AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

Quantitative Synthesis

Chapter 4. Quantifying, Testing, and Exploring Statistical Heterogeneity

Prepared for:
The Agency for Healthcare Research and Quality (AHRQ)
Training Modules for Systematic Reviews Methods Guide
www.ahrq.gov

Learning objective

- Distinguish between (1) clinical and methodological heterogeneity, and (2) statistical heterogeneity.

Statistical heterogeneity

- ***Statistical heterogeneity*** occurs when estimates across studies have greater variability than expected from chance.[1,2]
  - Statistical heterogeneity is expected and must be quantified in meta-analysis.[3]
  - Once a body of studies has been identified, statistical heterogeneity often remains even after accounting for clinical/methodological heterogeneity (Chapter 1).
- Graphical and quantitative exploration should be used in combination.[2]

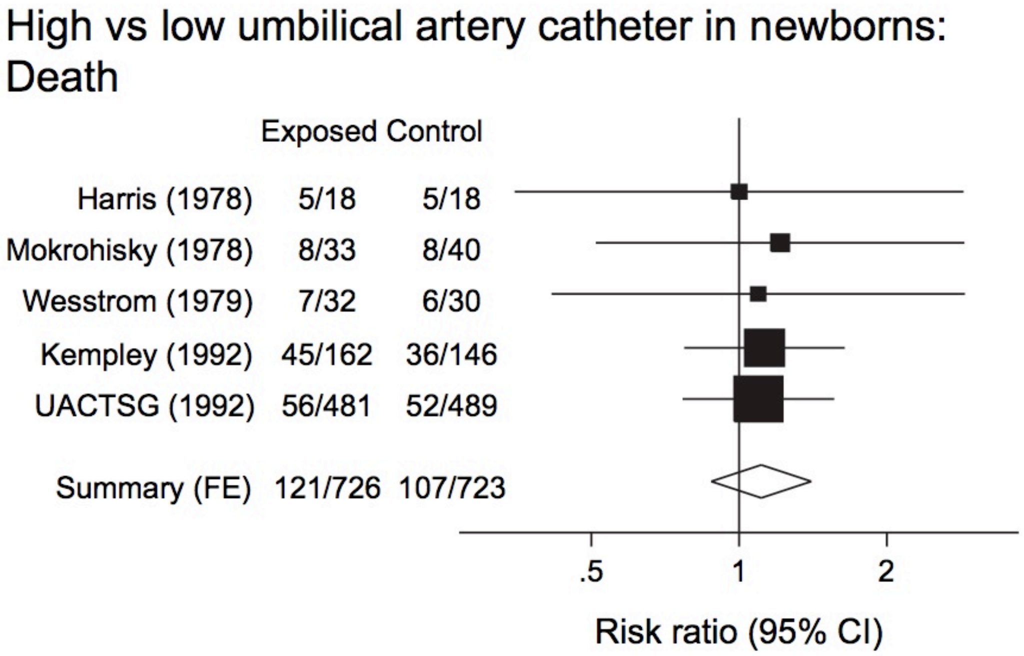
[1] Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002. http://dx.doi.org/10.1136/bmj.d4002 114.

[2] Langan D, Higgins JP, Simmonds M. An empirical comparison of heterogeneity variance estimators in 12,894 meta-analyses. Res Synth Methods. 2015;6(2):195-205. http://dx.doi.org/10.1002/jrsm.1140

[3] Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol. 2008;37(5):1158-60. https://doi.org/10.1093/ije/dyn204

Forest plots

- ***Forest plots*** are useful to identify sources of and extent of statistical heterogeneity.
- Meta-analysis with limited heterogeneity produces forest plots with a high degree of overlap of confidence internals and the summary estimate:[4]



High vs low umbilical artery catheter in newborns: Death

	Exposed	Control
Harris (1978)	5/18	5/18
Mokrohisky (1978)	8/33	8/40
Wesstrom (1979)	7/32	6/30
Kempley (1992)	45/162	36/146
UACTSG (1992)	56/481	52/489
Summary (FE)	121/726	107/723

