Ruprecht-Karls-Universität Heidelberg

Masterarbeit zur Erlangung des akademischen Grades
"Master of Science in Medical Biometry/Biostatistics"

Calculation of Uniform Effect Measures Based on Continuous Outcomes Reported Differently or Incompletely by Studies Included in a Meta-Analysis

vorgelegt von

Benjamin Kendziora, M.D., Ph.D.

2023

Calculation of Uniform Effect Measures Based on Continuous Outcomes Reported Differently or Incompletely by Studies Included in a Meta-Analysis

vorgelegt von:	Benjamin Kendziora, M.D., Ph.D.
Studiengang:	M.Sc. Medical Biometry/Biostatistics
Betreuer und Erstgutachter:	Lorenz Uhlmann, M.Sc., Ph.D. Novartis AG, Basel, Schweiz
Zweitgutachter:	Univ.-Prof. Peter Schlattmann, M.D., M.Sc. Universitätsklinikum Jena, Jena, Deutschland

Contents

Abstract

Differences and incompleteness in reporting continuous outcomes across studies included in a meta-analysis complicate the calculation of consistent effect measures before pooling the results and may lead to trade-offs that result in the exclusion of relevant studies. In this master thesis, conversion and imputation methods were compared in terms of bias and included in the newly created *uniform* R package. Bias was compared using simulation studies with different data generation scenarios. Methods resulting in the highest precision were set as default. Version 0.0.1 of the *uniform* R package allows to uniform a differently or incompletely reported continuous outcome across studies and to calculate effect measures for interventional controlled studies.

Chapter 1

Introduction

Evidence based medicine can be defined as the conscientious, judicious, and reasonable use of available evidence in decision making about the care of individual patients [1]. Most scientific evidence gets published in journal articels. The history of scientific journal articles goes back to the year 1665, and to date, more than 60 million articles have been published, with a strong upward trend [2]. This huge and growing amount of scientific evidence makes it increasingly difficult for researchers and clinicians to keep up to date. Systematic reviews and meta-analyses try to summarize and synthesize the available evidence and in this way inform clinical decision making [3].

The synthesis of evidence in meta-analyses is done by pooling effect measures from individual studies. Both an effect measure and a measure of dispersion are needed per study for pooling effects across studies. Effect measures are statistical constructs that summarize the strength of the link between an exposure and an outcome [4]. Most commonly, an outcome is compared between an exposed and a nonexposed group for calculating an effect measure [5]. A continuous outcome for patient groups can be reported differently, with different measures of central tendency (e.g., mean or median per group) and dispersion (e.g., standard deviation vs. interquartile range). In addition, some studies do not report a measure of dispersion at all or

transform the continuous variable into a categorical one. These differences in reporting considerably complicate or prevent the calculation of consistent effect measures for pooling the results of individual studies in a meta-analysis and may lead to trade-offs that result in the exclusion of relevant studies. Different methods for the conversion of different measures of central tendency and dispersion, imputation of missing measures of dispersion, and approximate retransformation of categorically presented continuous outcomes have been introduced [5]. The computation of uniform effect measures for studies included in a meta-analysis frequently requires the combination of these methods. A clear guidance on which methods should be used is still missing.

This master's thesis was conducted with the goal of comparing different methods for the conversion and imputation of differently or incompletely reported continuous outcomes, setting up a guidance on which methods should be used, and creating an R package for the conversion and imputation of continuous outcomes and calculation of consistent effect measures.

Chapter 2

Notation

The following notation is used throughout the thesis.

Greek letters refer to population parameters. Roman letters are used for sample attributes. E.g., σ refers to a (theoretical) population standard deviation, while s refers to the standard deviation of a sample.

The subsripts $ugtp$ are used for most parameters. E.g., x_{ugtp} refers to a single value of an outcome measured in study u, in participant group g, at time point t, and in participant p. The subscript $*$ symbolizes that a parameter refers to several studies u, participant groups g, time points t, and/or participants p. E.g., $\overline{x}_{ugt*}$ refers to the mean of an outcome in participant group g within study u at time point t across participants p.

Chapter 3

Data Extraction for Continuous Outcomes

Depending on the measured outcome, data can be structured differently: dichotomous data, continuous data, ordinal data, counts and rates, and time-to-event data. These types of outcome data and especially continuous outcome data can be presented differently, which is summarized in Appendix 1.

Depending on the design of the studies included in a meta-analysis, measurements performed at one or more time points should be extracted. E.g., for a meta-analysis on the effect of an intervention on a continuous outcome, the preintervention/baseline measurement of the continuous outcome, postintervention measurement, and change from baseline should be extracted from each study for each group of observations with the corresponding number of observations, if available.

As mentioned previously, included studies commonly report these measurements differently or incompletely with different measures of central tendency and dispersion, with a measure of central tendency but without measure of dispersion, or even dichotomized. In the following chapter, previously introduced conversion and imputation methods for the conversion and

imputation of differently or incompletely reported continuous outcomes are described. Usage of these methods may enable the subsequent calculation of consistent effect measures for studies included in a meta-analysis, which is described in Chapter 5.

Chapter 4

Conversion and Imputation Methods

4.1 Obtaining Mean and Standard Deviation from Mean and Standard Error by Algebraic Conversion

The standard deviation s_{ugt*} of an independent continuous outcome measured in a study u, in a participant group g, at a time point t, and across participants p can be computed from the standard error se_{ugt*} by

$$s_{ugt*} = se_{ugt*} \times \sqrt{n_{ugt*}}, \tag{4.1}$$

where n_{ugt*} refers to sample size of the patient group g within the study u at a time point t [6].

4.2 Obtaining Mean and Standard Deviation from Mean and Confidence Interval by Algebraic Conversion

The standard deviation of independent observations can be computed from the 95% confidence interval by

$$s_{ugt*} = \frac{\sqrt{n_{ugt*}} \times (ul_{ugt*} - ll_{ugt*})}{3.92}, \tag{4.2}$$

where n_{ugt*} refers to the sample size of a participant group g within a study u at a time point t, ul_{ugt*} to the upper limit of the confidence interval, ll_{ugt*} to the lower limit of the confidence interval, and 3.92 to the value of the inverse cumulative density function of the standard normal distribution for 0.975 [6]. If the 90% confidence interval or the 99% confidence interval is given instead of the 95% confidence interval, 3.92 should be exchanged for 3.29 or 5.15, the values of the inverse cumulative density function of the standard normal distribution for 0.95 and 0.995.

4.3 Obtaining Mean and Standard Deviation from Mean and Range

4.3.1 The "Range"-Method

In case of normally distributed values of individual patients, 95% of values approximately lie within two standard deviations either side of the mean [5,6]. The standard deviation of a continuous outcome in a participant group g within a study u at a time point t may thus be estimated by

$$\hat{s}_{ugt*} = \frac{b_{ugt*} - a_{ugt*}}{4}, \tag{4.3}$$

where a_{ugt*} refers to the minimum and b_{ugt*} to the maximum.

4.3.2 Method by Walter and Yao

If the sample size is larger, it is more likely that more extreme values are observed, which corresponds to a larger range. The method of Walter and Yao [7] is based on this fact. If the data is assumed to be normally distributed, Walter and Yao proposed to estimate the standard deviation by

$$\hat{s}_{ugt*} = f \times (b_{ugt*} - a_{ugt*}) \,, \tag{4.4}$$

where f is a conversion factor that decreases with increasing sample size. It decreases quickly within a range of small sample sizes and slowly for large sample sizes. The range for different sample sizes taken from normally distributed data was estimated first by Tippett in 1925 [8], and extensive tabulations of $1/f$ are provided by Pearson and Hartley [9].

4.4 Obtaining Mean and Standard Deviation from Median and Range

4.4.1 Method by Hozo et al.

Based on the results of simulation studies with variously distributed data (normal distribution, log-normal distribution, beta distribution, exponential distribution, and Weibull distribution), Hozo et al. proposed different formulas for estimating the mean and standard deviation from the median m and the range depending on the sample size [10].

For estimating the mean, Hozo et al. proposed

$$\hat{\bar{x}}_{ugt*} = \begin{cases} \frac{a_{ugt*}+2\times m_{ugt*}+b_{ugt*}}{4} & n_{ugt*} \leq 25 \\ m_{ugt*} & n_{ugt*} > 25 \end{cases} \tag{4.5}$$

where m_{ugt*} refers to the median of a continuous outcome in a participant group g within a study u at a time point t, a_{ugt*} to the minimum, b_{ugt*} to the maximum, and n_{ugt*} to the respective sample size. As can be seen, the median is used as an estimator of the mean with a sample size exceeding 25.

For estimating the standard deviation, the following formulas where proposed by Hozo et al.:

$$\hat{s}_{ugt*} = \begin{cases} \sqrt{\frac{1}{12}\left(\frac{(a_{ugt*}-2\times m_{ugt*}+b_{ugt*})^2}{4} + (b_{ugt*}-a_{ugt*})^2\right)} & n_{ugt*} \leq 15 \\ \frac{b_{ugt*}-a_{ugt*}}{4} & 15 < n_{ugt*} \leq 70 \\ \frac{b_{ugt*}-a_{ugt*}}{6} & n_{ugt*} > 70. \end{cases} \tag{4.6}$$

4.4.2 Method by Wan et al.

Wan et al. [11] stated that the method by Hozo et al. [10] applies somewhat arbitrarily threshold values, contradicts the assumption that the standard deviation is a finite value, and assumes positivity for minimum and maximum, which is quite restrictive.

