WORKSHOP

Meta-Analysis for the Synthesis of Evidence in Agriculture

Instructor: Larry Madden

Assistant: Felipe Dalla Lana

THE OHIO STATE UNIVERSITY
COLLEGE OF FOOD, AGRICULTURAL, AND ENVIRONMENTAL SCIENCES

Genesis of Meta-Analysis

- The psychotherapy debate (1952-1977)
- Glass (1976); Smith & Glass (1977)
  - "**META-ANALYSIS**"
- Rosenthal; Rosenthal & Rubin (1978)
- Schmidt & Hunter (1977)
- Precursors:
  - Pearson (1904): correlations
  - Fisher (1932): P values
  - Yates & Cochran (1938, …): Agricultural experiments
- Medical research (1980s-): heart disease, cancer, etc. – ubiquitous since the 1990s
  - *"It is obvious that the new scientific discipline of meta-analysis is here to stay"* -- Chalmers & Lau (1993)

Social sciences: psychology, education, Employment testing, personnel evaluation, etc,

Meta-Analysis

- "The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" -- Glass (1976)
- "Averaging results across studies" -- Hunter & Schmidt (2004)
- "…the combination of results from multiple independent studies" -- Sutton & Higgins (2008)
- "[combination of the] results of previous research in order to arrive at summary conclusions to resolve uncertainty about the underlying medical question"-- Mittlbock & Heinzl (2006)
- Our definition (applicable in general):
  - ***Analysis of results from multiple independent studies***
- Note: the notion of "*independent* studies" is debatable
  - Newer studies are often conducted based on the outcome of, and experiences from, older studies
  - However, Higgins et al. (J. Roy. Stat. Soc. A [2009]) argue that it is reasonable to assume that the study *effects* are independent

Meta-Analysis

- Basic concept:
  - Science is a cumulative process, where individual studies contribute to the overall total knowledge base
  - Individual studies contribute something, but it is the *collection* of results from many sources that matter in moving science forward
  - ***"…a single study will not resolve a major issue. Indeed, a small sample study will not even resolve a minor issue. Thus, the foundation of science is the culmination of knowledge from the results of many studies."*** -- Hunter & Schmidt (2004)

- Meta-analysis has always been controversial
  - "an exercise in mega-silliness" -- Eysenck (1978)
  - The problem of '*garbage-in, garbage-out*' : empirical data in certain studies may be untrustworthy
  - The problem of *'mixing apples and oranges'* : studies may differ too much from each other (methodology, treatments, measured responses, etc.), making synthesis problematic
  - Publication bias: only the '*good* ' results get published

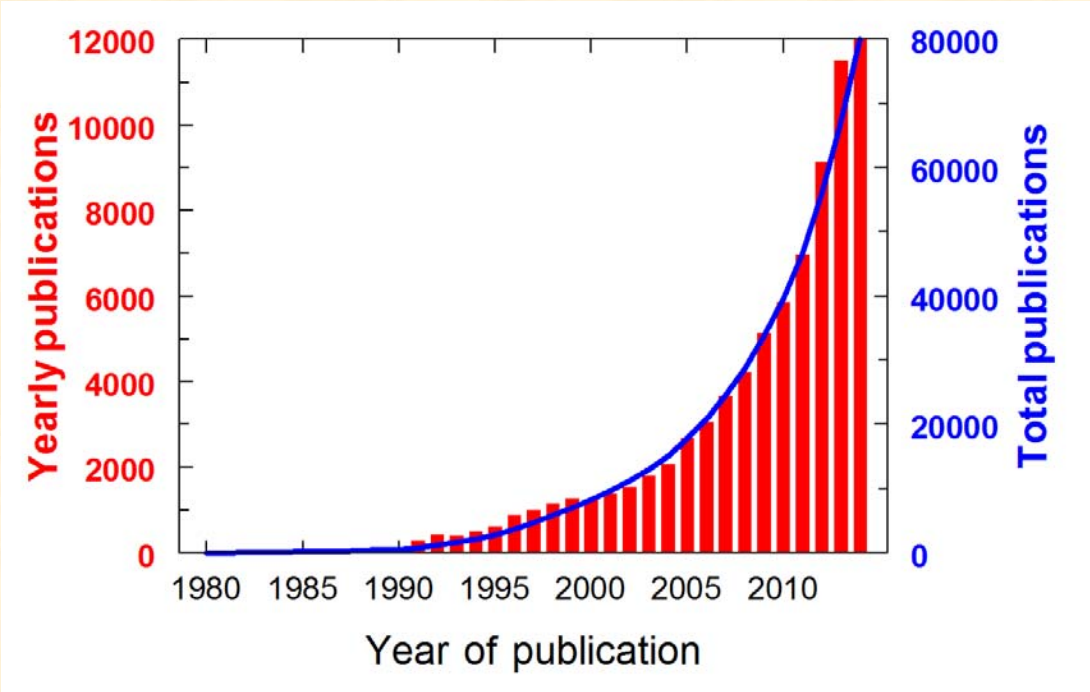
Meta-Analysis

- Controversies (continued)
  - Concerns about '*garbage-in, garbage-out'* and *'mixing apples and oranges'* can be (mostly) nullified by ***following strict criteria for the selection of studies included in the meta-analysis***
    - Textbooks often commit major space to these issues
  - There are many proposed methods to address publication bias (see optional slides in part II of presentation)

- Quantitative research synthesis (study selection):
  - Is the study replicated, with randomization, and sufficient observations?
  - Use only published studies?
  - Appropriate experimental units (e.g., plot size, soil type, tillage) and methods for treatment application (timing, formulation) and data collection?
  - Appropriate response variable? (continuous vs. ordinal vs discrete)
  - Appropriate analysis in primary study? Reported measure of variability in the study? (standard error, variance?)
  - Many other factors….
- See chapters 4-7 in: *The Handbook of Research Synthesis and Meta-Analysis, 2nd edition*. H. Cooper, L.V. Hedges, and J.C. Valentine, editors. Sage Foundation, NY.

Meta-Analysis

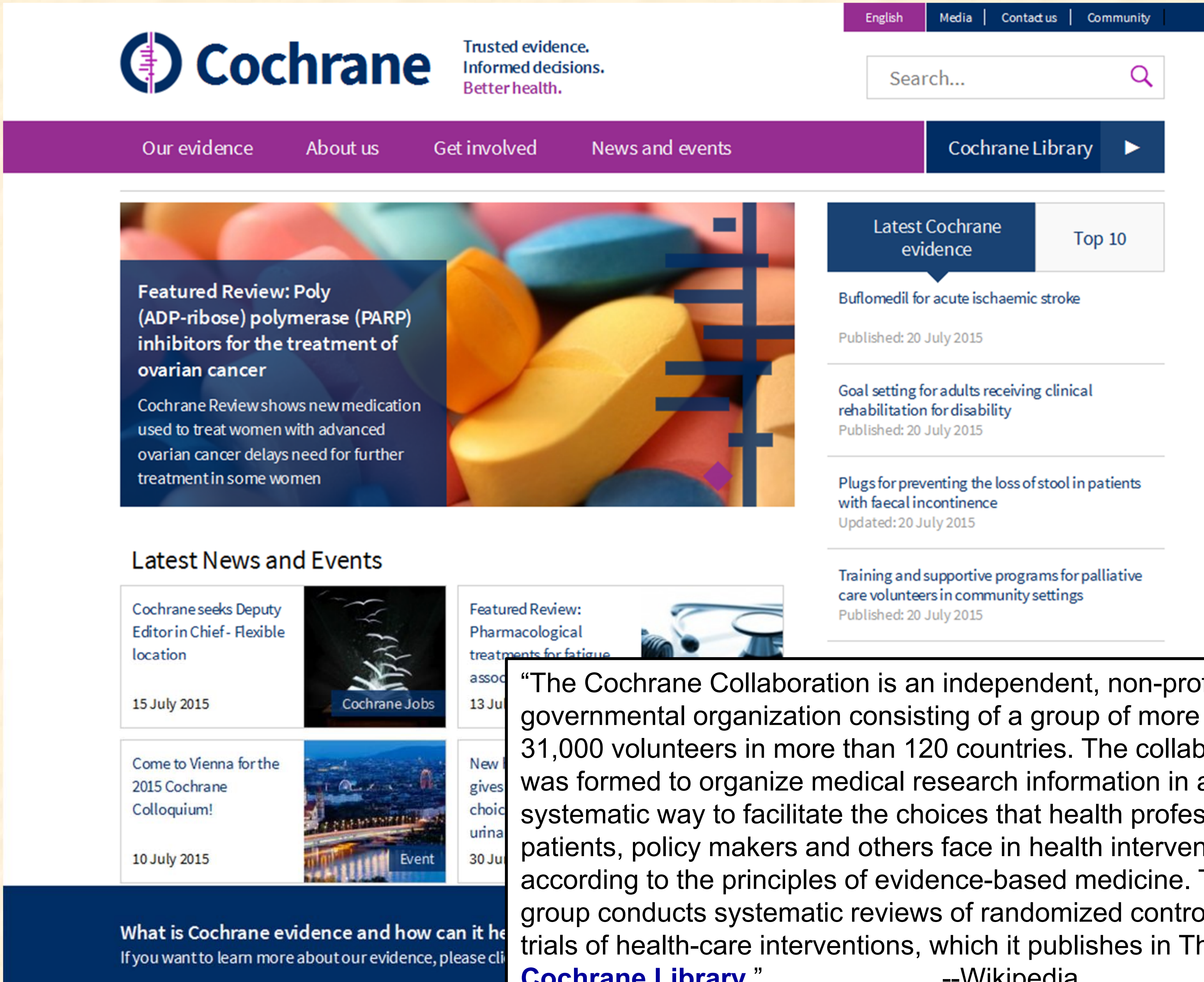
- Despite concerns, **meta-analysis has become the standard for evidence synthesis in many disciplines**
- There has been a tremendous growth in the number of papers using the method



Madden (Ohio State Univ.)

Study *results* vs. individual observations

- As reflected in the definitions, meta-analysis is traditionally thought to be based only on the ***summary results*** from each study, and not on the original observations within each study
  - Typically, the original observations (raw data; replications) are not available at the time of the meta-analysis
- However, it is now becoming more common to conduct a meta-analysis on the original observations from the studies (when data are available)
  - Known as Individual Patient Data (**IPD**) meta-analysis or Individual Participant Data (**IPD**) meta-analysis
  - Analysis with IPD can then be (almost) equivalent to multi-trial analyses in agriculture and medicine
    - **Multi-location, multi-location-year, multi-environment variety trials**
    - **Multi-center medical trials**
    - **GxE**
- In some disciplines (i.e., medicine), researchers assume that their primary studies will become part of a meta-analysis, so they now frequently make their observations to the wider community
  - See **Cochrane Collaboration**



"The Cochrane Collaboration is an independent, non-profit, non-governmental organization consisting of a group of more than 31,000 volunteers in more than 120 countries. The collaboration was formed to organize medical research information in a systematic way to facilitate the choices that health professionals, patients, policy makers and others face in health interventions according to the principles of evidence-based medicine. The group conducts systematic reviews of randomized controlled trials of health-care interventions, which it publishes in The **Cochrane Library**." --Wikipedia

Study *results* vs. individual observations

- Sometimes original observations are available from only some studies
  - One can just use the *results* from the individual studies (ignoring additional information in the original observations) – most common approach
    - Piepho et al. (Biom. J. [2012]) has shown that **one can recover most information** from the two-stage approach (where the analysis of the summary results is stage two)
  - Alternatively, Riley (Stat. Med. [2009]) and others have developed methods for combining original observations from some studies with results from other studies in a single simultaneous analysis (for continuous data)
    - May not be worth the effort
- IPD meta-analysis is most beneficial when
  - Individual studies are small, which means that the summary results are imprecise (especially the standard errors)
  - When one is focusing on sub-groups (individuals) within studies, and the variables (covariates) associated with the individuals
- IPD meta-analysis cannot be easily done when the experimental and treatment designs vary among the studies (a common occurrence)
- Even with the availability of the original (primary) observations, meta-analysis may still be most practical based on the summary results

Meta-analysis workshop : Outline

- Basic concepts (with case studies)
  - A little history and the goals of meta-analysis
  - The concept of **effect size** (explained through the case studies)
  - Graphical appraisal of the effect sizes
  - Models for fixed- vs. random-effect meta-analysis
  - Parameter estimation, and interpretation
  - Heterogeneity in effect sizes among studies - interpretation
  - Confidence intervals, prediction intervals
- Introduction to some key topics
  - **Moderator variables in a meta-analysis**
  - **Multiple-treatment meta-analysis**
- Not covered (but program code and slides are given):
  - More on graphical appraisals
  - Probability of effect size in future new study
  - Power of meta-analysis
  - Fallacy of counting *P* values!
  - Publication bias, and how to assess

Analyses demonstrated using SAS (macros and procedures), although use of an R package is summarized at end of PowerPoint file

An illustration: an individual study

An investigation of the effect of treatment T on severity of crop disease. Example:

- 2+ treatments or factor levels (T, C [=control], …)
- n replications of each treatment
- Response variable: y (e.g., disease severity)
- Conduct appropriate analysis for this study and estimate the **Effect Size** of interest:
  - **Estimated parameter, or function of estimated parameters, from an individual study.** Examples:
  - Difference in mean disease for T and C

$$D = \hat{\mu}_C - \hat{\mu}_T$$

  - Or, '*percent control*', C (relative reduction in disease compared to the control; a ratio)

$$C = 100(\hat{\mu}_C - \hat{\mu}_T)/\hat{\mu}_C = 100(1 - \hat{\mu}_T/\hat{\mu}_C)$$

  - Or, transformation of the above for statistical reasons (e.g., log-response ratio): $L = \ln(\hat{\mu}_T/\hat{\mu}_C)$

L is especially useful when the mean in the control could be small or large --e.g., D=3 is large when the control mean is 5 (**C = 100·3/5 = 60%**), but small when the control mean is 50 (**C = 100·3/50 = 6%**)

An illustration, *continued*

- Use z as a generic symbol for the *estimated* effect size (D, L, $\ldots$)
  - z is an estimate of a parameter υ ([true] expected effect size)
    - Sometimes simply called the '**true effect size**'
- Record the estimated effect size (z) of interest (e.g., difference of two treatment means), <u>and</u> also the estimated variance of z (label this s^2; known as the **sampling or within-study variance**)
  - For the subsequent analysis, s^2 is considered <u>known</u> and <u>fixed</u> (obviously, unrealistic, but standard)
- Meta-analysts often obtain the estimated effect sizes from published articles and other reports
  - z is easy to obtain, but s^2 is often not reported, or a measure of variation is reported that is related to s^2
  - A great deal of effort usually goes into determining s^2 from the available information
    - Multiple chapters in meta-analysis books deal with this issue
    - In a sense, working backwards from the reported statistics
  - Imputation may also be needed when no information is given on variability

Determining sampling variance

- Suppose that the estimated effect size is the difference of two means (i.e., $z = D$), then s^2 is the square of the *standard error of the difference* of means ($s^2 = \mathrm{SE}(D)^2$)
  - Let V be the residual variance (mean square error) from an ANOVA, and n represent the number of replicates of each treatment. With independent treatments,

$$s^2 = \mathrm{SE}(D)^2 = 2 \cdot V/n$$

  - Use alternative formula when sample sizes are not equal (n_T and n_C)

$$s^2 = V \cdot (1/n_T + 1/n_C)$$

  - Use alternative formula for variance heterogeneity
  - Alternative formula for correlated means
- It is very common to present Fisher's least significant difference (LSD) in some disciplines (e.g., agricultural sciences)
  - If two means are greater than LSD apart, then they are declared significantly different
  - If $t_{1\text{-}0.05/2,df}$ is the critical value for a Student *t* distribution at the 5% (α) significance level (within a single study), then $\mathrm{LSD} = t_{1\text{-}0.05/2,df} \cdot \mathrm{SE}(D)$
  - So,

$$s^2 = (\mathrm{LSD}/t_{1\text{-}0.05/2,df})^2 \approx (\mathrm{LSD}/2)^2 \text{, if } df \text{ is large (and } \alpha{=}0.05)$$

Determining sampling variance, *continued*

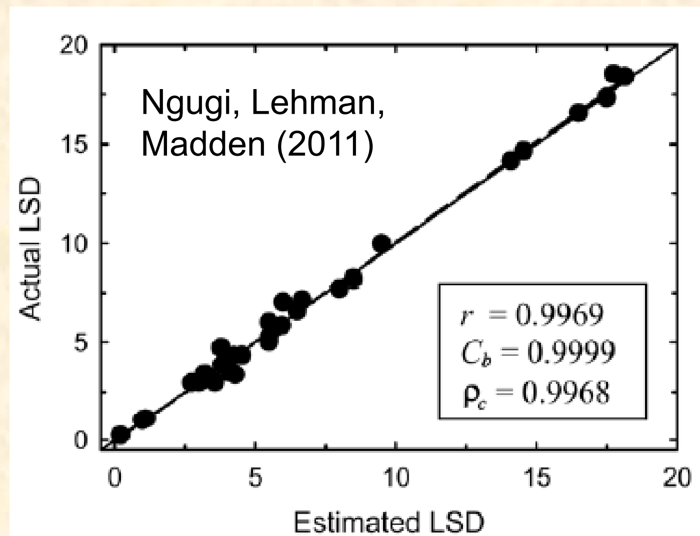
- Suppose one only has multiple comparison "line" ("letter") display of means, and the effect size was the difference between the first and last treatment mean

Treat	Mean
A	20 a
B	16 ab
C	15 b
D	10 c
E	7 c

The LSD is
- greater than the largest *nonsignificant* difference between the means, and
- smaller than the smallest *significant* difference
- as approximation, LSD is half-way between these two values

- Nonsignificant differences (*find largest*):
  20-16 = 4, 16-15 = 1, 10-7 = 3
- Significant differences (*find smallest*):
  20-15 = 5, 20-10=10, 20-7 = 13,
  16-10 = 6, 16-7 = 9, 15-10 = 5, 15-7 = 8

- LSD ≈ (4+5)/2 = 4.5
  - $s^2 \approx (\text{LSD}/2)^2 = (4.5/2)^2 = 5.06$



Meta-analysis

- There are many possible effect sizes

- For a meta-analysis, each study must contribute a <u>pair</u> of statistics, (z, s^2)

- From single to multiple studies:
  - Suppose there are K studies
  - Label the individual studies with i ($i = 1, \ldots, K$)
  - The pair (z_i, s_i^2) becomes a "*data point*" for a meta-analysis, and the unknown (true) expected effect size (a parameter) for study i is υ_i
    - υ_i is often called the '*true effect size'* for study i

- Usually assume that z_i has a normal distribution (original observations may have many different distributions)
  - This distributional assumption can be relaxed

Some effect sizes: Continuous data

- Mean (completely randomized) $\hat{\mu}_i$ $s_i^2 = V_i / n_i$
- Mean (RCBD) $\hat{\mu}_i$ $s_i^2 = (V_{b\,i} + V_i)/n_i$

V_i: i-th residual variance
$V_{b\,i}$: i-th block variance

- Difference in means (D_i)
  (completely randomized or RCBD [$V_{b\,i}$ cancels])
  $D_i = \hat{\mu}_{Ci} - \hat{\mu}_{Ti}$ $s_i^2 = 2V_i / n_i$
- Log ratio (L_i), or percent control…
  - Valuable when *relative* changes matter
  - *May* be useful when the response variable is not (quite) the same for all studies (different scales)

  $L_i = \ln(\hat{\mu}_{Ti} / \hat{\mu}_{Ci})$ $s_i^2 = \frac{V_i}{n_i}\left(\frac{1}{\hat{\mu}_{Ci}^2} + \frac{1}{\hat{\mu}_{Ti}^2}\right)$
- Standardized mean difference (d_i)
  - Attributed to Cohen, Hedges
  - Very common in social sciences
  - Advocated when the response variable differs among studies

  $d_i = \frac{(\hat{\mu}_{Ci} - \hat{\mu}_{Ti})}{S_i}$ $s_i^2 \approx \frac{2}{n_i} + \frac{d_i^2}{4n_i} \approx 2/n_i$

$S_i = \sqrt{V_i}$, within-study standard deviation

Corrections used to reduce the bias in d_i (and s_i^2)

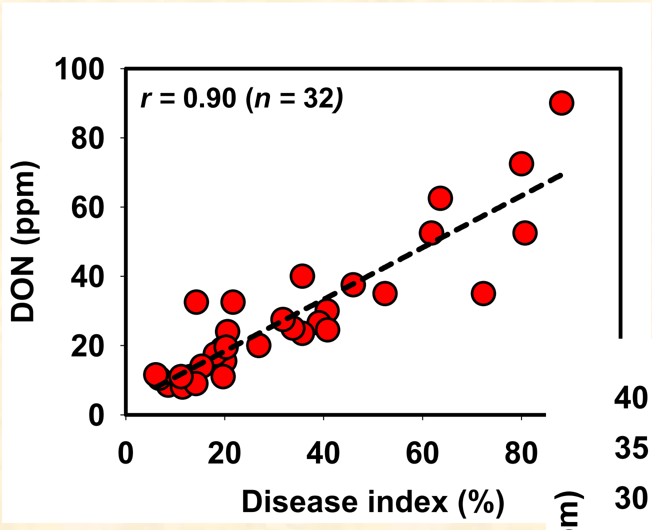
n_i: number of replicates in each group for study i, with generalizations for unequal sample size

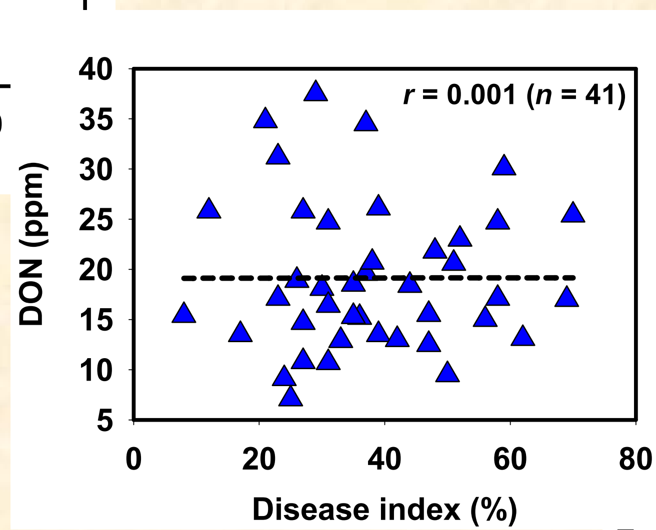
Effect sizes for relationships or associations

Correlation coefficient (r_i) or the Fisher transformation (Z_{ri}).

