [B]. These studies are still included in this analysis of the mean effect.

- Proceed to the regression module
- Create a model that includes latitude, and tick latitude, as in Figure 172 [C]
The results are shown in Figure 173. Note that the analysis is based on 10 studies [D] rather than 13, since three studies have been excluded.
• To see which studies have been excluded click More results > All data
• The program displays a line for every study in the database, and missing data points are highlighted in red (Figure 174).

Figure 174 | Table of missing data

This is a good way to identify the missing data and also to identify patterns of missing data.

• If data is missing primarily for one covariate across a lot of studies you may decide to remove that covariate from the analysis.
• If data is missing primarily for a few studies across many covariates you may decide to remove those studies and keep the covariates.

Of course, the decision to adopt one of these approaches or some other will depend on a host of factors, with attention paid to avoiding bias. However, the ability to identify the patterns of missing data is a crucial first step in this process.

In some cases, you may want to use another approach for missing data. For example, you may want to replace missing data with the mean, or a value imputed in some other way. You can do this by returning to the main data-entry screen and simply entering the desired value in place of the missing value.

In a more sophisticated version of this scheme you can create several variables based on the same variable, but using different approached to missing data. For example, suppose the initial variable is dose. You can create one variable called DoseA that replaces missing data with the mean, and another variable called DoseB that replaces missing data with another imputed score. Then, in any given model you would use one or the other, but not both.
Important

Missing data is based on all covariates in the list, and not only those that are checked. The program works this way to ensure that if you define two or more prediction models, the same studies will be used in all the models.

In Figure 175 we have un-checked Latitude [F], but the three studies are still excluded from the analysis as we can see in Figure 176 [G].

Figure 175 | Setup | Latitude in list, unchecked
In this example the studies are being excluded because they are missing a value for latitude. How can we include these studies in the analysis?

- If we want to use latitude as a predictor, the only option is to return to the data-entry screen and enter a value for latitude for each of these studies.

- If we are willing to run the regression without latitude, then we need to remove latitude from the list of covariates on the main screen.
It is not sufficient to simply un-tick latitude. Rather, we must remove it as shown in Figure 177.

- Highlight “Latitude” [G]
- Click “Remove covariates” [H]
We now run the analysis again (Figure 179) and see that the number of studies \([I]\) has returned to 13.
Note.

There are more sophisticated methods for handling missing data, such as multiple imputation and selection models for non-ignorable missing data. While these are beyond the scope of this manual, these and other methods can be implemented using CMA. You would use an external program to determine the data value for each study, and then input this value via the data-entry screen.

In this chapter we assumed that the each study has been entered into the database as one row, and the data for that row is either present or absent.

In Part 18: Complex data structures we show how to create data by combining data across subgroups, outcomes, or time-points, and how missing data is handled in that case.
PART 16: FILTER STUDIES

In some cases you may want to run a regression using a subset of the data. For example, you may want to limit the analysis to studies that employed acceptable methods for randomization and double-blinding. Or, you may want to limit the analysis to studies that were performed within the past ten years, or to studies that employed specific variants of the intervention or that enrolled persons from specific populations.

This procedure is called “Filtering”, in that we create a filter, and only studies that pass through the filter are submitted for the regression. The process is actually very simple, but requires that you understand the relationship among three distinct modules in the program. This is shown schematically in the following three figures.

1. In the data-entry module we enter data for all studies (Figure 180)
2. In the main analysis module we can create filters (Figure 181)
3. Studies that pass through the filters are submitted to the regression module (Figure 182)
We provide a few examples of filtering.

Example 1

Suppose you want to exclude two specific studies (Aaronson, Stein & Aaronson) by name.

On the main analysis screen (Figure 183), right-click on the names and select “Select by Study name”.

In Figure 184, un-tick the two studies and click [Ok].
The studies main analysis is now based on the remaining 11 studies. When you proceed to meta-regression, only these studies will be transmitted.
Example 2

The process of excluding studies by name works well if you only need to exclude a few studies, but becomes tedious and error-prone if you have a large database and need to exclude many studies. In this case, there are better options.

Suppose that you want to run a series of analyses using a specific subset of the studies. Create a categorical moderator (let’s call this Set-A) and code each study as belonging to this set (or not) as in Figure 185 [A].

![Figure 185 | Create a moderator for filtering](image)
On the main analysis screen

- Click Computation options > Select by
- Click the tab for Moderator
- Select Set-A
- De-select “No”

The analysis is now based on the eight studies that had been coded “Yes”. The mean effect size is -0.5894 [B].

Figure 186 | Filter by moderator
Click Analysis > Meta-regression 2

Run the regression

Figure 187 | Regression using a filter

The regression is based on these eight studies [C]. The mean effect size (the intercept) is −0.5894 [D].

You may create a series of these “Sets”, and easily switch between them. Run the regression for Set-A, and then for Set-B.
Example 3

You can filter studies based on existing moderators. For example, suppose you wanted to run an analysis using studies in a Hot climate that employed either Systematic or Random allocation.

- Click Computational options > Select by >
- Select Climate
- Tick [Hot]

Figure 188 | Select by moderator
• Click Add Filter
• Select Allocation
• Tick “Random” and “Systematic”

The analysis is now based on studies that meet both criterion, as shown in Figure 189 [E].

To ensure that things are working as intended, after running the regression click on All studies to see which have been used in the regression (Figure 190).

The studies in the regression are the same ones that were included in the main analysis when we applied these filters (Figure 189).
PART 17: DEFINING SEVERAL MODELS

Typically you will create one prediction model, which is the list of covariates to be included in the analysis. Then, you might try another model, and another, working one model at a time.

The program offers another option – to define a number of prediction models at once, and then run them all simultaneously. Here, for example, the user has defined one model that includes only the intercept [A], a second that adds latitude [B], and a third that adds year [C].
Then, when she runs the analysis the program runs all of the models as shown in Figure 193. The user can switch among them by using tabs at the bottom of the screen [A].

In Figure 193 the user has clicked on the tab labeled “Intercept” [A]. The screen displays the statistics for the analysis based on intercept alone.
In Figure 194 the user has clicked on the tab labeled “+Year” [B]. The screen displays the statistics for the analysis based on intercept and year.
In Figure 195 the user has clicked on the tab labeled “+Latitude” [C]. The screen displays the statistics for the analysis based on intercept, year and latitude.
Why would we want to define more than one model?

In the running example we defined three models on the main screen. Why take this approach, rather than simply working with one prediction model at a time? There are two reasons why this option may be useful.

First, the program summarizes the results of all models on one screen (Figure 196). To navigate to this screen click More results > [Compare models Detailed]
Second, the program displays a test for the difference in the explanatory power of the models (Figure 197). For example, the cell indicated by [A] compares the model that includes intercept and latitude with the one that adds year as well. This option is only available when one model is a subset of the other. When this is not the case, the corresponding cell will be left empty.

Figure 197 | Defining several models | Main-analysis | Intercept + year + latitude

How do we choose what covariates to include in each model? This depends on the questions we want to address.

For example, suppose the primary goal of the analysis is to assess the impact of treatment.

- One series of covariates such as mean age and location is seen as noise
- One series of covariates represents treatment condition
- One series (such as dose by treatment) represents potential interactions

We might define one model as “Nuisance”, a second as “Plus Treatment”, and a third as “Plus interactions”. Then, the summary screen provides a quick look at the three models while the comparison screen shows the statistical tests of the differences among them.
Working with multiple predictive models

As shown in Figure 198, to create a series of prediction models, use the toolbar [A] to

- Insert new models
- Delete models
- Rename models
- Move models left or right

Figure 198 | Defining several models | Setup
Two common scenarios for multiple models are a diagonal sequence and an incremental sequence. The program can generate these kinds of series automatically.

The incremental sequence is shown in Figure 199. The first model includes the intercept only, and then one covariate is added at each step. The name for each variable is “plus” that variable, since the variable has been added to the covariates.

To create this sequence click Generate sequence > Incremental sequence [A]

The diagonal sequence is Figure 200. Each model includes the intercept plus one covariate. The name for each variable is that variable alone.

To create this sequence click Generate sequence > Diagonal sequence [B]
Note.

When you run multiple models, always be sure to select the desired tab at the bottom when studying the results. These tabs control all tables that are model-specific.

In the earlier example we created models called “Intercept”, “+Year”, and “+Latitude”. Then we click on “Scatterplot”. We need to select a tab at the bottom of the screen to select the model for the scatterplot.

In Figure 201 we’ve clicked “+Year” [B] and the scatterplot is based on the intercept and latitude (as we can see from the prediction equation [C]).
By contrast, in Figure 202 we’ve clicked “+Latitude” [D] and the scatterplot is based on the intercept, latitude, and year, as we can see from the prediction equation [E].
The same holds true for the most screens. Specifically, as shown in Table 9

- Screens that present results for one predictive model will change as the user selects one or another model using the tabs at bottom.

- Screens that collate results for all one predictive models do not change as the user selects one or another model using the tabs at bottom.

<table>
<thead>
<tr>
<th>Screen</th>
<th>Varies by model</th>
<th>Identical for all models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main results</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Scatterplot</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>R2 graphic</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Covariance</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Diagnostics</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Valid studies</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Increments</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Models summary</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Compare models (detailed)</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Compare models (p-values)</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
PART 18: COMPLEX DATA STRUCTURES

On the main data-entry screen every row usually represents a single study. If there are 20 studies on the data-entry screen there will be 20 studies on the main analysis screen, and 20 studies transmitted to the meta-regression module.

However, in the case of complex data sets the situation is a bit more complicated. There are two kinds of complex data that we need to address.

- One is the case where we include two or more independent subgroups for some (or all) studies.
- The other is the case where we include two or more non-independent outcomes, time-points, or comparisons for some (or all) studies.

The program has a mechanism in place for dealing with these studies. The key to this mechanism is that all the data filtering and merging is performed on the main data-analysis screen. The rows displayed on this screen are the rows that will be transmitted to the regression module. In the regression module these rows will be treated as independent of each other.

INDEPENDENT SUBGROUPS WITHIN STUDIES

Consider the case where studies assess the impact of a drug, and report the results separately for males and for females. To record data for Independent subgroups (where each subject appears in one subgroup or the other, but not both) we use Insert > Column > for subgroups.

In this example we have five studies, and each reports the effect separately for males and females. For illustrative purposes we’ve set the effect size 0.20 points higher for males vs. females, and we’ve set the variance the same (0.10) for all subgroups for all studies.
We have two options.

- We can use the subgroup as the unit of analysis
- We can combined subgroups for each study, and use study as the unit of analysis

**Using subgroup as the unit of analysis**

We can use subgroup as the unit of analysis. In this case the variance of the summary effects should be 0.10/10, or 0.01, and the standard error should be 0.10.

On the data-analysis screen we right-click on the column labeled “Subgroup within study” and select “Use subgroup as the unit of analysis”
As expected, in Figure 206 the variance [A] is 0.01 and the standard error [B] is 0.10.

Now, we proceed to the meta-regression module.
For illustrative purposes we’ll run a regression with only the intercept.

In Figure 207 the number of studies in the analysis [C] is 10.

The standard error [D] is 0.10 (which implies a variance of 0.01).

These are the same numbers we saw in Figure 206.
Finally, we can click on More Results > All studies to display the data being used in the meta-regression (Figure 208). We see that the program is, in fact, working with the same 10 units that we had seen on the main analysis screen (Figure 206).

![Figure 208 | Regression | Subgroup within-study as unit of analysis](image)

### Using study as the unit of analysis

Immediately above we used subgroup as the unit of analysis. We also have the option of using study as the unit of analysis. In this case the program will merge the data for the subgroups within each study to yield study-level data. This data will then be used in the main analysis, and also in the meta-regression.

Since we are working with *independent* subgroups within each study, the variance of the estimate should be approximately the same regardless of whether we choose to use (a) subgroup or (b) study as the unit of analysis.

Above, we saw that when we used subgroups as the unit of analysis the variance of each subgroup was 0.10, and so the variance of the combined effect was 0.10/10=0.01.

When we use study as the unit of analysis the variance within each study is 0.10/2 (for two subgroups), or 0.05. Then, the variance of the combined effect is 0.05/5, which (as before) is 0.01.
To see this, let’s return to the main analysis screen and select Use Study as the unit of analysis.

In Figure 210, variance for each study is 0.05, the variance for the combined effect [A] is 0.01, and the standard error [B] is 0.10. These are the same numbers we saw in Figure 206.
Again, we can proceed to meta-regression and run the analysis with only the intercept.