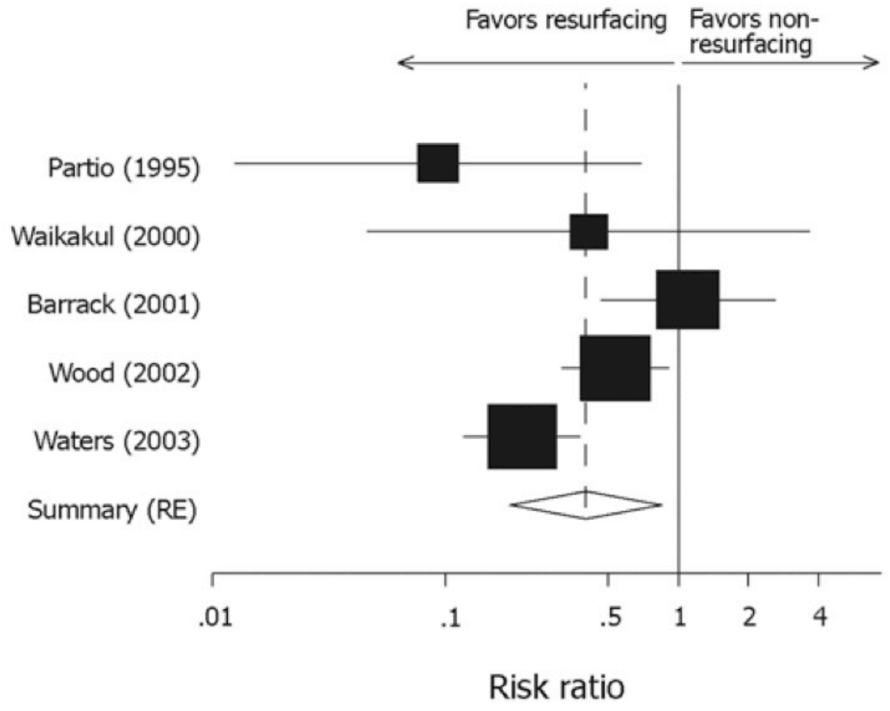
[4] Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: An overview with suggestions for practice. Res Synth Methods. 2010;1(1):66-80. http://dx.doi.org/10.1002/jrsm.6

[5] [Forest plot example taken from] Barrington KJ. Umbilical artery catheters in the newborn: effects of position of the catheter tip. Cochrane Database Syst Rev 2000;(2):CD000505. http://dx.doi.org/10.1002/14651858.CD000505

Forest plots

- Poor overlap between study confidence intervals and the summary estimate is a crude sign that statistical heterogeneity exists:[4]

Patellar Resurfacing in Total Knee Arthroplasty for Pain



RE = random effects model

6. [Forest plot example reprinted from] Pakos E, et al. Patellar resurfacing in total knee arthroplasty. A meta-analysis. J Bone Joint Surg Am 2005;87:1438-45 10.2106/JBJS.D.02422, with permission from Rockwater, Inc.

Random effects forest plots

- Graphically present between-study variance on forest plots of random effects meta-analyses using ***prediction intervals*** on the same scale as the outcome.[7]
  - The 95% prediction interval where the true estimate should lie for 95% of future studies.
  - When between-study variance exists, the prediction interval will cover a wider range than the confidence interval.
- The prediction interval should be represented using a rectangle at the bottom of the forest plot.
  - In contrast with the confidence interval, presented with a diamond.

[7] IntHout J, Ioannidis JP, Rovers MM, et al. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 2016;6(7):e010247. PMID: 27406637. http://dx.doi.org/10.1136/bmjopen-2015- 010247

Funnel plots

- **Funnel plots** aid in detecting heterogeneity (as well as publication bias).
  - They plot the effect sizes (X-axis) against the study precision (e.g., standard error, variance; Y-axis).
- A funnel plot without evidence of heterogeneity or publication bias resembles a symmetrical inverted funnel.[4]
  - The same underlying effect is estimated across levels of precision.
  - Heterogeneity and/or bias results in an asymmetrical scatter of studies around the summary effect size.

Asymmetry in funnel plots

- Asymmetry in funnel plots can be difficult to detect visually, and may owe to multiple contributing factors.[8]
  - Formal tests for funnel plot asymmetry exist, e.g., Egger's test.[9]
  - These should only be used in meta-analyses with ≥10 studies due to power issues.[1]
- Given these considerations, funnel plots should only be used to complement other approaches for detecting statistical heterogeneity.