Based on similar simulation studies as performed by Hozo et al. [10], Wan et al. proposed to use

$$\hat{\bar{x}}_{ugt*} = \frac{a_{ugt*} + 2 \times m_{ugt*} + b_{ugt*}}{4} \tag{4.7}$$

for estimating the mean independent of the sample size and estimate the standard deviation by

$$\hat{s}_{ugt*} = \frac{b_{ugt*} - a_{ugt*}}{2\Phi^{-1}\left(\frac{n_{ugt*}-0.375}{n_{ugt*}+0.25}\right)}, \tag{4.8}$$

where Φ^{-1} is the inverse cumulative density function of the standard normal distribution. According to Wan et al. [11], these formulas provided less biased results in comparison to the method by Hozo et al. [10].

4.5 Obtaining Mean and Standard Deviation from Median and Interquartile Range

4.5.1 Method by Wan et al.

Based on simulation studies with variously distributed data, Wan et al. [11] also proposed formulas for estimating the mean and standard deviation from the median and interquartile range. For estimating the mean, Wan et al. suggested

$$\hat{\bar{x}}_{ugt*} = \frac{lq_{ugt*} + m_{ugt*} + uq_{ugt*}}{3} \tag{4.9}$$

where lq_{ugt*} refers to the first quartile and uq_{ugt*} to the third quartile.

For estimating the standard deviation,

$$\hat{s}_{ugt*} = \frac{uq_{ugt*} - lq_{ugt*}}{2\Phi^{-1}\left(\frac{0.75n_{ugt*}-0.125}{n_{ugt*}+0.25}\right)} \tag{4.10}$$

was suggested with Φ^{-1} referring to the inverse cumulative density function of the standard normal distribution.

4.5.2 Cochrane Method

In the Cochrane Handbook for Systematic Reviews and Interventions [5], it was suggested to use the median as an estimator of the mean if the data is assumed to be symmetrically distributed. If the data is also assumed to be normally distributed, the authors of the handbook suggested to estimate the standard deviation from the interquartile range by

$$\hat{s}_{ugt*} = \frac{uq_{ugt*} - lq_{ugt*}}{1.35}. \tag{4.11}$$

4.6 Obtaining Mean and Standard Deviation from Median, Range, and Interquartile Range

4.6.1 Method by Bland

Bland [12] looked at a scenario in which the median, range, and interquartile range is known and aimed at improving the method by Hozo et al. for obtaining mean and standard deviation from median and range by incorporating the first and third quantiles. Backed up by simulation studies with normally and lognormally distributed data, Bland came up with the following formula for estimating the mean:

$$\hat{\bar{x}}_{ugt*} = \frac{(n_{ugt*}+3)\,a_{ugt*} + 2\,(n_{ugt*}-1)\,(lq_{ugt*}+m_{ugt*}+uq_{ugt*}) + (n_{ugt*}+3)\,b_{ugt*}}{8n_{ugt*}}. \tag{4.12}$$

For estimating the standard deviation, Bland proposed

$$\hat{s}_{ugt*} = \left(\begin{array}{l} \frac{1}{16}(a_{ugt*}^2 + 2\times lq_{ugt*}^2 + 2\times m_{ugt*}^2 + 2\times uq_{ugt*}^2 + b_{ugt*}^2) \\ +\frac{1}{8}\times(a_{ugt*}\times lq_{ugt*} + lq_{ugt*}\times m_{ugt*} + m_{ugt*}\times uq_{ugt*} + uq_{ugt*}\times b_{ugt*}) \\ -\frac{1}{64}(a_{ugt*} + 2\times lq_{ugt*} + 2\times m_{ugt*} + 2\times uq_{ugt*} + b_{ugt*})^2 \end{array} \right)^{\frac{1}{2}}. \tag{4.13}$$

4.6.2 Method by Wan et al.

Wan et al. [11] also looked at the scenario in which the median, range, and interquartile range are known. Based on the results of simulations studies with variously distributed data (normal distribution, log-normal distribution, beta distribution, exponential distribution, and Weibull distribution), the authors proposed a simplified version for estimating the mean:

$$\hat{\bar{x}}_{ugt*} = \frac{a_{ugt*} + 2\times lq_{ugt*} + 2\times m_{ugt*} + 2\times uq_{ugt*} + b_{ugt*}}{8}. \tag{4.14}$$

Compared to Bland, Wan et al. claimed to have improved the estimation

of the standard deviation by incorporating the sample size as following:

$$\hat{s}_{ugt*} = \frac{b_{ugt*} - a_{ugt*}}{4\Phi^{-1}\left(\frac{n_{ugt*}-0.375}{n_{ugt*}+0.25}\right)} + \frac{uq_{ugt*} - lq_{ugt*}}{4\Phi^{-1}\left(\frac{0.75n_{ugt*}-0.125}{n_{ugt*}+0.25}\right)}, \tag{4.15}$$

with Φ^{-1} referring to the inverse cumulative density function of the standard normal distribution.

4.7 Obtaining Mean and Standard Deviation from Mean and Missing Standard Deviation by Imputation

4.7.1 Method by Furukawa et al.

Furukawa et al. [13] suggested to impute missing standard deviations by pooling all available standard deviations from the studies included in a meta-analysis. How to deal with the situation where continuous outcomes can be measured at different time points is not explicitly mentioned. It is assumed that the available standard deviations are pooled separately for different time points. This results in the following equation for imputing a missing standard deviation in a patient group g within a study u at a time point t:

$$\hat{s}_{ugt*} = \sqrt{\frac{\sum_{u=1}^{U}\sum_{g=1}^{G}\left(n_{ugt*}-1\right)s_{ugt*}^2}{\sum_{u=1}^{U}\sum_{g=1}^{G}\left(n_{ugt*}-1\right)}}, \tag{4.16}$$

where s_{ugt*} refers to available standard deviations in patient groups g within studies u at a time point t.

Furukawa et al. [13] also suggested that missing standard deviations could be borrowed from a previous meta-analysis when the number of included studies is small or the number of studies with missing standard deviation

is large. The authors applied the two imputation methods to two real data examples assuming that some of the standard deviations were missing and reported the degree of concordance of the actual and imputed values as gratifying.

4.7.2 Method by Marinho et al.

Marinho et al. published a systematic review and meta-analysis on fluoride toothpastes for preventing dental caries in children and adolescents [14]. The authors imputed missing standard deviations of caries increments through linear regression of the logarithmized standard deviation of caries increments on the logarithmized mean of caries increments. The basis of the logarithm is not mentioned, but it can be assumed that the natural logarithm was used. In general terms, a missing standard deviation s_{ugt*} in a patient group g within a study u at a time point t can be imputed according to Marinho et al. by

$$\ln(\hat{s}_{ugt*}) = j \times \ln(\bar{x}_{ugt*}) + i, \tag{4.17}$$

where $\bar{x}_{ugt*}$ is the mean in a patient group g within a study u at a time point t. Parameters j and i refer to regression coefficients that are calculated based on available pairs of mean and standard deviation. The model assumes a linear relationship between the logarithmized standared deviation and logarithmized mean.

4.8 Obtaining the Standardized Mean Difference and Standard Error from a 2×2 Contingency Table for a Dichotomized Continuous Outcome

Some studies dichotomize a continuous outcome by applying a cutpoint on the continuous outcome. For example, studies on the effectiveness of a blood pressure medication can measure the decrease of the systolic blood pressure in mmHg or as a dichotomy (decrease of $\geq$ 10mmHg vs. $<$ 10mmHg). After dichotomizing a continuous variable in two groups (e.g., intervention vs. control), the results can be presented in a 2×2 contingency table (Table 4.1).

Outcome	Intervention	Control	Total
Success ($x_{ugtp} \geq y_{u*t*}$)	ns_{u1t*}	ns_{u2t*}	ns_{u*t*}
Failure ($x_{ugtp} < y_{u*t*}$)	nf_{u1t*}	nf_{u2t*}	nf_{u*t*}
Total	n_{u1t*}	n_{u2t*}	n_{u*t*}

Table 4.1: 2 x 2 Contingency table for a dichotomized continuous outcome. x_{ugt*} = measurement of a continuous outcome x in a study u, patient group g, time point t, and participant p; y_{u*t*} = cutpoint in a study u at time point t; ns_{u1t*} = number of participants with success ($x_{u1tp} \geq y_{u*t*}$) in study u in the intervention group ($g = 1$) at time point t; ns_{u1t*} = number of participants with success ($x_{u2tp} \geq y_{u*t*}$) in study u in the control group ($g = 2$) at time point t; ns_{u*t*} = number of participants with success ($x_{ugtp} \geq y_{u*t*}$) in study u at time point t; nf_{u1t*} = number of participants with failure ($x_{u1tp} < y_{u*t*}$) in study u in the intervention group ($g = 1$) at time point t; nf_{u2t*} = number of participants with failure ($x_{u1tp} < y_{u*t*}$) in study u in the control group ($g = 2$) at time point t; nf_{u*t*} = number of participants with failure ($x_{ugtp} < y_{u*t*}$) in study u at time point t; n_{u*t*} = number of participants in study u at time point t.