![Image of meta-analysis results](image)

Figure 211 | Regression | Study as unit of analysis

- In Figure 207 the number of studies in the analysis [C] is 5.
- The standard error [D] is 0.10 (which implies a variance of 0.01).

These are the same numbers we saw in Figure 210.
Finally, we can navigate to More Results > All Data. Figure 212 shows that we now have five studies rather than 10 subgroups, and the variance for each study is 0.05.

![Comprehensive meta analysis - [Form1]](image)

<table>
<thead>
<tr>
<th>Study name</th>
<th>ES</th>
<th>Variance</th>
<th>Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study A</td>
<td>0.60</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Study B</td>
<td>0.70</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Study C</td>
<td>0.50</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Study D</td>
<td>0.65</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Study E</td>
<td>0.55</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 212 | Regression | Study as unit of analysis**

The point is, when we have *independent* subgroups within studies, every subgroup yields independent information, and must be treated as such in the analysis. We may elect to use subgroup as the unit of analysis or we may elect to use study as the unit of analysis, but in either case the within-study variance for the combined effect should be approximately the same. It is in the main analysis, and it is in the regression.

If the two options yield the same result here, does it matter which one we use? Yes, it does.

In this example we created a homogeneous set of studies for illustrative purposes, and tau-squared was zero. When tau-squared is zero, the two approaches yield very similar (if not identical) results.

By contrast, in a real analysis the estimate of tau-squared will often be different if based on 10 subgroups as compared with 5 studies. Concretely, if the effect sizes tend to vary a lot from one study to the next, but to be relatively similar for the subgroups within a study, it follows that tau-squared based on studies will tend to be larger while tau-squared based on subgroups will tend to be smaller.

The decision to use on or the other depends on how you see the sampling frame, and the population to which you want to generalize. If you want to get a sense how the effects are distributed across studies, then it makes sense to use study as the unit of analysis. If you want to get a sense of how the effects are distributed across subgroups, then it makes sense to use subgroups as the unit of analysis.

As noted above, the precision for estimating the mean effect will likely be similar in the two cases. However, the estimate of the variance itself (and statistics that depend on this estimate) will differ.

This applies to both the main analysis and also to the regression.

For regression, there is an additional consideration, as follows.
Suppose that the subgroups are Male and Female. If you intend to use gender as a covariate, then the only option would be to use subgroup as the unit of analysis.

Even if you don’t intend to use gender as a covariate, using subgroup as the unit of analysis may allow you to work with a finer level of data. For example, suppose you have the mean age for each subgroup, and plan to use mean age as a covariate. If you use subgroup as the unit of analysis you can use the mean age for each subgroup. By contrast, if you use study as the unit of analysis you’ll need to use the mean age for the study, and any difference in age between subgroups will be lost.
MULTIPLE OUTCOMES OR TIME-POINTS

Above, we addressed the case where we have independent subgroups within studies. The key was that each subgroup represented independent information (a person was included in one subgroup or another, but not both) and was treated as such in the analysis.

Now, we turn to the case where we have multiple outcomes or time-points within studies. The key here is that the rows for each study are based (at least partly) on the same persons, and do not provide independent information. We'll use the example of multiple outcomes and then comment on the issue of multiple time-points below.

To highlight the difference between independent subgroups on the one hand, and multiple outcomes on the other, we'll use the same data as before. However, this time we'll identify the rows within each study as being for two outcomes (reading and math) rather than for two subgroups (male and female).

When creating the data file we use Insert > Column for > Outcome names (Figure 213).

The data are shown in Figure 214.

Proceed to the main analysis screen
The program initially shows an analysis for Math only (Figure 215).

Figure 215 | Basic analysis | Multiple outcomes | Select one outcome

Right-click on Outcome to see the following options (Figure 216)

- Use the mean of the selected outcomes
- Use all of the selected outcomes, assuming independence
- Use the first outcome, based on this sequence

Figure 216 | Basic analysis | Multiple outcomes | Use all outcomes, assuming independence

Here, we select “Use all of the selected outcomes, assuming independence”.
We present this option here to illustrate its impact, and not to suggest that this is generally a valid option. In fact, we would consider using this option only when (a) there is only minor overlap in the samples and/or (b) the correlation between outcomes is small. Otherwise, this option will underestimate the variance (over-estimate the precision) of the summary effect size.

By selecting this option we are treating the correlation between outcomes as zero, which is (in effect) what we did with two independent subgroups. It follows that the results should be the same as they had been before. In fact, in Figure 217 the combined effect has a variance \( A \) of .01 and a standard error \( B \) of .10, which are the same numbers we saw in Figure 206.

As before, we’ll run a meta-regression with only the intercept (Figure 218), to show that the regression will yield the same results as the traditional analysis.
The results (Figure 219) shows a standard error [C] of 0.1000, with an implied variance of 0.0100, the same numbers as we saw in Figure 217.
Finally, More results > All data displays the actual rows of data in the analysis (Figure 220), which are identical to those in Figure 217.

<table>
<thead>
<tr>
<th>Study name</th>
<th>ES</th>
<th>Variance</th>
<th>Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study A</td>
<td>0.50</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study A</td>
<td>0.70</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study B</td>
<td>0.60</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study B</td>
<td>0.80</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study C</td>
<td>0.40</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study C</td>
<td>0.60</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study D</td>
<td>0.55</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study D</td>
<td>0.75</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study E</td>
<td>0.45</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Study E</td>
<td>0.65</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>

Figure 220 | Multiple outcomes | Use all outcomes, assuming independence

The point of this exercise was to show that when we treat the outcomes as independent of each other, the impact is the same as when we are working with independent subgroups. The program treats every line of data as though it is providing new (unique) information. While this is usually appropriate for independent subgroups, it is rarely appropriate for multiple outcomes. We will usually want to treat the outcomes as dependent, and compute the variance accordingly, as shown here.

Back on the main analysis screen (Figure 221) select “Use the mean of the selected outcomes” [A]
In Figure 222 the program combines the data for math and reading to yield a “Combined” score for each study [B].

In computing the variance for each study the program assumes a correlation of 1.0 between reading and math. The variance for reading is 0.10, the variance for math is .10, and the variance for the study-level composite is also 0.10. This follows from our decision to treat the correlation between reading and math as 1.0. This means that the second outcome is providing no new information (and so has no impact on the variance). This approach is conservative, in the sense that the true correlation is usually
less than 1; the second outcome probably provides some new information; and the true variance is probably lower than the value we are using.

As always, once we have a variance for each study we can compute the variance for the combined effect size. In Figure 222 the variance of the combined effect[C] is 0.10/5, or 0.02. This is twice as large as the value in Figure 217, which treated the outcomes as independent of each other. Similarly, the standard error of the combined effect [D], computed as the square root if the variance, is now 0.1414, as compared with 0.1000 in Figure 217.

Now, we proceed to meta-regression (Error! Reference source not found.).
The results for the regression (Figure 223) are identical to those for the traditional analysis (Figure 222). In particular, the standard error [E] is 0.1414, which implies a variance of 0.0200.

![Table showing regression results](image)

**Main results for Model 1, Random effects (MM), Z-Distribution**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.6000</td>
<td>0.1414</td>
<td>0.3228</td>
<td>0.8772</td>
<td>4.2426</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Statistics for Model 1**

- **Test of the model:** Simultaneous test that all coefficients (excluding intercept) are zero
  - $Q = 0.0000$, df = 0, $p = 1.0000$
- **Goodness of fit:** Test that unexplained variance is zero
  - $\tau^2 = 0.0000$, $\tau = 0.0000$, $I^2 = \%0.00$, $Q = 0.2500$, df = 4, $p = 0.9928$

**Comparison of Model 1 with the null model**

- **Total between-study variance (intercept only)**
  - $\tau^2 = 0.0000$, $\tau = 0.0000$, $I^2 = \%0.00$, $Q = 0.2500$, df = 4, $p = 0.9928$
- **Proportion of total between-study variance explained by Model 1**
  - $R^2$ analog = 0.00

**Number of studies in the analysis**

5
Finally, we can click More results > All studies to see data rows that are being transmitted to the regression (Figure 224). There are five rows, and the variance for each is 0.10, precisely as we had seen in Figure 222.

Other options

Our goal here was to show how the transformations and filtering specified in the main analysis are carried over to the regression. For this purpose we outlined two options, to treat the outcomes as independent or to combine them into a composite score assuming a correlation of 1.0.

The program includes a number of other options for working with multiple outcomes. These are not discussed here, but could be used in the main analysis screen and would carry over to the regression. These include selecting one outcome in preference to the others, and taking the mean of some outcomes while excluding others. While the program always assumes a correlation of 1 when it automatically generates composite scores, you can create composite scores manually using any correlation.
How covariates are merged

To this point we’ve shown what happens with the effect size and variance when we are working with multiple subgroups or outcomes within a study. Now, we need to consider what happens with the covariates. For example, suppose the subgroups are male and female, and each has a value for “Age”. If we merge the data from the two subgroups, what do we use for “Age”?

Independent subgroups within a study

The two options are

- To use subgroup as the unit of analysis
- To use study as the unit of analysis

If we use subgroup as the unit of analysis then each subgroup has its own value for each covariate, and this value gets transmitted along with the effect size.

However, if we use study as the unit of analysis then we need one value for the study, and we need to consider how this value is created. The rules are as follows.

- If both subgroups have the identical value (for example, age is coded as 40) then this same value is assigned to the study.
- If the value differs from one subgroup to the next within a study (one subgroup has age 40 and another has age 50) the covariate is transmitted as missing, and the study will be excluded from any regression that uses this covariate (see section on missing data).

You can use the “All studies” table to see where data are missing. If appropriate, you can return to the data-entry screen and modify the data (perhaps use age 45 for both).

The same idea is applied to multiple outcomes.

- If we use all outcomes then each outcome has its own value for each covariate, and this value gets transmitted along with the effect size.
- If we use the mean of outcomes then we need one value for the study. If both outcomes have the identical value (for example, age is coded as 40) then this same value is assigned to the study. If the value differs (one outcome has age 40 and another has age 50) the covariate is transmitted as missing, and the study will be excluded from any regression that uses this covariate (see section on missing data).

You can use the “All studies” table to see where data are missing. If appropriate, you can return to the data-entry screen and modify the data (perhaps use age 45 for both).

Multiple time-points

The case of multiple time-points works the same as multiple outcomes.
MULTIPLE COMPARISONS

Here, we use the term “multiple comparisons” to refer to a study where the same control group serves as the comparator for a series of Drugs.

If you will look at one drug at a time, then the Control-group data appears only once in each analysis, in which case one could argue that each analysis is valid on its own. However, if you will look at analysis that compares all treatments vs. the control, then we need to address the fact that the same control group will appear more than once in the data.

There are several options that can be taken

- One option is to use the same approach discussed for multiple outcomes.
- Another option is to divide the control group into segments when entering the data, and assign each segment to one treated group. For example, if the Control group has 20 events over 100 people, treat is as being two control groups with 10 events and 50 people in each. This option is explored in “Introduction to Meta-Analysis”.
PART 19: SOME CAVEATS
STATISTICAL POWER FOR META-REGRESSION

Statistical power is the likelihood that a test of significance will reject the null hypothesis. In the case of meta-regression it is the likelihood that the $Z$-test of a single covariate or the $Q$-test of a set of covariates will yield a statistically significant $p$-value.

Power depends primarily on two factors,

- The magnitude of the effect
- The precision with which we measure the effect

In a fixed-effect analysis precision is driven primarily by the total number of individual subjects across all studies. In a random-effects analysis precision is driven by the total number of individual subjects across all studies, and also by the variance in treatment effects and the total number of studies. (cite Hedges and Pigott, 2004?)

While there is a general perception that power for testing the main effect is consistently high in meta-analysis, this perception is not correct, and certainly does not extend to tests of effects in meta-regression. (Hedges and Pigott, 2001) The failure to find a statistically significant $p$-value in meta-regression could mean that the effect (if any) is quite small, but could also mean that the analysis had poor power to detect even a large effect. One should never use a non-significant finding to conclude that a covariate (or a set of covariates) is not related to effect size.
MULTIPLE COMPARISONS

In primary studies researchers often need to address the issue of multiple comparisons. The basic problem is that if we conduct a series of analyses with alpha set at 0.05 for each, then the overall likelihood of a type I error (assuming that the null is actually true) will exceed 5%. In the context of regression, this could be a problem if we test a series of covariates.