[8] Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005;58(9):894-901. https://doi.org/10.1016/j.jclinepi.2005.01.006

[9] Egger M, Smith GD, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34. PMID: 9310563.

Quantifying heterogeneity

- For statistical heterogeneity in meta-analysis[10],
  - ► The ***null hypothesis*** is that all studies estimate the same effect.
  - ► The ***alternative hypothesis*** is that at least one study has an effect different from the summary effect.
- A common statistical heterogeneity tests statistical for statistical heterogeneity in meta-analysis is *Q*:[11]
  - ► *Q* is computed as the sum of squared deviations from each study's estimate from the summary estimate.
  - ► Each deviation is weighted in the same manner as the meta-analysis (e.g., inverse variance weighting).

[10] Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60. PMID: 12958120. http://dx.doi.org/10.1136/bmj.327.7414.557

[11] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88. https://doi.org/10.1016/0197-2456(86)90046-2

Q statistic of stat. heterogeneity

- The *Q* statistic is computed as:[12]

$$Q = \sum_{i=1}^{k} w_i(xi - \hat{x}_w)^2$$

Where:

w_i = study weight (inverse variance)
x_i = observed effect size of each trial
$\hat{x}_w$ = summary estimate (fixed effects)

- When P-value of *Q* statistic is low (typically P<0.1), the hull hypothesis of homogeneity can be rejected.[3,13]
- Do not interpret the *Q* statistic in isolation due to low power (especially with few studies)
- Non-significant *Q* statistic cannot be interpreted as evidence of heterogeneity.

[12] Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods. 2016;7(1):55- 79. http://dx.doi.org/10.1002/jrsm.1164
[13] Hoaglin DC. Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. Stat Med. 2016;35(4):485-95. http://dx.doi.org/10.1002/sim.6632

Between-study variance: τ^2

- **DerSimonian and Laird's $\boldsymbol{\tau}^2$** is another way of assessing between-study variance:[11]

$$\tau_{DL}^2 = \frac{Q-(k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$$

Where:

Q = heterogeneity statistic

k - 1 = degrees of freedom

w = study weight (inverse variance)

- Variance cannot be <0, so if τ^2<0, it is set to 0.
- τ^2 value is incorporated into random-effects meta-analysis weights.

Between-study inconsistency: *I*²

- According to the 2019 Cochrane handbook, ***I*²** describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance):[14]

$$I^2 = \frac{Q - (k-1)}{Q} * 100$$

Where:
Q = heterogeneity statistic
k - 1 = degrees of freedom

- Heterogeneity will always exist whether or not we happen to be able to detect it using a statistical test.
- Range of *I*² is 0% (no heterogeneity) to 100% (all variance attributable to heterogeneity).
- Although *I*² is easily interpreted, it is sample-size dependent (increases as study precision increases).
- *I*² is calculated based on *Q* or τ^2, so any issues affecting these statistics will also affect *I*².
- *I*² should be calculated even when *Q* is not statistically significant.

[14] Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

Interpreting I^2

- Thresholds for the interpretation of the I^2 statistic can be misleading, since the importance of inconsistency depends on several factors.[14]
- The Cochrane Handbook recommends this rough guide to evaluating I^2 and the degree of heterogeneity:
  - 0%-40% = might not be important
  - 30%-60% = may represent moderate heterogeneity*
  - 50%-90% = may represent substantial heterogeneity*
  - 75%-100% = considerable heterogeneity*

  *Importance of the observed value of I^2 depends on (1) magnitude and direction of effects, and (2) strength of evidence for heterogeneity (e.g. P value from the Chi^2 test, or a confidence interval for I^2)
- Uncertainty in the value of I^2 is substantial when the number of studies is small.
- I^2 = 0 cannot be interpreted as absence of heterogeneity; confidence limits must also be considered.

