

AGENCY FOR HEALTHCARE RESEARCH AND QUALITY

AHRQ

Quantitative Synthesis

Chapter 2. Presenting Different Effects for Different Data Types

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

Learning Objective

- Recognize common measures of association for meta-analysis (e.g., risk difference, odds ratio).

Outcomes and effect measures

- Methods for meta-analysis depend on the outcome type:
  - ***Binary outcomes*** take 2 levels (e.g., alive or dead; hospital admission or not)
  - ***Continuous outcomes**** can take any value in the outcome range (e.g., weight, systolic blood pressure).
- Here we focus on these 2 most common types and not rarer ones (e.g., ordinal, time-to-event).

*More information on continuous outcomes can be found in Definitions and Background

Binary outcomes

- Ideally, the data necessary to calculate binary effect sizes will be published and the data can be analyzed using statistical analysis software.

- The data needed for analysis can be presented in this way:

Table 1. Example binary data for effect size computation

	Treatment Events in Treatment Group	Treatment n	Events in Control Group	Control n
Study X	5	25	6	25
Study Y	23	194	21	189

Binary outcomes

- Every effort should be made to obtain all data in Table 1, and efforts should be described in the write-up, per PRISMA guidance.[1]
  - Trials are required to report events and sample sizes for all study arms.
  - But in some instances, only one measure (e.g., odds ratio, risk difference) will be available in publications.
  - In this case, meta-analysis will be conducted using available data.

1. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. PMID: 19621072 http://dx.doi.org/10.1371/journal.pmed.1000097

Choosing an effect measure

- The 3 most common effect measures in meta-analysis are:
  - Risk difference (RD, absolute measure)
  - Risk ratio* (RR, relative measure)
  - Odds ratio* (OR, relative measure)

*Formulas for calculating RR and OR can be found in Definitions and Background

	Benefits	Disadvantages	Conditions For Use
Risk Difference (RD)	- More easily interpretable among lay audiences - On the familiar percentage scale - Can be converted to NNT or NNH for clinical interpretability - Can address zero-event studies	- Not consistent between studies with differing baseline risks. - Not commonly reported in individual trials. - Not preferred when there is heterogeneity between studies in duration and incident rates	- Preferred when outcome incidence is similar across studies - Less consistent across studies compared to RR and OR - Avoid the RD when outcome is rare or differs substantially across studies
Risk Ratio (RR)	- Easily interpretable - Commonly reported in individual trials considered in meta-analyses - More likely to be consistent even with differing baseline risks	- Values of “death” and “survival” are not reciprocals of each other as would be intuitively expected. - Dependent on arbitrary definition of event versus no event.	- More easily interpretable than OR
Odds Ratio (OR)	- More likely to be consistent even with differing baseline risks - Commonly reported in individual trials considered in meta-analyses	- Not easily interpretable - Can be misleading when interpreted like relative risks - Widespread use in meta-analyses may be because of convenience and history rather than an assessment of appropriateness	- Preferred if switching from event and non-event (e.g. death and survival) and raw data are not available

RR versus OR

- When considering these relative effect measures, note that:
  - If outcome is rare enough (<10%), then the RR and OR are effectively equivalent.
  - The RR is more easily interpreted than the OR.
  - The OR is non-collapsible: if adjustment is made using different covariates, then effect modification will appear to be different as well.
  - *These factors favor the RR in many circumstances.*

RR versus OR

- The odds ratio has some beneficial properties as well:
  - The OR for an event is reciprocal to the OR for a non-event (i.e., $OR_{death} = 1/OR_{survival}$)
  - The RR does not share this property of being reciprocal for events and non-events.
  - Thus, if switching from event and non-event (e.g., from death to survival) is necessary for meta-analysis and raw data are not available, *the OR is preferred*.
- The selection of the preferred relative effect measure depends on the specific meta-analysis.

Number needed to treat

- **Number needed to treat (NNT)*** is the number of patients that would need to be treated for one to benefit.[2]
  - This straightforward statistic is broadly understood by clinicians and lay stakeholders.
  - Wald method[3] is most common to calculate NNT confidence intervals, but ***the Wilson method*** is preferred due to superior coverage properties.[4]

*Formula for calculating NNT can be found in Definitions and Background

2. Schulzer M, Mancini GJ. 'Unqualified Success' and 'Unmitigated Failure' Number-Needed-to-Treat-Related Concepts for Assessing Treatment Efficacy in the Presence of Treatment-Induced Adverse Events. Int J Epidemiol. 1996;25(4):704-12.
3. Altman DG. Confidence intervals for the number needed to treat. BMJ. 1998;317(7168):1309. PMID: 9804726
4. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 1998;17(8):873-90. PMID: 9595617.