Some suggest simply allowing for a 5% error rate for each covariate. To use this approach we can simply work with the $p$-values on the main results screen. In this case it’s a good idea to evaluate the data in context. For example, one significant $p$-value in forty tests would be suspect.

Some suggest conducting an omnibus test that asks if there are any non-zero effects, and then proceeding to look at pair-wise comparisons only if the initial test meets the criterion for significance. To implement this approach we could first look at the $p$-value for the model. Or, if some of the covariates are nuisance variables (potential confounds that we want to hold constant) while others are of interest (say, a set of covariates that represent the treatment condition) we could test the increment for this set and use this as the gateway $p$-value.

Others suggest going straight to the individual covariates but using a stricter criterion for significance (for example, a criterion alpha of 0.01 rather than 0.05 for five tests).

Others (e.g., Rothman 1990) suggest that there are many cases where we can safely ignore the problem of multiple comparisons.

The same issues exist in meta-analysis and the approaches outlined above for primary studies can be applied to meta-analysis as well. (e.g., Hedges and Olkin, 1985).
PART 20: TECHNICAL APPENDIX
APPENDIX 1: THE DATASET

The motivating example in this book is the BCG dataset.

- The Excel™ version is called BCG.xls
- The CMA version is called BCG.cma

See Part 1: Data files and downloads for location of files

The original dataset includes only a few moderators (Latitude, Year, Allocation) as shown in Figure 225. We needed to create some additional moderators for purposes of this book. Some of these are shown in Figure 226 and all of these are listed in Table 10.

Table 10 also shows how we can create the new moderators using Excel. If we start with the dataset in CMA, it is possible to create new moderators directly in CMA, but the process can become tedious. An easier approach is to copy the relevant block of data from CMA to Excel™, add the new columns in Excel™, and then copy these new columns back into CMA. Then, in CMA, we just need to assign a name to each new column, identify it as a moderator, and assign a type (Categorical, Integer, or Decimal).

Table 10 shows the formulas that were employed in Excel™ to assign values to each study for the new variables.
<table>
<thead>
<tr>
<th>A</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>W</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Study name</td>
<td>Latitude</td>
<td>Year</td>
<td>Allocation</td>
<td>Latitude</td>
<td>Hot</td>
<td>Year</td>
<td>Time</td>
<td>Early</td>
<td>Recent</td>
<td>Hot x Year</td>
<td>Recent</td>
<td>Latitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Integer</td>
<td>Integer</td>
<td>Catagorical</td>
<td>Decimal</td>
<td></td>
<td>Integer</td>
<td>Integer</td>
<td>Catagorical</td>
<td>Integer</td>
<td>Integer</td>
<td>Integer</td>
<td>Decimal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ferguson &amp; Simex, 1969</td>
<td>151.53868</td>
<td>11.0595325</td>
<td>Cold</td>
<td>0</td>
<td>1</td>
<td>-15.0769 Early</td>
<td>1</td>
<td>0</td>
<td>0.0000</td>
<td>0</td>
<td>-13.7337</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Aronson, 1948</td>
<td>110.53846</td>
<td>111.0595172</td>
<td>Cold</td>
<td>0</td>
<td>1</td>
<td>-13.0769 Early</td>
<td>1</td>
<td>0</td>
<td>0.0000</td>
<td>0</td>
<td>-137.107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Stein &amp; Aronson, 1953</td>
<td>10.53846</td>
<td>111.0595172</td>
<td>Cold</td>
<td>0</td>
<td>1</td>
<td>-13.0769 Early</td>
<td>1</td>
<td>0</td>
<td>0.0000</td>
<td>0</td>
<td>-137.107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Rosenthal et al, 1960</td>
<td>1.53846</td>
<td>72.0595124</td>
<td>Cold</td>
<td>0</td>
<td>1</td>
<td>-11.0769 Early</td>
<td>1</td>
<td>0</td>
<td>0.0000</td>
<td>0</td>
<td>94.1799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Rosenthal et al, 1961</td>
<td>1.53846</td>
<td>72.0595124</td>
<td>Cold</td>
<td>0</td>
<td>1</td>
<td>-7.0769 Early</td>
<td>1</td>
<td>0</td>
<td>0.0000</td>
<td>0</td>
<td>90.4260</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Cornstock and Webstar, 1969</td>
<td>0.46154</td>
<td>0.21027175</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>1.0769 Recent</td>
<td>0</td>
<td>1</td>
<td>1.0769</td>
<td>1</td>
<td>0.4976</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Cornstock et al, 1974</td>
<td>13.04615</td>
<td>239.0595172</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>0.9207 Recent</td>
<td>0</td>
<td>1</td>
<td>0.9207</td>
<td>1</td>
<td>14.1722</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Frisoddi-Moller et al, 1973</td>
<td>12.0013</td>
<td>418.074306</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>1.9207 Recent</td>
<td>0</td>
<td>1</td>
<td>1.9207</td>
<td>1</td>
<td>35.3491</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cornstock et al, 1976</td>
<td>13.0013</td>
<td>418.074306</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>1.9207 Recent</td>
<td>0</td>
<td>1</td>
<td>1.9207</td>
<td>1</td>
<td>0.9876</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Hart &amp; Sutherland, 1977</td>
<td>13.0013</td>
<td>418.074306</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>1.9207 Recent</td>
<td>0</td>
<td>1</td>
<td>1.9207</td>
<td>1</td>
<td>35.3491</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Vandoni et al, 1973</td>
<td>19.0650</td>
<td>209.136095</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>16.9208 Recent</td>
<td>0</td>
<td>1</td>
<td>16.9208</td>
<td>1</td>
<td>344.7337</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Coetze &amp; Bejak, 1968</td>
<td>27.0650</td>
<td>41.7514793</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>16.9208 Recent</td>
<td>0</td>
<td>1</td>
<td>16.9208</td>
<td>1</td>
<td>195.1491</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>TB Prevention trial, 1980</td>
<td>13.0650</td>
<td>418.074306</td>
<td>Hot</td>
<td>1</td>
<td>0</td>
<td>16.9208 Recent</td>
<td>0</td>
<td>1</td>
<td>16.9208</td>
<td>1</td>
<td>407.6568</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 226 | BCG Data in Excel™
<table>
<thead>
<tr>
<th>Column</th>
<th>Variable</th>
<th>Type</th>
<th>Formula for Row 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>Latitude</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>Year</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Allocation</td>
<td>Categorical</td>
<td></td>
</tr>
</tbody>
</table>

The following are the original variables

<table>
<thead>
<tr>
<th>M</th>
<th>LatitudeC</th>
<th>Integer</th>
<th>=K4-$J$18</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>LatitudeC2</td>
<td>Integer</td>
<td>=M4^2</td>
</tr>
<tr>
<td>O</td>
<td>Climate</td>
<td>Categorical</td>
<td>=IF(J4&lt;40,&quot;Hot&quot;,&quot;Cold&quot;)</td>
</tr>
<tr>
<td>P</td>
<td>Hot</td>
<td>Integer</td>
<td>=IF($O4=&quot;Hot&quot;,1,0)</td>
</tr>
<tr>
<td>Q</td>
<td>Cold</td>
<td>Integer</td>
<td>=IF($O4=&quot;Cold&quot;,1,0)</td>
</tr>
</tbody>
</table>

Variables related to latitude
We compute the mean Latitude, which is 33.46154, in cell J18
LatitudeC is Latitude minus the mean
LatitudeC2 is the square of LatitudeC
Climate is a categorical variable, coded either Hot (latitude < 40) or Cold.
Hot and Cold are dummy variables based on Climate.

<table>
<thead>
<tr>
<th>R</th>
<th>YearC</th>
<th>Decimal</th>
<th>=K4-$K$18</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Time</td>
<td>Categorical</td>
<td>=IF(K4&lt;1945,&quot;Early&quot;,&quot;Recent&quot;)</td>
</tr>
<tr>
<td>T</td>
<td>Early</td>
<td>Integer</td>
<td>=IF($S4=&quot;Early&quot;,1,0)</td>
</tr>
<tr>
<td>U</td>
<td>Recent</td>
<td>Integer</td>
<td>=IF($S4=&quot;Recent&quot;,1,0)</td>
</tr>
</tbody>
</table>

Variables related to Year
We compute the mean Year, which is 1948.0769, in cell L18
YearC is Year minus the mean
Time is a categorical variable, coded either Early (Year < 1945) or Recent.
Early and Recent are dummy variables based on Climate.

Interactions
Hot x YearC is the interaction of Hot by Year (centered)
Hot x Recent is the interaction of Hot by Recent
LatitudeC x YearC is the interaction of Latitude (centered) by Year (centered)

<table>
<thead>
<tr>
<th>V</th>
<th>Hot x YearC</th>
<th>Decimal</th>
<th>=P4*R4</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>Hot x Recent</td>
<td>Integer</td>
<td>=P4*U4</td>
</tr>
<tr>
<td>X</td>
<td>LatitudeC x YearC</td>
<td>Decimal</td>
<td>=M4*R4</td>
</tr>
</tbody>
</table>

The “IF” formulas in Excel™ take the source column and recode it. Alternatively, you can simply type in the new values.

Normally, you do not need to create the dummy variables such as Hot and Cold. Rather, you can enter categorical variables into the regression and the program will create the dummy variables on the fly. You do need to create the dummy variables (as shown here) if you will be running regression without the intercept, or if you plan to use them in interactions.
In Figure 226, row 2 indicates if each variable is categorical, integer, or decimal. In CMA, when you define a variable as a moderator, you must also specify that it is one of these types.

A note on Categorical and Dummy variables

A categorical variable is a variable where the codes represent distinct groups, and it makes sense to say that the groups are different from each other, but not that one is “more than” another. A common example is Gender. Usually, each study is assigned a text value (Male or Female) for the variable.

A categorical variable cannot be used directly in the regression. Rather, we create dummy variables that contain the same information as the categorical variables, and use these in the analysis.

In most cases, the program is able to create the dummy variables on the fly. Concretely, if you enter a categorical variable into the analysis, the program will create one (or more) dummy variables, assign a code for each study, and enter these dummy variables into the analysis.

When this approach can be used, it offers the advantage of speed and simplicity. However, this approach can only be used in simple regressions. This approach cannot be used if

- We want to omit the intercept from the prediction model
- We want to use effects-coding, contrast-coding, or some other scheme
- We want to create interaction terms that include the categorical variable

In any of these cases, we need to create the dummy codes manually and enter the data as we would for any other study. Note that the dummy variables must be identified as integer or decimal, and not as categorical. For example,

- Climate is categorical (the values are Cold and Hot). However, the variable called Cold is an integer variable (the values are 0 and 1) and the variable called Hot is an integer variable (the values are 0 and 1).
- Time is categorical (the values are Early and Late). However, the variable called Early is an integer variable (the values are 0 and 1) and the variable called Recent is an integer variable (the values are 0 and 1).
APPENDIX 2: UNDERSTANDING Q

Introduction

In a simple meta-analysis we compute $Q$, a number that reflects the dispersion of all effect sizes about the mean effect size. Then we use $Q$ to compute an array of indices such as $T$, $T^2$, and $I^2$, that address specific aspects of dispersion. We briefly review the computation of $Q$ and its relationship to these indices. Then we proceed to the primary focus of this appendix, to show that the computation of $Q$ and the other indices is fundamentally the same for a simple analysis (with one group of studies), for a subgroups analysis, and for a regression. In all three cases we compute $Q$ by working with the deviation of effect sizes from the predicted effect size. In a simple analysis the predicted effect size is the mean of all studies, in a subgroups analysis it is the mean of the studies within the subgroup, and in the regression it is the relevant point on the regression line.

Background

$Q$ is a measure of dispersion on a standardized scale. We use $Q$ to address the question “Do all studies share a common effect size, $Y$”. To understand how $Q$ works, consider the following.

Suppose that we have a single study and want to ask, “Is the true effect size for this study equal to $Y$?” We could compute a $Z$-value for the study using

$$Z = \frac{X - Y}{SE_X}.$$  \hspace{1cm} (1.34)

where $X$ is the observed effect size, $Y$ is the predicted effect size, and $SE_X$ is the standard error of the effect size for this study.

To test the hypothesis that the true effect for the study is $Y$, and that the observed deviation $(X-Y)$ is due to sampling error, we could compare the observed $Z$ to the $Z$ distribution. If the observed $Z$ exceeds 1.96 (or −1.96) we might conclude that the true effect size is probably not $Y$.