To compare a binary outcome between groups of participants, three effect measures are calculated most commonly: the risk difference, the risk ratio, and the odds ratio. These effect measures cannot be pooled with effect measures that are calculated for comparing a continuous outcome between groups of participants, such as the mean difference or the standardized mean difference. Sánchez-Meca et al. [15] compared the performance of seven methods that convert a dichotomized continuous outcome in two participants groups into the standardized mean difference with standard error. The concept of these methods differs from the concept of the conversion methods described above. The above methods aim at obtaining the mean and standard deviation within a participant group. In a second step, the mean and standard deviation can be compared between participant groups within a study by calculating an effect measure such as the mean difference or standardized mean difference. The methods summarized by Sánchez-Meca et al. [15] use the data of two participant groups and directly obtain the standardized mean difference comparing these groups.

Sánchez-Meca et al. [15] set up simulation studies to compare the seven methods for converting a dichotomized continuous outcome in two participant groups into the standardized mean difference with standard error. The authors assumed that the two populations were normally distributed but also checked the robustness of the methods under several conditions representing nonnormal distributions. Three methods performed best. These three methods are described in the following.

4.8.1 Method by Cox

In 1970, Cox [16] proposed an index that is based on the odds ratio and its logit transformation. The standardized mean difference smd_{u*t*} in a study

u at a time point t is estimated by

$$\hat{smd}_{u*t*} = \frac{\ln\left(\frac{ns_{u1t*} \times nf_{u2t*}}{ns_{u2t*} \times nf_{u1t*}}\right)}{1.65}, \tag{4.18}$$

which is the natural logarithm of the odds ratio divided by 1.65.

The standard error $se_{smd_{u*t*}}$ is estimated by

$$\hat{se}_{smd_{u*t*}} = \sqrt{0.367 \times \left(\frac{1}{ns_{u1t*}} + \frac{1}{nf_{u1t*}} + \frac{1}{ns_{u2t*}} + \frac{1}{nf_{u2t*}}\right)}. \tag{4.19}$$

4.8.2 Method by Glas et al.

In 1981, Glas et al. [17] proposed an index that is based on the normal distribution assumption. The previously described method by Cox assumes logistic distributions, however, most primary studies assume normal distribution in the underlying populations. According to Glas et al. [17], the standardized mean difference smd_{u*t*} in a study u at a time point t can be estimated by

$$\hat{smd}_{u*t*} = \Phi^{-1}\left(\frac{ns_{u1t*}}{n_{u1t*}}\right) - \Phi^{-1}\left(\frac{ns_{u2t*}}{n_{u2t*}}\right), \tag{4.20}$$

with Φ^{-1} referring to the inverse cumulative density function of the standard normal distribution, and $\frac{ns_{u1t*}}{n_{u1t*}}$ and $\frac{ns_{u2t*}}{n_{u2t*}}$ being the success proportions in the intervention and control groups.

The standard error $se_{smd_{u*t*}}$ is estimated by

$$\hat{se}_{smd_{u*t*}} = \left(\frac{\frac{2 \times \pi \times \frac{ns_{u1t*}}{n_{u1t*}}(1 - \frac{ns_{u1t*}}{n_{u1t*}}) \times e^{(\Phi^{-1}(\frac{ns_{u1t*}}{n_{u1t*}}))^2}}{n_{u1t*}}}{+ \frac{2 \times \pi \times \frac{ns_{u2t*}}{n_{u2t*}}(1 - \frac{ns_{u2t*}}{n_{u2t*}}) \times e^{(\Phi^{-1}(\frac{ns_{u2t*}}{n_{u2t*}}))^2}}{n_{u2t*}}} \right)^{\frac{1}{2}}. \tag{4.21}$$

4.8.3 Method by Hedges and Olkin

Another method that assumes normal distribution in the underlying populations is performed by calculating the biserial-phi correlation coefficient ϕ_{bis} based on the phi coefficient ϕ and transforming the biserial-phi coefficient, which was published by Thorndike in 1949 [18], into the standardized mean difference [19].

The phi coefficient ϕ is calculated from the 2×2 contingency table by

$$\phi = \frac{ns_{u1t*} \times nf_{u2t*} - nf_{u1t*} \times ns_{u2t*}}{\sqrt{n_{u1t*} \times n_{u2t*} \times ns_{u*t*} \times nf_{u*t*}}} \tag{4.22}$$

The biserial-phi correlation coefficient is calculated by

$$\phi_{bis} = \frac{\sqrt{\frac{ns_{u*t*}}{n_{u*t*}} \times \left(1 - \frac{ns_{u*t*}}{n_{u*t*}}\right)}}{\Phi\left(\frac{ns_{u*t*}}{n_{u*t*}}\right)} \times \phi, \tag{4.23}$$

where $\frac{ns_{u*t*}}{n_{u*t*}}$ is the global success proportion and Φ the cumulative density function of the standard normal distribution.

Based on the biserial-phi correlation coefficient, the standardized mean difference smd_{u*t*} in a study u at a time point t can be estimated by

$$\hat{smd}_{u*t*} = \frac{\phi_{bis}}{\sqrt{1 - \phi_{bis}^2}} \times \sqrt{\frac{(n_{u*t*} - 2) \times n_{u*t*}}{n_{u1t*} \times n_{u2t}}}, \tag{4.24}$$

with $n_{u*t*} - 2$ being the degrees of freedom.

The standard error $se_{smd_{u*t*}}$ is estimated by

$$\hat{se}_{smd_{u*t*}} = \sqrt{\frac{\frac{ns_{u*t*}}{n_{u*t*}} \times \left(1 - \frac{ns_{u*t*}}{n_{u*t*}}\right) \times (1 - \phi^2) \times (n_{u*t*})}{\left(\Phi\left(\frac{ns_{u*t*}}{n_{u*t*}}\right)\right)^2 \times n_{u1t*} \times n_{u2t*} \times (1 - \phi_{bis}^2)^3}}, \tag{4.25}$$

where Φ refers to the cumulative density function of the standard normal

distribution.

4.9 Obtaining Mean Change from Baseline with Standard Deviation from Mean Baseline, Mean Postintervention, Mean Baseline Standard Deviation, and Mean Postintervention Standard Deviation with a Method by Follmann et al.

For interventional studies, there is an additional method that is useful for obtaining the mean change from baseline with standard deviation from information on baseline and postintervention measurements [20].

The mean change from baseline $\bar{x}_{ug3*}$ for a participant group g within a study u can simply be calculated by substracting the baseline mean $\bar{x}_{ug1*}$ from the postintervention mean $\bar{x}_{ug2*}$:

$$\bar{x}_{ug3*} = \bar{x}_{ug2*} - \bar{x}_{ug1*}. \tag{4.26}$$

The standard deviation of the change from baseline s_{ug3*} can be calculated by

$$s_{ug3*} = \sqrt{s^2_{ug1*} + s^2_{ug2*} - (2 \times r_{ug**} \times s_{ug1*} \times s_{ug2*})}, \tag{4.27}$$

where s_{ug1*} refers to the baseline standard deviation, s_{ug2*} to the postintervention standard deviation, and r_{ug**} to the correlation coefficient between baseline and postintervention measurements in a participant group g within a study u.

This correlation coefficient r_{ug**} is usually unknown, however, it can be substituted by a reasonable guess or estimated based on information from

other included studies for whom mean and standard deviation are available for baseline measurements, postintervention measurements, and the change from baseline.

If a study u provides the latter information for a participant group g, the correlation coefficient r_{ug**} can be calculated by

$$r_{ug**} = \frac{s^2_{ug1*} + s^2_{ug2*} - s^2_{ug3*}}{2 \times s_{ug1*} \times s_{ug2*}}. \tag{4.28}$$

For participant groups for whom the standard deviation of the change from baseline is missing, it makes intuitively sense to take the weighted average across all participant groups for whom the above mentioned information was provided to estimate missing r_{ug**} values. This is, however, not explicitly mentioned by Follmann et al. [20]. The correlation coefficient $\hat{r}_{*g**}$ for a participant group g in studies in which the mean change from baseline with standard deviation is missing is then estimated by

$$\hat{r}_{*g**} = \frac{\sum_{u=1}^{U} \sum_{g=1}^{G} n_{ug**} \times r_{ug**}}{\sum_{u=1}^{U} \sum_{g=1}^{G} n_{ug**}}, \tag{4.29}$$

where r_{ug**} refers to correlation coefficients that can be calculated according to equation 3.28 and n_{ug**} to the number of patients in a patient group g within a study u corresponding to r_{ug**}. If it is assumed that correlation coefficients differ between intervention and control groups, the weighted average should be calculated separately for intervention and control groups.

Chapter 5

Effect Measure Calculation

5.1 Choice of Effect Measure

To summarize the strength of the link between an exposure and an outcome, effect measures are calculated [4]. Most commonly, an outcome is compared between an exposed and a nonexposed group for this purpose [5]. Effect measures can then either be ratio or difference measures, and the types of effect measures with corresponding types of outcome data needed are listed in Appendix 2. A continuous outcome measured in two groups of participants can be compared between these groups by calculating the absolute difference in means, the standardized difference in means, or the ratio of means.