Meta-regression

- ***Meta-regression*** examines the degree to which study-level factors explain statistical heterogeneity.[15]
  - Random effects meta-regression is recommended over fixed-effects meta regression as it allows for residual heterogeneity in the model.[16]
  - A *t* distribution should be used to estimate statistical precision instead of a standard normal distribution (and is default in several statistical packages).[17]
    - This helps mitigate false-positives common in meta-regression.
- There should be at least 10 studies per characteristic modeled in a meta-regression.[14]

[15] Thompson SG, Higgins J. How should meta- regression analyses be undertaken and interpreted? Stat Med. 2002;21(11):1559-73. http://dx.doi.org/10.1002/sim.1187

[16] Berkey CS, Hoaglin DC, Mosteller F, et al. A random-effects regression model for meta-analysis. Stat Med 1995;14(4):395- 411. PMID: 7746979

[17] Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003;22(17):2693-710. http://dx.doi.org/10.1002/sim.1482

Meta-regression considerations

- Conceptual considerations must be taken into account:[14,18]
  - Study-level characteristics studied with meta-regression should be pre-specified and hypothesis-driven.
  - Not all study-level features that may modify treatment effects may be identified or measured; investigators should focus on plausible ones.
  - The ***ecological bias*** must be remembered and taken into account.[19]
    - Associations across studies may be different from associations within a study.
- ***Multiple meta-regression*** is also an option, and requires consideration of these factors as well (e.g., ≥10 studies per factor; pre-specified variables).[10,14,16]

[18] Gagnier JJ, Morgenstern H, Altman DG, et al. Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. BMC Med Res Methodol. 2013;13(1):106. http://dx.doi.org/10.1186/1471-2288-13-106

[19] Berlin JA, Santanna J, Schmid CH, et al. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med. 2002;21(3):371-87. http://dx.doi.org/10.1002/sim.1023

Subgroup analysis

- ***Subgroup analysis*** is a form of meta-regression with a categorical study-level factor.
  - Like other meta-regressions approaches, the subgroups should be pre-specified, hypothesis-driven, and scientifically defensible.[14,18]
- Subgroup analysis also has a high false-positive rate, especially with few studies.[10]
- Two approaches exist to subgroup analysis:
  - **1.** It is common to perform separate meta-analyses within subgroups, without statistical comparison.[14]
  - **2.** It is recommended to incorporate the subgroup factor into a meta-regression framework.[20]
    - This enables formal statistical testing, allows for quantification of heterogeneity and testing for residual heterogeneity.

[20] Borenstein M, Higgins JPT. Meta-Analysis and Subgroups. Prev Sci. 2013;14(2):134- 43. http://dx.doi.org/10.1007/s11121-013- 0377-7

Detecting outliers

- Identifying outlier studies that add bias may justify their removal from meta-analysis.[14]
  - Visual inspection of forest, funnel, normal probability, and Baujat plots may be helpful.
  - Another approach is to calculate summary effects without each study (i.e., ***one study removed***), to identify influential studies.
  - ***Cumulative meta-analysis*** graphs the accumulation of evidence across trials, incorporating information up to and including each trial.[21]

21. Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327(4):248-54. http://dx.doi.org/10.1056/NEJM199207233270406

Baseline risk meta-regression

- The baseline risk (or the “control rate”) is the proportion of the control group experiencing the outcome.[22,23]
  - This baseline risk can serve as a surrogate for covariate differences between studies.
  - Baseline risk is affected by illness severity, other treatments, follow-up duration, and other between-study differences.
  - Thus baseline risk can be used to assess for potential interactions with treatment.
- To examine for an interaction between baseline risk and treatment effects:[23,24]
  - Create scatter plot of treatment effect (RR or OR) and baseline risk, to graphically assess for association.
  - Then, use hierarchical meta-regression or Bayesian meta-regression to formally test for interaction between baseline risk and treatment effects.

22. M.W. M. The population risk as an explanatory variable in research synthesis of clinical trials. Stat Med. 1996;15(16):1713- 28. http://dx.doi.org/10.1002/(SICI)10970258(19960830)15:163.0.CO;2-D144
23. Schmid CH, Lau J, McIntosh MW, et al. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med. 1998;17(17):1923-42.
24. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med. 1997;16(23):2741-58.