Alternatively, we could have performed the same analysis by working with $Z^2$. The square of $Z$ is called $Q$, and the observed $Q$ value is evaluated with reference to the chi-squared distribution with 1 df. If the observed $Q$ exceeds 3.84 (that is, $1.96^2$) we might conclude that the true effect size is probably not $Y$.

In this example we wanted to test the hypothesis that the true value for one study is $Y$. The same logic can be extended to multiple studies, and this is exactly what happens when we test the assumption of homogeneity. For example, suppose that the meta-analysis includes five studies, and we want to test the assumption of homogeneity, that the true effect size for each of the five studies is $Y$. We compute $Z^2$ for each study and sum these values over all studies to yield $Q$. We then evaluate $Q$ with reference to the chi-squared distribution with $df$ equal to the number of studies minus one (here, $df = 4$)

We’ve elected to present $Q$ in this way in order to highlight its relationship to $Z$. Researchers understand that $Z$ is on a standardized scale, and that, if all studies share a common true effect size,
then observed effect sizes with deviations of 1.96 or 2.58 on this scale occur in 5% or 1% of cases, respectively. From this perspective, it’s easy to understand how \( Q \) can serve the same function.

In most texts (including this one) \( Q \) is also described as a “Weighted sum of squares”. This description highlights the fact that the deviations are weighted, and emphasizes the fact that the deviation of \( X \) from \( Y \) (the observed effect from the predicted effect) gets more weight in a more precise study (typically a larger study) than it does in a less precise (smaller) study.

To compute a weighted sum of squares we would compute each deviation and then weight it by the inverse variance. Thus, for any single study

\[
Q = (X - Y)^2 \times \frac{1}{V}
\]  

(1.35)

That is, we start with the deviation (squared) and then weight that value by the inverse of the within-study variance.

This equation (1.35) highlights the fact that studies with lower variance (typically the larger studies) are assigned greater weight. The earlier equation (1.34) highlighted the fact that the \( Q \)-values are on a standardized scale (squared). But in fact, the two formulas are algebraically equivalent.

To see this, note that (1.35) is in squared values. If we take the square root for both sides of the equation we get

\[
Z = (X - Y) \times \frac{1}{SE_x}
\]  

(1.36)

or

\[
Z = \frac{(X - Y)}{SE_x}.
\]  

(1.37)

which is identical to (1.34).

In either case (whether we use (1.34) or (1.35)) we need to sum the \( Q \) values over all studies. Hence the term “WSS”, or weighted sum of squares.

Once we have computed \( Q \), we can use it for a number of purposes.

First, we can test the \( Q \) value for statistical significance. If we are working with the fixed-effect model, the model requires that all studies share a common effect size. If the \( Q \)-value is statistically significant, we conclude that the studies do not share a common effect size, and this assumption has been violated. We probably should not be using the fixed-effect model.
Second, we can use $Q$ to estimate the amount of between-study variance and related indexes. Since $Q$ follows the chi-squared distribution, the expected value of $Q$ is equal to $df$. It follows that if the observed value of $Q$ exceeds the $df$, the excess can be attributed to between-study variance. This value $(Q - df)$, is the basis for the following.

Variance and standard deviation of true effects

To compute $T^2$, the between-studies variance, we start with $(Q - df)$. We multiply this difference by a factor (called $C$) that puts it into the metric of the effect size, squared. If $T^2$ is the variance of effect sizes between studies then its square root, $T$, is the standard deviation of effect sizes between studies.

Ratio of observed variance to true variance

To compute $I^2$, the ratio of true to total variance, we start with $(Q - df)$. If $Q$ reflects the total WSS and $(Q - df)$ reflects the WSS between studies, then the ratio $(Q - df)/Q$ is by definition the ratio of true/total variance, called $I^2$. By convention we multiple $I^2$ by 100 and express it as a percentage (0% to 100%).

Proportion of true variance explained by the predictors

In primary studies, $R^2$ is the proportion of variance explained by predictors. In meta-regression the analog to $R^2$ is the proportion of true variance explained by predictors, based on the explained variance as a proportion of the original variance.

Important

It is important to note that $Q$ is computed in exactly the same way (using weights based solely on within-study variance) regardless of whether we are working with a fixed-effect or a random-effects model. The only difference between the models is in how we interpret $Q$ and the associated statistics.
• Under the fixed-effect model, $Q$ addresses the statistical model. If $Q$ is statistically significant the statistical model is not valid.

• Under the random-effects model, $Q$ addresses the prediction model. If $Q$ is statistically significant the prediction model is incomplete, meaning that some of the true variance in effects is unexplained.

All of these statistics are based on $Q$, which is based on the deviation of observed from predicted effects. Below, we show how this works for the three general cases we called (A) one group of studies, (B) two subgroups of studies, and (C) a regression. The key point we want to emphasize with these examples is that these cases are all fundamentally the same.

Case A

The computation of $Q$ and derivative statistics for the simple case (the overall mean) is shown here

First, we run the regression using fixed-effect weights Figure 228.

![Main results for Model 1, Fixed effect, Z-Distribution](image)

This yields the prediction equation

$$Y = -0.4303.$$  \hspace{1cm} (1.38)
In Figure 229 we use (1.38) to compute the weighted squared deviation for each study and sum these to yield $Q$, which is 152.2329 (allowing for rounding error).

Working with $Q$ we then compute $T^2$, the variance of true effect sizes in the sample. It is given by

$$T^2 = \frac{Q - df}{C} = \frac{152.2329 - 12}{454.1808} = 0.3088$$

(1.39)

In this equation $df$ is the number of studies minus the number of predictors (the intercept), and $C$ is a conversion factor that allows us to move from the standardized scale of $Q$ to the metric of the effect size. This estimate of $T^2$ is then employed to assign random-effect weights to all studies and re-run the regression. The result is shown in Figure 230, where the prediction equation is

$$Y = -0.7141.$$  

(1.40)
The standard deviation of true effect sizes, \( T \), given by

\[
T = \sqrt{T^2} = \sqrt{0.3088} = 0.5557
\]  

(1.41)

If we assume that the true effects are normally distributed about their predicted value (and that the predicted value is correct), then 95% of all studies would have true effects in the range given the predicted value plus/minus 1.96 \( T \).

Under the random-effects model the predicted effect for all studies is \(-0.714\), and so the true effects would fall in the range of

\[
LL = -0.7141 - 1.96 \times 0.5557 = -1.8032 \quad \text{(1.42)}
\]

\[
UL = -0.7141 + 1.96 \times 0.5557 = 0.3750 \quad \text{(1.43)}
\]

These are the values we use to create the graphic. The normal curve is centered at \(-0.7141\) and extends roughly from 0.3750 to \(-1.8032\), intended to reflect the true effect size in some 95% of relevant studies. (This simplified example assumes that the mean is known. To actually compute prediction intervals we would take account of the error in estimating the mean.)
We can also compute $I^2$, the ratio of true variance to total variance. This is given by

$$I^2 = \frac{Q - df}{Q} = \frac{152.33 - 12}{152.33} = 92.12\%$$ \hspace{1cm} (1.44) \]

Typically we can also compute $R^2$, the proportion of variance explained by the covariates. In this case there are no covariates, and so this is not applicable.
Case B

When we turn to case B (two subgroups) all of the formulas remain the same. The only thing that changes is that the predicted value is now different for studies in each subgroup. (Also, the conversion factor C is based on a different formula).

First, we run the regression using fixed-effect weights Figure 228.

**Main results for Model 1, Fixed effect, Z-Distribution**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>Z-sided P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.9986</td>
<td>0.0676</td>
<td>-1.1310</td>
<td>-0.8662</td>
<td>-14.7807</td>
<td>0.0000</td>
<td>2.783</td>
</tr>
<tr>
<td>Climate: Hot</td>
<td>0.8870</td>
<td>0.0844</td>
<td>0.7216</td>
<td>1.0525</td>
<td>10.5092</td>
<td>0.0000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Analysis of variance**

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>df</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>110.4436</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>42.7894</td>
<td>11.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>152.2330</td>
<td>12.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Number of studies in the analysis 13

**Figure 232 | Case-B | Main results | Fixed-effect weights**

This yields the prediction equation

\[
Y = -0.9986 + 0.8870 \times \text{Hot} , \tag{1.45}
\]

where Hot is 0 for cold climates and 1 for hot climates.
In Figure 229 we use (1.38) to compute the weighted squared deviation for each study and sum these to yield $Q$, which is 152.2329 (allowing for rounding error).

Working with $Q$ we then compute $\tau^2$, the variance of true effect sizes in the sample. It is given by

$$\tau^2 = \frac{Q - df}{C} = \frac{41.7894 - 11}{319.4579} = 0.0964$$  \hspace{1cm} (1.46)

In this equation $df$ is the number of studies minus the number of predictors (the intercept and climate), and $C$ is a conversion factor that allows us to move from the standardized scale of $Q$ to the metric of the effect size. This estimate of $\tau^2$ is then employed to assign random-effect weights to all studies and re-run the regression.
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.1987</td>
<td>0.1769</td>
<td>-1.5455</td>
<td>-0.8518</td>
<td>-6.7740</td>
<td>0.0000</td>
<td>2.351</td>
</tr>
<tr>
<td>Climate: Hot</td>
<td>0.9203</td>
<td>0.2334</td>
<td>0.4628</td>
<td>1.3777</td>
<td>3.9426</td>
<td>0.0001</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
$Q = 15.5445$, df = 1, p = 0.0001

Goodness of fit: Test that unexplained variance is zero
$\tau^2 = 0.0964$, $SE = 0.0711$, $I-sq = 73.6775\%$, $Q = 41.7894$, df = 11, p = 0.0000

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
$\tau^2 = 0.3088$, $SE = 0.2295$, $I-sq = 92.1173\%$, $Q = 152.2330$, df = 12, p = 0.0000

Proportion of total between-study variance explained by Model 1
$R^2$ analog = 68.7848%

Number of studies in the analysis: 13

Figure 234 | Case-B | Main results | Random-effects

The result is shown in Figure 234, where the prediction equation for cold studies is

$$Y = -1.1987 + 0 \times 0.9203 = -1.1987$$ (1.47)

and for hot studies is

$$Y = -1.1987 + 1 \times 0.9203 = -0.1116$$ (1.48)

The standard deviation of true effect sizes, $T$, given by

$$T = \sqrt{T^2} = \sqrt{0.0964} = 0.3105$$ (1.49)

If we assume that the true effects are normally distributed about their predicted value (and that the predicted value is correct), then 95% of all studies would have true effects in the range given the predicted value plus/minus 1.96 $T$.

For cold climates, 95% of true effects would fall in the range of

$$LL = -0.9986 - 1.96 \times 0.3105 = -1.6071$$ (1.50)

$$UL = -0.9986 - 1.96 \times 0.3105 = -0.3901$$ (1.51)
For hot climates, 95% of true effects would fall in the range of

\[
LL = -0.1116 - 1.96 \times 0.3105 = -0.7201
\]

\[
UL = -0.1116 + 1.96 \times 0.3105 = +0.4969
\]

These are the values we use to create the graphic shown in Figure 237.

For cold studies the normal curve is centered at −0.9996 and extends roughly from −1.6071 to −0.3901, intended to reflect the true effect size in some 95% of relevant studies. For hot studies the normal curve is centered at −0.1116 and extends roughly from −0.7201 to +0.4969, intended to reflect the true effect size in some 95% of relevant studies. (This simplified example assumes that the mean is known. To actually compute prediction intervals we would take account of the error in estimating the mean.)

We can also compute \( \hat{I}^2 \), the ratio of true variance to total variance. This is given by

\[
\hat{I}^2 = \frac{Q - df}{Q} = \frac{41.7894 - 11}{41.7894} \times 100 = 73.68\%
\]

We can also compute \( R^2 \), the proportion of variance explained by the covariates. The unexplained variance with only the intercept is 0.3088 and the unexplained variance with the intercept plus climate is 0.0964, which means that the variance explained by the intercept is

\[
T^2_{\text{Explained}} = 0.3088 - 0.0964 = 0.2124
\]

Then, \( R^2 \) is the explained variance as a proportion of the total variance, or
or .6878.

Case C

Finally, we can apply exactly the same approach to compute $Q$ for the case where we use latitude to predict the effect size.

First, we run the regression using fixed-effect weights (Figure 236).