The ratio of means is rarely used. It is relatively challenging to interpret and cannot be applied for comparing the change from baseline since the change from baseline can be 0 or positive for one participant group and negative for the other. The ratio of means can be used to pool the results of studies that measured the same outcome but with different tools or on different scales. For example, the severity of atopic dermatitis can be measured using different scores. It is important to mention that the lower limits must be comparable when using the ratio of means. Because of its scarce use and described disadvantages, the ratio of means is not further covered in the

following.

The absolute difference in means is easy to interpret and has the advantage that it can be used to pool both postinterventional measurements and change from baseline measurements in one analysis. It is a common problem that some studies included in a meta-analysis published postintervention measurements while others published change from baseline measurements. Above described methods can in many cases be used to obtain the change from baseline and postintervention measurements, but in some cases, this may not be possible. The main disadvantage of choosing the absolute difference in means as effect measure is that the results of studies that measured the same outcome with different tools or on different scales must be analyzed separately.

Choosing the standardized difference allows to pool the results of studies that measured the same outcome with different tools or on different scales, even if the lower limits are not comparable. Furthermore, studies that dichotomized a continuous outcome can be included in the analysis since previously described methods for the retransformation of a dichotomized continuous outcome result in an estimation for the standardized difference in means. On the downside, the interpretation of the standardized difference in means can be more challenging than the interpretation of the absolute difference in means, and the usage of the standardized difference in means does not allow to pool both postinterventional measurements and change from baseline measurements in one analysis.

In summary, the absolute difference in means should be used for easy interpretation and to pool both postinterventional measurements and change from baseline measurements in one analysis. The standardized difference in means should be used used to pool the results of studies that measured the same outcome with different tools or on different scales and for including studies that dichotomized a continuous outcome.

5.2 Calculation of the Absolute Difference in Means

Let's assume that a normally distributed endpoint exists in two separate populations (intervention vs. control). The absolute difference in means δ_{**t*} at a time point t can then be defined by

$$\delta_{**t*} = \mu_{*1t*} - \mu_{*2t*}, \tag{5.1}$$

with μ_{*1t*} referring to the population mean in the intervention group at a time point t and μ_{*2t*} referring to the population mean in the control group at a time point t.

A meta-analysis is usually performed to estimate parameters of a large population. The absolute difference in means between two populations δ_{**t*} at a time point t is estimated in a meta-analysis by pooling the absolute differences in means at time point t of studies included in the meta-analysis. The absolute difference in means d_{u*t*} in a single included study u at a time point t can be calculated by

$$d_{u*t*} = \overline{x}_{u1t*} - \overline{x}_{u2t*}, \tag{5.2}$$

with $\overline{x}_{u1t*}$ referring to the mean of outcome measurements in the intervention group at a time point t and $\overline{x}_{u2t*}$ referring to the mean of outcome measurements in the control group at a time point t.

With the assumption of equal standard deviations in the intervention and control group σ^2_{*1t*} and σ^2_{*2t*}, the standard error of d_{u*t*} can be calculated by

$$se_{d_{u*t*}} = \sqrt{\frac{n_{u1t*} + n_{u2t*}}{n_{u1t*} \times n_{u2t*}} \times s^2_{u*t*}}, \tag{5.3}$$

where n_{u1t*} refers to the sample size in the intervention group at a time point t, n_{u2t*} to the sample size in the control group at a time point t, and s_{u*t*} to

the pooled standard deviation in a study u at a time point t [21]. The pooled standard deviation can be calculated by weighting the standard deviations in the groups by the sample size of each group. Consequently, s_{u*t*} is calculated by

$$s_{u*t*} = \sqrt{\frac{(n_{u1t*} - 1) \times s^2_{u1t*} + (n_{u2t*} - 1)s^2_{u2t*}}{n_{u1t*} + n_{u2t*} - 2}}, \tag{5.4}$$

where s_{u1t*} refers to the standard deviation in the intervention group at a time point t and s_{u2t*} to the standard deviation in the control group at a time point t [21]. Combining Formula 5.4 and Formula 5.3, the standard error of d_{u*t*} can be calculated by

$$se_{d_{u*t*}} = \sqrt{\frac{n_{u1t*} + n_{u2t*}}{n_{u1t*} \times n_{u2t*}} \times \frac{(n_{u1t*} - 1) \times s^2_{u1t*} + (n_{u2t*} - 1)s^2_{u2t*}}{n_{u1t*} + n_{u2t*} - 2}}. \tag{5.5}$$

5.3 Calculation of the Standardized Difference in Means

The standardized mean difference shows the effect measured in a study in relation to the between-participant variability in the outcome. Depending on the chosen measure of between-participant variability, several modifications of the standardized mean difference can be calculated. Hedges' g is the most popular modification, perhaps also because of its use in Cochrane reviews [5].

Let's again assume that a normally distributed endpoint exists in two separate populations (intervention vs. control). The standardized difference in means can then be defined by

$$\lambda_{**t*} = \frac{\mu_{*1t*} - \mu_{*2t*}}{\sigma_{**t*}}, \tag{5.6}$$

with μ_{*1t*} referring to the population mean in the intervention group at a time point t, μ_{*2t*} referring to the population mean in the control group

at a time point t, and σ_{**t*} referring to the common population standard deviation at a time point t [15, 19].

The population standardized difference in means λ_{**t*} at a time point t can be estimated in a meta-analysis by pooling the standardized absolute differences in means of included studies at time point t (Chapter 6). The standardized difference in means smd_{u*t*} in a single study u at a time point t can be calculated by

$$smd_{u*t*} = \frac{\overline{x}_{u1t*} - \overline{x}_{u2t*}}{s_{u*t*}}, \tag{5.7}$$

with $\overline{x}_{u1t*}$ referring to the mean of outcome measurements in the intervention group at a time point t, $\overline{x}_{u2t*}$ referring to the mean of outcome measurements in the control group at a time point t, s_{u*t*} referring to the pooled standard deviation in the study u at a time point t, with $x_{u1tp} \sim N(\overline{x}_{u1t*}, s^2_{u1t*})$, and with $x_{u2tp} \sim N(\overline{x}_{u2t*}, s^2_{u2t*})$.

By substituting the pooled standard deviation s_{u*t*} in formula 5.7 with formula 5.4, the standardized difference in means smd_{u*t*} in a study u at a time point t can be calculated by

$$smd_{u*t*} = \frac{\overline{x}_{u1t*} - \overline{x}_{u2t*}}{\sqrt{\frac{(n_{u1t*}-1)\times s^2_{u1t*} + (n_{u2t*}-1)s^2_{u2t*}}{n_{u1t*}+n_{u2t*}-2}}}, \tag{5.8}$$

which is called Hedges' g [19].

Hedges' g shows a positive bias for small sample sizes [19]. To account for this bias, Hedges and Olkin proposed the correction factor $1 - \frac{3}{4\times n_{u*t*}-1}$ (Equation 10, page 81). Its application reveals a value commonly referred to as Hedges' adjusted g:

$$smd_{u*t*} = \frac{\overline{x}_{u1t*} - \overline{x}_{u2t*}}{\sqrt{\frac{(n_{u1t*}-1)\times s^2_{u1t*} + (n_{u2t*}-1)s^2_{u2t*}}{n_{u1t*}+n_{u2t*}-2}}} \times (1 - \frac{3}{4 \times n_{u*t*} - 1}). \tag{5.9}$$

The standard error of Hedges' adjusted g is calculated by

$$se_{smd_{u*t*}} = \sqrt{\frac{n_{u1t*} + n_{u2t*}}{n_{u1t*} \times n_{u2t*}} + \frac{smd^2_{u*t*}}{2(n_{u1t*} + n_{u2t*})}}. \tag{5.10}$$

Chapter 6

Pooling Effect Measures

The final synthesis of evidence in meta-analyses is done by pooling effect measures from individual studies. The pooled effect $\hat{\theta}_{**t*}$ at a time point t, an estimate of the population effect θ_{**t*}, is calculated by the weighted mean of effect measures across all studies

$$\hat{\theta}_{**t*} = \frac{\sum_{u=1}^{U} w_{u*t*}\hat{\theta}_{u*t*}}{\sum_{u=1}^{U} w_{u*t*}}, \tag{6.1}$$

where w_{u*t*} is the weight of an individual study u at time point t and $\hat{\theta}_{u*t*}$ the effect estimate of an individual study u at time point t [22].

How the weight of an individual study w_{u*t*} is calculated differs depending on whether a fixed-effect or random-effects model is applied.

In the fixed-effect model, one true or common effect size underlying all included individual studies is assumed. Applying the inverse variance method, where more weight is given to studies reporting a more precise effect estimate, the weight in the fixed-effect model can be calculated by

$$w_{fixed_{u*t*}} = \frac{1}{se^2_{\hat{\theta}_{u*t*}}}, \tag{6.2}$$

with $se^2_{\hat{\theta}_{u*t*}}$ referring to the square of the standard error of the effect estimate

of an individual study u at time point t.