Recommendations

- Expect, visually inspect, quantify, and sufficiently address statistical heterogeneity in all meta-analyses.
- Include prediction intervals in all forest plots.
- Consider evaluating multiple metrics of heterogeneity, between-study variance, and inconsistency (i.e., Q, τ^2 and I^2 along with their respective confidence intervals when possible).
- A non-significant Q should not be interpreted as the absence of heterogeneity, and there are nuances to the interpretation of Q that carry over to the interpretation of τ^2 and I^2.
- Random effects is the preferred method for meta-regression that should be used under consideration of low power associated with limited studies (i.e., <10 studies per study-level factor) and the potential for ecological bias.
- A simplified two-step approach to control-rate meta-regression that involves scatter plotting and then hierarchical or Bayesian meta-regression is recommend.
- Routine use of multivariate meta-analysis not recommended.

Author

- This presentation was prepared by Jonathan Snowden, Ph.D.
- The presentation is based on the chapter entitled “Quantifying, Testing, and Exploring Statistical Heterogeneity” in the Methods Guide for Comparative Effectiveness Reviews (available at: https://doi.org/10.23970/AHRQEPCMETHGUIDE3

References

1. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002. http://dx.doi.org/10.1136/bmj.d4002 114.

2. Langan D, Higgins JP, Simmonds M. An empirical comparison of heterogeneity variance estimators in 12,894 meta-analyses. Res Synth Methods. 2015;6(2):195-205. http://dx.doi.org/10.1002/jrsm.1140

3. Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol. 2008;37(5):1158-60. https://doi.org/10.1093/ije/dyn204

4. Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: An overview with suggestions for practice. Res Synth Methods. 2010;1(1):66-80. http://dx.doi.org/10.1002/jrsm.6

5. [Forest plot example taken from] Barrington KJ. Umbilical artery catheters in the newborn: effects of position of the catheter tip. Cochrane Database Syst Rev 2000;(2):CD000505. http://dx.doi.org/10.1002/14651858.CD000505

6. [Forest plot example reprinted from] Pakos E, et al. Patellar resurfacing in total knee arthroplasty. A meta-analysis. Bone Joint Surg Am 2005;87:1438-45 10.2106/JBJS.D.02422, with permission from Rockwater, Inc.

7. IntHout J, Ioannidis JP, Rovers MM, et al. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 2016;6(7):e010247. PMID: 27406637. http://dx.doi.org/10.1136/bmjopen-2015- 010247

8. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J Clin Epidemiol. 2005;58(9):894-901. https://doi.org/10.1016/j.jclinepi.2005.01.006

9. Egger M, Smith GD, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629- 34. PMID: 9310563.

10. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60. PMID: 12958120. http://dx.doi.org/10.1136/bmj.327.7414.557

11. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88. https://doi.org/10.1016/0197- 2456(86)90046-2

12. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods. 2016;7(1):55- 79. http://dx.doi.org/10.1002/jrsm.1164

13. Hoaglin DC. Misunderstandings about Q and 'Cochran's Q test' in meta-analysis. Stat Med. 2016;35(4):485-95. http://dx.doi.org/10.1002/sim.6632

14. Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.

15. Thompson SG, Higgins J. How should meta- regression analyses be undertaken and interpreted? Stat Med. 2002;21(11):1559-73. http://dx.doi.org/10.1002/sim.1187

16. Berkey CS, Hoaglin DC, Mosteller F, et al. A random-effects regression model for meta-analysis. Stat Med 1995;14(4):395- 411. PMID: 7746979

17. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003;22(17):2693-710. http://dx.doi.org/10.1002/sim.1482

18. Gagnier JJ, Morgenstern H, Altman DG, et al. Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. BMC Med Res Methodol. 2013;13(1):106. http://dx.doi.org/10.1186/1471-2288-13-106

19. Berlin JA, Santanna J, Schmid CH, et al. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med. 2002;21(3):371-87. http://dx.doi.org/10.1002/sim.1023

20. Borenstein M, Higgins JPT. Meta-Analysis and Subgroups. Prev Sci. 2013;14(2):134- 43. http://dx.doi.org/10.1007/s11121-013- 0377-7

21. Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327(4):248-54. http://dx.doi.org/10.1056/NEJM199207233270406
22. M.W. M. The population risk as an explanatory variable in research synthesis of clinical trials. Stat Med. 1996;15(16):1713- 28. http://dx.doi.org/10.1002/(SICI)10970258(19960830)15:163.0.CO;2-D144
23. Schmid CH, Lau J, McIntosh MW, et al. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med. 1998;17(17):1923-42.