**Main results for Model 1, Fixed effect, Z-Distribution**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3436</td>
<td>0.0810</td>
<td>0.1847</td>
<td>0.5024</td>
<td>4.2390</td>
<td>0.0000</td>
<td>4.005</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0292</td>
<td>0.0027</td>
<td>-0.0344</td>
<td>-0.0240</td>
<td>-11.0227</td>
<td>0.0000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Analysis of variance

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>121.4999</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>30.7331</td>
<td>11.0000</td>
<td>0.0012</td>
</tr>
<tr>
<td>Total</td>
<td>152.2330</td>
<td>12.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Number of studies in the analysis 13

Figure 236 | Case-C | Main results | Fixed-effect weights

This yields the prediction equation

$$ Y = 0.3436 - 0.0292 \times \text{Latitude}, $$

(1.57)

where Latitude is an absolute value.
In Figure 237 we use (1.38) to compute the weighted squared deviation for each study and sum these to yield $Q$, which is 30.7339 (allowing for rounding error).

Working with $Q$ we then compute $\tau^2$, the variance of true effect sizes in the sample. It is given by

$$\tau^2 = \frac{Q - df}{C} = \frac{30.7339 - 11}{311.7368} = 0.0633$$

In this equation $df$ is the number of studies minus the number of predictors (the intercept and climate), and $C$ is a conversion factor that allows us to move from the standardized scale of $Q$ to the metric of the effect size. This estimate of $\tau^2$ is then employed to assign random-effect weights to all studies and re-run the regression.
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.2595</td>
<td>0.2323</td>
<td>-0.358</td>
<td>0.7149</td>
<td>1.1172</td>
<td>0.2639</td>
<td>5.390</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0292</td>
<td>0.0067</td>
<td>-0.0424</td>
<td>-0.0160</td>
<td>-4.3411</td>
<td>0.0000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 18.8452, df = 1, p = 0.0000

Goodness of fit: Test that unexplained variance is zero
Tau-sq = 0.0633, SE = 0.0548, I-sq = 64.2080%, Q = 30.7331, df = 11, p = 0.0012

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau-sq = 0.3088, SE = 0.2299, I-sq = 92.1173%, Q = 152.2330, df = 12, p = 0.0000
Proportion of total between-study variance explained by Model 1
R-sq analog = 79.4985%

Number of studies in the analysis 13

Figure 238 | Case-C | Main results | Random-effects

The result is shown in Figure 239, where the prediction equation for all studies is

\[ Y = 0.2595 - 0.0292 \times \text{Latitude} \]  

(1.59)

The standard deviation of true effect sizes, \( T \), given by

\[ T = \sqrt{T^2} = \sqrt{0.0633} = 0.2516 \]  

(1.60)

If we assume that the true effects are normally distributed about their predicted value (and that the predicted value is correct), then 95% of all studies would have true effects in the range given the predicted value plus/minus 1.96 \( T \).
In the figure we show the range of effects for three arbitrary points on the regression line. For studies at a latitude of 20, the predicted effect size is

\[ Y = 0.2595 - 0.0292 \times 20 = -0.2404 \] (1.61)

Then, 95% of true effects would fall in the range of

\[ LL = -0.2404 - 1.96 \times 0.2516 = -0.7335 \] (1.62)
\[ UL = -0.2404 + 1.96 \times 0.2516 = 0.2527 \] (1.63)

Thus, for studies at a latitude of 20 the normal curve is centered at −0.2404 and extends roughly from −0.7335 to 0.2527, intended to reflect the true effect size in some 95% of relevant studies. (This simplified example assumes that the mean is known. To actually compute prediction intervals we would take account of the error in estimating the mean.)

We can also compute \( I^2 \), the ratio of true variance to total variance. This is given by

\[ I^2 = \frac{Q - df}{Q} = \frac{30.72339 - 11}{30.7339} \times 100 = 64.21\% \] (1.64)

We can also compute \( R^2 \), the proportion of variance explained by the covariates. The unexplained variance with only the intercept is 0.3088 and the unexplained variance with the intercept plus climate is 0.0964, which means that the variance explained by the intercept is

\[ T_{Explained}^2 = 0.3088 - 0.0633 = 0.2455 \] (1.65)
Then, $R^2$ is the explained variance as a proportion of the total variance, or

$$R^2 = \frac{T^2_{\text{Explained}}}{T^2_{\text{Total}}} = \frac{0.2455}{0.3088} = 0.7950$$

or .7950.
APPENDIX 3: TESTS OF HETEROGENEITY

In any analysis, whether based on the fixed-effect model or the random-effects model, tests of heterogeneity are based on fixed-effect weights. Our goal in this section is to explain why. Consider Figure 240, which presents results for a simple analysis, with one set of studies.

At the left the program shows the analyses that address the effect size. One line presents results based on fixed-effect weights and the other presents results based on random-effects weights. Since the study weights on the first row are based on \( V \) while those on the second row are based on \( V + T^2 \), the effect size, the standard error, and all other statistics differ from one line to the next.

In the section on heterogeneity one might similarly expect to see two sets of statistics, one based on fixed-effect weights and the other on random-effects weights. In fact, though there is only one set of statistics, based on fixed-effect weights, that applies to both statistical models.

The reason has to do with the nature of the null hypothesis for heterogeneity, and the Q statistics that we use to test this hypothesis. Recall that Q is the WSS, or the weighed sum of squared deviations. That is,

1. We take the deviation of every effect size from the mean effect size
2. We square that deviation
3. We weight the squared deviation
4. We sum this value across all studies

Under the fixed-effect model,

In step 1 we compute the deviation of every effect size from \(-0.4303\). In step 3 we weight each squared deviation by \(1/V\). Computed in this way, \( Q \) is 152.2330. With 12 degrees of freedom the corresponding \( p \)-value is < 0.0001

Similarly, under the random-effects model,

In step 1 we compute the deviation of every effect size from \(-0.4303\). In step 3 we weight each squared deviation by \(1/V\). Computed in this way, \( Q \) is 152.2330. With 12 degrees of freedom the corresponding \( p \)-value is < 0.0001.
Thus, the test is identical under the two models because under the null there is no variation in true effects, which means that $T^2$ is zero. It follows that –

1. In step 1, when we compute the deviation of each effect size from the mean effect size, the mean effect size we need to use is −0.4303 (the fixed-effect estimate) rather than −0.7141 (the random-effects estimate). If $T^2$ is zero, the weighted mean would be −0.4303.
2. In step 3, when we weight each squared deviation to get the weighted sum of squares ($Q$), the weights must be based on $V$. Again if $T^2$ is zero then it has no impact on the weights.

Thus, the expected variation in effects under the null (that $T^2$ is zero) is identical whether we are using the fixed-effect model or the random-effects model. Therefore, the test of this null is identical under both models.

If it still seems odd that we’re using fixed-effect weights to test the null for a random-effects model, this might help. Rather than thinking of the test as using fixed-effect weights based on $V$, think of it as using random-effects weights based on $V + T^2$, where (under the null) $T^2$ happens to be zero.

Of course, the same logic holds if we use a regression approach.

Figure 241 | Heterogeneity statistics in regression
Figure 259 and Figure 260 show results for a fixed-effect analysis and a random-effects analysis using the same data as Figure 240.

- In the fixed-effect analysis the residual $Q$ is 152.2330 with $df = 12$ and $p < 0.0001$.
- In the random-effects analysis the Goodness of fit test shows $Q$ is 152.2330 with $df = 12$ and $p < 0.0001$.

These numbers are identical to each other.

While the test for heterogeneity is identical under the FE and the RE models, the way that we use the test results depends on the model.

- Under the FE model a significant p-value for heterogeneity tells us that the data are not consistent with the model. There is no reason to compute $T^2$ or $I^2$, since these are assumed to be zero. For this reason, regardless of whether or the p-value is statistically significant Figure 241 shows no estimate of $T^2$ nor of $I^2$.

- Under the RE model a significant p-value tells us that there is empirical evidence of heterogeneity, while a non-significant p-value tells us that this evidence is not present. Whether or not the p-value is statistically significant, we proceed to compute $T^2$ which is incorporated into the weighting scheme. For this reason, regardless of whether or the p-value is statistically significant, in Figure 260 the line that displays $Q$ also displays an estimate of $T^2$ and $I^2$.

To this point we’ve established that in a simple analysis, the test of the null that $T^2$ is zero is identical for fixed-effect and for random-effects models. The same logic applies when we move on to more complex analyses, such as analyses that involve subgroups or continuous covariates.

Figure 261 shows an analysis to assess the relationship between effect size and climate
As was true for the simple analysis, hypothesis tests that involve the relationship between subgroups and climate do depend on the statistical model.

Under the fixed-effect model the difference in subgroup means is based on computations where studies are weighted by $1/V$. The $Q$-value is 110.4436, with 1 $df$ and $p < 0.0001$.

Under the random-effects model the difference in subgroup means is based on computations where studies are weighted by $1/V + T^2$. The $Q$-value is 15.5445, with 1 $df$ and $p = 0.0001$.

By contrast, when we turn to the question of heterogeneity within subgroups, the program presents only one set of statistics. As before, these statistics are based on Fixed-effect weights and are displayed in the fixed-effect section.

The null hypothesis for heterogeneity is that studies within subgroups share the same effect size. To test the null that $T^2$ is zero we compute $Q$ and $df$ within subgroups and sum these values across subgroups. In this case, for the Cold and Hot subgroups $Q$ is 20.3464 plus 21.4431 for a total of 41.7894, while degrees of freedom are 5 plus 6 for a total of 11. A $Q$ value of 41.7894 with 11 $df$ yields a $p$-value of < 0.0001, and we conclude that there probably is some variation in true effects, even within subgroups.

As was true for the simple analysis, the computation of $Q$ is identical under the FE and the RE models since in both cases $T^2$ under the null is zero. When $T^2$ is zero the mean effect size in each group is based on study weights of $V$ rather than $V + T^2$, and the weights assigned to compute the weighed sum of squares are based on $V$ rather than $V + T^2$.

To this point we’ve shown that in a simple analysis and also in a subgroups analysis, the test of homogeneity is identical for the fixed-effect and the random-effects models. The same applies also to a regression analysis with continuous covariates.

Figure 262 and Figure 263 show a regression analysis using latitude as a covariate, for the fixed-effect model and the random-effects model, respectively.
Under the null hypothesis for homogeneity, \( T^2 \) is zero for all studies with the same predicted effect size (here for studies at the same latitude). And, as before, this if the null is true then study weights are based on \( 1/V \) for both models. This means that under the null hypothesis for heterogeneity, the regression line for the RE model will be identical to the regression line under the FE model and so the squared deviation of each study from the regression line will be the same. And, the weights employed to compute WSS will be the same under the two models. So, the \( Q \) value will be identical under the two models.

For that reason the statistics for the residual \( Q \) in Figure 244 and the statistics for Goodness of fit in Figure 263 are identical to each other.
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.1957</td>
<td>0.1769</td>
<td>-1.5435</td>
<td>-0.8518</td>
<td>-6.7740</td>
<td>0.0000</td>
</tr>
<tr>
<td>Climate: Hot</td>
<td>0.9203</td>
<td>0.2334</td>
<td>0.4628</td>
<td>1.3777</td>
<td>3.9426</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 15.5445, df = 1, \( p = 0.0001 \)

Goodness of fit: Test that unexplained variance is zero
\( \tau^2 = 0.0964, \tau = 0.3105, \hat{I}^2 = 73.68\%, Q = 41.7894, df = 11, \ p = 0.0000 \)

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
\( \tau^2 = 0.3088, \tau = 0.5557, \hat{I}^2 = 92.12\%, Q = 152.2330, df = 12, \ p = 0.0000 \)

Proportion of total between-study variance explained by Model 1
\( R^2 \) analog = 0.69

Number of studies in the analysis \( \quad 13 \)
### Main results for Model 1, Fixed effect, Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.3436</td>
<td>0.0810</td>
<td>0.1847</td>
<td>0.5024</td>
<td>4.2390</td>
<td>0.0000</td>
</tr>
<tr>
<td>Latitude</td>
<td>-0.0292</td>
<td>0.0027</td>
<td>-0.0344</td>
<td>-0.0240</td>
<td>-11.0227</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Analysis of variance**

<table>
<thead>
<tr>
<th></th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>121.4999</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>30.7331</td>
<td>11.0000</td>
<td>0.0012</td>
</tr>
<tr>
<td>Total</td>
<td>152.2330</td>
<td>12.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Number of studies in the analysis**: 13

---

**Figure 246 | Heterogeneity statistics with continuous covariate**
To avoid confusion we should emphasize that this section deals only with the issue of heterogeneity, and not with the issue of effect sizes.