In contrary, the true effect size is assumed to vary between individual studies in the random-effects model. Different true effect sizes may be the result of differences in study populations and applied methods between included studies. The difference in the true effect size between studies results in an additional component of variance that has to be included in the inverse variance method for calculating the weight w_{u*t*} of an individual study u at time point t, the between-study variance τ^2_{**t*} at time point t. The weight in the random-effects model is calculated by

$$w_{random_{u*t*}} = \frac{1}{se^2_{\hat{\theta}_{u*t*}} + \tau^2_{**t*}}, \tag{6.3}$$

where $se^2_{\hat{\theta}_{u*t*}}$ is the square of the standard error of the effect estimate of an individual study u at time point t, similar to the fixed-effect model, and τ^2_{**t*} the between-study variance at time point t, which refers to the variance of the true effect size between studies. The between-study variance τ^2_{**t*} can conceptually be estimated in a three-step process [22]. First, the total study to-study variation is calculated. Second, it is estimated how much the effect estimates of the individual studies would differ given that the true effect sizes would be the same across studies. Third, the between-study variance in the true effect sizes is estimated by calculating the difference in the two latter variances. One method following this concept is the DerSimonian and Laird method, also referred to as the weighted method of moments [23] First, the q statistic, a standardized and weighted sum of squares of the effect size estimates about the fixed-effect estimate, is calculated by

$$q = \sum_{U}^{u=1} w_{u*t*}(\hat{\theta}_{u*t*} - \hat{\theta}_{**t*})^2 = \sum_{u=1}^{U} \frac{(\hat{\theta}_{u*t*} - \hat{\theta}_{fixed_{**t*}})^2}{se^2_{\hat{\theta}_{u*t*}}}. \tag{6.4}$$

Second, the degrees of freedom df, which represent the expected value of the

q statistic in case all studies share the same effect size, are calculated by

$$df = U - 1, \tag{6.5}$$

where U refers to the number of included studies. Last, the between-study variance τ^2_{**t*} is estimated by

$$\tau^2_{**t*} = \frac{Q - df}{c}, \tag{6.6}$$

where dividing by

$$c = \sum_{u=1}^{U} w_{fixed_{u*t*}} - \frac{\sum_{u=1}^{U} w^2_{fixed_{u*t*}}}{\sum_{u=1}^{U} w_{fixed_{u*t*}}} \tag{6.7}$$

cancels the standardization. Besides the DerSimonian and Laird method, other methods for estimating the between-study variance τ^2 such as maximum likelihood estimation or restricted maximum likelihood estimation can be applied [22].

Since the inverse of the variance is used for weighting, the inverse of the weight can be used to estimate the uncertainty in the pooled effect. More specifically, the variance in the pooled effect, which is the squared standard error of the pooled effect, is estimated by

$$se^2_{\hat{\theta}^2_{**t*}} = \frac{1}{\sum_{u=1}^{U} w_{u*t*}}, \tag{6.8}$$

where the weights of the individual studies u are summed across studies. Consequently, the square root can be used to estimate the standard error by

$$se_{\hat{\theta}_{**t*}} = \sqrt{se^2_{\hat{\theta}^2_{**t*}}}. \tag{6.9}$$

The 95% confidence interval can be estimated by

$$ll_{\hat{\theta}_{**t*}} = \hat{\theta}_{**t*} - 1.96 \times se_{\hat{\theta}_{**t*}} \tag{6.10}$$

and

$$ul_{\hat{\theta}_{**t*}} = \hat{\theta}_{**t*} + 1.96 \times se_{\hat{\theta}_{**t*}}, \tag{6.11}$$

where $ll_{\hat{\theta}}$ refers to the lower limit and $ul_{\hat{\theta}}$ to the upper limit [22].

For testing the null hypothesis that the pooled effect does not differ from a value h, the z statistic can be calculated by

$$z = \frac{\hat{\theta}_{**t*} - h}{se_{\hat{\theta}}}. \tag{6.12}$$

The p-value for a two-tailed test can afterwards be obtained by

$$p = 2[1 - (\Phi(|z|))], \tag{6.13}$$

where $\Phi(|z|)$ refers to the standard normal cumulative distribution [22].

Chapter 7

Evaluation of Bias in Simulation Studies

In 2017, Morris, White, and Crowther published a structered approach for planning, conducting, and reporting simulation studies named the ADEMP approach [24]. ADEMP is an acronym for: Aims, Data-generating mechanisms, Methods, Estimands, and Performance measures. The simulation studies of this thesis followed the ADEMP approach.

7.1 Methods

7.1.1 Aims

Different conversion and imputation methods for differently or incompletely reported continuous outcomes were described in Chapter 3. The simulation studies of this thesis aimed at comparing these methods in terms of precision when applied for computing uniform effect measures that are pooled in a meta-analysis. The term uniform effect measures refers to effect measures that are of the same form and can therefore be pooled.

7.1.2 Data-Generation Mechanisms

The data of 1000 meta-analyses were simulated per data generation scenario. Each meta-analysis included a random number of interventional controlled studies that was uniformly distributed between 10 and 30. The number of 1000 meta-analyses per data generation scenario was a compromise between time expense and accuracy.

Simulation of Individual Patient Data

Continuous outcome values x_{ugtp} for a patient p at time point t in group g in study u were generated by performing parametric draws from known models. The following models were applied for baseline ($t = 1$), postintervention ($t = 2$), and change from baseline values ($t = 3$):

$$x_{ug1p} = \overline{x}_{u*1*} + e_{ug1p} \tag{7.1}$$

$$x_{ug2p} = x_{ug1p} + d_{u*2*} \times k_{ugtp} + e_{ug2p} \tag{7.2}$$

$$x_{ug3p} = x_{ug2p} - x_{ug1p} \tag{7.3}$$

The factors included in these models are described in the following.

- $\overline{x}_{u*1*}$ refers to the mean outcome in study u at baseline ($t = 1$). It was assumed to be normally distributed around 100 with a standard deviation of 5, $\overline{x}_{u*1*} \sim N(100, 25)$.

- e_{ug1p} refers to an individual baseline residual. It could vary across data-generation mechanisms in two ways.

  - Symmetric distribution of the individual baseline residual: The individual baseline residual was assumed to be normally distributed around 0 with a variance of s^2_{u***}, $e_{ug1p} \sim N(0, s^2_{u***})$. The standard deviation of the individual baseline residual s_{u***} might differ across studies u since patients may be more similar in some

studies compared to others. The standard deviation of the individual baseline residual cannot be negative and was assumed to be gamma distributed with shape parameter $\alpha = 10$ and scale parameter $\beta = 2$, $s_{u***} \sim Gamma(10, 2)$.

- Skewed distribution of the individual baseline residual: For generating skewed individual baseline residuals, a gamma distribution with shape parameter $\alpha = 6$ and scale parameter β_{u***} was used. Since a gamma distribution only includes positive values, the expected value of the used gamma distribution $\alpha \times \beta_{u***}$ was subtracted from each value drawn from the gamma distribution to obtain positive and negative individual baseline residuals. Consequently, the individual baseline residual plus the expected value of the gamma distribution was gamma distributed, $e_{ug1p} + \alpha \times \beta_{u***} \sim Gamma(6, \beta_{u***})$. The scale parameter β_{u***} was assumed to differ across studies u because patients may be more similar in some studies compared to others. It was assumed to be gamma distributed with shape parameter $\alpha = 4$ and scale parameter $\beta = 2$, $\beta_{u***} \sim Gamma(4, 2)$.

- d_{u*2*} refers to the treatment effect in study u. It could vary across data-generation mechanisms in two ways:

  - Small treatment effect size in the studies: The treatment effect size in the studies was assumed to be normally distributed around the population treatment effect δ_{**2*} of 10 with a standard deviation of 4, $d_{u*2*} \sim N(10, 16)$.
  - Large treatment effect size in the studies: The treatment effect size in the studies was assumed to be normally distributed around the population treatment effect δ_{**2*} of 20 with a standard deviation of 4, $d_{u*2*} \sim N(20, 16)$.

- k_{ugtp} in Formula 7.2 is a treatment indicator. It is $k_{u1tp} = 1$ in the intervention group ($g = 1$) and $k_{u2tp} = 0$ in the control group ($g = 2$). One study included n_{u***} patients with $n_{ug**} = \frac{n_{u***}}{2}$ patients per group. The sample size per group n_{ug**} could vary across data-generation mechanisms in two ways.
  - Small sample size per group: The sample size per group was assumed to be uniformly distributed between 10 and 30, $n_{ug**} \sim U(10, 30)$.
  - Large sample size per group: The sample size per group was assumed to be uniformly distributed between 30 and 100, $n_{ug**} \sim U(30, 100)$.
- e_{ug2p} refers to an individual postintervention residual. It could vary across data generation mechanisms in two ways.
  - Symmetric distribution of the individual postintervention residual: The individual postintervention residual was assumed to be normally distributed around 0 with a variance of s^2_{u***}, $e_{ug2p} \sim N(0, s^2_{u***})$. The standard deviation of the individual postintervention residual s_{u***} might differ across studies u since patients may be more similar in some studies compared to others. The standard deviation of the individual postintervention residual cannot be negative and was assumed to be gamma distributed with shape parameter $\alpha = 10$ and scale parameter $\beta = 2$, $s_{u***} \sim Gamma(10, 2)$.
  - Skewed distribution of the individual postintervention residual: For generating skewed individual postintervention residuals, a gamma distribution with shape parameter $\alpha = 6$ and scale parameter β_{u***} was used. Since a gamma distribution only includes positive values, the expected value of the used gamma distribution

$\alpha \times \beta_{u***}$ was subtracted from each value drawn from the gamma distribution to obtain positive and negative individual baseline residuals. Consequently, the individual baseline residual plus the expected value of the gamma distribution was gamma distributed, $e_{ug1p} + \alpha \times \beta_{u***} \sim Gamma(6, \beta_{u***})$. The scale parameter β_{u***} was assumed to differ across studies u because patients may be more similar in some studies compared to others. It was assumed to be gamma distributed with shape parameter $\alpha = 4$ and scale parameter $\beta = 2$, $\beta_{u***} \sim Gamma(4, 2)$.