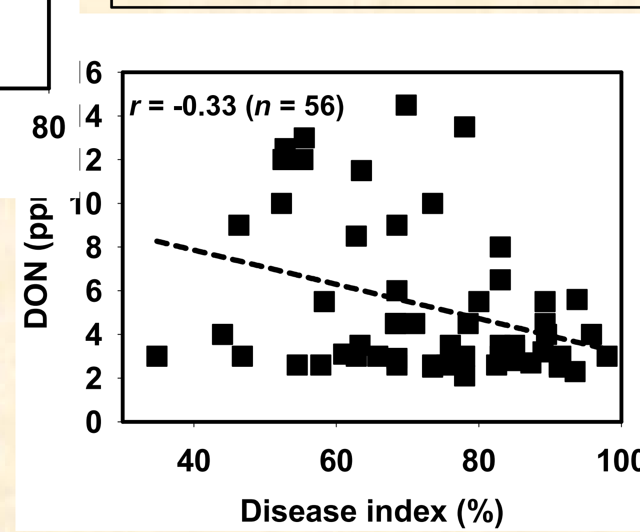
Sampling variance of Z_{ri} is s_i^2 = 1/(n_i-3)

$$Z_{r_i} = \frac{1}{2}\log\left(\frac{1+r_i}{1-r_i}\right)$$





Three individual studies



Slope (b_i) and/or intercept (a_i) of model fitted to the data for each study (i = 1, …, K).
Sampling variance (s_i^2): square of estimated standard error of slope or intercept

Effect Sizes: Discrete (binary) data and survival data (not covered)

- Many possible effect sizes, such as:
  - Proportion, or its transformation
  - Odds or **log odds**
  - Difference of proportions (risk difference)
  - Relative risk (ratio of proportions) or log of relative risk
  - Odds ratio or **log odds ratio**

  - Hazard ratio or **log hazard ratio**
- Analysis proceeds in the same manner as with continuous data, with the usual assumption that the estimated effect sizes (not the original observations) have a normal distribution
- If individual participant data (IPD) are available, can use generalized linear mixed models (GLMMs), with binomial, Poisson, or other conditional (within-study) distributions

Case study 1:
Reduction in toxin concentration in harvested wheat grain

- Wheat (one of the most economically important crops in the world) is often affected by the disease known as Fusarium head blight (FHB)
  - A mammalian toxin—DON (deoxynivalenol)—is often produced in infected wheat seeds (grain)
  - One control practice is to treat the wheat with a fungicide (pesticide) in the field at a particular date
- Studies on disease control were conducted for more than a decade at multiple U.S. locations, using standardized experimental protocols
- Here, we use results for the effect of the fungicide Folicur (=tebuconazole) on DON toxin concentration (ppm) in harvested grain
- The studies are analogous to **clinical trials** in medicine, but typically only the results are available

Case study 1: Reduction in toxin concentration in harvested wheat grain

- Details:
  - Each study consisted of 4-6 replicates
  - Two or more treatments in each study (only two treatments used in the meta-analysis shown here)
    - T: Folicur (applied at a single wheat growth stage [Feekes 10.5.1])
    - C: Check (control; “placebo”)
  - μ_T , μ_C : mean DON toxin concentration (ppm) in wheat grain
  - Data from individual studies were analyzed with ANOVA or mixed models
    - Among other things, estimate treatment means and standard errors, and residual variance (V)
- There were K = 101 studies in the meta-analysis
- Primary interest in percent control (C): $100 \cdot (1 - \mu_T/\mu_C)$
  - So, log response ratio used as the estimated effect size (z_i; $i = 1, \ldots 101$)

$$z_i = L_i = \ln(\hat{\mu}_{Ti} / \hat{\mu}_{Ci}) \qquad s_i^{\,2} = \frac{V_i}{n_i}\left(\frac{1}{\hat{\mu}_{Ci}^2} + \frac{1}{\hat{\mu}_{Ti}^2}\right)$$

Paul et al. *Phytopathology* 97: 211-220 [2007]; Madden & Paul Phytopathology 101: 16-30 [2011].

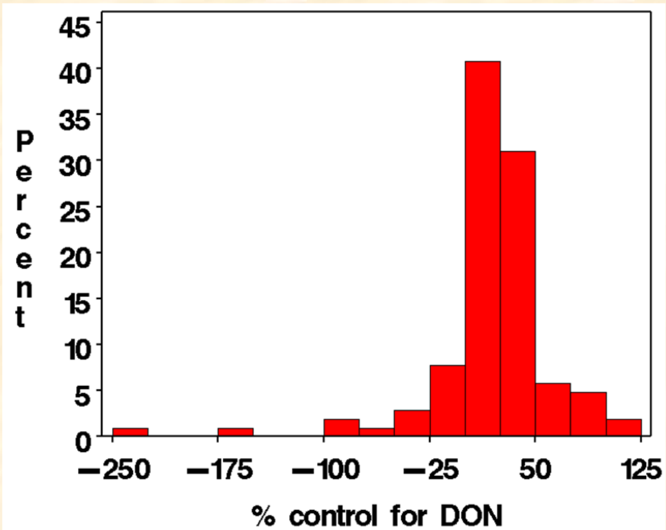
Case study 1:
The meta-analytical data set

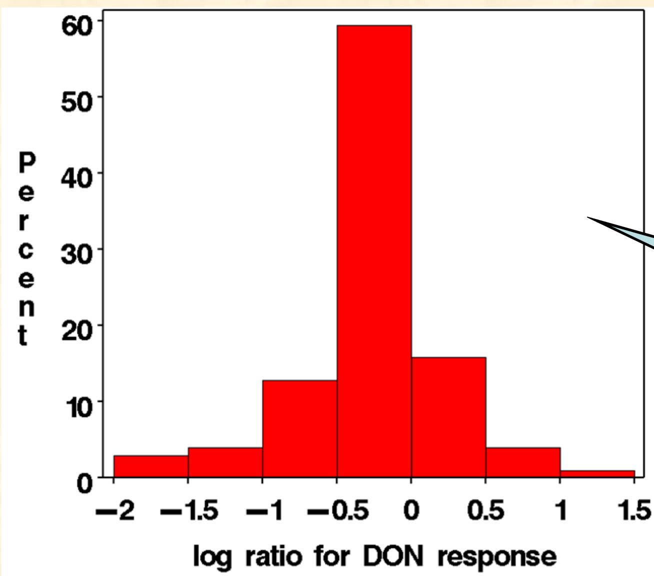
i

z_i

Study	$\hat{\mu}_{C\,i}$	$\hat{\mu}_{T\,i}$	C_i	$\ln(\hat{\mu}_{Ti} / \hat{\mu}_{Ci})$	s_i^2
1	10.3	4.8	53.4	-0.764	0.029
2	8.0	4.4	45.0	-0.598	0.017
3	3.9	3.8	2.8	-0.029	0.011
4	5.3	2.7	49.1	-0.674	0.036
5	8.2	7.1	13.4	-0.144	0.019
...					

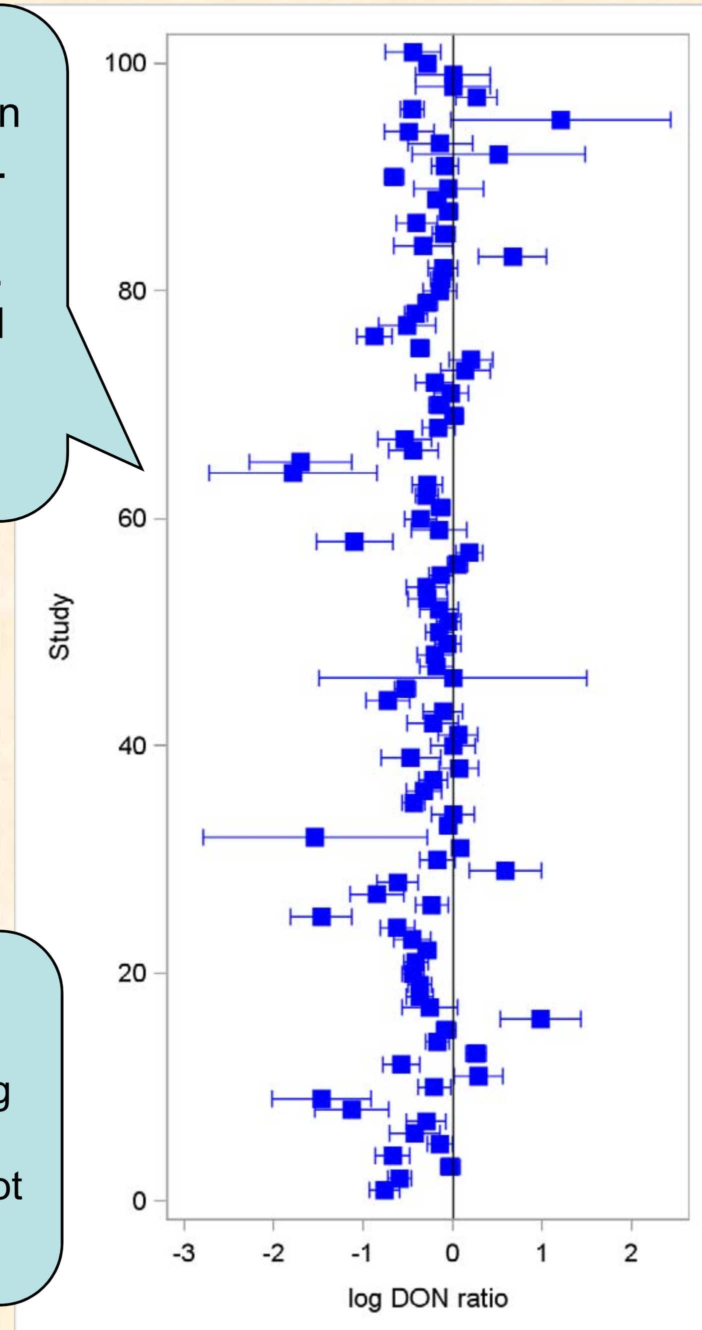
K=101 studies:
Each study becomes an
“observation” in the new dataset





"Forest Plot" (very common in meta-analysis).

Study ID vs. z_i. Show standard errors or confidence intervals.



Histogram of z_i. Because of unequal "sampling variances", histograms may not be overly useful

Meta-analytical random-effects model

Estimated effect size for study i (often assume normal)

Expected effect size, across studies

Random effect of study i on the effect size (an among-study effect)

$$z_i = \zeta + u_i + \varepsilon_i$$

Residual or within-study *"sampling variation"* term

Distributional assumptions:

$$u_i \sim N(0, \sigma^2)$$

σ^2 *:* among-study variance (many authors use τ^2)

$$\varepsilon_i \sim N(0, s_i^2)$$

s_i^2: sampling variance (separate for each study; **treat as a known parameter for each study**). Assume u and ε are independent.

$$z_i \sim N(\zeta, \sigma^2 + s_i^2)$$

One estimates ζ and σ^2

Meta-analytical random-effects model

$$z_i = \zeta + u_i + \varepsilon_i \qquad u_i \sim N(0, \sigma^2)$$

$$\varepsilon_i \sim N(0, s_i^2)$$

Equivalent model formulation (emphasizing the hierarchy):

$$z_i = \upsilon_i + \varepsilon_i$$

$$\upsilon_i = \zeta + u_i$$

υ_i is the '*true effect size*' for study i, assumed to vary among studies

$$z_i \mid \upsilon_i \sim N(\upsilon_i, s_i^2),\ \upsilon_i \sim N(\zeta, \sigma^2)$$

$$z_i \sim N(\zeta, \sigma^2 + s_i^2)$$

Generalizations:

One can relax distributional assumptions (although most do not consider the latter).

Meta-analysis models

- **Random-effects model** (explicit consideration of among-study variability in effect size)

$$z_i = \zeta + u_i + \varepsilon_i$$

  - $\sigma^2 \geq 0$

  A positive σ^2 indicates *heterogeneity in (true) effect sizes*

$$u_i \sim N(0, \sigma^2)$$

$$\varepsilon_i \sim N(0, s_i^2)$$

- **Fixed-effects model** (assume that there is no random variation in the effect size per study) – "old-fashioned" approach

$$z_i = \zeta + \varepsilon_i$$

$$\varepsilon_i \sim N(0, s_i^2)$$

  - i.e., $u_i = 0$ (i=1,…,K), or $\sigma^2 = 0$
  - In this case, think of ζ as a *common* (not expected) effect for all the studies
    - Some call this the *common-effect model*
    - Estimate only a single parameter
    - *Homogeneity in effect sizes*
  - Note that meta-analysts use 'fixed effects' differently from others

Meta-Analysis: Model Fitting

- Parameter estimation for ζ and σ^2 (most common approaches):
  - **Method of moments** (the classic meta-analytical approach, but *may* not be the most general)
    - Several approaches, but one method is most common ("**DL**")
  - **Maximum likelihood (ML)** and restricted (residual) maximum likelihood (**REML**)
    - Iterative and more computer-intensive
  - **Bayesian analysis**
- *In general, an investigator uses <u>one</u> estimation method (but we demonstrate several here, for instructional purposes)*
- Meta-analysis: a method of obtaining weighted averages of estimated effect sizes

$$\hat{\zeta} = \frac{\sum w_i z_i}{\sum w_i} \;, \quad w_i = \frac{1}{\sigma^2 + s_i^2} \,, \qquad \mathrm{SE}(\hat{\zeta}) = \left(\sum w_i\right)^{-1/2}$$

One substitutes the estimate of σ^2 in the formulae

"Method of Moments"

- Meta-analysts use the term **"Method of moments"** in a specific way
- One first fits the fixed-effects model (no u_i term [σ^2=0], which means that $w_i = s_i^{-2}$)
- Calculate Cochran's Q statistic, which has a chi-squared distribution with K-1 df when $\sigma^2 = 0$

$$Q = \sum s_i^{-2}\left(z_i - \hat{\zeta}^{(FIX)}\right)^2$$

- Many use Q to formally test for non-zero among-study variability, but the test has very low power
- But one can equate Q to its expected value for nonzero σ^2, and solve for σ^2. This is known as the **DerSimonian and Laird (1986) method (DL)**; a non-iterative approach.
  - With the estimate of σ^2, estimate w_i and then estimate the random-effects expected value, ζ
  - Probably still the most common meta-analytical method performed (although it is not implemented in standard general-purpose statistical software)

$$\hat{\sigma}^2 = \frac{Q-(K-1)}{c}$$

$$c = \sum s_i^{-2} - \frac{\sum s_i^{-4}}{\sum s_i^{-2}}$$

$$\hat{\zeta} = \frac{\sum \hat{w}_i z_i}{\sum \hat{w}_i},$$

$$\hat{w}_i = \frac{1}{\hat{\sigma}^2 + s_i^2}$$

$$\mathrm{S\hat{E}}(\hat{\zeta}) = \left(\sum \hat{w}_i\right)^{-1/2}$$

"Method of Moments"

- The DerSimonian and Laird (DL) approach is the basis for several specialized computer programs or specialized macros and packages
  - e.g., **COMPREHENSIVE META-ANALYSIS (CMA)** (Biostat, Inc.)
    - Programs may be quite expensive, but they often are window-driven and customized for this type of analysis, providing many of the specialized features that meta-analysts expect (especially for graphs)—ideal for non-statisticians
  - Macros have been written for commercial software (e.g., in `SAS` and `STATA`)
  - `R` packages and functions are also available (e.g., `metafor`); these will have a steeper learning curve for the non-statistician, but are very powerful
- Some modern textbooks in meta-analysis are primarily based on the DL approach (e.g., Borenstein et al. [2009], an excellent introductory text).
- There are several variations of this method of moments not discussed here
- Several advantages, including:
  - Method does not explicitly depend on normality
  - May be much faster than iterative approaches for large and complex data sets
  - Performs reasonably well in terms of confidence interval coverage, efficiency, bias, power, for the expected effect size (may perform less well for the among-study variance)

ML and REML

- It is straightforward to fit the random-effects meta-analytical model using maximum likelihood (**ML**) or restricted (residual) maximum likelihood (**REML**)
  - In fact, DerSimonian and Laird (1986) also proposed both of these approaches in addition to their namesake "method of moments"
  - Because likelihood-based mixed-model software was not generally available in the 1980s, the DL moment-method became entrenched with meta-analysts
  - The simplicity argument for the moment method is much less compelling today, given the speed of personal computers and the sophisticated general-purpose software available for ML and REML estimation of mixed models
- Arguments in favor of ML (or REML) include:
  - Many good statistical properties of the parameter estimates, and the direct ability to formally compare nested models
  - Calculation of EBLUPs for random study effects
  - ***Generality* of the approach for a wide range of possible random and mixed-effects models, including many expansions of the model presented here**
  - Availability of commercial (`SAS, STATA`) and free (`R`) software for fitting models
- The iterative ML and REML methodology is standard (not described here)
  - **Care is required in using mixed-model software because of the fixed sampling variances (s_i^2; a known parameter [constant] for each study); "tricks" may be needed to prevent the estimation of a residual variance**

Case Study 1: ML Estimation (K =101)

Typical statistics from a meta-analysis:

$\hat{\zeta}$	SE($\hat{\zeta}$)	95% CI for $\hat{\zeta}$	t = $\hat{\zeta}$ / SE($\hat{\zeta}$)	p value	Control % (C)	95% CI for C
–0.24	**0.028**	**–0.30 - –0.19**	**–8.85**	**<0.001**	**21.6%**	**17.2% - 25.8%**

Median Percent Control: $\hat{C} = 100 \cdot (1 - \exp(\hat{\zeta}))$

H_0: $\zeta = 0$ (i.e., expected log response ratio (L) = 0)
H_a: $\zeta \neq 0$

- Significance determined with Wald statistic (with an assumed Student t distribution under H_0)
- The choice of degrees of freedom (df) is not resolved (not relevant in this example with large K)
- Confidence intervals based on assumed distribution (normal or t) of estimated ζ

$$t = \frac{\hat{\zeta} - \zeta}{S\hat{E}(\hat{\zeta})}$$

Estimation methods

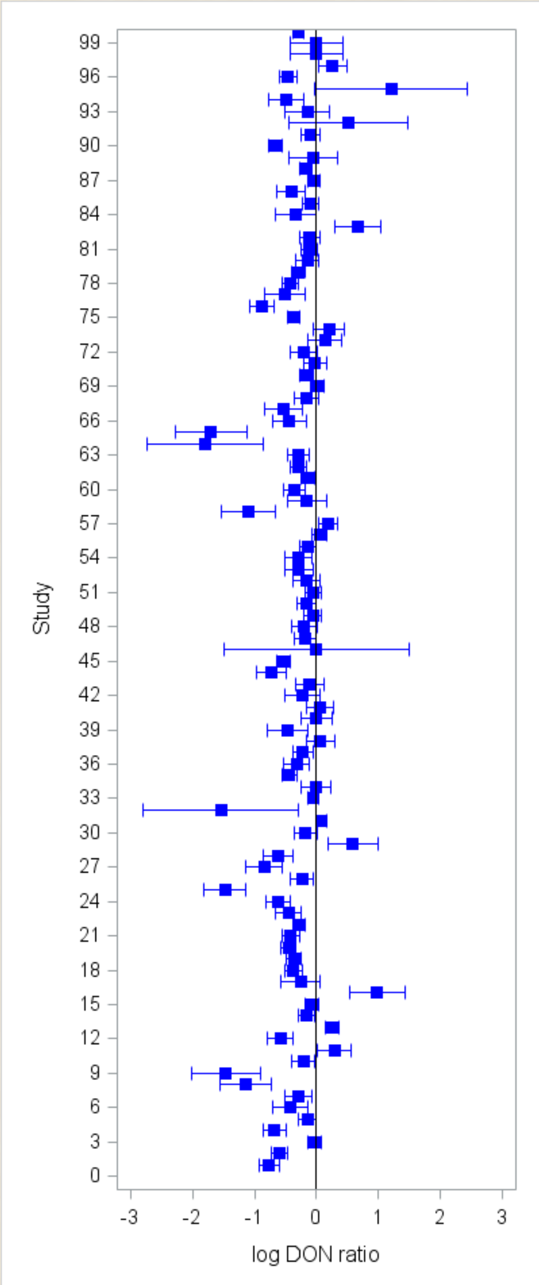
ML and REML give very similar results here because of large K

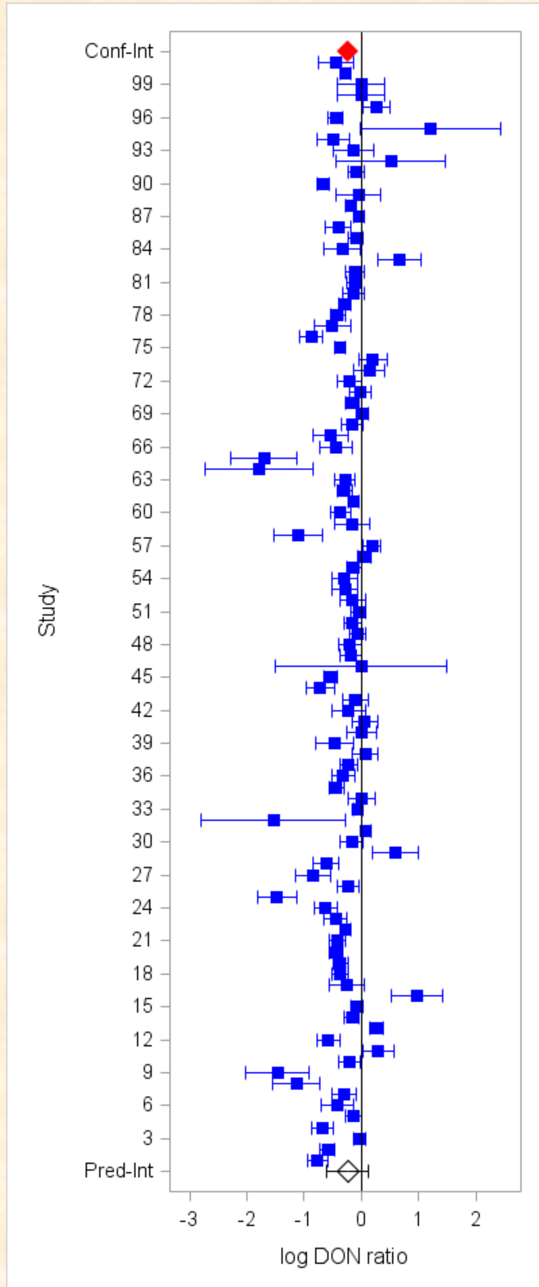
Moment Method of DerSimonian and Laird (DL)

Fixed-effect estimates (**assumes σ^2=0**). Common historicially; should <u>not</u> be used, in general.

Method	$\hat{\zeta}$ (SE)	Confidence Interval (95%)
ML	**-0.244 (.0276)**	**-0.299 – -0.189**
REML	**-0.244 (.0278)**	**-0.299 – -0.189**
Moment	**-0.245 (.0285)**	**-0.301 – -0.188**
Fixed	***-0.223 (.0163)***	***-0.255 – -0.192***
Bayesian	**-0.242 (.0281)**	**-0.298 – -0.184**

Noninformative priors were used. Bayesian results show the mean and standard deviation of the posterior distribution, and the 95% **credible interval** (equal tails version).
Bayesian approach accounts for the uncertainty of the variance parameters





Forest plot, with addition of estimated ζ, together with **confidence interval** (red diamond), and **prediction interval** (open black diamond)

One could also show the fixed-effects results ($\sigma^2 = 0$). Here, one cannot see any confidence interval (too narrow).