- When we are focusing on the effect size itself, the test of the null does depend on the statistical model. Under the FE model we pose the null that the effect size is zero and test this null assuming a variance of $V$. Under the RE model we pose the null that the effect size is zero, and test this null assuming a variance of $V + T^2$. (The same idea carries through to the more complex cases).

- By contrast, when we are focusing on $T^2$ itself, as we do in this chapter, the test of the null is the test that $T^2$ is zero, and as such is identical for both statistical models.

Finally, while the computation of $Q$ is identical for the FE and the RE model, the implications of the $Q$ statistic, and what we do with this statistic, do depend on our assumptions about the statistical model.
As explained earlier, the $Q$-value can be used to test the null hypothesis for statistical significance. It can also be used to compute $T^2$ (the variance of true effect sizes) and to compute $I^2$ (the proportion of observed variance that represent differences in real effects rather than sampling error).

If we have adopted the fixed-effect model, a significant $Q$-value tells us that the data are not consistent with the model. Whether or not the $Q$-value is statistically significant we do not compute $T^2$ nor $I^2$, since under the FE model these are defined as being zero.

If we have adopted the random-effects model, then the fact that $Q$ is (or is not) statistically significant has little real bearing. The model allows for variance in true effects and is valid whether or not this variance exists. Whether or not the $Q$-value is statistically significant we do compute $T^2$ (which is employed to assign weights) and $I^2$ (which helps us to describe the distribution of effects).
APPENDIX 4: COMPUTING $\tau^2$ IN THE PRESENCE OF SUBGROUPS

When we’re working with subgroups, it’s clear that we need to estimate the between-study variance $T_2$ within subgroups rather than for the full set of studies. In our example, when we’re estimating the mean effect for Hot studies and for Cold studies, the between-study variance that we need to assign weights and to discuss the unexplained variance is clearly the variance within subgroups.

However, there are two ways to estimate $\tau^2$ within subgroups.

One option is to compute one estimate of $\tau^2$ for the Hot studies, and a separate estimate for the Cold studies. Then, we would use each estimate for the corresponding set of studies.

The other option is to compute one estimate of $\tau^2$ for the Hot studies and a separate estimate for the Cold studies. Then we would pool the two estimates and use the pooled estimate for both sets of studies.

The logic for choosing the second option is that estimate of $\tau^2$ are not reliable unless they’re based on a large number of studies. Unless we have good reason to believe that $\tau^2$ will differ substantially from one subgroup to the next, it’s often a better idea to assume that the true value of $\tau^2$ is comparable for each subgroup, and we’ll get a better estimate of this common value by pooling the within-subgroups estimates.

This is the option that we use with meta-regression since (at least when we’re using continuous covariates) the first option is not tenable. For purposes of comparing the subgroups analyses with the regression, we must select this option for subgroups as well.

On the analysis screen

- Select Computational options > Mixed and random effects options
- Select the option to Assume a common among-study variance across subgroups
- Select the option to Combine subgroup using a fixed-effect model

Figure 248 | Computing $\tau^2$ in the presence of subgroups
In this example we would pool the within-subgroup estimates of $T_2$ and apply the pooled estimate to both subgroups. The program does not display the pooled estimate on this screen. To see the pooled estimate click Next table and Calculations. The pooled estimate is 0.0964.
For those who are interested, we show how to actually pool the estimates of $\tau^2$ to get the pooled estimate, and how to display the pooled estimate. The mechanism for pooling is not to take the mean of the two estimates, but rather to pool the underlying statistics and then computed $\tau^2$ from the combined data.

Recall that

$$T^2 = \frac{Q - df}{C} \quad (1.67)$$

and so to compute a pooled value we use

$$T^2 = \frac{\sum Q - \sum df}{\sum C} \quad (1.68)$$

where values are summed across all subgroups. To get the within-subgroup values prior to pooling

- Select “Do not assume a common among-study variance component”.
- Click Fixed
- Click Calculations
In Figure 252

- Column C shows within-subgroup values of 110.9413 and 208.5166
- Column Q shows within-subgroup values of 20.3464 and 21.4431
- Column Q df shows within-subgroup values of 5 plus 6

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>Q</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold</td>
<td>110.9413</td>
<td>20.3464</td>
<td>5</td>
</tr>
<tr>
<td>Hot</td>
<td>208.5166</td>
<td>21.4431</td>
<td>6</td>
</tr>
<tr>
<td>Sum</td>
<td>319.4579</td>
<td>41.7895</td>
<td>11</td>
</tr>
</tbody>
</table>

Figure 252 | Computing $\tau^2$ in the presence of subgroups

$$T^2 = \frac{41.7895 - 1}{319.4579} = 0.0964,$$  

(1.69)

which is the same value we saw in Figure 251.

Be sure to switch the option back to “Assume a common value” and to re-select the tab for Random.
APPENDIX 5: CREATING VARIABLES FOR INTERACTIONS

See Part 1: Data files and downloads for location of files

In general, it’s a good idea to create all the variables you’ll need on the data-entry screen before proceeding to the analysis. However, if you’re in the middle of the analyses and discover that you need to create additional variables, you can easily return to the data-entry screen to do so.

In some cases it’s easiest to enter data for the new variables manually. This may be the case if you have only a few studies and the computation of the new data points is simple (for example, each study is coded 0 or 1).

In other cases, however, it’s easier to copy the data out to Excel, create the data for the new variables, and then copy the data back into CMA. We show that process here.

Suppose we’re working with the data set shown in Figure 253, which includes the variable Latitude. At some point we realize that we need a variable called Latitude-C (centered) and one called Latitude-C2 (centered, then squared) for the analyses.

Figure 253 | Creating variables for interactions
• Return to this screen.

Figure 254 | Creating variables for interactions

Insert a column called Latitude-C and define it as Moderator > Decimal
Insert a column called Latitude-C2 and define it as Moderator > Decimal

Figure 255 | Creating variables for interactions

• Click on the Latitude Column
• Click Edit > Copy with Header
Open Excel™

- Paste the column into Column A
- Define Cell A18 as =AVERAGE(A3:A15)
- Define Cell B3 as =A3-A18 and copy to other rows
- Define Cell C3 as =B3^2 and copy to other rows
- Copy columns B and C (rows 3 to 15) to the clipboard

Figure 256 | Creating variables for interactions

Return to CMA

- Click on Row-1 in the Latitude-C column
- Click CTRL-V to paste the data
- Save the file with the new data

Figure 257 | Creating variables for interactions
APPENDIX 6: PLOTTING A CURVILINEAR RELATIONSHIP

The spreadsheet used in this example is [Plot of curvilinear relationship.xlsx]
See Part 1: Data files and downloads for location of files

In chapter Part 14: Interaction we showed how to use Latitude and Latitude^2 to predict effect size. The program will not plot a curvilinear relationship, so we need to create the plot in Excel™.

In this example we used Latitude-C (centered) and Latitude-C^2 as covariates. Figure 258 shows the results of the analysis. To create the plot we need the coefficients [A] from this figure.

In Excel™, create columns as shown in Table 11 and Figure 259.

Table 11

<table>
<thead>
<tr>
<th>Column</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column A is a constant</td>
<td>All studies have a value of 1</td>
</tr>
<tr>
<td>Column B is the intercept (B0)</td>
<td>The value of -0.5501 comes from Figure 258</td>
</tr>
<tr>
<td>Column C is Latitude-C</td>
<td>Values range from -20 to +20</td>
</tr>
<tr>
<td>Column D is coefficient for Latitude-C</td>
<td>The value of -0.0316 comes from Figure 258</td>
</tr>
<tr>
<td>Column E is Latitude-C, squared</td>
<td>This is the square of column C</td>
</tr>
<tr>
<td>Column F is coefficient for Latitude-C^2</td>
<td>The value of -0.0008 comes from Figure 258</td>
</tr>
<tr>
<td>Column G is blank</td>
<td></td>
</tr>
<tr>
<td>Column H is latitude, un-centered</td>
<td>This is column C plus the mean latitude, 33.46</td>
</tr>
<tr>
<td>Column I is the predicted effect size</td>
<td>For row 3 this is $A3<em>B3+C3</em>D3+E3*F3$</td>
</tr>
</tbody>
</table>
In rows 3-43 we enter data for a sequence of possible studies. There are 41 rows, where the value of Latitude-C (centered) ranges from -20 to +20. Note that these are not the actual studies in our analysis, but rather the approximate range of values for Latitude-C in our analysis.
At this point column H holds the latitudes while Column I holds the predicted values. We can use these two columns to create a plot.

The instructions here are for Microsoft Excel™ 2010. The specific commands may vary slightly for other versions.

- Highlight the two columns from row 3 to the bottom (Figure 260)
- Select Insert > Scatter > Scatter with smooth lines
- The program creates this plot (Figure 261)
You may want to customize the graph as follows

Layout
Legend
None

Axes
Primary vertical axis
More options > Vertical axis crosses > Axis value −1.8
Primary vertical axis
More options > Number > Number > Decimal places > 2

Chart title
Above chart
Effect size as a function of Latitude and Latitude²

Axis title
Primary horizontal axis title
Title below Axis > Latitude
Primary vertical axis title
Rotated Title > Log Risk Ratio

At this point the plot should look like Figure 262

![Effect size as a function of Latitude and Latitude²](image)

Figure 262 | Plotting a curvilinear relationship

Notes

The key to plotting a non-linear relationship is that we create a row for every value of Latitude-C. Then we computed the corresponding value of the square, and the predicted value. The same idea can be extended for higher-order relationships as well.

We could have plotted effect size as a function of Latitude-C, in which case the X-axis would run from −20 to +20. Instead, we un-centered the predictor (adding the mean latitude to each centered value), and so the axis runs from 0 to 60. Critically, this edit takes place only at the last step. The prediction equations are based on the centered scores.
APPENDIX 7: PLOTTING INTERACTIONS

In this appendix we show how to plot three kinds of interactions

- The interaction of two categorical covariates
- The interaction of one categorical and one continuous covariate
- The interaction of two continuous covariates

All three follow basically the same format in Excel™. The only important difference is that for the first two we select the Line graph, and for the third we select the Scatter graph.
**Plotting the interaction of two categorical covariates**

In chapter we ran an analysis to assess the interaction of Climate by Time. Here, we show how to plot this interaction using Excel™.

See Part 1: Data files and downloads for location of files
The spreadsheet used in this example is [Plot of hot x time.xlsx]

The original data set includes the covariates Latitude and Year, both of which are continuous. For purposes of this discussion we need two categorical covariates, and so we create them by dichotomizing Latitude and Year.

- Hot is coded 1 if the latitude is 34 or less, and is coded 0 if the latitude exceeds 34.
- Recent is coded 1 if the Year is 1945 or later, and is coded 0 if the Year is earlier than 1945.
- Hot x Recent is created by multiplying Hot x Recent.

The results of the analysis are shown in Figure 264. To create the plot we’ll need the covariate names and coefficients from this figure. Copy these to Excel™ columns B and C as shown in Figure 265.

![Figure 263](image)

**Figure 263 | Plotting interaction of two categorical covariates**
To create the plot we need two points (Early and Recent) for the Cold studies

- Column E: Cold and Early. The X values are Hot (0), Recent (0), Interaction (0).
- Column F: Cold and Recent. The X values are Hot (0), Recent (1), Interaction (0).

To create the plot we need two points (Early and Recent) for the Hot studies

- Column H: Hot and Early. The X values are Hot (1), Recent (0), Interaction (0).
- Column I: Hot and Recent. The X values are Hot (1), Recent (1), Interaction (1).
The predicted value for each column (E, F, H, I) is given by multiplying each X value by the corresponding coefficient in column C, and then summing across the four rows.

For example, the formula for E10 is =E5*$C$5+E6*$C$6+E7*$C$7+E8*$C$8

To create the plot we need to identify the cells E10 and F10 as endpoints for the Cold studies, and H10 and I10 as endpoints for the Hot studies.