As mentioned, four factors could vary across data-generation mechanisms in two ways, respectively: the individual baseline residual, the treatment effect size, the sample size per group, and the individual postintervention residual.

If the individual baseline residual was assumed to be normally distributed, the individual postintervention residual was also assumed to be normally distributed. If the individual baseline residual followed a gamma distribution, the individual postintervention residual did so as well. Otherwise, a fully factorial design was executed.

Computation of Study Level Data

For each group g in each study u, the following measures were calculated for baseline, postintervention, and change from baseline values, respectively:

- Mean $\overline{x}_{ugt*}$
- Standard deviation s_{ugt*}
- Standard error of the mean se_{ugt*}
- 95% confidence interval of the mean with lower limit ll_{ugt*} and upper limit ul_{ugt*}

- Median m_{ugt*}
- Range with smallest value a_{ugt*} and largest value b_{ugt*}
- Interquartile range with first quartile lq_{ugt*} and third quartile uq_{ugt*}
- Numbers of patients after dichotomization with success ns_{ugt*} and failure ns_{ugt*}

For dichotomization, the numbers of patients with values greater than or equal to (success) and less than (failure) a threshold were calculated. The threshold y_{u*t*} in a study u at time point t was assumed to be normally distributed with parameters dependent on time point and effect size (Table 7.1).

Time Point	Effect Size	Mean	Standard Deviation
Baseline	Small	110	2
Baseline	Large	120	2
Postintervention	Small	110	2
Postintervention	Large	120	2
Change from Baseline	Small	10	2
Change from Baseline	Large	20	2

Table 7.1: Parameters of the normal distribution assumed for the threshold used for dichotomizing.

For comparing methods for imputing missing standard deviation, it was assumed that the standard deviation for a study was missing with a probability of 0.25.

For the evaluation of the method by Follmann et al. for obtaining mean change from baseline with standard deviation from mean baseline with standard deviation and mean postintervention with standard deviation, it was assumed that the standard deviation of the mean change from baseline was

missing with a probability of 0.25 while mean baseline standard deviation and mean postintervention standard deviation were given.

7.1.3 Estimands

Simulation studies typically evaluate or compare methods for estimating population quantities [24]. Morris et al. [24] referred to these as estimands, which is retained here. The following estimands were defined:

- Population absolute difference in means between the change from baseline values of the intervention and control group δ_{**3*}
- Population standardized difference in means between the change from baseline values of the intervention and control group λ_{**3*}

7.1.4 Methods

In Chapter 3, the following conversion and imputation methods for the conversion and imputation of differently or incompletely reported continuous outcomes were described.

- Obtaining Mean and Standard Deviation from Mean and Standard Error by Algebraic Conversion
- Obtaining Mean and Standard Deviation from Mean and Confidence Interval by Algebraic Conversion
- Obtaining Mean and Standard Deviation from Mean and Range
  - The "Range"-Method
  - Method by Walter and Yao
- Obtaining Mean and Standard Deviation form Median and Range
  - Method by Hozo et al.

  - Method by Wan et al.

- Obtaining Mean and Standard Deviation from Median and Interquartile Range
  - Method by Wan et al.
  - Cochrane Method

- Obtaining Mean and Standard Deviation form Median, Range, and Interquartile Range
  - Method by Bland
  - Method by Wan et al.

- Obtaining Mean and Standard Deviation from Mean and Missing Standard Deviation by Imputation
  - Method by Furukawa et al.
  - Method by Marinho et al.

- Obtaining the Standardized Mean Difference and Standard Error from a 2 x 2 Contingency Table for a Dichotomized Continuous Outcome
  - Method by Cox
  - Method by Glas et al.
  - Method by Hedges and Olkin

- Obtaining Mean Change from Baseline with Standard Deviation from Mean Baseline with Standard Deviation and Mean Postintervention with Standard Deviation with a Method by Follmann et al.

All of these methods except the algebraic conversion methods were evaluated in terms of causing bias when applied for computing uniform effect measures for interventional studies that are pooled in a meta-analysis.

For methods applied to obtain the standardized mean difference and standard error from a 2 x 2 contingency table for a dichotomized continuous outcome, only the population standardized difference in means was applicable as estimand.

7.1.5 Performance Measures

Mean bias of the pooled effect estimate and coverage of the 95% confidence interval with the corresponding monte carlo standard errors were used as performance measures.

7.2 Results

7.2.1 Obtaining Mean and Standard Deviation from Mean and Range

The detailed results for the "range"-method and the method by Walter and Yao are listed in Table 7.2 and Table 7.3.

After application of these methods for obtaining mean and standard deviation from mean and range of a continuous outcome measured in intervention and control groups included in studies that are included in meta-analyses, there are no major differences in bias and coverage between the meta-analytically pooled treatment effects, except in case of skewed distributions of individual residuals, large treatment effects, as well as large samples. In this case, the "range" method performed better and resulted in less bias and higher coverage.

The "Range"-Method

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.035	0.054	0.930	0.008
Skewed	Small	Small	0.084	0.052	0.919	0.009
Symmetric	Large	Small	0.004	0.055	0.928	0.008
Skewed	Large	Small	0.055	0.052	0.917	0.009
Symmetric	Small	Large	0.037	0.041	0.933	0.008
Skewed	Small	Large	-0.009	0.040	0.921	0.009
Symmetric	Large	Large	-0.022	0.040	0.924	0.008
Skewed	Large	Large	0.049	0.040	0.923	0.008
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.072	0.004	0.894	0.010
Skewed	Small	Small	0.194	0.008	0.731	0.014
Symmetric	Large	Small	0.147	0.007	0.821	0.012
Skewed	Large	Small	0.380	0.014	0.519	0.016
Symmetric	Small	Large	-0.031	0.002	0.877	0.010
Skewed	Small	Large	0.070	0.004	0.879	0.010
Symmetric	Large	Large	-0.063	0.004	0.813	0.012
Skewed	Large	Large	0.130	0.007	0.840	0.012

Table 7.2: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Method by Walter and Yao

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.054	0.055	0.921	0.009
Skewed	Small	Small	0.091	0.052	0.916	0.009
Symmetric	Large	Small	0.001	0.055	0.922	0.008
Skewed	Large	Small	0.054	0.052	0.912	0.009
Symmetric	Small	Large	0.040	0.041	0.930	0.008
Skewed	Small	Large	-0.014	0.039	0.923	0.008
Symmetric	Large	Large	-0.019	0.040	0.926	0.008
Skewed	Large	Large	0.044	0.039	0.931	0.008
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.031	0.003	0.926	0.008
Skewed	Small	Small	0.143	0.006	0.817	0.012
Symmetric	Large	Small	0.064	0.005	0.910	0.009
Skewed	Large	Small	0.278	0.011	0.689	0.015
Symmetric	Small	Large	0.049	0.003	0.894	0.010
Skewed	Small	Large	0.166	0.007	0.719	0.014
Symmetric	Large	Large	0.093	0.005	0.864	0.011
Skewed	Large	Large	0.319	0.012	0.553	0.016

Table 7.3: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

7.2.2 Obtaining Mean and Standard Deviation from Median and Range

Bias and coverage after application of the method by Hozo et al. and the method by Wan et al. for obtaining mean and standard deviation from median and range can be found in Table 7.4 and Table 7.5.

Overall, the method by Wan et al. performs better. There is an overestimation of the population standardized difference in means after application of both methods, especially when the distributions of individual residuals are skewed positively.