Diagnostics

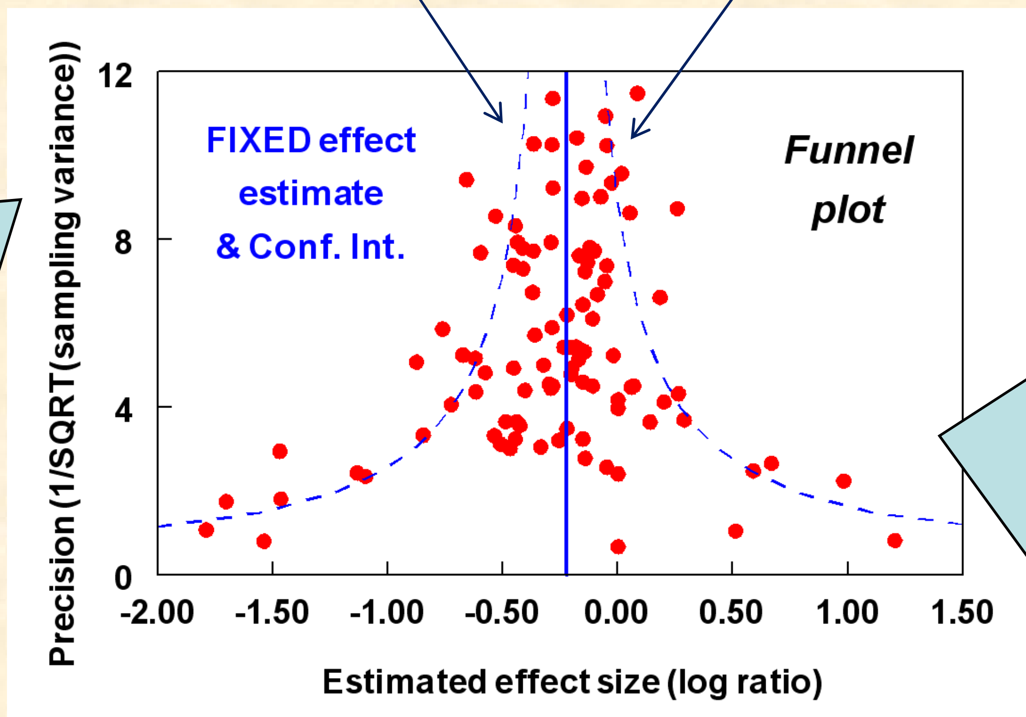
- Model assessment (criticism) in meta-analysis has unique issues
- The usual residual plots (residual versus predicted values) *may* not be of much value for the simple random effects model, because of the unequal sampling variances
  - The unique s_i^2 for each z_i makes interpretation difficult
  - For the simplest model, the "predicted" value is a single number (ζ) (thus, no range of the x-axis for a graph)
- Meta-analysts have developed some specialized graphs that are not typically seen in other applications
  - In addition to the Forest plot, so-called **funnel** and **radial** plots
    - These can help assess the need for a random-effects or a fixed-effects model, and explore the possibility of publication bias
- Moreover, versions of diagnostic plots from the broader field of mixed-model analysis have value (but are *much* less reported). *Not* covered here.
  - Studentized deleted residual versus study ID, PRESS statistics versus study ID,…
  - Cook's Distance for the fixed effect (ζ) and the variance (σ^2) versus study ID
    - Measures the influence of observations (studies) on parameter estimates
      - A scaled measure of the squared distance between parameter estimates based on the full dataset and the estimates when each observation (study) is deleted (with mixed models, the model is refitted with each observation deleted)

Special graphical views of effect sizes

- These graphs *can* simultaneously be used to determine: if a fixed-effects (common-effects) analysis is warranted; and if there is bias due to missing studies (*publication bias*)
  - Requires moderate-to-large K
- **Funnel plot** (Light and Pillemer 1984; Egger et al. 1997): Graph of "precision" $(1/s_i)$ vs. z_i, fixed-effects estimated effect size, and pseudo-confidence interval

$\hat{\zeta}^{(FIX)} - 2 \cdot s_i$ $\hat{\zeta}^{(FIX)} + 2 \cdot s_i$

If among-study variance is 0 (justifying fixed-effects), almost all points should be inside the dashed lines). Evidence here is for random effects



If not upside-down funnel, and not symmetrical, then selective reporting of results *may* be occurring. (No obvious bias here).

Sterne et al. (BMJ [2011]) questions tests of asymmetry and interpretations of funnel plots.

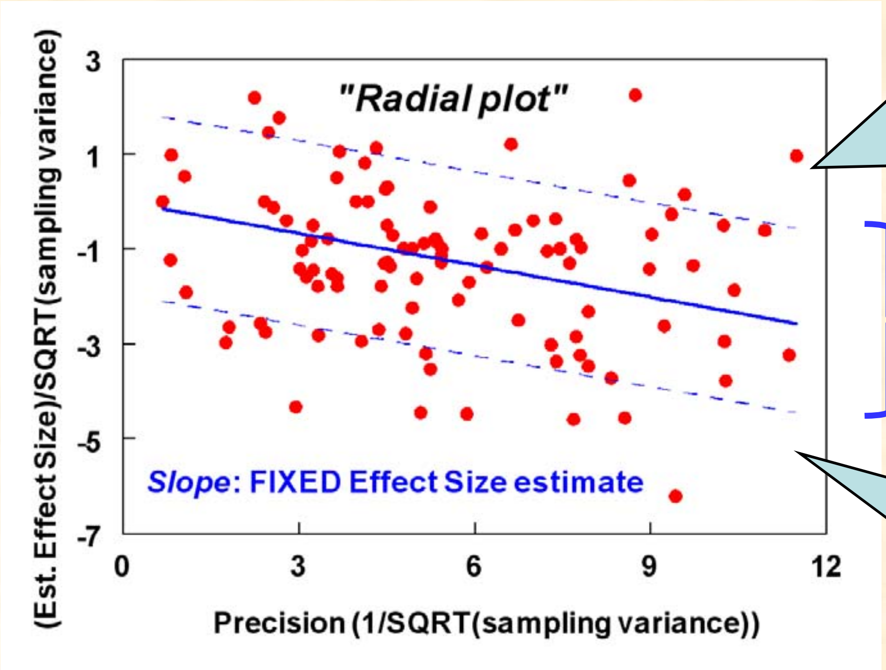
– In addition to **funnel plot**, a so-called "Radial plot" or "Galbraith plot" (1988) may be useful

  – **Radial plot:** Graph of "*standardized* estimated effect size" versus "precision" $(1/s_i)$

    • z_i / s_i vs. $1/s_i$

  • Slope of the zero-intercept (fixed effects) regression line is $\hat{\zeta}^{(FIX)}$ (when residual variance is fixed at 1)

If among-study variance is 0 (justifying fixed-effects), almost all points should be inside the dashed lines). Evidence here is for random effects

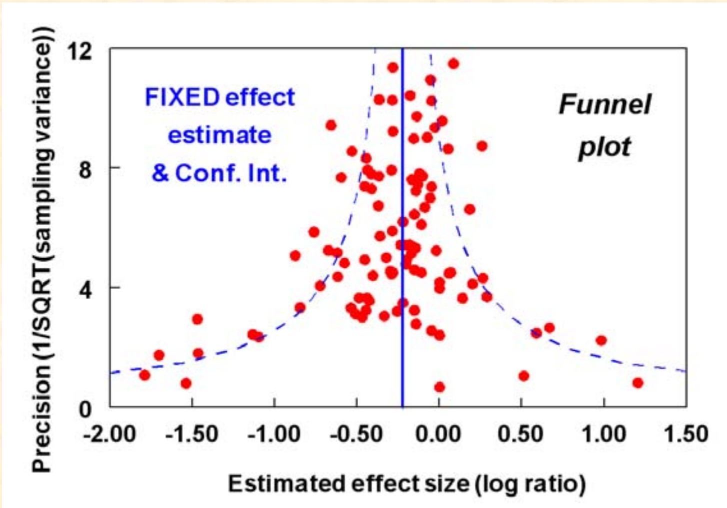


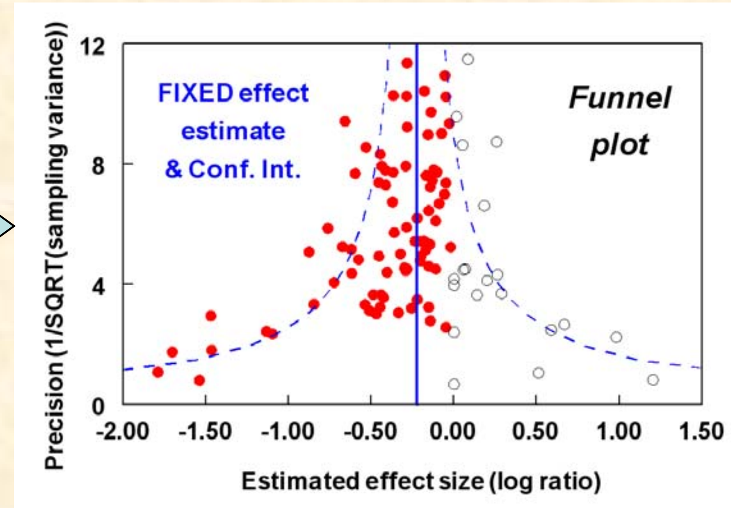
If no bias, there should be a random scatter around the line (no gaps at certain precisions or at certain effect sizes)

$$\frac{\hat{\zeta}^{(FIX)}}{s_i} \pm 2$$

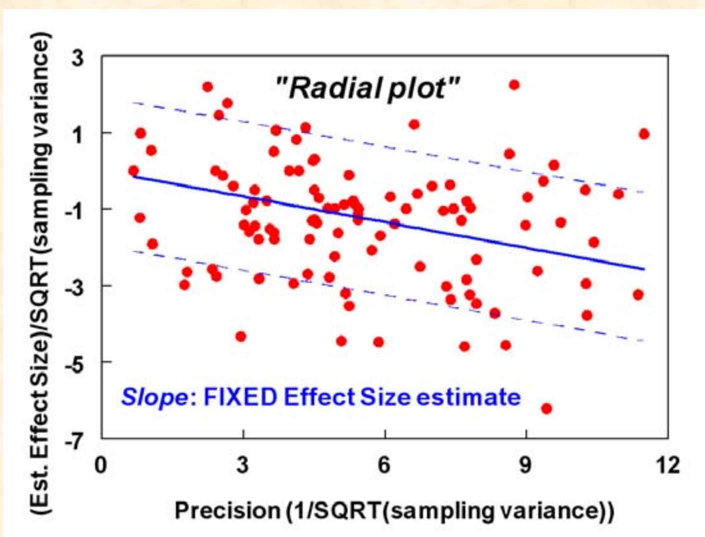
Use as a guide only (especially with small K) . I think the graph can be hard to interpret.

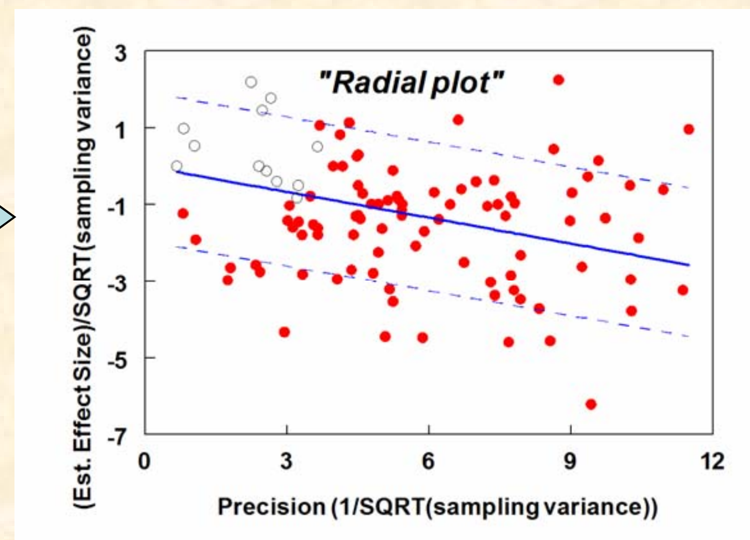
Publication bias: Plots *may* help





If no bias, there should be a random scatter around the line (no gaps at certain precisions or at certain effect sizes) – a (rough) guide only (especially with small K)





Case study 2: Slope of a linear relation for yield loss of maize (corn) in relation to disease

- One often wants to know the relationship between symptoms of a crop disease ("disease severity"—degree of infection) and the reduction in yield (yield or crop loss)
- Shah & Dillard (Plant Disease 90: 1413-1418 [2006]) described yield loss in sweet corn (y) in relation to the severity of rust disease at a single growth stage (x) in K = 20 studies
  - A zero-intercept linear regression model was used for each study (when disease is not present, there is no reduction in yield)
- **Effect size:** slope of the regression model (i.e., $z_i = b_i$, where b_i is the estimated slope for the i-th study ($i = 1, \ldots, 20$)

$$z_i = b_i = \zeta + u_i + \varepsilon_i \qquad u_i \sim N(0, \sigma^2),\ \varepsilon_i \sim N(0, s_i^2)$$

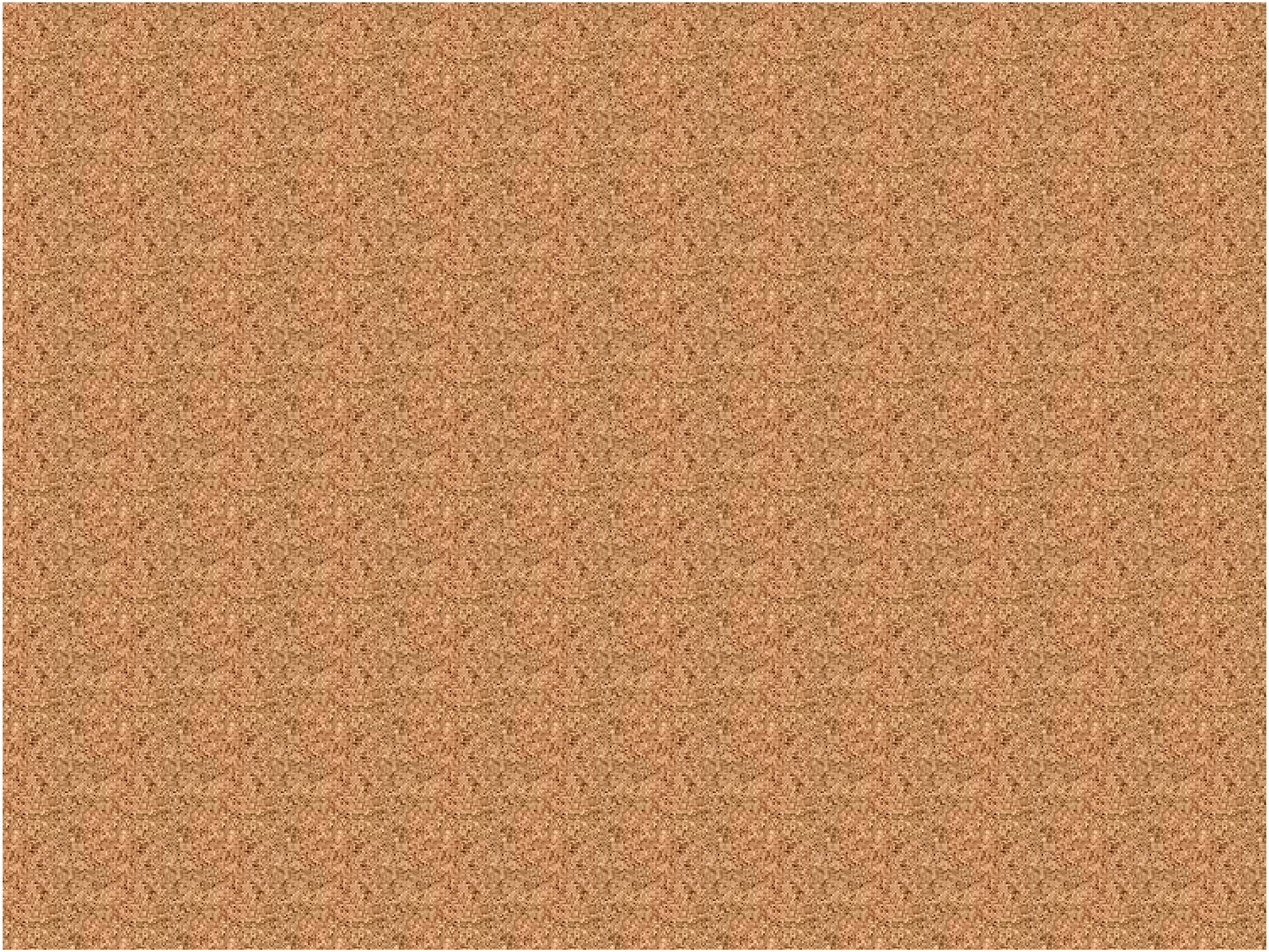
File: meta-analysis Shah slope example.sas

Obs	study	slope	SE	sampvar	wgt	State	Variety	Year	MeanD
1	1	2.13607	2.18587	4.77803	0.21	NY	Zenith	1999	0.39
2	2	0.47288	0.18572	0.03449	28.99	NY	Zenith	1998	3.05
3	8	1.17716	0.14263	0.02034	49.16	NY	Squeen	2000	6.58
4	11	0.33829	0.05394	0.00291	343.70	NY	Bold	2001	45.91
5	22a	0.78559	0.24789	0.06145	16.27	NY	Zenith	1997	8.10
6	22b	-0.50102	0.42602	0.18149	5.51	NY	Rival	1997	3.73
7	31	0.31888	0.01484	0.00022	4540.80	MI	HMX83865	1993	18.73
8	35	0.14559	0.12224	0.01494	66.92	MI	YBelle	1992	22.96
9	50	-0.79923	0.31152	0.09704	10.30	NY	Jubilee	1984	11.73
10	59	0.63267	0.02300	0.00053	1890.36	IL	FSSweet	1984	38.07
11	60	1.08843	0.06339	0.00402	248.86	IL	FSSweet	1985	26.51
12	61	0.66299	0.25871	0.06693	14.94	IL	FSSweet	1986	9.54
13	62	0.85302	0.04194	0.00176	568.52	IL	Gcup	1984	39.79
14	63	0.55555	0.03612	0.00130	766.49	IL	Gcup	1985	27.63
15	64	0.15786	0.28335	0.08029	12.46	IL	Gcup	1986	6.22
16	65	0.62416	0.01996	0.00040	2510.03	IL	Stylepak	1984	38.94
17	66	0.37280	0.02808	0.00079	1268.25	IL	Stylepak	1985	25.70
18	67	0.78146	0.14652	0.02147	46.58	IL	Stylepak	1986	6.72
19	70	0.59867	0.01999	0.00040	2502.50	IL	SnowWhit	2001	36.95
20	71	0.40333	0.01810	0.00033	3052.41	IL	Sterling	2001	35.81

Data set for Shah & Dillard (2006)

File: meta-analysis Shah slope example.sas

Go to SAS file…



Madden (Ohio State Univ.)

Heterogeneity of effect sizes

- The among-study variance (σ^2) is of value for:
  – Estimating the expected effect size and its standard error
  – Assessing the *magnitude* of effect-size heterogeneity (i.e., "Is there heterogeneity of (true) effect sizes?", How much heterogeneity?) and possibly the *impact* of the heterogeneity
    - If $\sigma^2 = 0$:

      One could use fixed-effect analysis, but there is really no reason to do so (random-effect analysis is just as easy, which automatically takes care of the among-study variability [if present])
  – Specialized post-model fitting analyses (alternative to confidence interval):
    - **<u>Prediction interval</u>: interval in which a randomly selected *future* (true) effect size will fall (υ_{new}), with associated probability (e.g., 0.95)**
    - The probability that the effect size in a randomly selected future study will be *less than* (*greater than*) a constant (ϑ) of interest, e.g., Prob($\upsilon_{new} < \vartheta$)
      » See Madden & Paul (2011), or later material (if there is time)
- One can test for significance of σ^2 in several ways, including with a **likelihood ratio test** (for MLE and REML), or with Cochran's Q statistic
- Confidence intervals based on **profile likelihoods** (MLE and REML) or based on properties of the Q statistic (moment method -- *specialized*)

Heterogeneity of effect sizes, *continued*

- Often, one wants to know the *relative* magnitude of among-study heterogeneity of the effect sizes (eliminates units of the Effect Size)
  - Higgins & Thompson (Stat. Med. [2002]) proposed three (interrelated) relative indices, primarily to ascertain the *impact* of heterogeneity on the results
    - $H^2 = Q/(K-1)$, **total variability relative to variability under homogeneity**
    - "R^2": Square of ratio of the width of the confidence interval (or SE) for estimated effect size (ζ) for a random-effect and fixed-effect analysis (loosely analogous to a *design effect* in survey sampling)
      - Larger then ~ 2 means that among-study variation is having a substantial impact on the results

$$"R^2" = \left(\frac{SE(\hat{\zeta})^{(RAN)}}{SE(\hat{\zeta})^{(FIX)}} \right)^2$$

    - I^2: **Percentage of total variability that is due to among-study heterogeneity**, defined directly from Q and related statistics (loosely analogous to an intra-class correlation):

$$I^2 = 100 \cdot (H^2 - 1)/H^2 = 100 \cdot [Q - (K-1)]/Q$$

      - Heavily reported by those using the method of moments (*an extremely popular statistic*)

Heterogeneity of effect sizes, *continued*

- I^2: Percentage of total variability that is due to among-study heterogeneity (defined directly from Q and related statistics):

$$I^2 = 100 \cdot \frac{H^2 - 1}{H^2} = 100 \cdot \frac{Q - (K-1)}{Q} = 100 \cdot \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + s^2}$$

- The last term *only* applies with non-varying sampling variances
  - That is, if $s_i^2 = s^2$ for all i (identical known sampling variances across all studies), then I^2 can be written directly in terms of within- and among-study variances
    - Some authors incorrectly substitute a simple average of the s_i^2 values
    - There are ways of estimating a "weighted average" of the s_i^2 values ("*typical* within-study variance"), so that the last term holds as an approximation
- With **Case Study 1**:
    - DL est.: $Q = 250.4$, K-1 = 100, $H^2 = 2.5$, $R^2 = 3.0$, $I^2 = 60\%$
    - REML: $H^2 = 2.3$, $R^2 = 2.9$, $I^2 = 57\%$

$$H^2 = \frac{\left(\sum s_i^{-2} - \left(\sum s_i^{-4} / \sum s_i^{-2}\right)\right) \cdot \sigma^{2(ML)}}{K-1} + 1$$

If ML or REML is used (no Q statistic)

Meta-analysis: Among-study variability (case study 1)

ML estimation for log response ratio data (Case Study 1);
Profile likelihood CI method

p value based on likelihood-ratio statistic (LRS) (difference of log-likelihoods between the random and fixed effects models)

$\hat{\sigma}^2$	95% CI for $\hat{\sigma}^2$	p value	"R^2"
0.036	**0.020 - 0.063**	**<0.001**	**2.9**

Relative impact of heterogeneity (> 2 is high)

H_0: $\sigma^2 = 0$ (i.e., no heterogeneity in the [true] effect size among studies; υ_i varies among studies)