Recommendations for Chapter 2. Different effects for different data types

- **For binary outcomes:**
  - Consider carefully which binary measure to analyze.
  - Risk difference is the preferred measure if conversion to NNT or NNH is sought.
  - Risk ratio and odds ratio are likely to be more consistent than the risk difference when studies differ in baseline risk.
  - Risk difference is not the preferred measure when the event is rare.
  - Risk ratio is not the preferred measure if switching between occurrence and non-occurrence of the event is important to the meta-analysis.
  - The odds ratio can be misleading.

Recommendations for Chapter 2. Different effects for different data types

- **For continuous outcomes:**
  - When studies use the same metric, mean difference is preferred measure.*
  - When calculating standardized mean difference, Hedges' *g* is preferred over Cohen's *d* due to the reduction in bias.*
- **General:**
  - If baseline values are unbalanced, perform an ANCOVA analysis. If ANCOVA cannot be performed and the correlation is greater than 0.5, change from baseline values should be used to compute the mean difference.
    - If the correlation is less than or equal to 0.5, follow-up values should be used.
  - Data from clustered randomized trials should be adjusted for the design effect.*

*More information can be found in Definitions and Background

DEFINITIONS & BACKGROUND

Concepts reviewed in Definitions & Background

- Calculations for RR, OR, and NNT
- Continuous outcomes
- Mean difference
- Standardized mean difference (SMD)
- Cohen's *d* for SMD calculation
- Hedges' *g* for SMD calculation
- Ratio of means
- Crossover trials
- Cluster-randomized trials

Risk ratio

- The risk ratio is calculated using this formula:

Table 2: Organizing binary data for effect size computation			
	Events	*No Events*	*N*
Treatment	A	B	n_1
Control	C	D	n_2

- $RR = \frac{A/n_1}{C/n_2}$
- Metrics of dispersion are calculated on the log scale and then transformed back to RR scale:
  - $V_{LN(RR)} = \frac{1}{A} + \frac{1}{C} - \frac{1}{n_1} - \frac{1}{n_2}$
  - $SE_{LN(RR)} = \sqrt{V_{LNRR}}$

Where:

RR = risk ratio

$V_{LN(RR)}$ = variance of risk ratio

$SE_{LN(RR)}$ = standard error of RR

Odds ratio

- The odds ratio is calculated using this formula:

Table 2: Organizing binary data for effect size computation			
	Events	*No Events*	*N*
Treatment	A	B	n_1
Control	C	D	n_2

- $OR = \frac{AD}{BC}$
- Metrics of dispersion are calculated on the log scale and then transformed back to OR scale:
  - $V_{LN(OR)} = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$
  - $SE_{LN(OR)} = \sqrt{V_{LNOR}}$

Where:

OR = odds ratio

$V_{LN(OR)}$ = variance of odds ratio

$SE_{LN(OR)}$ = standard error of OR

Number needed to treat

- Number needed to treat (NNT) – number of patients that would need to be treated for one to benefit – is inverse of risk difference (RD):

$$NNT = \frac{1}{|RD|}$$

Continuous outcomes

- Required data for effect size computation:
  - Estimated ***differences*** between comparison groups
  - Estimated ***standard errors*** of those differences
- Differences may be obtained in various ways:
  - May be reported in the original study, as standardized mean difference or ratio of means.
  - May be computed, if means for each group are reported.
- Standard errors may be obtained:
  - May be reported by the original study
  - May be computed, if precision measures are reported (e.g., confidence interval, P-value, z-statistics).

Mean difference

- **Mean difference** is the most common measure for summarizing a continuous outcome.
  - Previously called the “weighted mean difference”
  - Mean difference at a given time point vs net mean difference
  - May be computed when all studies in the meta-analysis report outcome on the same scale or scales that can be easily converted (e.g., kg or lbs; mm Hg)
  - It is not possible to pool results of studies that measure outcome on scales that cannot be converted (e.g., differing Quality of Life measurement scales)
- Calculation of the mean difference is straightforward and requires means, sample sizes, and standard errors.