Method by Hozo et al.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.094	0.060	0.936	0.008
Skewed	Small	Small	0.086	0.057	0.925	0.008
Symmetric	Large	Small	-0.108	0.062	0.919	0.009
Skewed	Large	Small	-0.069	0.059	0.928	0.008
Symmetric	Small	Large	0.020	0.047	0.919	0.009
Skewed	Small	Large	-0.014	0.043	0.918	0.009
Symmetric	Large	Large	0.019	0.044	0.930	0.008
Skewed	Large	Large	0.049	0.043	0.914	0.009
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.045	0.004	0.912	0.009
Skewed	Small	Small	0.156	0.007	0.802	0.013
Symmetric	Large	Small	0.073	0.005	0.903	0.009
Skewed	Large	Small	0.262	0.011	0.709	0.014
Symmetric	Small	Large	0.071	0.004	0.861	0.011
Skewed	Small	Large	0.190	0.008	0.688	0.015
Symmetric	Large	Large	0.135	0.006	0.800	0.013
Skewed	Large	Large	0.366	0.014	0.502	0.016

Table 7.4: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Method by Wan et al.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.030	0.060	0.932	0.008
Skewed	Small	Small	0.112	0.058	0.913	0.009
Symmetric	Large	Small	0.031	0.061	0.915	0.009
Skewed	Large	Small	0.043	0.059	0.927	0.008
Symmetric	Small	Large	0.053	0.049	0.926	0.008
Skewed	Small	Large	0.028	0.050	0.909	0.009
Symmetric	Large	Large	-0.077	0.049	0.927	0.008
Skewed	Large	Large	0.020	0.051	0.914	0.009
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.033	0.004	0.926	0.008
Skewed	Small	Small	0.146	0.007	0.823	0.012
Symmetric	Large	Small	0.067	0.005	0.904	0.009
Skewed	Large	Small	0.282	0.012	0.694	0.015
Symmetric	Small	Large	0.047	0.003	0.905	0.009
Skewed	Small	Large	0.166	0.007	0.744	0.014
Symmetric	Large	Large	0.086	0.005	0.885	0.010
Skewed	Large	Large	0.314	0.012	0.588	0.016

Table 7.5: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

7.2.3 Obtaining Mean and Standard Deviation from Median and Interquartile Range

Performance measures after application of the method by Wan et al. and the Cochrane method for obtaining mean and standard deviation from median and interquartile range can be found in Table 7.6 and Table 7.7.

The application of the method by Wan et al. leads to lower overall bias and higher coverage than the application of the Cochrane method. After applying both methods, there is an overestimation of the standardized population mean difference, especially if the distributions of the individual residuals are positively skewed.

Method by Wan et al.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.029	0.058	0.932	0.008
Skewed	Small	Small	0.078	0.055	0.901	0.009
Symmetric	Large	Small	-0.036	0.059	0.924	0.008
Skewed	Large	Small	0.050	0.053	0.921	0.009
Symmetric	Small	Large	0.010	0.042	0.930	0.008
Skewed	Small	Large	-0.002	0.040	0.932	0.008
Symmetric	Large	Large	-0.009	0.041	0.926	0.008
Skewed	Large	Large	0.038	0.040	0.924	0.008
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.034	0.004	0.930	0.008
Skewed	Small	Small	0.157	0.007	0.807	0.012
Symmetric	Large	Small	0.066	0.005	0.914	0.009
Skewed	Large	Small	0.306	0.012	0.649	0.015
Symmetric	Small	Large	0.047	0.003	0.898	0.010
Skewed	Small	Large	0.182	0.007	0.677	0.015
Symmetric	Large	Large	0.092	0.005	0.851	0.011
Skewed	Large	Large	0.350	0.013	0.481	0.016

Table 7.6: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Cochrane Method

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.017	0.064	0.931	0.008
Skewed	Small	Small	0.117	0.059	0.910	0.009
Symmetric	Large	Small	-0.085	0.064	0.928	0.008
Skewed	Large	Small	0.024	0.058	0.930	0.008
Symmetric	Small	Large	0.006	0.047	0.920	0.009
Skewed	Small	Large	0.000	0.042	0.925	0.008
Symmetric	Large	Large	0.010	0.044	0.934	0.008
Skewed	Large	Large	0.045	0.043	0.919	0.009
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.080	0.005	0.884	0.010
Skewed	Small	Small	0.212	0.009	0.712	0.014
Symmetric	Large	Small	0.149	0.007	0.843	0.012
Skewed	Large	Small	0.412	0.015	0.503	0.016
Symmetric	Small	Large	0.061	0.004	0.867	0.011
Skewed	Small	Large	0.200	0.008	0.637	0.015
Symmetric	Large	Large	0.122	0.006	0.803	0.013
Skewed	Large	Large	0.385	0.014	0.431	0.016

Table 7.7: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

7.2.4 Obtaining Mean and Standard Deviation from Median, Range, and Interquartile Range

Performance measures after application of the method by Bland and the method by Wand et al. for obtaining mean and standard deviation from median, range, and interquartile range can be found in Table 7.8 and Table 7.9.

There is no larger difference in bias and coverage for the standardized difference in means. The method by Wan et al. results in less bias and higher coverage of estimates of the absolute difference in means. Application of both methods tends to result in an overestimation of both the absolute and standardized difference in means.

Method by Bland

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.372	0.059	0.925	0.008
Skewed	Small	Small	0.489	0.058	0.901	0.009
Symmetric	Large	Small	0.823	0.065	0.906	0.009
Skewed	Large	Small	0.857	0.062	0.892	0.010
Symmetric	Small	Large	0.157	0.043	0.930	0.008
Skewed	Small	Large	0.137	0.042	0.923	0.008
Symmetric	Large	Large	0.203	0.043	0.928	0.008
Skewed	Large	Large	0.279	0.043	0.930	0.008
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.104	0.005	0.848	0.011
Skewed	Small	Small	0.230	0.009	0.644	0.015
Symmetric	Large	Small	0.211	0.008	0.698	0.015
Skewed	Large	Small	0.451	0.016	0.410	0.016
Symmetric	Small	Large	0.009	0.003	0.925	0.008
Skewed	Small	Large	0.115	0.006	0.821	0.012
Symmetric	Large	Large	0.015	0.004	0.932	0.008
Skewed	Large	Large	0.216	0.009	0.734	0.014

Table 7.8: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Method by Wan et al.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.036	0.056	0.930	0.008
Skewed	Small	Small	0.083	0.053	0.911	0.009
Symmetric	Large	Small	0.012	0.057	0.920	0.009
Skewed	Large	Small	0.056	0.053	0.922	0.008
Symmetric	Small	Large	0.033	0.042	0.933	0.008
Skewed	Small	Large	0.014	0.041	0.924	0.008
Symmetric	Large	Large	-0.045	0.042	0.927	0.008
Skewed	Large	Large	0.026	0.042	0.928	0.008
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.030	0.003	0.925	0.008
Skewed	Small	Small	0.146	0.006	0.807	0.012
Symmetric	Large	Small	0.062	0.005	0.915	0.009
Skewed	Large	Small	0.286	0.012	0.668	0.015
Symmetric	Small	Large	0.046	0.003	0.902	0.009
Skewed	Small	Large	0.172	0.007	0.712	0.014
Symmetric	Large	Large	0.086	0.005	0.866	0.011
Skewed	Large	Large	0.327	0.013	0.528	0.016

Table 7.9: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

7.2.5 Obtaining Mean and Standard Deviation from Mean and Missing Standard Deviation by Imputation

Bias and coverage after application of the method by Furukawa et al. and the method by Marinho et al. for the imputation of missing standard deviation can be found in Table 7.10 and Table 7.11.

There is no larger difference in bias and coverage for the standardized difference in means. The method by Furukawa et al. results in less bias and higher coverage of estimates of the standardized difference in means. There is an overestimation of the standardized difference in means after application of both methods.

Method by Furukawa et al.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.026	0.056	0.944	0.007
Skewed	Small	Small	-0.023	0.056	0.915	0.009
Symmetric	Large	Small	-0.004	0.057	0.929	0.008
Skewed	Large	Small	0.021	0.051	0.937	0.008
Symmetric	Small	Large	0.056	0.039	0.944	0.007
Skewed	Small	Large	-0.069	0.038	0.940	0.008
Symmetric	Large	Large	-0.025	0.040	0.922	0.008
Skewed	Large	Large	-0.016	0.041	0.912	0.009
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.021	0.003	0.928	0.008
Skewed	Small	Small	0.080	0.005	0.868	0.011
Symmetric	Large	Small	0.037	0.004	0.890	0.010
Skewed	Large	Small	0.151	0.008	0.770	0.013
Symmetric	Small	Large	0.032	0.003	0.909	0.009
Skewed	Small	Large	0.089	0.005	0.838	0.012
Symmetric	Large	Large	0.061	0.004	0.862	0.011
Skewed	Large	Large	0.177	0.009	0.680	0.015

Table 7.10: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Method by Marinho et al.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.013	0.057	0.938	0.008
Skewed	Small	Small	-0.049	0.055	0.918	0.009
Symmetric	Large	Small	0.008	0.057	0.929	0.008
Skewed	Large	Small	-0.010	0.050	0.939	0.008
Symmetric	Small	Large	0.062	0.039	0.940	0.008
Skewed	Small	Large	-0.061	0.038	0.930	0.008
Symmetric	Large	Large	-0.019	0.040	0.920	0.009
Skewed	Large	Large	-0.022	0.041	0.902	0.009
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.030	0.004	0.917	0.009
Skewed	Small	Small	0.105	0.006	0.835	0.012
Symmetric	Large	Small	0.056	0.005	0.867	0.011
Skewed	Large	Small	0.200	0.009	0.712	0.014
Symmetric	Small	Large	0.040	0.003	0.898	0.010
Skewed	Small	Large	0.113	0.005	0.795	0.013
Symmetric	Large	Large	0.078	0.005	0.846	0.011
Skewed	Large	Large	0.224	0.010	0.615	0.015

Table 7.11: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

7.2.6 Obtaining the Standardized Mean Difference and Standard Error from a 2 x 2 Contingency Table for a Dichotomized Continuous Outcome

Bias and coverage after application of the method by Cox, the method by Glas et al. and the method by Hedges and Olkin for obtaining the standardized mean difference and standard error from a 2 x 2 contingency table for a dichotomized continuous outcome can be found in Table 7.12, Table 7.13 and Table 7.14.