H_a: $\sigma^2 > 0$ (i.e., heterogeneity in the [true] effect size)

Higgins and Thompson metric:

$$"R^2" = \left(\frac{SE(\hat{\zeta})^{(RAN)}}{SE(\hat{\zeta})^{(FIX)}} \right)^2 = \left(\frac{.0276}{.0163} \right)^2 = 2.9$$

Confidence Interval (for expected value), case study 1:

$$\hat{\zeta} \pm t_{1\text{-}0.05/2,df} \cdot \mathrm{SE}(\hat{\zeta})$$

$\hat{\zeta}$	SE($\hat{\zeta}$)	**95% CI**	t = $\hat{\zeta}$ / SE($\hat{\zeta}$)	P value	Control % (C%)	**95% CI for *C%***
–0.24	**0.028**	**–0.30 – –0.19**	**–8.85**	**<0.001**	**21.6%**	**17.2% – 25.8%**

$$\text{Median}: \hat{C} = 100 \cdot (1 - \exp(\hat{\zeta}))$$

Prediction Interval (for new-individual-study effect size):

$$\hat{\zeta} \pm t_{1-0.05/2,df} \cdot \left(\mathrm{SE}(\hat{\zeta})^2 + \hat{\sigma}^2\right)^{0.5}$$

$\hat{\zeta}$	SE($\hat{\zeta}$)	**95% Pred. Int.**	t = $\hat{\zeta}$ / SE($\hat{\zeta}$)	P value	Control % (C%)	**95% Pred. Int. for *C%***
–0.24	**0.028**	**–0.62 – 0.13**	**–8.85**	**<0.001**	**21.6%**	**-14.3% – 46.3%**

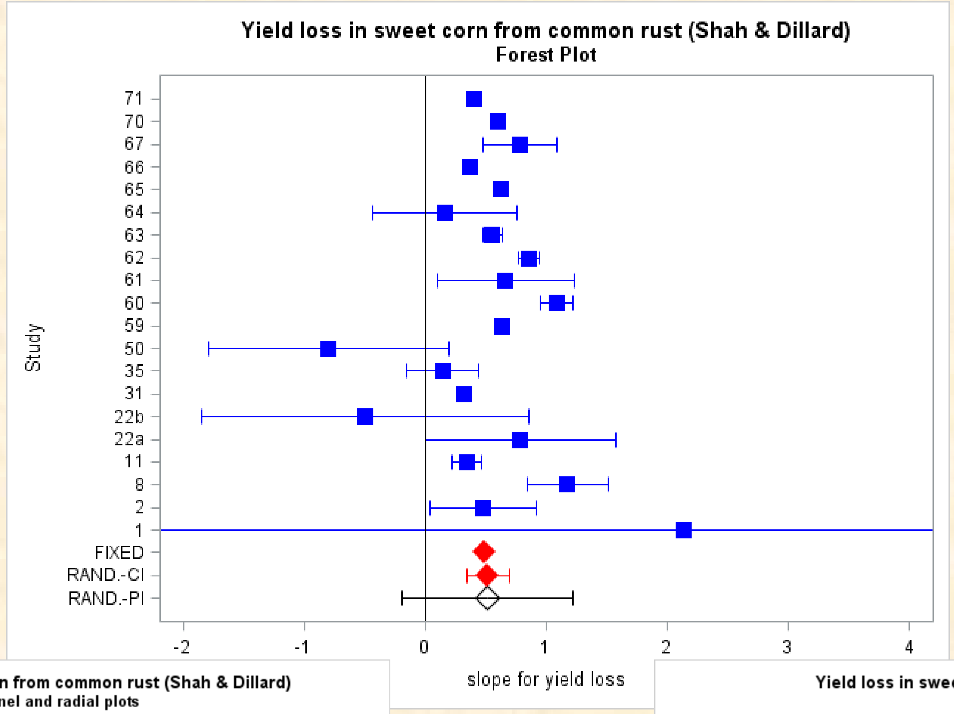
$$SE(\hat{\zeta}) = 0.0276, \hat{\sigma}^2 = 0.0365, \ \hat{\sigma} = 0.191$$
$$\sqrt{SE(\hat{\zeta})^2 + \hat{\sigma}^2} = 0.193 \approx \hat{\sigma}$$

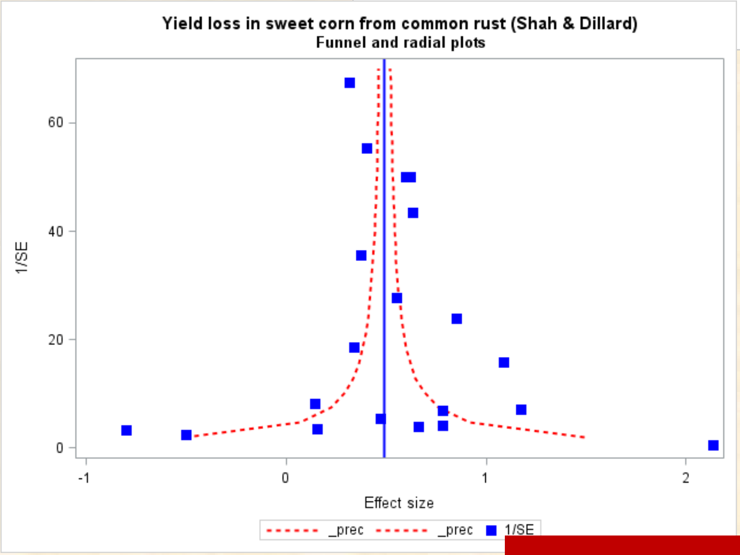
See Madden & Paul (2011) for estimating Prob($\upsilon_{new} < \vartheta$)

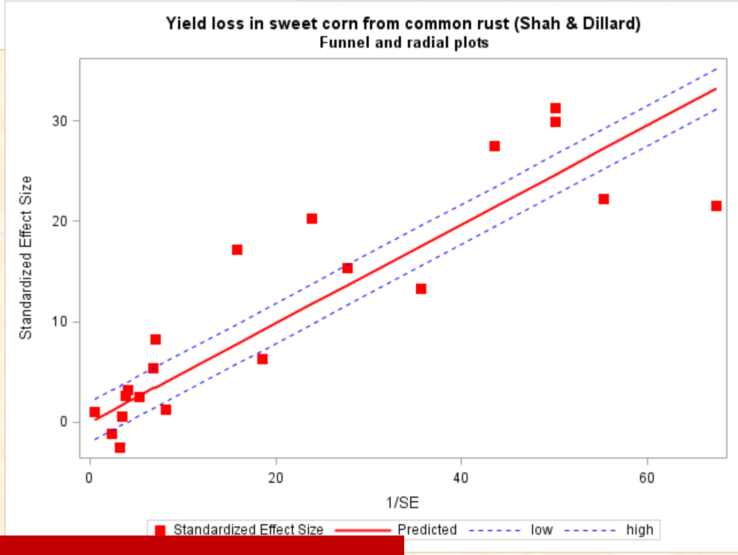
Case study 2, *continued*

meta-analysis Shah slope example.sas

Case study 2, *continued*



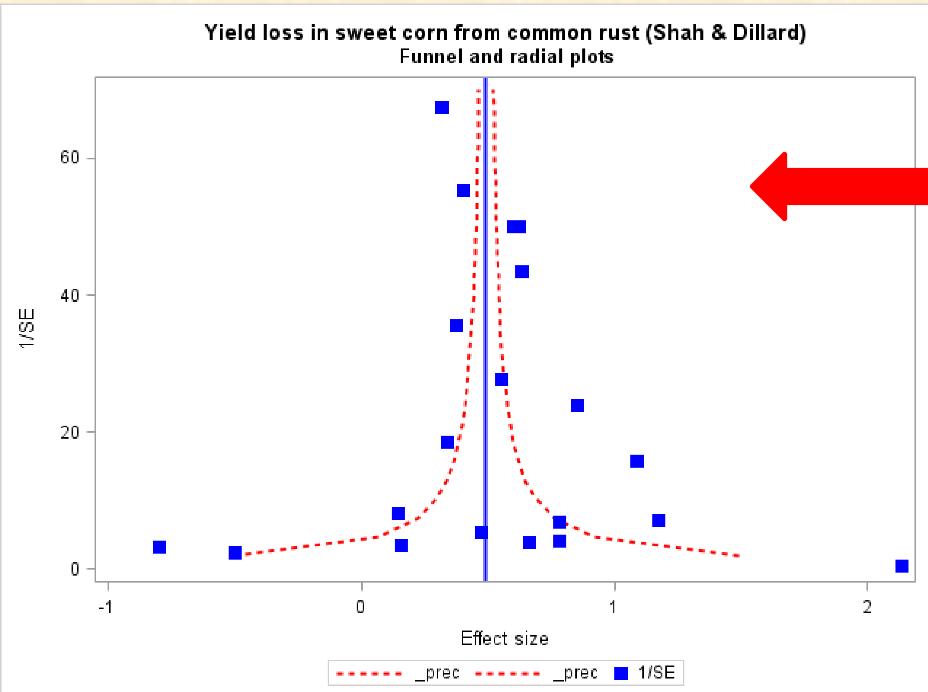


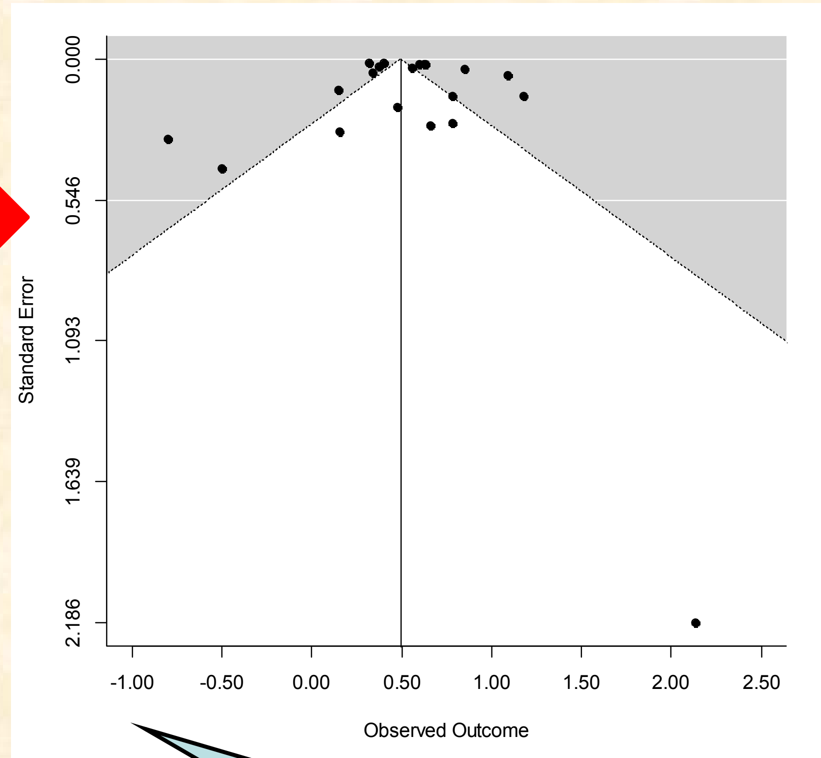


meta-analysis Shah slope example.sas

Madden (Ohio State Univ.)

Case study 2, *continued*





Optional funnel plot using `metafor` R package (note the inverse and reversed y-axis)

Estimates (Fixed effects model, or random effects model: DL method of moments or REML)							
	Mean	SE	Among study Variance	-95%CI	+95%CI	t	*p* value
Fixed	0.493	0.0075	0	0.478	0.508	66.012	<0.0001
Moment (DL)	0.547	0.0496	0.033	0.450	0.644	11.015	<0.0001
REML	0.516	0.0856	0.118 (se=0.046)	0.337	0.696	6.03	<0.0001

$Q^2 = 523.1$, df=19,

$H^{2(DL)} = 27.5$, $I^{2(DL)} = 96.4\%$, $R^{2(DL)} = 44.1\ (= (.0496/.007465)^2)$,

$H^{2(REML)} = 95.6$, $I^{2(REML)} = 98.95\%$, $R^{2(REML)} = 131.5\ (= (.0856/.007465)^2)$

Study heterogeneity ($\sigma^2 > 0$), *continued*

- Causes include:
  - Differences in study conditions (experimental methods, data collection approaches, etc.)
  - Environment (broad sense)
- Study conditions/environment can be accounted for in the meta-analysis through the incorporation of **moderator variables** in the model
  - **Moderator variable**: *study-level characteristics (continuous or categorical variables) that can affect the magnitude of the effect size*
    - Examples for case study 1: wheat variety, local weather, baseline disease incidence, measurement methods for the toxin, etc.
  - Moderator variables are fixed effects in the model. Thus, moderator-variable analysis involves a **mixed model**
  - Accounting for moderator variables can increase our understanding of the phenomenon under investigation, and possibly lower the estimated among-study variance and the standard error of the estimated effect sizes

Meta-Analysis

Effect size for study i

$$z_i = \zeta_0 + u_i + \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i$$

Within-study random effect term; residual or *"sampling variation".*

Intercept constant

Random effect of study i on the effect size.

Effect of moderator variable(s) for the i-th study.
$\mathbf{X}_i$: a row vector of l different continuous moderator variables, or "dummy variables" to indicate categories or class levels ($1 \times l$)
Can put in form of categorical effects (with extra subscript)
β: vector of effects of the moderator variables on the effect size ($l \times 1$)

Expected value:
$$E(z_i) = \zeta = \zeta_0 + \mathbf{X}\boldsymbol{\beta}$$

$$\left\{ \begin{array}{l} u_i \sim N(0, \sigma^2) \\ \varepsilon_i \sim N(0, s_i^2) \end{array} \right.$$

σ^2: among-study variance

s_i^2: sampling (residual) variance (separate for each study; assume known)

$$z_i \sim N(\zeta_0 + \mathbf{X}\boldsymbol{\beta}, \sigma^2 + s_i^2)$$

Moderator variables

$$z_i = \zeta_0 + u_i + \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i \qquad z_i \sim N(\zeta_0 + \mathbf{X}\boldsymbol{\beta}, \sigma^2 + s_i^2)$$

Specific cases:

One continuous moderator (e.g., mean disease severity [X_i] for case study 2)

$$z_i = \zeta_0 + u_i + X_i \beta + \varepsilon_i \qquad z_i \sim N(\zeta_0 + X\beta, \sigma^2 + s_i^2)$$

$$\hat{\zeta} = \hat{\zeta}_0 + X_i \hat{\beta} = 0.4350 + X_i \cdot 0.003305$$

$$\hat{\zeta}_{X=10\%} = 0.4350 + 10 \cdot 0.003305 = 0.4680$$

X_j: Mean disease severity in study (case study 2)

One categorical (factor) moderator variable (case study 2) – the two-subscript syntax (same as using X_i = 0,1 dummy variable)

$$z_{ij} = \zeta_0 + u_i + M_j + \varepsilon_i \qquad z_{ij} \sim N(\zeta_0 + M_j, \sigma^2 + s_i^2)$$

$$\hat{\zeta} =$$

$$\hat{\zeta}_0 + \hat{M}_1 = 0.5587 - 0.1122 = 0.4465$$

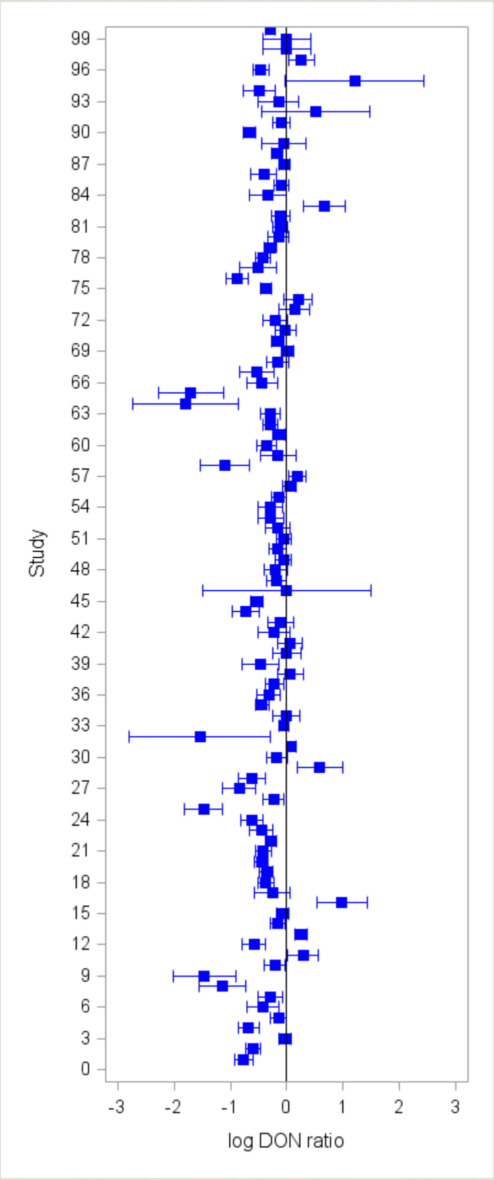
$$\hat{\zeta}_0 + \hat{M}_2 = 0.5587 + 0 = 0.5587$$

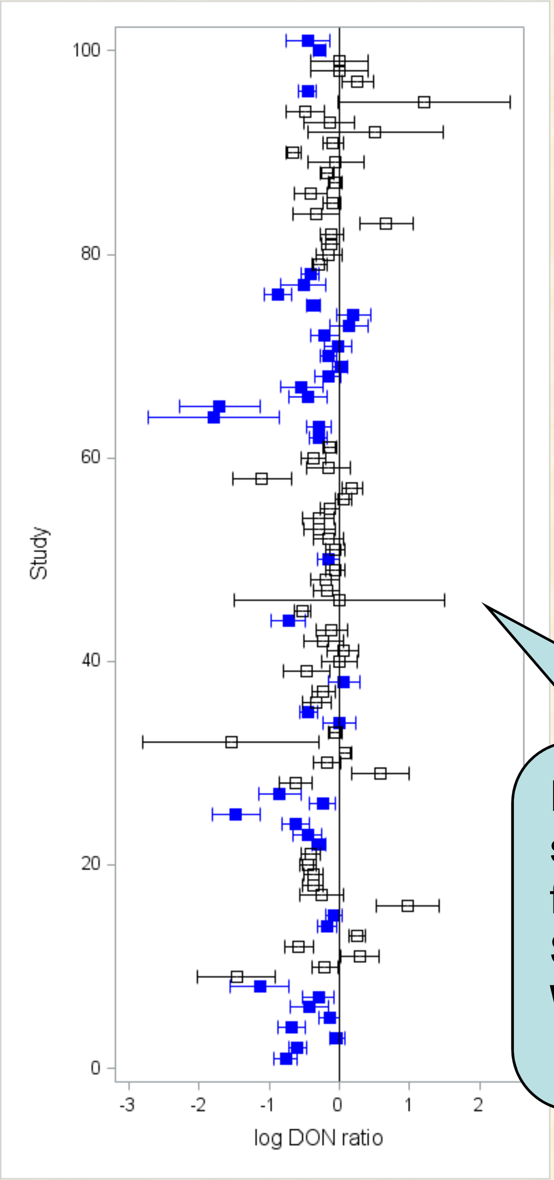
M_j: Effect of "base disease severity"
1: low (max. severity $\leq$ 50%)
2: high (max. severity > 50%)
(case study 2)

Meta-analysis with moderator variables

- Analysis proceeds in the same fashion as with the simpler analysis
- One can visualize results with **expanded Forest plots** (different symbols for different levels of a factor) or with x-y bubble graphs for continuous variables (with bubble size proportional to $1/s_i^2$ or to s_i^2)
- **Funnel and radial plots** should be based on the *residuals* from a fixed-effect model with moderator variables
- One can use moment (DL) or likelihood (ML or REML) based model-fitting methods
  - `metareg`, `metaf` SAS macros; **PROC GLIMMIX/MIXED**; `metafor` R package
- Tests of the effects of moderator variables can be based on Wald statistics (chi-squared or F), likelihood ratios, or more specialized (robust) statistics (**Wald test is more common**)
  - With large K (number of studies), choices will not matter too much
  - With small-to-moderate K, and high variation in the s_i^2 values, there is no consensus on the best testing approach (see Hartung et al. [2008] book)
    - **Adjustments to the usual F or chi-squared distributions (or corresponding df) to account for the variation in s_i^2 and estimated σ^2**
    - **I recommend Kenward-Roger (KR) adjustment with GLIMMIX or MIXED**

Case Study 1, *continued*





Forest plot with different symbols for two levels of factor “*wheat type*”: **Spring** (blue, solid); and **Winter** (black, open)

Moderator Variable Example (Case study 1):

Wheat Type (Winter [W] or Spring [S])

Wheat type	$\hat{\zeta}$	SE($\hat{\zeta}$)	95% CI for $\hat{\zeta}$	t = $\hat{\zeta}$ / SE($\hat{\zeta}$)	p value	Control % (C)	95% CI for C
W	**–0.17**	**0.035**	**–0.24 – –0.11**	**–4.9**	**<0.001**	**16%**	**10% – 21%**
S	**-0.33**	**0.041**	**–0.42 – –0.25**	**–8.2**	**<0.001**	**28%**	**22% – 34%**