<table>
<thead>
<tr>
<th>Insert &gt; Graph &gt; Line &gt; With markers</th>
<th>Design &gt; Select data &gt; Add Series</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Series name</td>
</tr>
<tr>
<td></td>
<td>Series X-values</td>
</tr>
<tr>
<td></td>
<td>Series Y-Values</td>
</tr>
<tr>
<td></td>
<td>Design &gt; Select data &gt; Add Series</td>
</tr>
<tr>
<td></td>
<td>Series name</td>
</tr>
<tr>
<td></td>
<td>Series X-values</td>
</tr>
<tr>
<td></td>
<td>Series Y-Values</td>
</tr>
<tr>
<td></td>
<td>Layout &gt; Axis</td>
</tr>
<tr>
<td></td>
<td>Primary vertical axis</td>
</tr>
<tr>
<td></td>
<td>Primary vertical axis</td>
</tr>
<tr>
<td></td>
<td>Primary vertical axis</td>
</tr>
<tr>
<td></td>
<td>Primary vertical axis</td>
</tr>
<tr>
<td></td>
<td>Layout &gt; Chart Title</td>
</tr>
<tr>
<td></td>
<td>Above chart</td>
</tr>
<tr>
<td></td>
<td>Format</td>
</tr>
<tr>
<td></td>
<td>Series &gt; Hot</td>
</tr>
<tr>
<td></td>
<td>Series &gt; Cold</td>
</tr>
</tbody>
</table>
Effect size as a function of time and climate

Figure 266 | Plotting interaction of two categorical covariates
Plotting the interaction of a categorical covariate by a continuous covariate

The spreadsheet used in this example is [Plot of hot x year-c.xlsx]

In Part 14: Interaction we ran an analysis to assess the interaction of Year-C by Climate.

Since Climate is a categorical variable, the actual analysis covariates are Year-C and Hot, and the interaction is Year-C by Hot. Following our convention, Hot is coded 0 for Cold studies and 1 for Hot studies. The interaction is the product of the two. The results of the analysis are shown in Figure 267.

To create the plot we’ll need the covariate names and coefficients from Figure 267. In Excel™, copy these to columns B and C as shown in Figure 268.

![Table showing main results for Model 1, Random effects (MM), Z-Distribution](image)

Figure 267 | Plotting interaction of categorical by continuous covariates
Figure 268 | Plotting interaction of categorical by continuous covariates

To create the plot we need two points (1933 and 1968) for the Cold studies

- Column E: Cold and 1933. The X values are Hot (0), Year-C (−15), Interaction (0).
- Column F: Cold and 1968. The X values are Hot (0), Recent (20), Interaction (0).

To create the plot we need two points (1933 and 1968) for the Hot studies

- Column H: Hot and 1933. The X values are Hot (1), Year-C (−15), Interaction (−15).
- Column I: Hot and 1968. The X values are Hot (1), Year-C (20), Interaction (20).

The predicted value for each column (E, F, H, I) is given by multiplying each X value by the corresponding coefficient in column C, and then summing across the four rows.

For example, the formula for E10 is =E5*$C$5+E6*$C$6+E7*$C$7+E8*$C$8
Note

The columns are labeled 1933 and 1968, since these are the values we want to display in the plot. Critically, these are not the values we use to compute the effect size. Rather, the X values in the data columns are the corresponding centered values, −15 and 20.

To create the plot we need to identify the cells E10 and F10 as endpoints for the Cold studies, and H10 and I10 as endpoints for the Hot studies.

<table>
<thead>
<tr>
<th>Insert</th>
<th>Graph</th>
<th>Line</th>
<th>With markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design</td>
<td>Select data</td>
<td>Add Series</td>
<td></td>
</tr>
<tr>
<td>Series name</td>
<td>H2</td>
<td>Hot</td>
<td></td>
</tr>
<tr>
<td>Series X-values</td>
<td>H3:I3</td>
<td>1933, 1968</td>
<td></td>
</tr>
<tr>
<td>Series Y-Values</td>
<td>H10:I10</td>
<td>−0.0275, −0.3929</td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>Select data</td>
<td>Add Series</td>
<td></td>
</tr>
<tr>
<td>Series name</td>
<td>E2</td>
<td>Cold</td>
<td></td>
</tr>
<tr>
<td>Series X-values</td>
<td>E3:F3</td>
<td>1933, 1968</td>
<td></td>
</tr>
<tr>
<td>Series Y-Values</td>
<td>E10:F10</td>
<td>−1.0184, −1.9286</td>
<td></td>
</tr>
<tr>
<td>Layout</td>
<td>Axis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary vertical axis</td>
<td>Axis options &gt; Minimum &gt; Fixed &gt; −2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary vertical axis</td>
<td>Axis options &gt; Maximum &gt; Fixed &gt; 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary vertical axis</td>
<td>Horizontal Axis Crosses &gt; −2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary vertical axis</td>
<td>Number &gt; Number &gt; Decimal places &gt; 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layout</td>
<td>Axis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Axis options &gt; Minimum &gt; Fixed &gt; 1930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Axis options &gt; Maximum &gt; Fixed &gt; 1970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Vertical Axis Crosses &gt; 1930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Number &gt; 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layout</td>
<td>Chart Title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above chart</td>
<td>Log risk ratio as a function of climate, year, interaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series &gt; Hot</td>
<td>Format selection &gt; Line Style &gt; Dash type &gt; Solid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Series &gt; Cold</td>
<td>Format selection &gt; Line Style &gt; Dash type &gt; Dashed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 269 | Plotting interaction of categorical by continuous covariates
Plotting the interaction of two continuous covariates

In this example we assess the impact of Year-C, Latitude-C, and Year-C x Latitude-C.

See Part 1: Data files and downloads for location of files
The spreadsheet used in this example is [Plot of latitude-c x year-c.xlsx]

- Year-C is the study year, centered
- Latitude-C is the latitude, centered
- Year-C x Latitude-C is the interaction

The results of the analysis are shown in Figure 270

To create the plot we’ll need the covariate names and coefficients from Figure 270. In Excel™, copy these to columns B and C as shown in Figure 271.
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.683397</td>
<td>0.166884</td>
<td>-1.015484</td>
<td>-0.361311</td>
<td>-4.1250</td>
<td>0.000037</td>
</tr>
<tr>
<td>LatitudeC</td>
<td>-0.032197</td>
<td>0.011854</td>
<td>-0.055431</td>
<td>-0.008963</td>
<td>-2.7161</td>
<td>0.006506</td>
</tr>
<tr>
<td>YearC</td>
<td>-0.004755</td>
<td>0.015522</td>
<td>-0.035177</td>
<td>0.025686</td>
<td>-0.3693</td>
<td>0.759363</td>
</tr>
<tr>
<td>LatitudeC x YearC</td>
<td>0.000350</td>
<td>0.000968</td>
<td>-0.001548</td>
<td>0.002247</td>
<td>0.3810</td>
<td>0.718066</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 11.3936, df = 3, p = 0.009777

Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0.1241, Tau = 0.3523, I^2 = 64.58%, Q = 25.4100, df = 9, p = 0.002550

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
Tau^2 = 0.3088, Tau = 0.5557, I^2 = 92.12%, Q = 152.2330, df = 12, p = 0.000000
Proportion of total between-study variance explained by Model 1
R^2 analog = 0.60

Number of studies in the analysis: 13

Figure 270 | Plotting interaction of continuous covariates

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.683397</td>
</tr>
<tr>
<td>Year-C</td>
<td>-0.004755</td>
</tr>
<tr>
<td>Latitude-C</td>
<td>-0.032197</td>
</tr>
<tr>
<td>Latitude-C x YearC</td>
<td>0.000350</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lat-13</th>
<th>Lat-13</th>
<th>Lat-33</th>
<th>Lat-33</th>
<th>Lat-55</th>
<th>Lat-55</th>
</tr>
</thead>
<tbody>
<tr>
<td>1933</td>
<td>1908</td>
<td>1933</td>
<td>1908</td>
<td>1933</td>
<td>1908</td>
</tr>
<tr>
<td>300</td>
<td>-400</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-315</td>
<td>420</td>
<td>-21</td>
<td>21</td>
<td>-15</td>
<td>20</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-15</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 271 | Plotting interaction of continuous covariates

348
We will plot effect size for three latitudes (13, 33, 55). For each of these latitudes, we need the effect size at two years (1933, 1968)

To create the plot we need two points (1933 and 1968) for Latitude 13

- Column E: Latitude=13, 1933. The X values are Latitude-C (~20), Year-C (~15), Interaction (300).
- Column F: Latitude=13, 1968. The X values are Latitude-C (~20), Year-C (20), Interaction (~400).

To create the plot we need two points (1933 and 1968) for Latitude 33

- Column E: Latitude=33, 1933. The X values are Latitude-C (0), Year-C (~15), Interaction (0).
- Column F: Latitude=33, 1968. The X values are Latitude-C (0), Year-C (20), Interaction (0).

To create the plot we need two points (1933 and 1968) for Latitude 55

- Column E: Latitude=55, 1933. The X values are Latitude-C (21), Year-C (~15), Interaction (~315).
- Column F: Latitude=55, 1968. The X values are Latitude-C (21), Year-C (20), Interaction (~420).

The predicted value for each column (E, F, H, I, K, L) is given by multiplying each X value by the corresponding coefficient in column C, and then summing across the four rows.

For example, the formula for E10 is =E5*$C$5+E6*$C$6+E7*$C$7+E8*$C$8

Note

The columns are labeled with the years 1933 and 1968, since these are the values we want to display in the plot. Critically, these are not the values we use to compute the effect size. Rather, the X values in the data columns are the corresponding centered values, ~15 and 20.

Similarly, the columns are labeled with the latitudes 13, 33, and 55, since these are the values we want to display in the plot. Critically, these are not the values we use to compute the effect size. Rather, the X values in the data columns are the corresponding centered values, ~20, 0, and 21.

To create the plot we need to identify the cells E10 and F10 as endpoints for the Latitude=13 studies, H10 and I10 as endpoints for the Latitude=33 studies, and K10 and L10 as endpoints for the Latitude=55 studies.
**Insert > Scatter > Smooth lines**

**Design > Select data > Add Series**

<table>
<thead>
<tr>
<th>Series name</th>
<th>E2</th>
<th>Lat=13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series X-values</td>
<td>E3:F3</td>
<td>1933, 1968</td>
</tr>
<tr>
<td>Series Y-Values</td>
<td>E10:F10</td>
<td>0.1318,−0.2796</td>
</tr>
</tbody>
</table>

**Design > Select data > Add Series**

<table>
<thead>
<tr>
<th>Series name</th>
<th>H2</th>
<th>Lat=33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series X-values</td>
<td>H3:I3</td>
<td>1933, 1968</td>
</tr>
<tr>
<td>Series Y-Values</td>
<td>H10:I10</td>
<td>−0.6171,−0.7835</td>
</tr>
</tbody>
</table>

**Design > Select data > Add Series**

<table>
<thead>
<tr>
<th>Series name</th>
<th>K2</th>
<th>Lat=55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series X-values</td>
<td>K3:L3</td>
<td>1933, 1968</td>
</tr>
<tr>
<td>Series Y-Values</td>
<td>K10:L10</td>
<td>−1.4035,−1.3126</td>
</tr>
</tbody>
</table>

**Layout > Axis**

<table>
<thead>
<tr>
<th>Primary vertical axis</th>
<th>Axis options &gt; Minimum &gt; Fixed &gt; −2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary vertical axis</td>
<td>Axis options &gt; Maximum &gt; Fixed &gt; 0.5</td>
</tr>
<tr>
<td>Primary vertical axis</td>
<td>Horizontal Axis Crosses &gt; −2.5</td>
</tr>
<tr>
<td>Primary vertical axis</td>
<td>Number &gt; Number &gt; Decimal places &gt; 2</td>
</tr>
</tbody>
</table>

**Layout > Axis**

<table>
<thead>
<tr>
<th>Primary horizontal axis</th>
<th>Axis options &gt; Minimum &gt; Fixed &gt; 1930</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary horizontal axis</td>
<td>Axis options &gt; Maximum &gt; Fixed &gt; 1970</td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Vertical Axis Crosses &gt; 1930</td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Number &gt; 0</td>
</tr>
<tr>
<td>Primary horizontal axis</td>
<td>Number &gt; Use 1000 separator &gt; No</td>
</tr>
</tbody>
</table>

**Layout > Chart Title**

- Above chart: Log risk ratio as a function of climate, year, interaction

**Format**

<table>
<thead>
<tr>
<th>Series &gt; Lat=13</th>
<th>Format selection &gt; Line Style &gt; Dash type &gt; Solid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series &gt; Lat=33</td>
<td>Format selection &gt; Line Style &gt; Dash type &gt; Dashed</td>
</tr>
<tr>
<td>Series &gt; Lat=55</td>
<td>Format selection &gt; Line Style &gt; Dash type &gt; Dashed</td>
</tr>
</tbody>
</table>
Figure 272 | Plotting interaction of continuous covariates
APPENDIX 8: INTERPRETING REGRESSION COEFFICIENTS

Here, we are working with studies where the treatment is associated with lower risk. Given the coding, most risk ratios are less than 1.0 and (equivalently) most log risk ratios are negative. Therefore, a negative coefficient implies that a higher score is associated with a lower risk (the log risk ratio becomes more negative, and the risk ratio moves further from 1.0).