**This slide primarily refers to net mean difference, or the difference in change between two study arms. Investigators may instead consider the mean difference at a given time point (eg, mean final value with intervention vs. mean final value without intervention), ignoring baseline.*

Standardized mean difference

- **Standardized mean difference (SMD)** is used for outcomes measured on different scales that cannot be converted to a common scale.
  - SMD is computed as the mean effect divided by a pooled estimate of the standard error, putting all estimates on the same scale (i.e., standard deviations).
  - SMDs may then be pooled across studies.
  - If mean difference may be computed, it is preferred to SMD, because it is more easily interpretable.
- Several methods exist to calculate SMD, including Cohen's *d* and Hedges' *g*.

Cohen's *d* for SMD calculation

- **Cohen's *d*** is the simplest SMD computation.[5]
  - Defined as mean difference divided by the pooled standard deviation of the treatment and control groups:

$$d = \frac{m_T - mC}{S_{pooled}}$$

Where:

m_T = mean of treatment group

m_C = mean of control group

S_{pooled} = pooled standard error

  - This estimate is biased in estimating the true population SMD in small sample sizes.
  - Therefore, Hedges' *g* is more commonly used.

5. Fu R, Vandermeer BW, Shamliyan TA, et al. Handling Continuous Outcomes in Quantitative Synthesis Agency for Healthcare Research and Quality. Rockville, MD: 2013.

Hedges' *g* for SMD calculation

- **Hedges' *g*** is a transformation of Cohen's *d* that adjusts for small sample size bias.[5]
  - Transformation is achieved by multiplying Cohen's *d* by a function of the total sample size:

$$g = d\left(1 - \frac{3}{4N - 9}\right)$$

Where:
N = total trial sample size

  - For very large sample sizes Cohen's *d* and Hedges' *g* will be similar.

Ratio of means

- **Ratio of means (RoM)** is an alternative to the SMD.[5]
  - Also called the “response ratio.”
  - Calculated as the mean of the intervention group divided by the mean of the control group.
  - Interpreted as the percentage change in outcome in the treatment group as compared to control.
  - Assumes that the relative effect of treatment will be the same, regardless of the scale used to measure it.
- Using the RoM to pool requires 2 key assumptions:
  - All values of the outcome scale must be positive (or negative).
  - A value of zero must truly represent zero.

Special topics: Crossover trials

- In a ***crossover trial*** all participants receive both the treatment and control interventions in sequence.
- The correlation between the two study arms must be considered when computing standard errors.[5]
  - Generally the correlation is positive, resulting in smaller SEs than in a standard parallel trial.
- If correlation is available the pooled SE can be computed:

$$SE_p = \sqrt{SE_T^2 + SE_C^2 + 2rSE_TSE_C}$$

Where:

r = within-person correlation

$SE_{T/C/P}$ = standard error (treatment/control/pooled)

Special topics: Cluster-randomized trials

- ***Cluster randomized trials*** are those where participants are assigned to treatment or control in groups rather than individually.
- If units within clusters are correlated, then precision is lower than for an individually randomized trial of the same size.
- The intra-class correlation coefficient is defined as the proportion of total variance due to between-cluster variance, and quantifies cluster-related variability.[5]
- The Design Effect (DE) of cluster RCT is the multiplier needed to adjust the standard error to account for this precision loss:

$$DE = 1 + (M - 1)ICC$$

Where:

M = average cluster size

ICC = intra-class correlation coefficient

Author

- This presentation was prepared by Jonathan Snowden, Ph.D.
- The presentation is based on the chapter entitled “Optimizing Use of Effect Size Data” in the Methods Guide for Comparative Effectiveness Reviews (available at: https://doi.org/10.23970/AHRQEPCMETHGUIDE3

References

1. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. PMID: 19621072 http://dx.doi.org/10.1371/journal.pmed.1000097

2. Schulzer M, Mancini GJ. 'Unqualified Success' and 'Unmitigated Failure' Number-Needed-to-Treat-Related Concepts for Assessing Treatment Efficacy in the Presence of Treatment-Induced Adverse Events. Int J Epidemiol. 1996;25(4):704-12.

3. Altman DG. Confidence intervals for the number needed to treat. BMJ. 1998;317(7168):1309. PMID: 9804726

4. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 1998;17(8):873-90. PMID: 9595617.

5. Fu R, Vandermeer BW, Shamliyan TA, et al. Handling Continuous Outcomes in Quantitative Synthesis Agency for Healthcare Research and Quality. Rockville, MD: 2013.