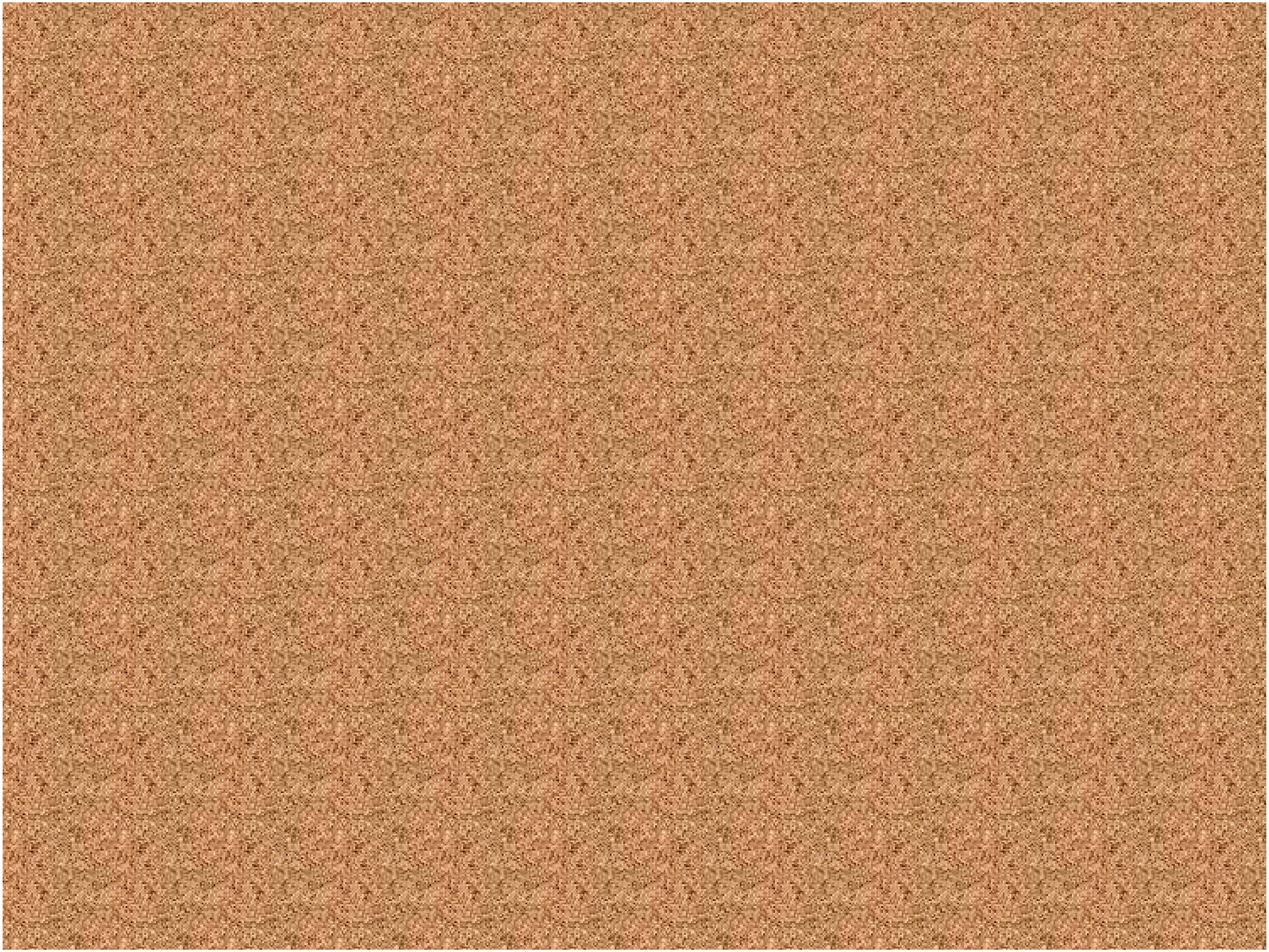
F test indicated a highly significant effect of wheat type. The estimated among-study variance, however, was only slightly decreased (from 0.036 to 0.032)

Effect	***df***	***F***	***P***
Wheat type	1,99	8.77	.0038

Between-within df method

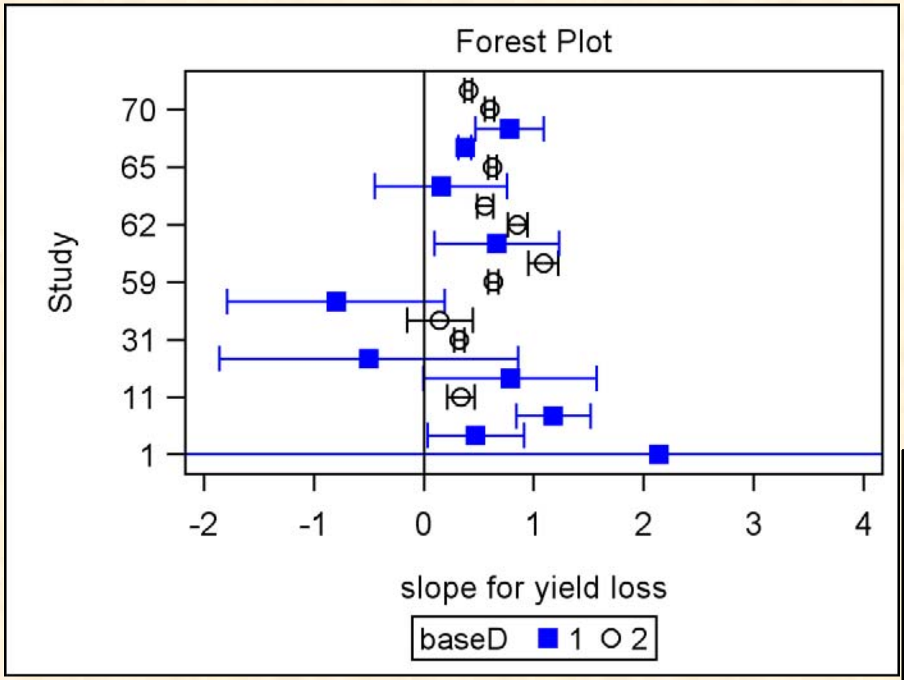
Effect	***df***	***F***	***P***
Wheat type	1,73.2	8.72	.0042

Kenward-Roger df method



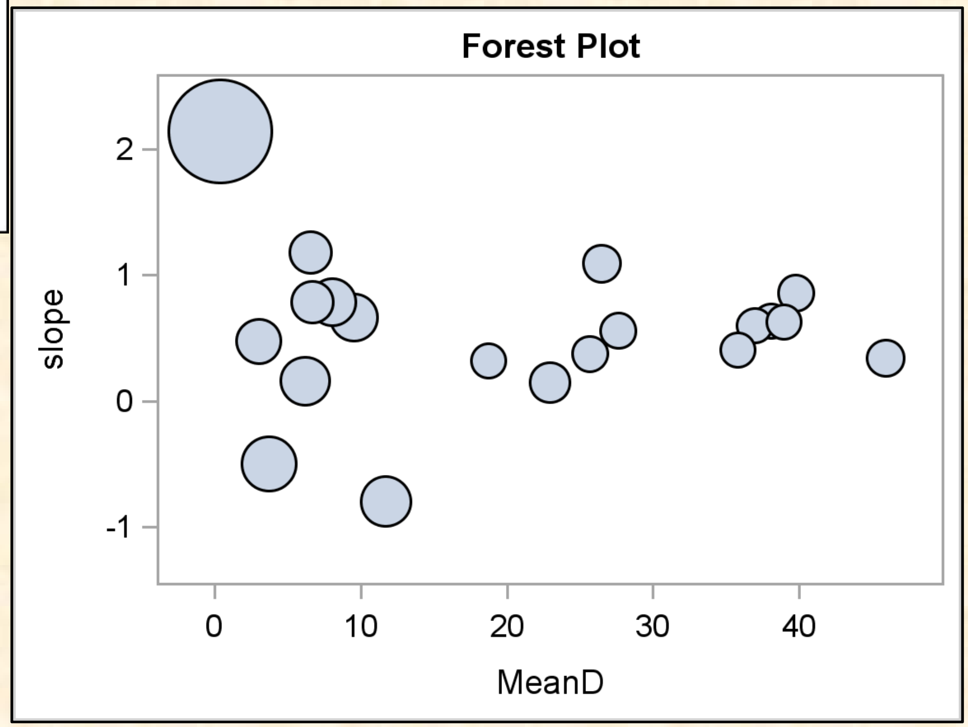
Madden (Ohio State Univ.)

Case study 2, *continued*



$$\hat{\mu} = \hat{\zeta} + X_i\hat{\beta} = 0.4350 + X_i \cdot 0.003305$$
$$\hat{\mu}_{10\%} = 0.4350 + 10 \cdot 0.003305 = 0.4680$$

$$\hat{\zeta} =$$
$$\hat{\zeta}_0 + \hat{M}_1 = 0.5587 - 0.1122 = 0.4465$$
$$\hat{\zeta}_0 + \hat{M}_2 = 0.5587 + 0 = 0.5587$$



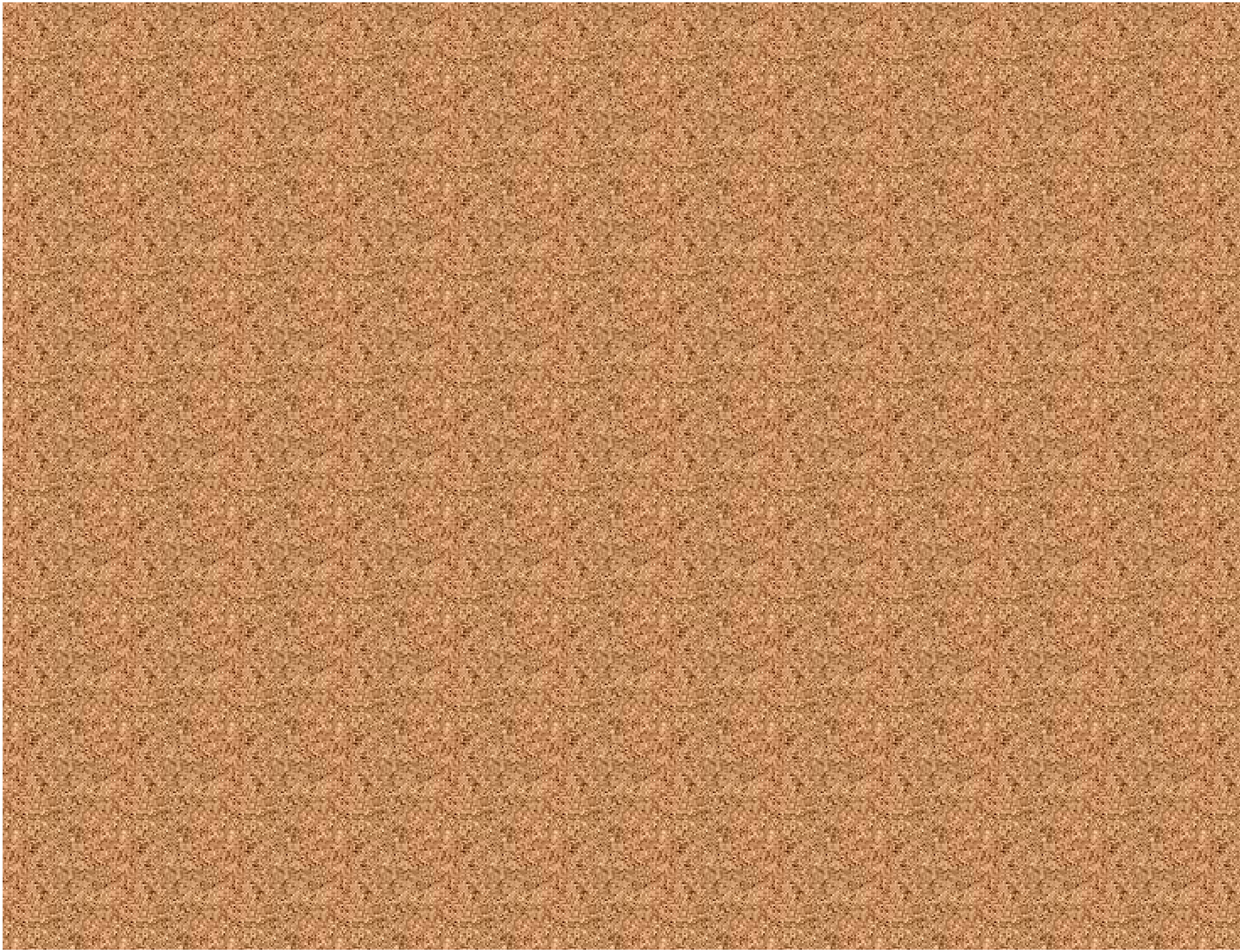
Moderator: baseD	Mean estimate	
high	0.5587	a
low	0.4465	a

File: meta-analysis Shah slope example.sas

Case study 3: Yield of corn in relation to fungicide treatment (even in absence of disease)

- Strobilurin fungicides are being marketed for “plant health” benefits, such as increased yield, even when plant disease is not present
- Paul and colleagues investigated the yield response of corn (maize) hybrids when treated once with a strobilurin fungicide between growth stages VT (tassel emergence) and R1 (silk emergence)
  - Paul, Madden, Bradley, et al. (2011). Phytopathology 101: 1122-1132.
    - See paper for study selection criteria and how the literature was searched
- Four different fungicides were evaluated in separate meta-analyses; results for Quilt (azoxystrobin + propiconazole) are used here.
- K = 61 studies
- We will work through this example directly in SAS (*no output summaries in PowerPoint*)

File: meta_quilt.sas



Madden (Ohio State Univ.)

Multiple effect sizes

- There may be more than one estimated effect size in each study
  - Multiple (q) endpoints (i.e., response variables), repeated measures, or possibly estimated parameter estimates for relationships between variables (e.g., intercept and slope) for each trial
  - Multiple treatments for each study, where an individual study could have between 1 and q treatments
  - Multiple treatments *and* endpoints
- Many data analysts ignore the multiple effect-size nature of the studies and carry out several univariate analyses
  - Often, only one effect size is of interest from each study
  - With multiple effect sizes, the univariate approaches ignore the correlations within and among studies, and *can* therefore be misleading
- The meta-analytical fixed or random-effects models can be expanded for q random variables per study
  - All studies do not have to contain all q effect sizes [$q(i)$]
  - We focus on the multiple treatment (**multi-treatment**) problem
  - We consider only normal distributions (for estimated effect sizes)

Multiple treatments (groups) per study

- There are many approaches to meta-analysis with $q \geq 2$ treatments
  - For demonstration purposes, assume there are three treatments

$$y_{i1} = \hat{\mu}_{i1}$$
$$y_{i2} = \hat{\mu}_{i2}$$
$$y_{i3} = \hat{\mu}_{i3}$$

Estimated expected values (means) for the three treatments (j = 1, 2, 3) for the i-th study (y is a mean across all reps or blocks within a study, not an individual observation)

  - If j = 3 is the control (for example), then one may be interested in the mean difference (contrast) as the effect size

$$z_{i1} = y_{i1} - y_{i3}$$
$$z_{i2} = y_{i2} - y_{i3}$$

- Separate meta-analysis for each contrast:
  - Methods described previously are applied to each effect size
  - This approach ignores the correlations of the estimated effect sizes due to the presence of a common treatment mean in each contrast

$$z_{i1} = \zeta + u_i + \varepsilon_i \qquad z_{i2} = \zeta + u_i + \varepsilon_i$$

Multiple treatments (groups) per study

- A more elaborate approach is to conduct a multivariate multi-treatment meta-analysis based on the vector of contrasts ($\mathbf{z}_i$) for each study

$$\mathbf{z_i} = \begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix} = \begin{pmatrix} y_{i1} - y_{i3} \\ y_{i2} - y_{i3} \end{pmatrix}$$

  - **A study does *not* have to include all treatments to be used in the analysis**
    - In contrast, with the univariate approach, one can only use a study if treatments 1 *and* 3 (z_{i1}) were included; or treatments 2 *and* 3 were included (z_{i2})
  - Because each effect size includes a common (control) mean, the effect sizes *must* be correlated *within* studies
  - Effect sizes *may* also be correlated *between* studies
- This is the general (multivariate) approach of Gleser and Olkin (2009) – available in `metafor` R package
- The basis for so-called ***network meta-analysis*** (**Mixed Treatment Comparisons [MTC**], **multi-treatment**) approach of Lu and Ades (*J. Am. Stat. Assoc.* 101:447-459 [2006]; *Stat. Med*. 23:3105-3124 [2004])
    - The Lu and Ades methodology is actually more complex (not covered here)
    - Lu and Ades take a Bayesian approach, but a frequentist analysis is also possible (Piepho, Williams, Madden, *Biometrics* 68: 1269-1277 [2012])

Multiple treatments (groups), *continued*

- The Lu & Ades approach has been used heavily in medical statistics, especially with IPD analyses. It is quite effective.
- This is a *"non-standard"* model for mixed-model analysis, requiring much more of the analyst (especially with IPD). Because all treatments do not occur in all studies, great care must be taken in:
  - Constructing the fixed-effects portion of the meta-analytical model
  - Constructing the within-study covariance matrix for each study to account for the correlation of the different contrasts
  - Constructing the among-study covariance matrix
- Instead of analyzing contrasts ($\mathbf{z}_i$) of the means, one can conduct the analysis directly on treatment means ($\mathbf{y}_i$) for each study, <u>and calculate contrasts post-model fitting</u> based on expected values

$$\mathbf{z_i} = \begin{pmatrix} y_{i1} - y_{i3} \\ y_{i2} - y_{i3} \end{pmatrix} \Longrightarrow \mathbf{y_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix}$$

Multiple treatments (groups), *continued*

$$\mathbf{z_i} = \begin{pmatrix} y_{i1} - y_{i3} \\ y_{i2} - y_{i3} \end{pmatrix} \Longleftrightarrow \mathbf{y_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix}$$

- For a certain class of (variance component) models, analysis of $\mathbf{y}_i$ gives *identical* results to an analysis of $\mathbf{z}_i$ for each study (Piepho, Williams, Madden, *Biometrics* [2012]), when REML is used for model fitting
- See Piepho (*BMC Med. Res. Meth.* 14:61- [2014]) for more on the equivalence of the two approaches (and lots of hints on the analysis)
- Thus, one can readily use standard mixed-model software without *too many* additional steps (always some extra work with meta-analysis!)
- For both approaches, **direct** and **indirect** information is utilized
  - Suppose one is interested in the expected difference in means for treatments 1 and 2
  - Direct evidence (from studies with treatments 1 <u>and</u> 2): μ_1-μ_2
  - Indirect evidence (from studies with 1 <u>or</u> 2): (μ_1-μ_3) – (μ_2-μ_3) = μ_1-μ_2
    - So, studies without μ_1 (or μ_2) still provide information on the expected difference of treatments 1 and 2

Multi-treatment meta-analysis

There are *several* possible models based on estimated means for each study (only a few examples are given) – start with univariate representation:

y_{ij} is the estimated mean for treatment j in the i-th study ($y_{ij} = \hat{\mu}_{ij}$)

(u_{ij}, σ^2, ε, s_{ij}^2 are all defined now in terms of the means, **not** the differences)

Response variable (estimated mean)

Fixed or random effect of study i

Fixed effect of treatment j

Random effect of study on the treatment effect (*interaction of study and treatment*)

$$y_{ij} = \beta_i + \tau_j + u_{ij} + \varepsilon_{ij}$$

Residual (error in estimating means within the study, held *fixed*); sampling variance

Interest is in the difference in treatment effects:

$$E(y_{ij} - y_{ij'}) = \tau_j - \tau_{j'} \qquad \text{e.g., } \tau_1 - \tau_3$$

τ_j - $\tau_{j'}$ plays the role of ζ in the univariate analysis

Study effects in model

- Many meta-analysts favor **fixed** rather than random study main effects (β_i), although either approach can be justified
- **Fixed study main effects:**
  - Analogous to an **incomplete block design with *fixed* block effects**
  - Fixed treatment effects are based on intra-study information only
    - *May* be important because no randomization is involved in the selection or design or locations of studies, even though there may be (or will be) randomization within each study
    - There is still a random effect of study on the treatment effect (interaction: u_{ij})
  - For q = 2 treatments, the multi-treatment model with *fixed* effect of study is *equivalent* to the "univariate" contrast model ($z_i = y_{i1} - y_{i2} = \zeta + u_i + \varepsilon_i$)
    - Within-study sampling variances and the among-study variance for z_i differences are double the values in the multi-treatment model for y_{ij} means (for independent groups within studies), but **one obtains *identical* estimates of the *difference* of expected values (ζ or τ_1-τ_2, and corresponding SE)**
  - The case of q > 2 is straightforward

Study effects in model, *continued*

- **Random study main effects:**
  - Analogous to an **incomplete block design with random block effects**
  - Treatment effects are based on *intra*- and *inter*-trial information
    - One recovers some information on treatment effects from the “other” studies, not just from within each study
    - Proponents include van Houwelingen, Arends, and others (e.g., 2002)
  - Although Senn (Biom. J. 2010) and some others (Riley et al. 2008) argue against a random main effect for study, Senn also believes the results often will be similar for fixed or random study effects (I agree!)
  - It is common in agriculture to consider the study main effect as random
- Model:

$$y_{ij} = \beta_i + \tau_j + u_{ij} + \varepsilon_{ij}$$

$$\beta_i \sim N(0, \sigma_\beta^2),\; u_{ij} \sim N(0, \sigma^2),\; \varepsilon_{ij} \sim N(0, s_{ij}^2)$$

$$\boxed{E(y_{ij}) = \tau_j \quad E(y_{i1} - y_{i2}) = \tau_1 - \tau_2}$$

Effect of study i

Effect of treatment j

Effect of study i on treatment effect j

Residual; within-study error. Held fixed in analysis.

Within-study sampling variance (s_{ij}^2)?