By contrast, if we were working with studies where the risk tended to be greater than 1 (for example the “risk” of living longer) then the log risk ratio would tend to be positive. Then, 0 is no effect, +1 is a large effect, and +2 is a very large effect. Therefore, a positive coefficients means that as the covariate gets larger the vaccine is more effective.

The coefficient for Year is 0.0235, which means that for every increase of one year the log risk ratio will increase by 0.0235 (the vaccine became less effective over time). The corresponding $p$-value is 0.1390.
APPENDIX 9: META-REGRESSION IN STATA

In this section we show the correspondence between results produced by CMA and those produced by the stata macro “metareg”
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Set</th>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allocation</td>
<td>Intercept</td>
<td>-29.23517</td>
<td>45.795202</td>
<td>-115.600704</td>
<td>60.526871</td>
<td>-0.638300</td>
<td>0.252246</td>
</tr>
<tr>
<td>Allocation</td>
<td>Allocation</td>
<td>0.485409</td>
<td>0.480857</td>
<td>0.459694</td>
<td>1.427931</td>
<td>1.009949</td>
<td>0.312092</td>
</tr>
<tr>
<td>Allocation</td>
<td>Systematic</td>
<td>0.457559</td>
<td>0.377773</td>
<td>-0.280371</td>
<td>1.197771</td>
<td>1.210849</td>
<td>0.226038</td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td>0.014831</td>
<td>0.023194</td>
<td>-0.035028</td>
<td>0.069095</td>
<td>0.002222</td>
<td>1.192149</td>
</tr>
<tr>
<td>Latitude</td>
<td></td>
<td>-0.001903</td>
<td>0.002545</td>
<td>-0.003205</td>
<td>0.002222</td>
<td>0.022222</td>
<td>1.192149</td>
</tr>
</tbody>
</table>

Test of the model: Simultaneous test that all coefficients (excluding intercept) are zero
Q = 13.174235, df = 4, p = 0.010953

Goodness of fit: Test that unexplained variance is zero
\( \tau^2 = 0.119410, \tau_u = 0.342057, I^2 = 66.686974\%, Q = 24.634414, df = 5, p = 0.002279 \)

Comparison of Model 1 with the null model

Total between-study variance (Intercept only)
\( \tau^2 = 0.308760, \tau_u = 0.550062, I^2 = 92.117544\%, Q = 152.230388, df = 12, p = 0.000000 \)

Proportion of total between-study variance explained by Model 1
R^2 analog = 0.613651

Number of studies in the analysis = 13

---

Figure 273 | CMA | Intercept + Year + Latitude + Allocation | Z | Method of moments

```
.meta reg logrr alternate systematic year lat, wse(e) beest(mm) tau2test z
```

Meta-regression

| Method of moments estimate of between-study variance | tau2 | .1194 |
| % residual variation due to heterogeneity | I-squared_res | 66.696 |
| Proportion of between-study variance explained | R-squared | 61.526 |
| Joint test for all covariates | Model chl2(4) | 13.17 |
| Without Knapp-Hartung modification | Prob > chl2 | 0.0105 |

| Coef. Std. Err. | z  | P>|z| | 95% Conf. Interval |
|-----------------|----|------|------------------|
| alternate | .4554685 | .4806573 | 1.01 | 0.313 | -1.456944 | 1.427931 |
| systematic | .45735 | .3777728 | 1.21 | 0.226 | -2.93071 | 1.197771 |
| year | .0148307 | .0231938 | 0.64 | 0.523 | -0.030283 | 0.060289 |
| lat | -0.0190132 | .0159451 | -1.19 | 0.253 | -0.050264 | 0.012365 |
| _cons | -29.23517 | 45.795202 | -0.64 | 0.523 | -119.0007 | 60.526871 |

Test for residual between-study variance (of tau2=0) Q_res (8 df) = 24.01
Prob > Q_res = 0.0022

Figure 274 | Metareg | Intercept + Year + Latitude + Allocation | Z | Method of moments
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold</td>
<td>-1.19861</td>
<td>0.17955</td>
<td>-1.545475</td>
<td>-0.851846</td>
<td>-6.774024</td>
<td>0.000000</td>
</tr>
<tr>
<td>Hot</td>
<td>-0.278335</td>
<td>0.132219</td>
<td>-0.570738</td>
<td>0.019949</td>
<td>-1.828912</td>
<td>0.067413</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients are zero
Q = 49.23233, df = 2, p = 0.000000

Goodness of fit: Test that unexplained variance is zero
Tau^2 = 0.09638, Tau = 0.310452, I^2 = 73.67548%, Q = 41.785426, df = 11, p = 0.000018

Comparison of Model 1 with the null model

Total between-study variance (Intercept only)
This is not defined when the intercept is excluded
Proportion of total between-study variance explained by Model 1
This is not defined when the intercept is excluded

Number of studies in the analysis: 13

Figure 275 | CMA | Allocation | Z | Method of moments

```
. metareg logrr cold hot, vce(sa) haest(mm) tau2 test s noconst

Meta-regression: Number of obs = 13
Method of moments estimate of between-study variance: tau2 = 0.09638
% residual variation due to heterogeneity: I-squared_res = 73.67548%
Joint test for all covariates: Model chI2(2) = 49.22
Without Knapp-Hartung modification: Prob > chI2 = 0.0000

| logrr | Coef. | Std. Err. | z    | P>|z|  | (95% Conf. Interval) |
|-------|-------|-----------|------|-----|---------------------|
| cold  | -1.19861 | 0.17955 | -6.77 | 0.000 | -1.545475  -0.851846 |
| hot   | -0.278335  | 0.132219 | -2.1214 | 0.035 | -0.549575  -0.007095 |
```

Test for residual between-study variance (of tau2=0): Q_res (11 df) = 41.78
Prob > Q_res = 0.0000

Figure 276 | Metareg | Allocation | Z | Method of moments
Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>2-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold</td>
<td>-1.355796</td>
<td>0.252226</td>
<td>-1.830069</td>
<td>-0.861353</td>
<td>-5.374968</td>
<td>0.000000</td>
</tr>
<tr>
<td>Hot</td>
<td>-0.1317542</td>
<td>0.288336</td>
<td>-0.660008</td>
<td>0.397980</td>
<td>-0.57609</td>
<td>0.564204</td>
</tr>
<tr>
<td>YearC</td>
<td>-0.017390</td>
<td>0.017570</td>
<td>-0.051827</td>
<td>0.021704</td>
<td>-0.989757</td>
<td>0.322293</td>
</tr>
</tbody>
</table>

Statistics for Model 1

Test of the model: Simultaneous test that all coefficients are zero
\[ Q = 31.235579, df = 3, \ p = 0.000000 \]

Goodness of fit: Test that unexplained variance is zero
\[ R^2 = 0.459728, \ \tau^2 = 0.369174, \ \rho^2 = 0.459545, \ Q = 41.77128, \ df = 10, \ p = 0.000000 \]

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
This is not defined when the intercept is excluded

Proportion of total between-study variance explained by Model 1
This is not defined when the intercept is excluded

Number of studies in the analysis: 13

Figure 277 | CMA | Allocation, Year | Z | Method of moments

```
. metareg logrr cold hot yearc, vce(se) cluster(mm) tau2test  nocoest
```

Meta-regression

Method of moments estimate of between-study variance
\[ \tau^2 = 0.1497 \]

Heterogeneity statistic
\[ I^2 = 76.08\% \]

Model with random effects

Joint test for all covariates
\[ \chi^2 (3) = 38.24 \]

Without Knapp-Hartung modification
\[ \text{Prob} > \chi^2 = 0.0000 \]

| logrr   | Coef. | Std. Err. | z     | P>|z|     | 95% Conf. Interval |
|---------|-------|-----------|-------|---------|-------------------|
| cold    | -1.355796 | 0.252226  | -5.37 | 0.000   | -1.65066   -1.061353 |
| hot     | -0.1317542 | 0.288336  | -0.56 | 0.564   | 0.605062   0.5239796 |
| yearc   | -0.017390 | 0.017570  | -0.99 | 0.322   | 0.0518272   0.0170467 |

Test for residual between-study variance (df = 0)
\[ Q_{res} (10 df) = 41.70 \]
\[ \text{Prob} > Q_{res} = 0.0000 \]

Figure 278 | Metareg | Allocation, Year | Z | Method of moments
## Main results for Model 1, Random effects (MM), Z-Distribution

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>95% Lower</th>
<th>95% Upper</th>
<th>Z-value</th>
<th>Z-sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.350081</td>
<td>0.186246</td>
<td>-0.515116</td>
<td>-0.185046</td>
<td>-2.953524</td>
<td>0.003342</td>
</tr>
<tr>
<td>LatitudeC</td>
<td>-0.001361</td>
<td>0.000762</td>
<td>-0.004082</td>
<td>-0.001835</td>
<td>-4.294460</td>
<td>0.000023</td>
</tr>
<tr>
<td>Latitude-C2</td>
<td>-0.000790</td>
<td>0.000702</td>
<td>-0.002166</td>
<td>0.000585</td>
<td>-1.126446</td>
<td>0.259977</td>
</tr>
</tbody>
</table>

### Statistics for Model 1

Test of the model: simultaneous test that all coefficients (excluding intercept) are zero
\[ Q = 17.899795, df = 2, p = 0.000130 \]

Goodness of fit: test that unexplained variance is zero
\[ Tu^2 = 0.074718, Tu^2 = 0.074718, I^2 = 64.967462\%, Q = 28.490606, df = 16, p = 0.001503 \]

Comparison of Model 1 with the null model

Total between-study variance (intercept only)
\[ Tu^2 = 0.052706, Tu^2 = 0.052706, I^2 = 22.177347\%, Q = 152.230668, df = 12, p = 0.000000 \]

Proportion of total between-study variance explained by Model 1
\[ R^2 analog = 0.758000 \]

Number of studies in the analysis: 13

---

### Figure 279 | CMA | Intercept, Year-C, Year-C2 | Z | Method of moments

```stata
.meta reg logrr latc latc2, wse(se) bse(tm) tau2test z
```

**Meta-regression**

<table>
<thead>
<tr>
<th>Method of moments estimate of between-study variance</th>
<th>Number of obs = 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>tau2</td>
<td>0.07472</td>
</tr>
<tr>
<td>* residual variation due to heterogeneity</td>
<td>I-squared_res = 64.91%</td>
</tr>
<tr>
<td>Proportion of between-study variance explained</td>
<td>Adj R-squared = 75.89%</td>
</tr>
<tr>
<td>Joint test for all covariates</td>
<td>Model chi2(2) = 17.90</td>
</tr>
<tr>
<td>Without Knapp-Hartung modification</td>
<td>Prob &gt; chi2 = 0.0001</td>
</tr>
</tbody>
</table>

| logrr | Coef. | Std. Err. | z     | P>|z<| | 95% Conf. Interval |
|-------|-------|-----------|-------|------|-------------------|
| latc  | -0.0015607 | 0.0074422 | -4.23 | 0.000 | -0.0461963 | -0.002604 |
| latc2 | -0.0007904 | 0.0007017 | -1.12 | 0.260 | -0.0026065 | 0.000094 |
| cons  | -0.5500882 | 0.1862457 | -2.95 | 0.003 | -0.9151161 | -0.185046 |

Test for residual between-study variance (of tau2 = 0)
\[ Q_{res} (10 df) = 20.50 \]
\[ Prob > Q_{res} = 0.0015 \]

---

### Figure 280 | Metareg | Intercept, Year-C, Year-C2 | Z | Method of moments
REFERENCES


Rothman, K.J. (1990). No Adjustments are Needed for Multiple Comparisons, *Epidemiology*, 1, 43-46.