- With many experimental designs, e.g., randomized complete block (RCBD), the within-study means (and not just the differences of means) are correlated
  - One needs to account for the correlation in the meta-analysis
- One can specify a within-study **variance-covariance matrix** based on residual and block variances
  - This is tricky, and very tedious to set up (but achievable with a lot of work!)
  - One may not know the block variance
- If one used the *actual* within-study sampling variance of the mean as s_{ij}^2, and *ignored* the correlation, one would obtain *incorrect* mean effect sizes (e.g., estimated τ_1-τ_3) and SE of mean effect sizes in meta-analysis
- However, still "ignoring" correlation, if one used **one-half of the variance of the *difference* of means** (½ of the square of the within-study SED) as s_{ij}^2, then one obtains the *correct* estimated mean effect sizes (e.g., estimated τ_1-τ_3) and SE of the estimated mean effect sizes in the meta-analysis! (SED ≈ LSD/2 often given)

Residual variance: V_i Block variance: V_{bi} Number of blocks: n_i

Variance of mean:	$(V_{bi} + V_i)/n_i$	
Covariance of two means:	V_{bi} / n_i	
Variance of **difference**:	$2V_i / n_i$	(block variance cancels out)

Use: $s_{ij}^2 = V_i / n_i$

See Möhring & Piepho (2009 Crop Sci.) for more on this concept

Multivariate meta-analysis: model fitting

- Parameter estimation:
  - **Method of moments** (expansion of DL method)
    - Requires specialized software, but is faster and less computer intensive than the alternatives (STATA and R code are available)
  - **ML** and **REML**
    - Iterative and more computer-intensive, but can be performed using standard linear mixed model software (if the within-study variance-covariance matrix can be fixed)
      - Allows for missing at random
      - Straightforward in SAS (MIXED or GLIMMIX)
  - **Bayesian analysis**
- The big issues of importance with univariate meta-analysis are important with multivariate analysis

Case study 3: Yield of corn, *continued*

- Effects of Quilt (azoxystrobin + propiconazole) fungicide
  - Treatments: Quilt and control ($q = 2$)
- Previously, analyzed $z_i = y_{i1} - y_{i2} = y_{i\mathrm{T}} - y_{i\mathrm{C}}$
- Now analyze as a multi-treatment (network or mixed treatment comparisons [MTC]) meta-analysis
  - Fixed and random main effect of study

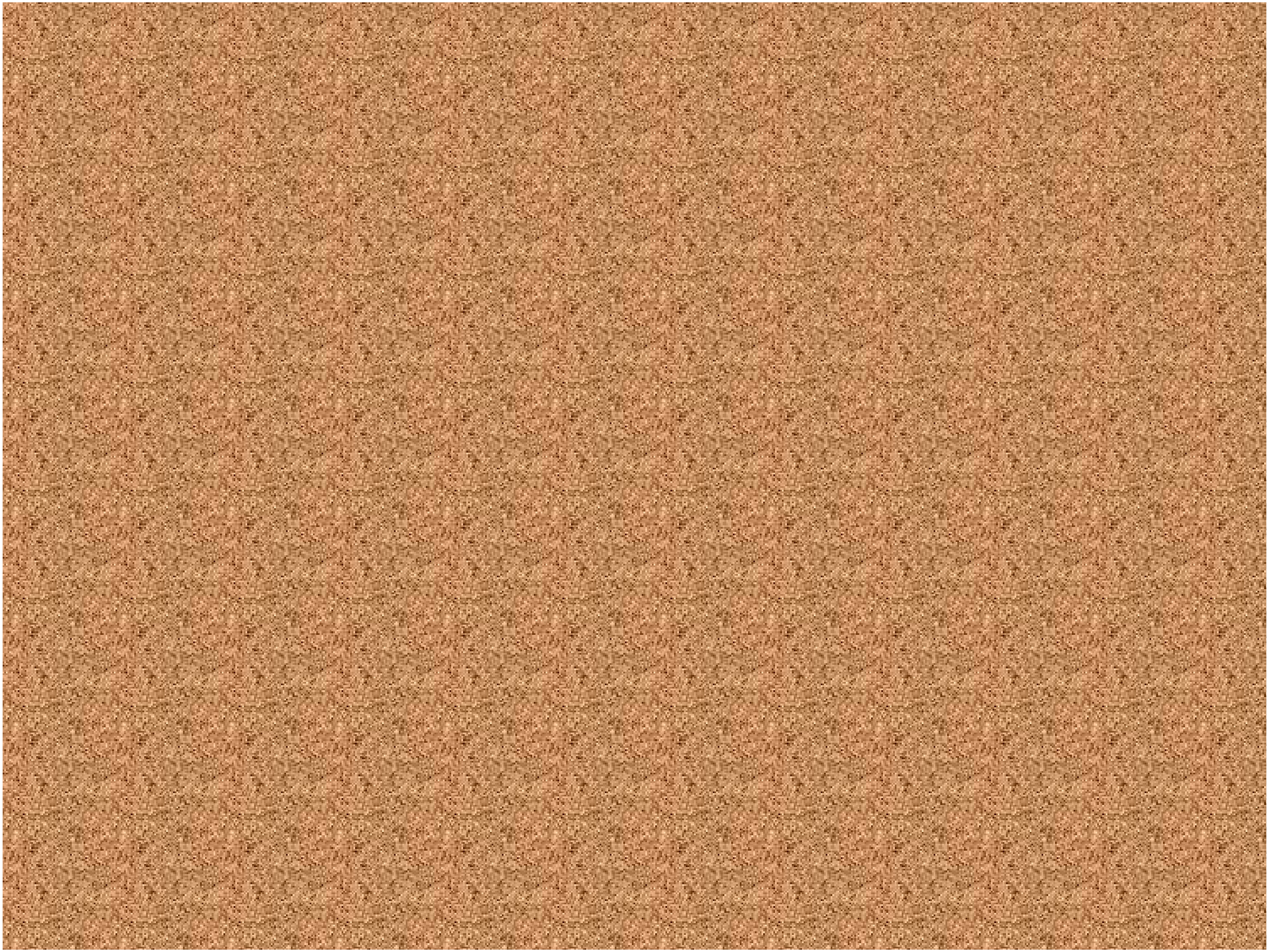
$$\mathbf{y_i} = \begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix}$$

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
mean diff	5.2909	1.4720	60	3.59	0.0007	0.05	2.3464	8.2353

Cov Parm (σ^2)	Estimate	Standard Error
trial	71.9081	24.5434

File: meta_quilt.sas

Go to SAS file



Madden (Ohio State Univ.)

Concluding comments

- With over 80,000 journal articles in print, and numerous textbooks, meta-analysis is here to stay
  - In many disciplines, it *is* the standard approach to quantitative research synthesis
  - For some regulatory government agencies, meta-analysis is virtually mandatory (e.g., approval of new drugs or treatments)
- After a slow start, meta-analysis is now gaining greater acceptance in the agricultural sciences
  - More agricultural scientists will need to become proficient in this area in order to review and understand the literature
- Meta-analysts continue to make advances with the statistical methodology, especially for network (multi-treatment) analysis, non-normal data, and for rare events
  - Advances with mixed models will play a large role in meta-analysis

Part II
ADDITIONAL MATERIAL
(not covered in workshop)

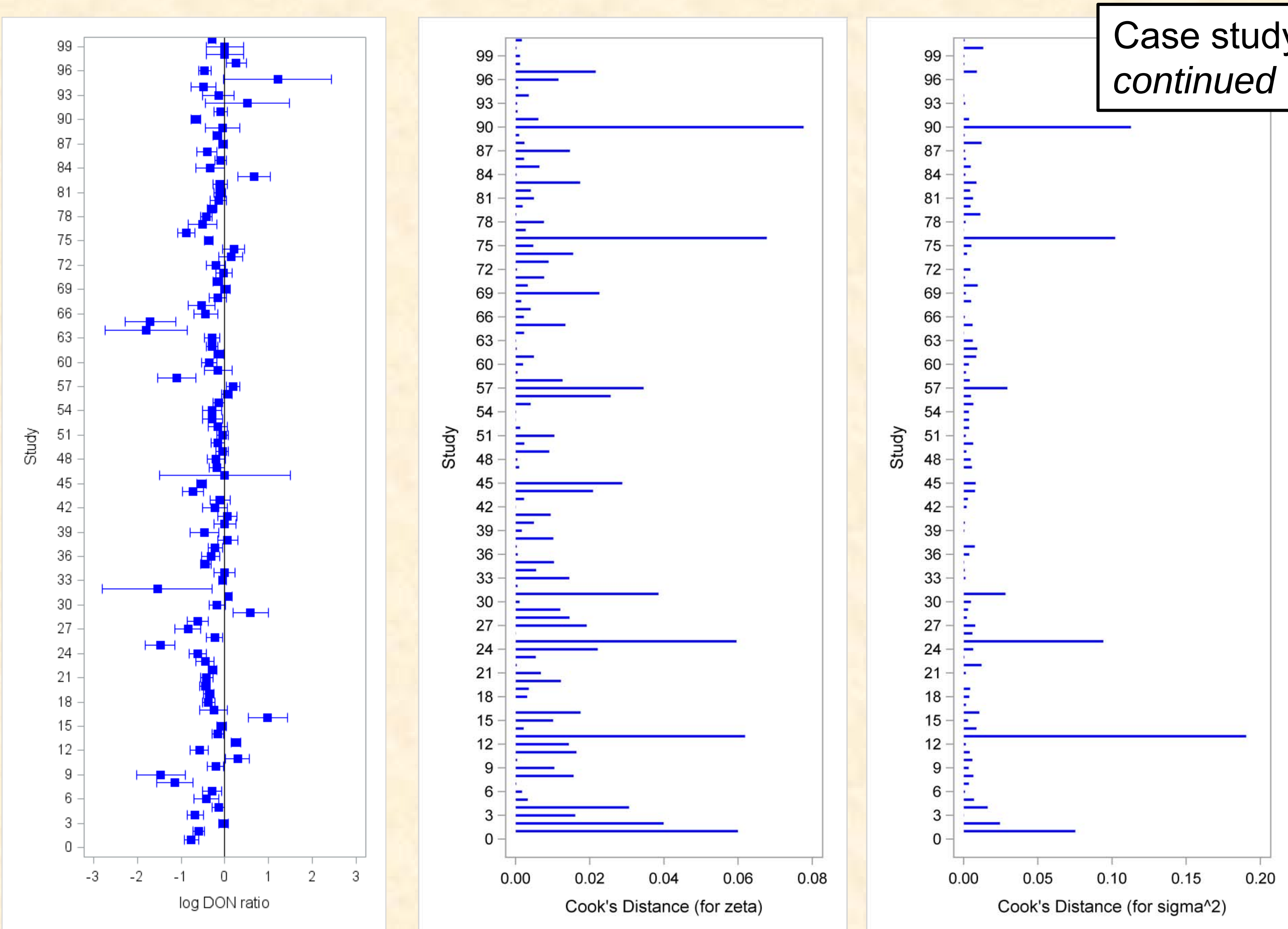
The SAS code performs all the analyses described in this additional material

Diagnostics, *continued (part 2)*

Previously

- Model assessment (criticism) in meta-analysis has unique issues
- The usual residual plot (residual, Studentized residual, Pearson residual, deleted Studentized residual, etc., versus predicted values) *may* not be of much value for the simple random effects model, because of the unequal sampling variances (patterns in the residual plot may not be a problem)
  - The unique s_i^2 for each z_i makes interpretation difficult
- Meta-analysts have developed some specialized graphs that are not typically seen in other applications
  - In addition to the Forest plot, so-called **funnel** and **radial** plots
    - These can help assess the need for a random-effects or a fixed-effects model, and explore the possibility of publication bias
- Moreover, versions of diagnostic plots from the broader field of mixed-model analysis have value (but are *much* less reported). These include:
  - Studentized deleted residual versus study ID, PRESS statistics versus study ID,…
  - **Cook's Distance** for the fixed effect (ζ) and the variance (σ^2) versus study ID
    - Measures the influence of observations (studies) on parameter estimates
      - A scaled measure of the squared distance between parameter estimates based on the full dataset and the estimates when each observation (study) is deleted (with mixed models, the model is refitted with each observation deleted)

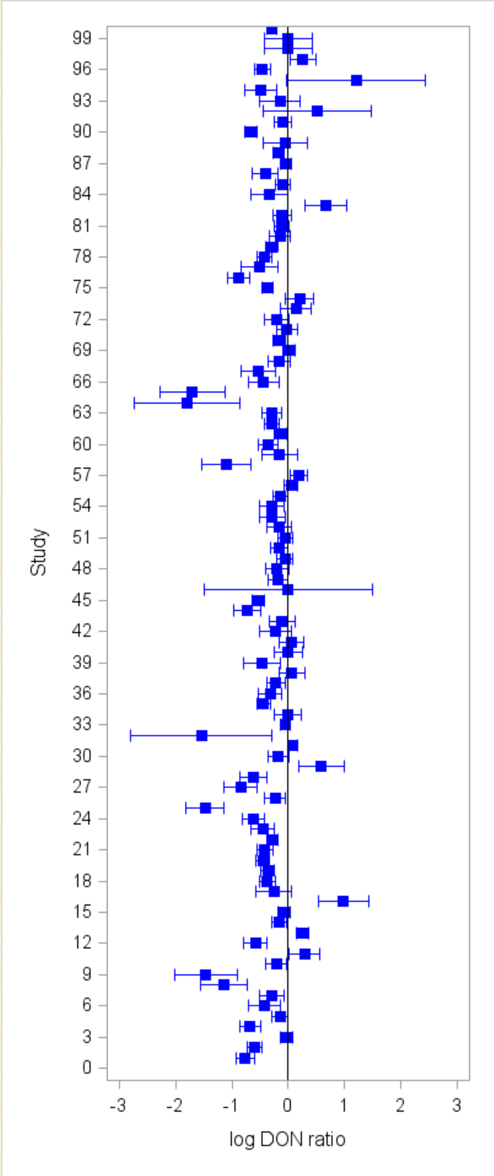
Case study 1, *continued*

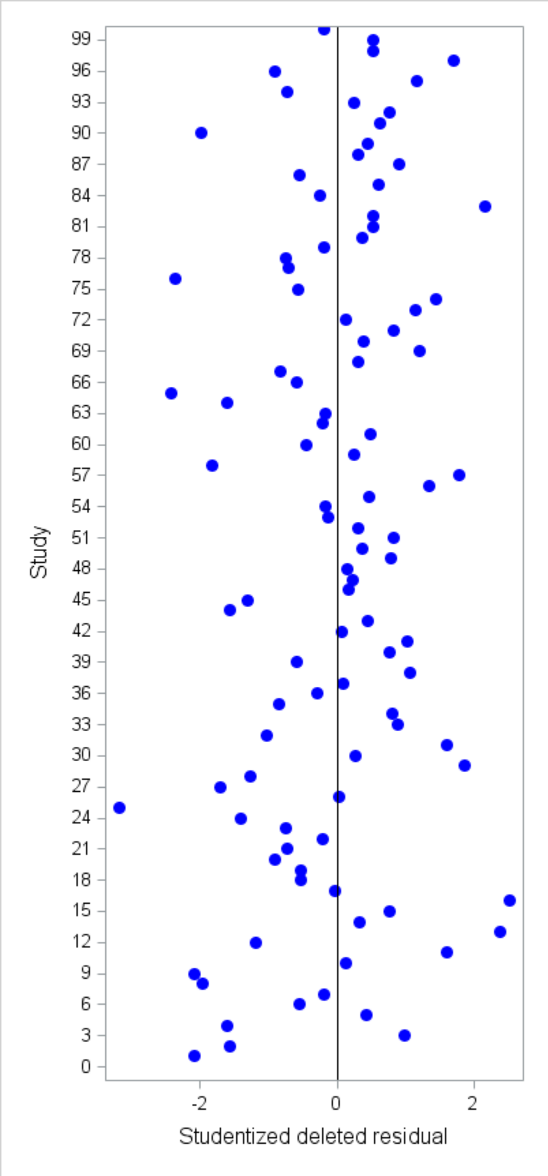


Diagnostic graphs for mixed models

As a general guide, values of **Cook's Distance** greater than $F_{0.5,1,df}$ ($\approx$ 0.45 here) may be considered large. No large values in this example. Can plot horizontally, also.

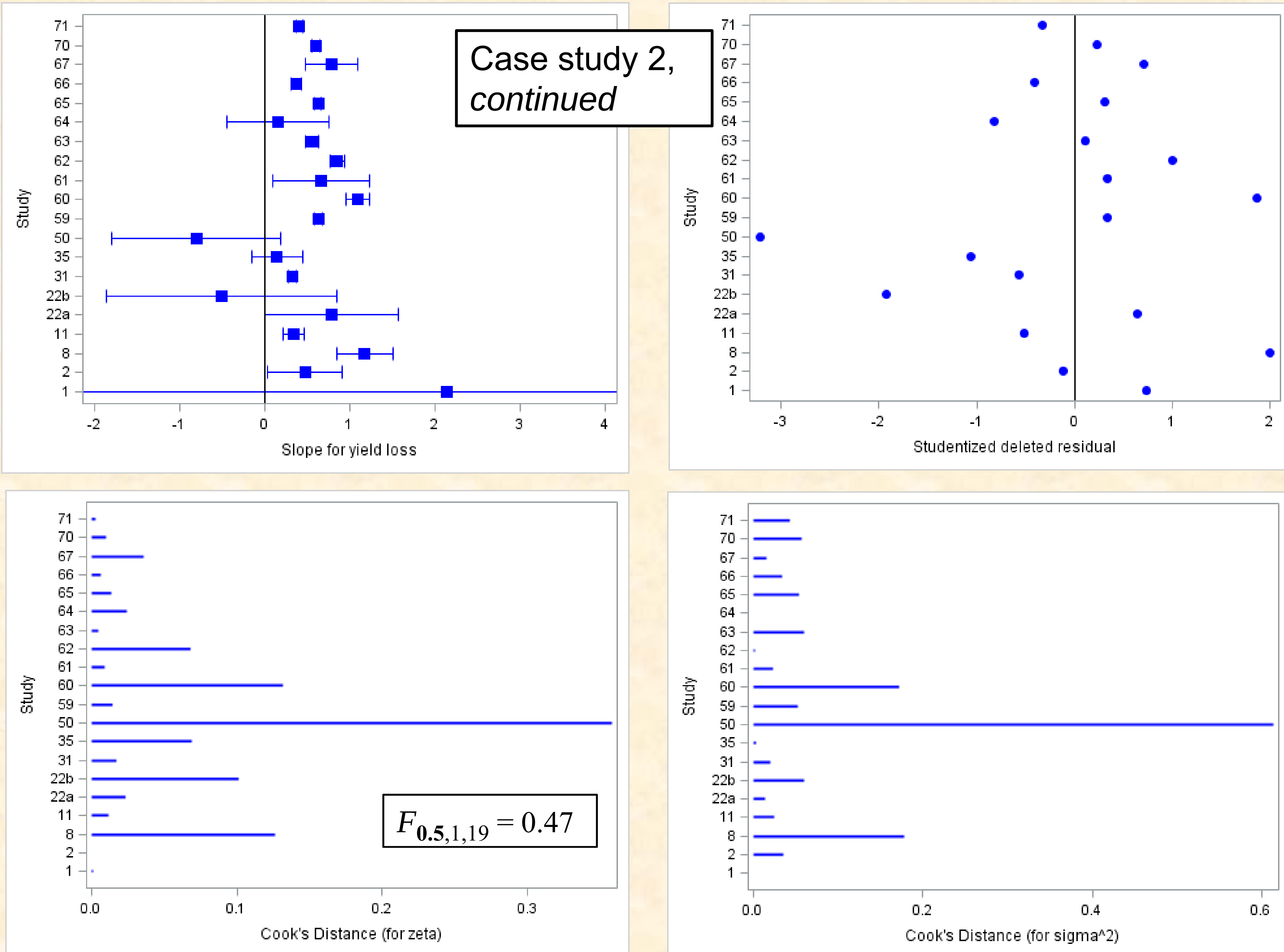
Case study 1, *continued*





Diagnostic graphs for mixed models:
Studentized deleted residual

Case study 2, *continued*



$F_{\mathbf{0.5},1,19} = 0.47$

Heterogeneity and risk probabilities

- The *mean* effect size and its confidence interval are of interest for determining the *expected* outcome in the *long run* (over many studies or over many fields [as in the case study])
- A prediction interval gives a sense of the variation (uncertainty) in individual (future) estimated effect sizes
- More directly, one can estimate the probability that the effect size in a randomly selected future study will be *less than* (or *greater than*) any constant of interest (ϑ) (see van Houwelingen et al., 2007)
  - For instance, with DON control for Fusarium head blight (case study), a grower might be most interested in knowing the probability that υ_{new} < 0 (percent control > 0%), $\Pr(\upsilon_i < 0)$, or maybe probability that υ_{new} < -0.69 (percent control > 50%)
  - Assuming that the standard error of the expected effect size is small, one can (*approximately*) estimate p_ϑ, assuming a normal distribution

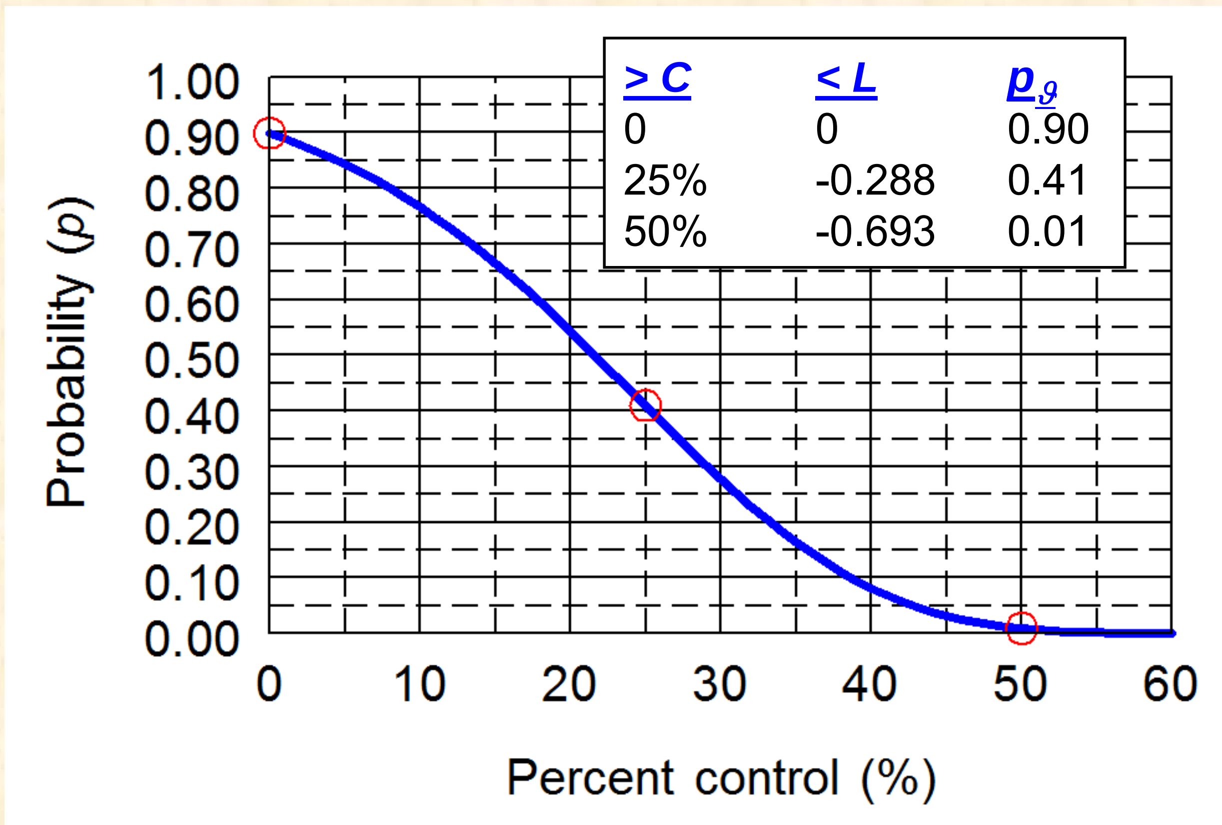
$$p_\vartheta = \Phi((\vartheta - \hat{\zeta}) / \hat{\sigma})$$

$\Phi(\bullet)$ is the cumulative normal distribution, use to obtain probability that effect size is less than ϑ

$$p_{\vartheta} = \Phi((\vartheta - \hat{\zeta}) / \hat{\sigma})$$

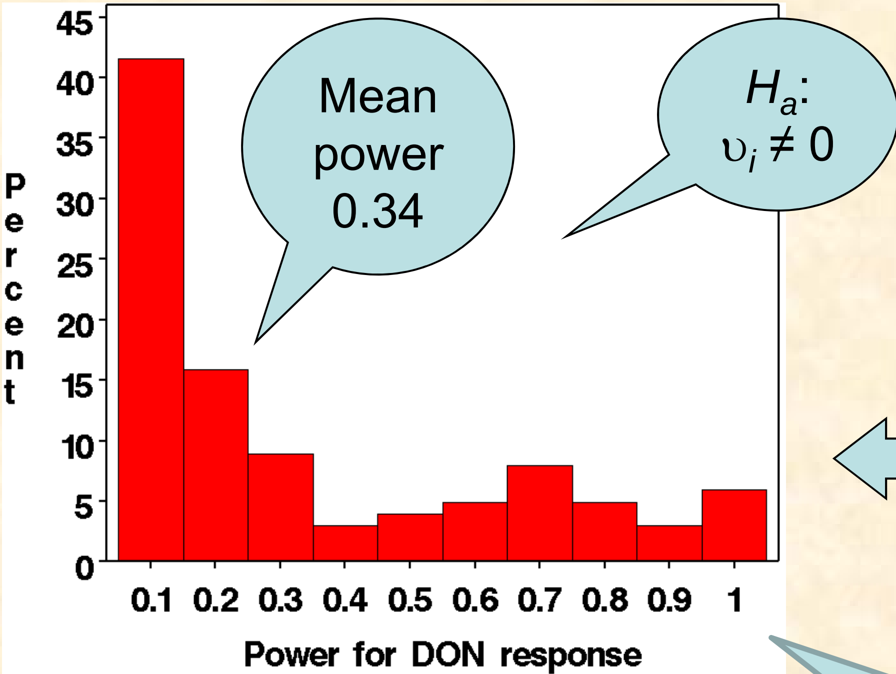
Risk probability for DON control: Determined for log ratio and than converted back to median percent control

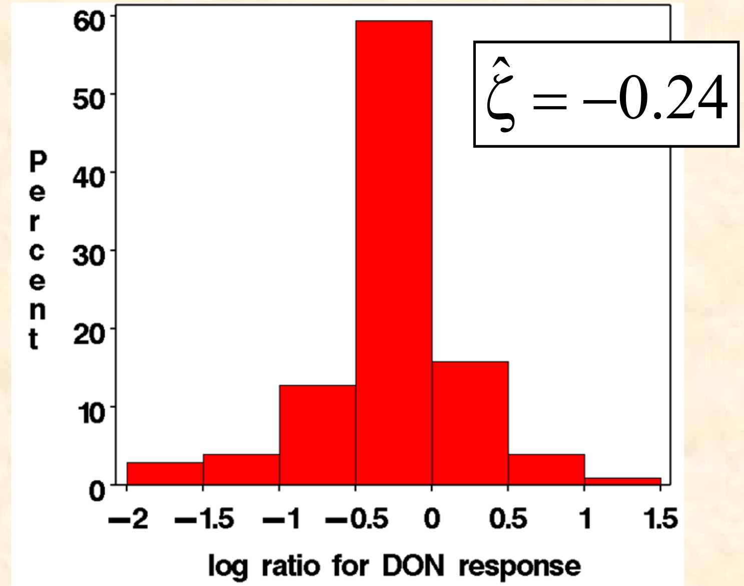
$$\hat{\zeta} = -0.244, \hat{\sigma} = \sqrt{0.036} = 0.19$$



Statistical power

- Individual studies in many disciplines are often under-powered for testing various hypotheses
- However, it is easy to show that meta-analysis of multiple studies can have very high power
  - It is possible that $p > 0.05$ for *each* individual study, and $p < 0.05$ for the meta-analysis (although H_0 involves υ_i with the former and ζ with the latter)
- Power could be the most compelling argument in favor of meta-analysis
- Assume H_a ($\zeta \neq 0$) is true (treatment is truly effective)
- Statistical power**:**
  - **Probability of rejecting H_0 when H_0 is false**
- Estimation of power can be done using:
  - Classical methods, such as using the non-centrality parameter and a shifted t or F distribution (although there are complications)
  - Simulation
- Fixed (and unequal) sampling variances (s_i^2) complicate the analyses. Thus, simulation approaches are probably best. **See Madden & Paul (2011).**
- To justify the use of meta-analysis, we can first estimate the power for *each* study (assuming H_a is true for *every* study ($\upsilon_i \neq 0$)
    - Hypothetical and **unrealistic** here, but useful for demonstration purposes – consider Case Study #1





In reality, this exercise works for ***another*** 101 (new) studies with the same statistical results as found with the actual studies (we do not determine power for studies already conducted)

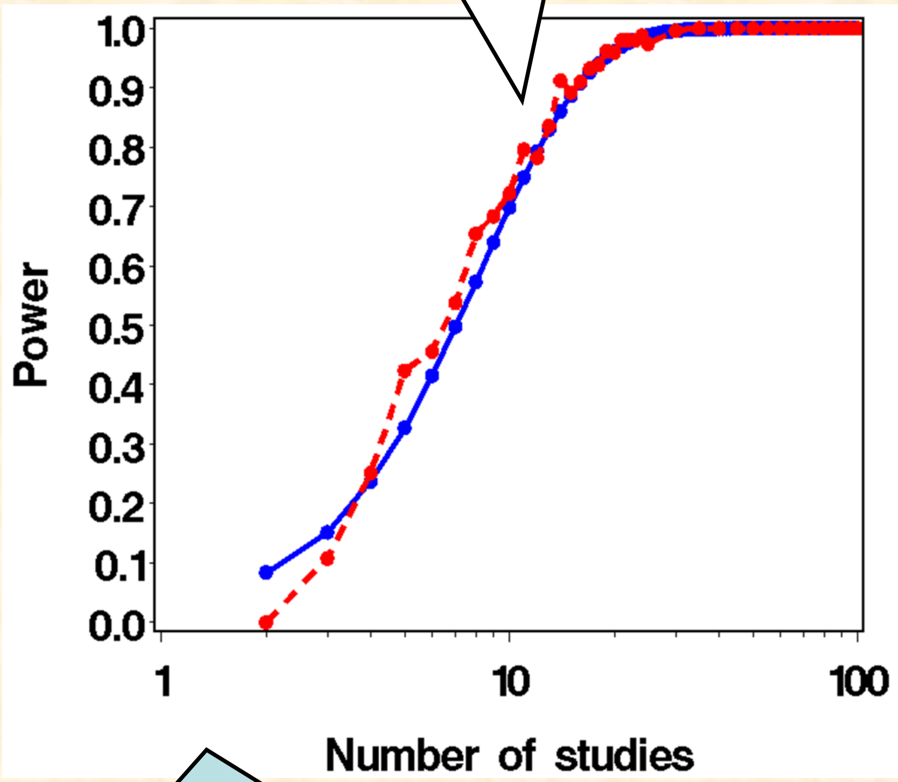
Power is low for the *individual* studies (in terms of treatment effect on DON). Note that we do not believe that the alternative hypothesis would be true for each study (this is an exercise to demonstrate concept)

We can estimate the power for a meta-analysis of a 101 new studies (we do not need to assume that H_a is true for every study, just that $\zeta \neq 0$): *Power* > 0.999

Power in meta-analysis

- One can determine power for any number (2→*K*) of randomly-selected (new) studies
- Details are in Madden & Paul (2011)
- Reminder:
  – Mean <u>individual</u> power (assuming that Folicur always has an effect) is 0.34
  – Meta-analysis ***Power* > 0.999** with 101 studies (no assumption about individual studies)
    - Even a *Power* of 0.8 could be reached with < 20 studies
- SAS macros are provided to conduct power analysis based on classical methods
  – Methods not given in workshop

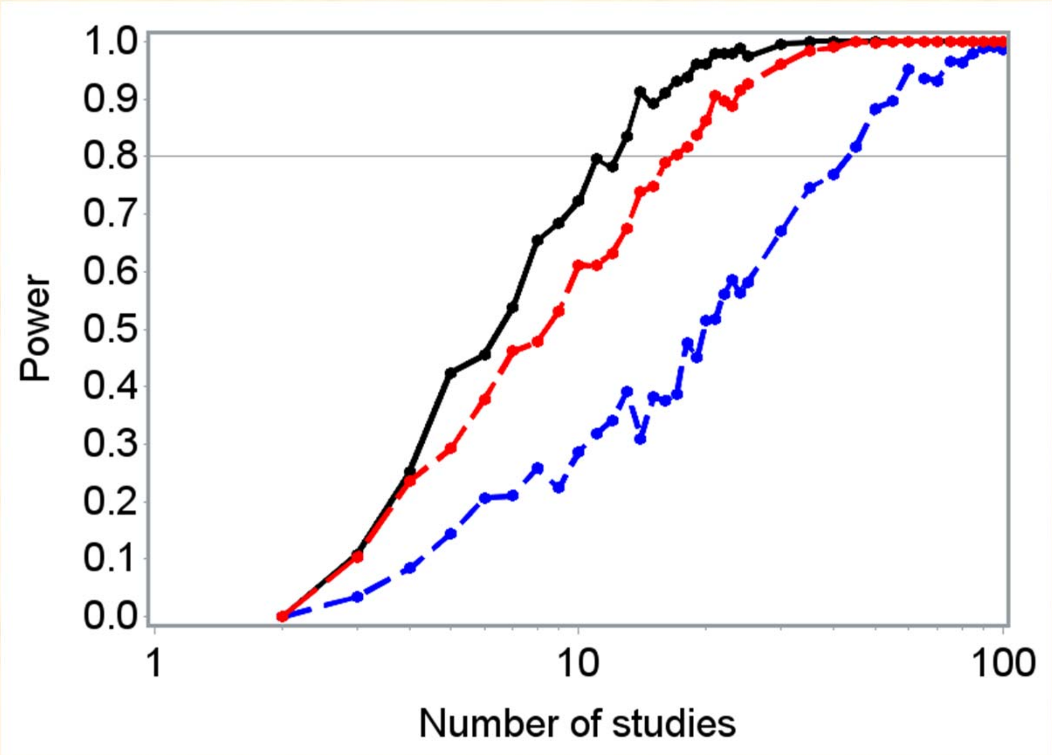
Blue: theoretical
Red: simulation



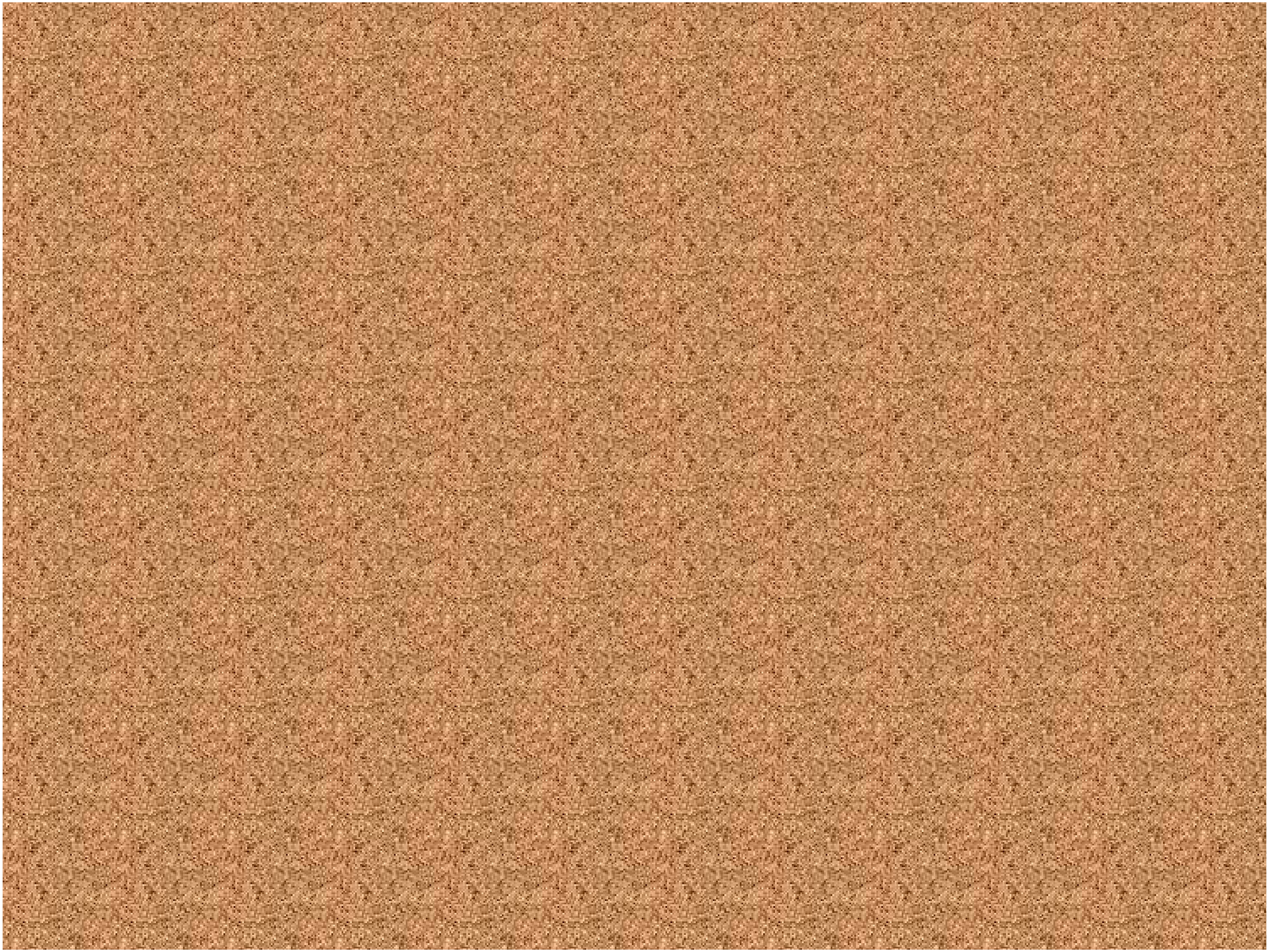
See Madden & Paul (2011) for details

Case study 1: Power for three scenarios

ζ = -0.244, σ^2 = 0.0365
ζ = -0.244, σ^2 = 0.072
ζ = -0.122, σ^2 = 0.0365



Separate simulations for each value of K



Madden (Ohio State Univ.)

The fallacy of counting *P* values (instead of doing a meta-analysis)

- *Suppose* K independent studies were conducted, and that there is *truly* a significant treatment effect (say, $\upsilon_i < 0$) in <u>every</u> study (i.e., H_a is always true) **-- returning to our hypothetical scenario**
- But also *suppose* that individual-study power is 0.40 (not a very high chance of detecting the true effect)
- A typical "qualitative" ("narrative") summary is to count the number of significant results (studies where $P \leq 0.05$): vote counting
  - **Conclude that the treatment is effective if *at least half* the studies are significant**
- With a large number of studies (say, K = 150), ~ 40% will have significant results (on average) with this power
  - **Thus, one would falsely conclude here that treatment was <u>not</u> effective, even though it was (truly) effective in <u>every</u> study.**

Fallacy of counting P values

- As the number of studies *increases*, it becomes *less and less* likely to every find 50+% of the studies with significant results (when individual power < 0.5).
  - **In fact, there is a higher chance of finding 50+% of the studies with significant results if fewer studies are considered (a major violation of good statistical practice)**
- Demonstration:
  - Chance of at least half the studies being significant ($P \leq 0.05$) when H_a is always true and individual-study power is **0.40** (low, but higher than in example)

Studies	*Prob*
10	0.17
20	0.13
30	0.10
50	0.06
100	0.02

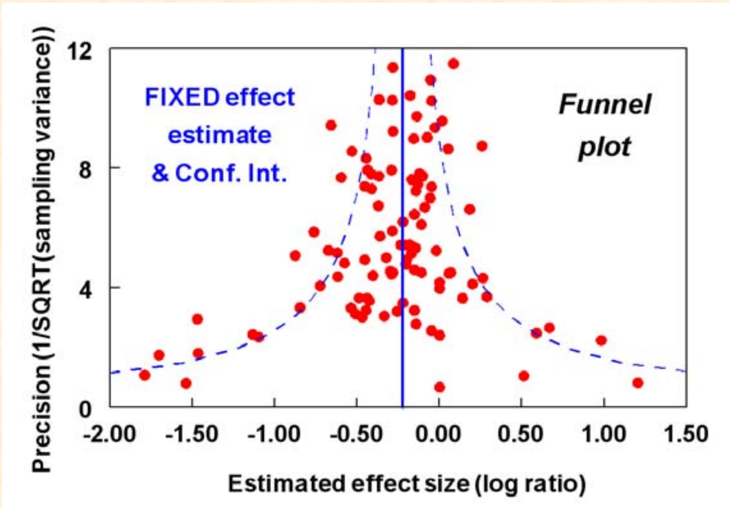
With a small number of studies, one actually has a better chance of finding half (or more) of the studies being significant

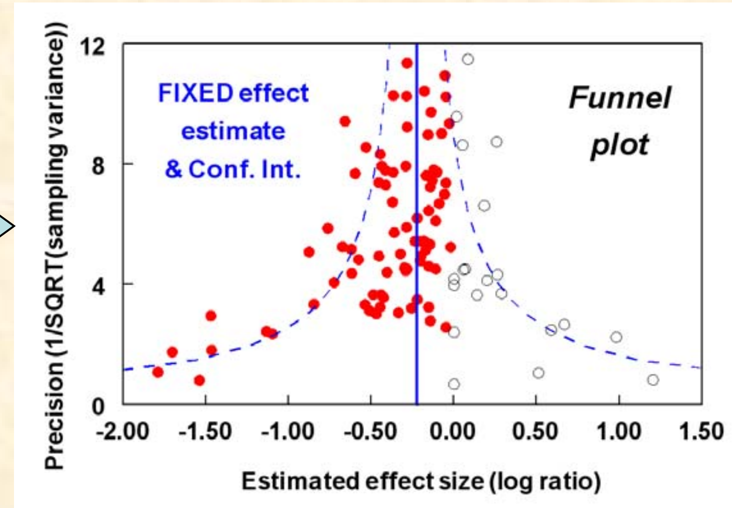
There are valid ways to combine *P* values to determine overall significance (going back to work by Fisher), but these are not discussed here. SAS macro written for this.

Publication bias

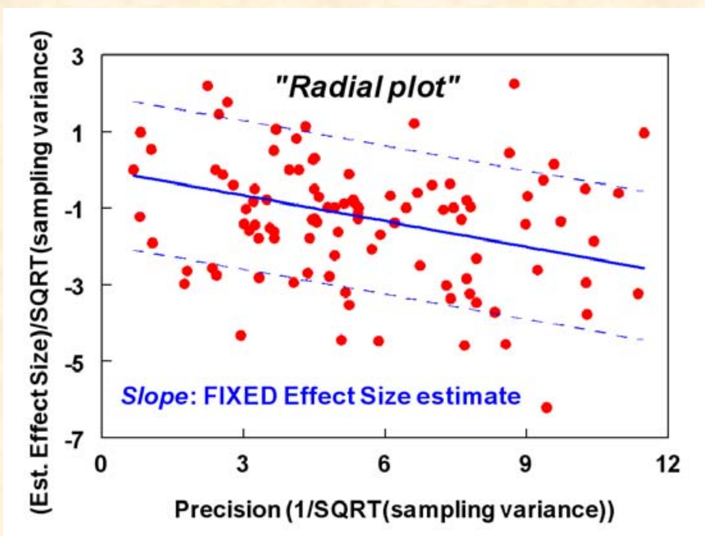
- Most meta-analyses make the tacit assumption that the studies under review are a random sample from a hypothetical population of possible studies, or that the study effects comprise a random sample from a hypothetical population of effects (Higgins et al. 2009) (i.e., that the studies are exchangeable)
  - Unlikely to be true, of course
  - It is likely that studies with significant results, or studies that support current dogma, or studies from famous laboratories, or studies from scientists trying to get tenure, have a higher probability of being published or being made available
  - The "*nightmare*" of meta-analysis (van Houwelingen, 1997).
- If inclusion of a study in the dataset depends on the realized effect size or p value, then the meta-analytical results (fixed effect parameters and variance-covariance parameters will be biased)
- Not of concern, for the most part, with Fusarium head blight case studies. The U.S. national initiative encouraged the 'publication' of all studies in proceedings and reports

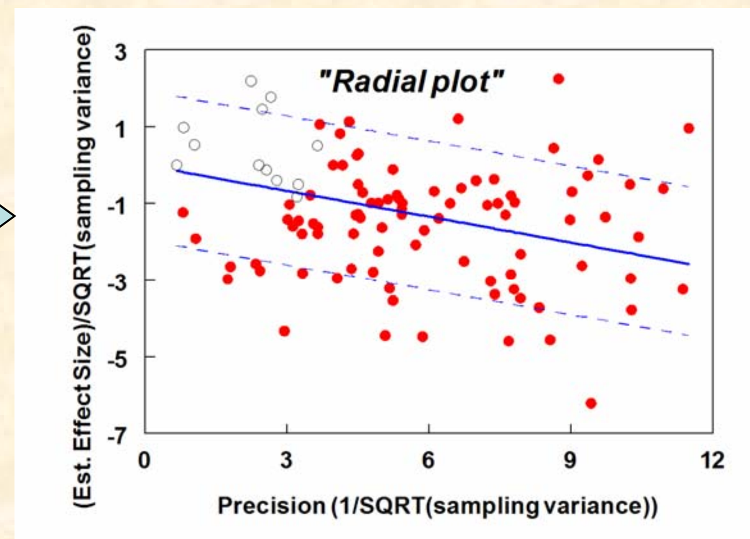
Publication bias: Plots *may* help





If no bias, there should be a random scatter around the line (no gaps at certain precisions or at certain effect sizes) – a (rough) guide only (especially with small K)





Publication bias: Solutions

- Ignore the “selection bias” of studies (usual “solution”)
- Use various analytical methods (including weighting of effect sizes and/or studies), based on various assumptions regarding the study selection process
  - Conduct sensitivity analysis to see consequences of different selection choices, which can lead to a bias adjustment
  - Many publications in this area (e.g., Sutton et al. 2000)
- However, *it is impossible to determine the study-selection mechanism from the available studies*
- A very interesting fairly new alternative is to determine the **upper bound on the bias** for any number of unpublished studies (i.e., for any study-selection probability)
  - Copas & Jackson (Biometrics [2004]) show that the absolute value of the bound for bias (for <u>any</u> selection mechanism) is straight-forward to calculate
  - Only assume that, on average, *lower precision studies cannot have a greater chance of selection than higher precision studies*

Copas and Jackson (2004):
Bounds for publication bias

$$|\,bias\ bound\,| = \frac{K+m}{K}\,\phi\left\{\Phi^{-1}\left(\frac{K}{K+m}\right)\right\}\frac{\sum_i^K\left(s_i^2+\sigma^2\right)^{-0.5}}{\sum_i^K\left(s_i^2+\sigma^2\right)^{-1}}$$

m	Index for the unobserved study ($m = 1, \ldots, M$) (possibly choose M to be $2K$)
$K+m$	The hypothetical total number of studies (with K being observed)
$K/(K+m)$	Study selection probability*
$\Phi^{-1}(.)$	Inverse standard normal cumulative distribution function
$\phi\{.\}$	Standard normal density function

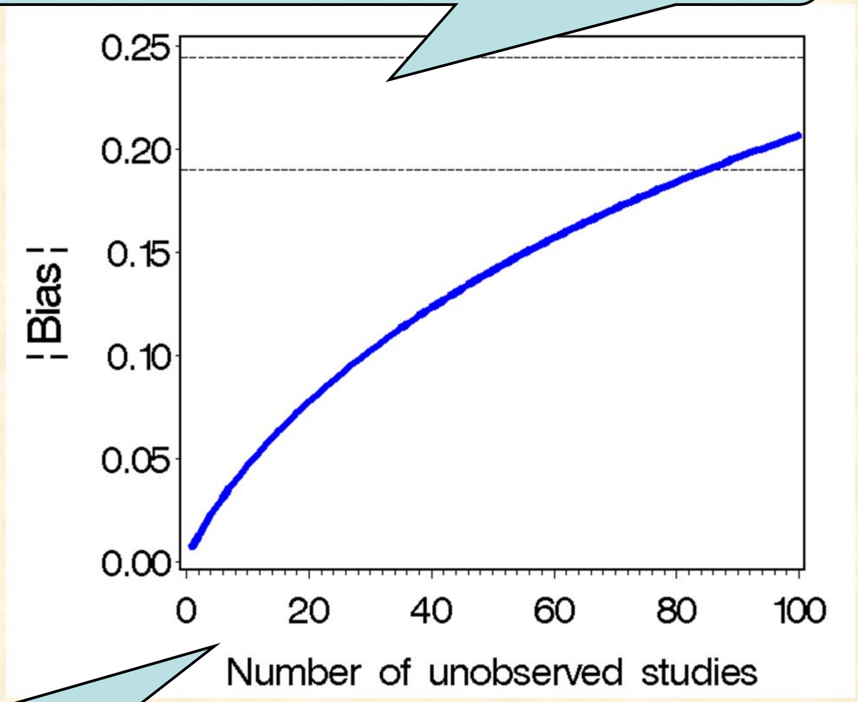
* One does not know the selection probability, but one determines the upper bound for bias for a range of possible selection probabilities: $K/(K+1)$, $K/(K+2)$, … $K/(K+M)$

See several articles by Copas and colleagues for extensions of this approach

Upper bound for bias: Case Study 1

- Fusarium head blight example (log response ratio)
  - K = 101 studies
  - Effect size: log ratio
    - **Mean = -0.24**
    - **Among-study variance = 0.0365**

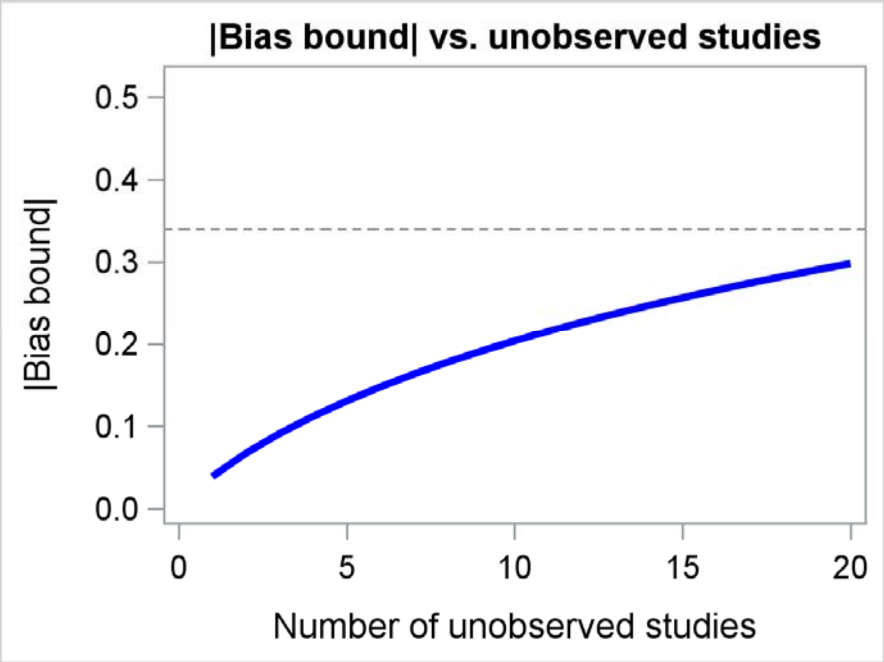
Compare **|Bias|** to **|mean|** from the published studies, or **|mean|-(2·SE)**



Example, if there are **20** unpublished studies, the total number of studies is 121 (not 101), with a selection probability of 101/121 = 0.84.
The mean effect size could be as *large* as -0.24**+0.077** (-0.163) or as *small* as -0.24**-0.077** (-0.317)

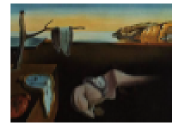
Upper bound for bias: Case Study 2

- Yield loss in relation to disease severity
  - K = 20 studies
  - Effect size: slope
    - **Mean = 0.52**
    - **Among-study variance = 0.118**



Several R packages, but **metafor** may be the most comprehensive. Actively supported, with updates and new features added periodically.

Will also calculate estimated effect sizes from original observations (for some situations).



Journal of Statistical Software

August 2010, Volume 36, Issue 3. *http://www.jstatsoft.org/*

Conducting Meta-Analyses in R with the metafor Package

Wolfgang Viechtbauer
Maastricht University

The metafor package
A meta-analysis package for R

Login

Recent changes Media Manager Sitemap

Trace: · news · features · forest_plot_with_subgroups · plots · funnel_plot_variations · **metafor**

Navigation

- **Main Page**
- Package Features
- News and Updates
- Download and Installation
- Documentation and Help
- Analysis Examples
- Plots and Figures
- List of Articles
- FAQs

metafor

The metafor package: A meta-analysis package for R

The metafor package is a free and open-source add-on for conducting meta-analyses with the statistical software environment R. The package consists of a collection of functions that allow the user to calculate various effect size or outcome measures, fit fixed-, random-, and mixed-effects models to such data, carry out moderator and meta-regression analyses, and create various types of meta-analytical plots.

On this website, you can find a more detailed description of the current and planned package features, news and updates concerning the package, instructions for downloading and installing the package, information on how to obtain further documentation and help when using the package, several analysis examples showing how to apply various meta-analytic models and methods described in the literature with the package, a little showcase of plots and figures that can be created with the package, an (incomplete) list of articles that have used the metafor package (or its predecessor, the 'mima' function) as part of their analyses, and a frequently asked questions section.

The metafor package was written by Wolfgang Viechtbauer. It is licensed under the GNU General Public License Version 2.

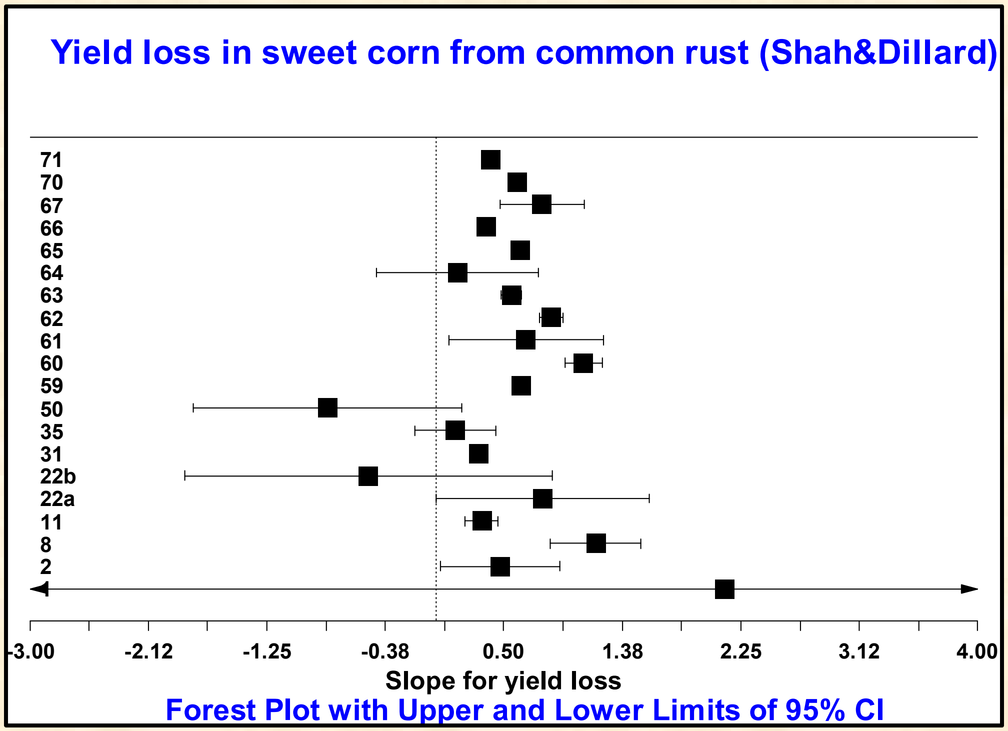
metafor.txt · Last modified: 2013/10/27 13:30 by wviechtb

```
> RustData<-read.table("C:/.../SweetCornRust.txt", header=TRUE)
```

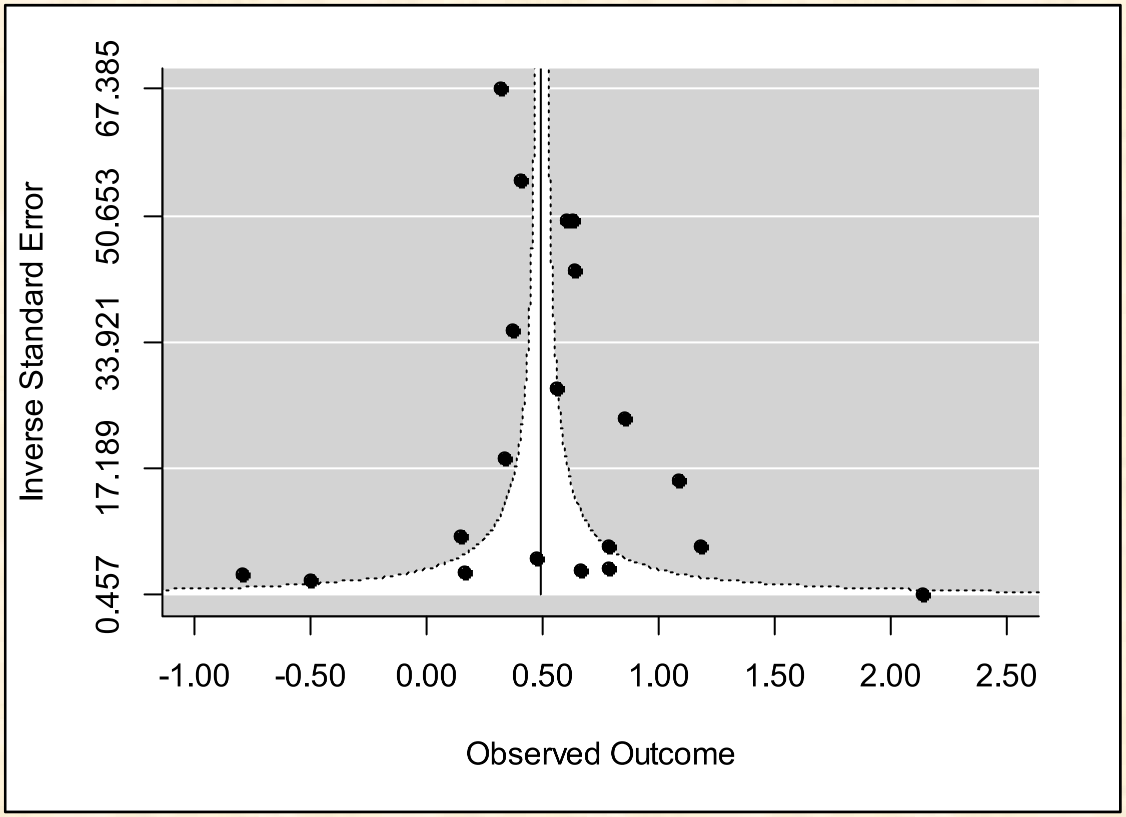
```
> forest(rev(RustData$slope), ci.lb=rev(RustData$lowerlimit), ci.ub=rev(RustData$upperlimit),
  annotate=FALSE, xlab="Slope for yield loss", font=2, slab=rev(RustData$Study), alim=c(-3, 4),
  cex.lab=1.5, pch=15, step=17, psize=2, cex=1.25, cex.axis=1.25,xlim=c(-3,4))

> title("Yield loss in sweet corn from common rust (Shah&Dillard)", sub = "Forest Plot with Upper and
Lower Limits of 95% CI", cex.main = 2, font.main= 2, col.main= "blue",cex.sub = 1.75, font.sub = 2,
col.sub = "blue")
```

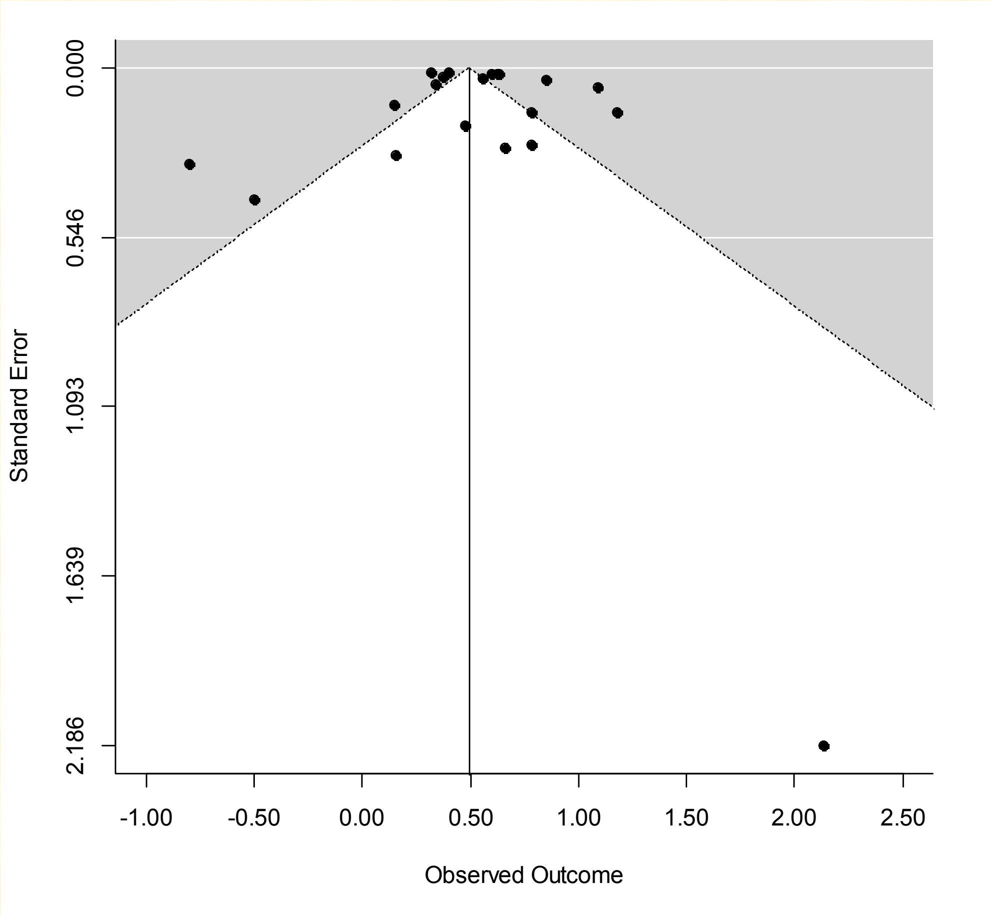
Most code is for annotation and labeling



```
> CornRust_Fixed<- rma(slope, SE^2, method="FE", data=RustData)
> funnel(CornRust_Fixed,yaxis="seinv",xlim=c(-1,2.5))
```



```
> CornRust_Fixed<- rma(slope, SE^2, method="FE", data=RustData)
> funnel(CornRust_Fixed,yaxis="sei",xlim=c(-1,2.5))
```



Several methods available, including “ML”, “DL”, others

```
> CornRust_Random_reml<- rma(slope, SE^2, method="REML", data=RustData)
> CornRust_Random_reml

Random-Effects Model (k = 20; tau^2 estimator: REML)
tau^2 (estimate of total amount of heterogeneity): 0.1181 SE = 0.0456)
tau (sqrt of the estimate of total heterogeneity): 0.3436
I^2 (% of total variability due to heterogeneity): 98.95%
H^2 (total variability / sampling variability):    95.63

Test for Heterogeneity:
Q(df = 19) = 523.0911, p-val < .0001

Model Results:
estimate      se     zval    pval    ci.lb    ci.ub
  0.5164   0.0856   6.0327   <.0001   0.3486   0.6842      ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

τ^2 is σ^2 (our notation)

I^2 and H^2 not determined from Q (be careful)

Q determined after a fixed-effects analysis (has no role in the REML analysis)

Many diagnostic plots (residuals, etc.) can also be produced. Also, EBLUPs.

Moderator variable analysis

```
rma(slope,SE^2, method="REML", data=RustData,mods=MeanD)

Mixed-Effects Model (k = 20; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.1297 (SE = 0.0511)
tau (square root of estimated tau^2 value):             0.3602
I^2 (residual heterogeneity / unaccounted variability): 98.87%
H^2 (unaccounted variability / sampling variability):   88.55
R^2 (amount of heterogeneity accounted for):            0.00%

Test for Residual Heterogeneity:
QE(df = 18) = 363.9921, p-val < .0001

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 0.2639, p-val = 0.6074

Model Results:

         estimate     se    zval    pval    ci.lb   ci.ub
intrcpt    0.4350 0.1778  2.4468  0.0144   0.0865  0.7834  *
mods       0.0033 0.0064  0.5137  0.6074  -0.0093  0.0159

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

R^2 here is the traditional statistic for explained variability (**not** the ratio of squared SEs)

Function	Description
print()	standard print method
summary()	alternative print method that also provides fit statistics
coef()	extracts the estimated model coefficients, corresponding standard errors, test statistics, p values, and confidence interval bounds
vcov()	extracts the variance-covariance matrix of the model coefficients
fitstats()	extracts the (restricted) log likelihood, deviance, AIC, and BIC
fitted()	fitted values
predict()	fitted/predicted values (with confidence intervals), also for new data
blup()	best linear unbiased predictions (BLUPs) of the true outcomes
residuals()	raw residuals
rstandard()	internally standardized residuals
rstudent()	externally standardized (studentized deleted) residuals
hatvalues()	extracts the diagonal elements of the hat matrix
weights()	extracts the weights used for model fitting
influence()	various case and deletion diagnostics
leave1out()	leave-one-out sensitivity analyses for fixed/random-effects models
forest()	forest plot
funnel()	funnel plot
radial()	radial (Galbraith) plot
qqnorm()	normal quantile-quantile plot
plot()	general plot function for model objects
addpoly()	function to add polygons to a forest plot
ranktest()	rank correlation test for funnel plot asymmetry
regtest()	regression tests for funnel plot asymmetry
trimfill()	trim and fill method
confint()	confidence interval for the amount of (residual) heterogeneity in random- and mixed-effects models (confidence intervals for the model coefficients can also be obtained)
cumul()	cumulative meta-analysis for fixed/random-effects models
anova()	model comparisons in terms of fit statistics and likelihoods
permutest()	permutation tests for model coefficients

Table 1: Functions and methods for fitted model objects created by the `rma.uni()` function.

Functions in `metafor` package

Multi-treatment analysis (random study effects)

$$y_{ij} = \beta_i + \tau_j + u_{ij} + \varepsilon_{ij} \qquad \beta_i \sim N(0, \sigma_\beta^2),\ u_{ij} \sim N(0, \sigma^2),\ \varepsilon_{ij} \sim N(0, s_{ij}^2)$$

$$y_{ij} = \tau_j + h_{ij} + \varepsilon_{ij}\ , \quad h_{ij} = \beta_i + u_{ij} \quad \boxed{E(y_{ij}) = \tau_j \quad E(y_{i1} - y_{i2}) = \tau_1 - \tau_2}$$

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iq})^T \quad \boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_q)^T \quad \mathbf{h}_i = (h_{i1}, h_{i2}, \ldots, h_{iq})^T \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iq})^T$$

Treatment effects

$$\mathbf{y}_i = \boldsymbol{\tau} + \mathbf{h}_i + \boldsymbol{\varepsilon}_i \qquad \mathbf{h}_i \sim N(\mathbf{0}, \mathbf{G}) \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i) \quad \mathbf{y}_i \sim N(\boldsymbol{\tau}, \mathbf{G} + \mathbf{R}_i)$$

Among-study variance-covariance matrix (generalization of σ^2)

Within-study variance-covariance matrix (generalization of s_i^2)

Compound Symmetry (CS), but other structures can be defined

$$\mathbf{G} = \begin{pmatrix} \sigma_\beta^2 + \sigma^2 & \sigma_\beta^2 & \cdots & \sigma_\beta^2 \\ \sigma_\beta^2 & \sigma_\beta^2 + \sigma^2 & & \\ \vdots & & \ddots & \\ \sigma_\beta^2 & & & \sigma_\beta^2 + \sigma^2 \end{pmatrix} \qquad \mathbf{R}_i = \begin{pmatrix} s_{i1}^2 & 0 & \cdots & 0 \\ 0 & s_{i2}^2 & & \\ \vdots & & \ddots & \\ 0 & & & s_{iq}^2 \end{pmatrix}$$

Within-study variance-covariance matrix

- It is computationally useful to use a <u>diagonal</u> $\mathbf{R}_i$ matrix (i.e., 0s for the covariances [off-diagonal elements])
  - Computationally, one can just use weights for each treatment within a study, while holding the residual variance fixed at 1
- With many experimental designs, means are correlated. For RCBD:

Residual variance: V_i Block variance: V_{bi} Number of blocks: n_i

Variance of mean:	$(V_{bi} + V_i)/n_i$	
Covariance of two means:	V_{bi} / n_i	
Variance of **difference**:	$2V_i / n_i$	(block variance cancels out)

Using just variances of means is easier, but one ends up with **incorrect <u>variances</u> of differences.**

$$\mathbf{R}_i = \begin{pmatrix} \frac{V_{bi}+V_i}{n_i} & \frac{V_{bi}}{n_i} & \cdots & \frac{V_{bi}}{n_i} \\ \frac{V_{bi}}{n_i} & \frac{V_{bi}+V_i}{n_i} & & \\ \vdots & & \ddots & \\ \frac{V_{bi}}{n_i} & & & \frac{V_{bi}+V_i}{n_i} \end{pmatrix}$$

Actual within-study var.-cov. matrix. Can be used, but tedious to fit. **May not know block variance**!

$$\mathbf{R}_i = \begin{pmatrix} \frac{V_{bi}+V_i}{n_i} & 0 & \cdots & 0 \\ 0 & \frac{V_{bi}+V_i}{n_i} & & \\ \vdots & & \ddots & \\ 0 & & & \frac{V_{bi}+V_i}{n_i} \end{pmatrix}$$

Resulting **incorrect** var. of diff.: $2\cdot(V_{bi} + V_i)/n_i$

Within-study variance-covariance matrix

- Alternative: use **½ of the variance of the difference** as the diagonal elements of the diagonal matrix
  - Variance of difference (balanced case): $2V_i / n_i$

$$\mathbf{R}_i = \begin{pmatrix} s_{i1}^2 & 0 & \cdots & 0 \\ 0 & s_{i2}^2 & & \\ \vdots & & \ddots & \\ 0 & & & s_{iq}^2 \end{pmatrix} \Longleftrightarrow \mathbf{R}_i = \begin{pmatrix} V_i/n_i & 0 & \cdots & 0 \\ 0 & V_i/n_i & & \\ \vdots & & \ddots & \\ 0 & & & V_i/n_i \end{pmatrix}$$

SE of the treatment **mean** would be incorrect, but interest is in the pairwise *differences* of means

Produces **correct** variance of difference within study: $2V_i \; / \; n_i$

Results in **correct** estimated mean effect size (treatment *difference*) and SE of the estimated mean effect size in the meta-analysis (e.g., estimated τ_1-τ_3)

- See Möhring & Piepho (2009 Crop Sci.) for other choices for a diagonal $\mathbf{R}$ matrix (including for unequal sampling variances)
  - Approach advocated here is most convenient for meta-analysis because the variance of the difference (derived from LSD, etc.) is often available
  - Their motivation was for two-stage analysis of variety trials, but work applies to meta-analysis