Both the method by Cox and the method by Glas et al. show better performance measures than the method by Hedges and Olkin. The methods of Cox and Glas et al. perform quite similarly, with Cox's method having a slight advantage in a majority of the created scenarios.

Method by Cox

Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.030	0.004	0.947	0.007
Skewed	Small	Small	0.042	0.004	0.935	0.008
Symmetric	Large	Small	-0.034	0.005	0.923	0.008
Skewed	Large	Small	-0.155	0.007	0.774	0.013
Symmetric	Small	Large	0.042	0.003	0.911	0.009
Skewed	Small	Large	0.080	0.005	0.863	0.011
Symmetric	Large	Large	0.064	0.004	0.889	0.010
Skewed	Large	Large	-0.015	0.004	0.904	0.009

Table 7.12: Performance measures for estimating the population standardized difference in means λ_3 between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Method by Glas et al.

Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.050	0.004	0.929	0.008
Skewed	Small	Small	0.059	0.005	0.927	0.008
Symmetric	Large	Small	-0.018	0.004	0.926	0.008
Skewed	Large	Small	-0.139	0.007	0.799	0.013
Symmetric	Small	Large	0.053	0.003	0.895	0.010
Skewed	Small	Large	0.087	0.005	0.861	0.011
Symmetric	Large	Large	0.058	0.004	0.886	0.010
Skewed	Large	Large	-0.024	0.004	0.908	0.009

Table 7.13: Performance measures for estimating the population standardized difference in means λ_3 between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Method by Hedges and Olkin

Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	-0.165	0.006	0.346	0.015
Skewed	Small	Small	-0.202	0.007	0.225	0.013
Symmetric	Large	Small	-0.438	0.014	0.007	0.003
Skewed	Large	Small	-0.563	0.018	0.001	0.001
Symmetric	Small	Large	-0.150	0.005	0.233	0.013
Skewed	Small	Large	-0.184	0.006	0.207	0.013
Symmetric	Large	Large	-0.389	0.012	0.003	0.002
Skewed	Large	Large	-0.505	0.016	0.002	0.001

Table 7.14: Performance measures for estimating the population standardized difference in means λ_3 between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

7.2.7 Obtaining Mean Change from Baseline with Standard Deviation from Mean Baseline with Standard Deviation and Mean Postintervention with Standard Deviation with a Method by Follmann et al.

Bias and coverage after application of the method by Follmann et al. for obtaining mean change from baseline with standard deviation from mean baseline, mean postintervention, mean baseline standard deviation, and mean postintervention standard deviation can be found in Table 7.15.

There is no method for comparison. There is an overestimation of the standardized difference in means, especially for skewed distributions of individual residuals.

Absolute difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.086	0.055	0.936	0.008
Skewed	Small	Small	0.009	0.053	0.911	0.009
Symmetric	Large	Small	0.042	0.053	0.940	0.008
Skewed	Large	Small	0.141	0.050	0.920	0.009
Symmetric	Small	Large	-0.017	0.040	0.926	0.008
Skewed	Small	Large	-0.035	0.040	0.923	0.008
Symmetric	Large	Large	-0.080	0.040	0.934	0.008
Skewed	Large	Large	0.064	0.040	0.901	0.009
Standardized difference in means						
Distribution	Effect	Sample	Bias	MCSE	Coverage	MCSE
Symmetric	Small	Small	0.042	0.004	0.924	0.008
Skewed	Small	Small	0.139	0.006	0.808	0.012
Symmetric	Large	Small	0.074	0.005	0.906	0.009
Skewed	Large	Small	0.264	0.011	0.695	0.015
Symmetric	Small	Large	0.047	0.003	0.899	0.010
Skewed	Small	Large	0.159	0.007	0.732	0.014
Symmetric	Large	Large	0.089	0.005	0.879	0.010
Skewed	Large	Large	0.313	0.012	0.542	0.016

Table 7.15: Performance measures for estimating the population absolute difference in means δ_{**3*} and population standardized difference in means λ_{**3*} between change from baseline values of the intervention and control group. Distribution = distribution of individual baseline residual e_{ug1p} and individual postintervention residual e_{ug2p}; effect = treatment effect d_{u*2*}; sample = sample size n_{ug**}; bias = mean bias of the absolute difference in means $bias_{\hat{\delta}_{**3*}}$ and mean bias of the standardized difference in means $bias_{\hat{\lambda}_{**3*}}$; coverage = coverage of absolute difference in means $cover_{\hat{\delta}_{**3*}}$ and coverage of the standardized difference in means $cover_{\hat{\lambda}_{**3*}}$; MCSE = monte carlo standard error.

Chapter 8

R Package

Several software packages have been developed to help researchers perform meta-analyses. These include packages for R. R is a programming language for statistical computing and creating figures and has gained popularity among researchers in recent years thanks to its open source software environment. The free software environment allows users to create packages that extend the features of the R language [25]. R software packages have been developed for pooling effect measures from individual studies in a meta-analysis [26, 27]. To the best of my knowledge, there is no R software package that allows the calculation of consistent effect measures from continuous outcome data presented in different ways.

The *uniform* package was created for this purpose and can be installed in R by executing the following R code:

```
library(devtools)
devtools::install_github(``bkendzio/uniform'')
library(uniform)
```

The install_github function of the devtools package [28] allows the installation of packages hosted on GitHub [29]. After installing the package, the vignette of the package can be accessed via:

```
vignette(``uniformVignette'',
package = ``uniform'')
```

This vignette, which includes a step-by-step guide for using the functions of the package, is shown on the next pages.

uniform: Unification of a Differently or Incompletely Reported Continuous Outcome across Studies Included in a Meta-Analysis

The goals of *uniform* are the unification of a differently or incompletely reported continuous outcome across studies included in a meta-analysis and calculation of consistent effect measures so that they can be pooled. Currently, *uniform* comprises the following functions:

- `co.co` uniforms a differently or incompletely reported continuous outcome across studies included in a meta-analysis.
- `em.co_interventionalControlled` calculates effect measures for continuous continuous outcome data given by interventional controlled studies and uniformed by the co.co function.

In the following, there is a step-by-step guide for using the functions of the package.

Installation

The package *uniform* can be installed from GitHub and loaded via:

```
library(devtools)
devtools::install_github("bkendzio/uniform")
library(uniform)
```

Challenge and Example Data Set

The synthesis of evidence in meta-analyses is done by pooling effect measures from individual studies. Both an effect measure and a measure of dispersion are needed per study for pooling effects across studies. Effect measures are statistical constructs that measure the strength of the relationship between two variables in a population, e.g. the strength of the link between exposure and outcome (Tripepi et al. 2007). Most commonly, an outcome is campared between an exposed and a nonexposed group for calculating an effect measure (Higgins, Li, and Deeks 2019). Continuous outcomes for groups of observations can be reported differently, with different measures of central tendency (e.g., mean or median per group) and dispersion (e.g., standard deviation vs. interquartile range). In addition, some studies do not report a measure of dispersion at all. These differences in reporting considerably complicate or prevent the calculation of consistent effect measures for pooling the results of individual studies in a meta-analysis and may lead to trade-offs that result in the exclusion of relevant studies.

The artificially created example data set (`dataCoRaw`), which is included in the package, illustrates such a situation. The data set contains information of 13 interventional controlled studies that shall be included in a meta-analysis. The studies measured a continuous outcome at baseline (t=1) and postintervention (t=2) and calculated the change from baseline (t=3) for an intervention (group=1) and a control group (group=2), respectively. As it is common in reality, the studies reported the continuous outcome differently, with different measures of central tendency (e.g., mean or median) and dispersion (e.g., standard deviation or interquartile range). In addition, some measurements were not reported. The reported measures of central tendency and dispersion of the outcome of the first 6 patient groups can be viewed after running the following code:

```
data(dataCoRaw)
head(dataCoRaw)
#>   study group t  n      mean        m         s       se      ll95     ul95
#> 1     1     1 1 26  98.25075       NA        NA       NA  91.94109 104.5604
#> 2     2     1 1 26        NA 87.78617 27.255960 5.345334        NA 106.1816
#> 3     3     1 1 12        NA 89.57264        NA 2.752580  82.99739  93.7873
#> 4     4     1 1 25  97.29694       NA 24.909225 4.981845        NA 107.0612
#> 5     5     1 1 21 101.81822 98.64052 24.043639 5.246752  91.53477       NA
#> 6     6     1 1 11 105.54045       NA  8.127579 2.450557 100.73745       NA
#>       ll90      ul90     ll99     ul99        a        b       lq        uq
#> 1 92.95551        NA 89.95844 106.5431 66.87363 121.4857 86.01877 113.37884
#> 2 86.91268 104.49727       NA 109.4736 58.26235 146.5305       NA 113.26442
#> 3 83.86476  92.91994 81.30217       NA 71.93112 105.7785       NA  90.72357
#> 4 89.10253 105.49134 84.46456 110.1293 51.52891       NA 86.33158 108.94238
#> 5 93.18808 110.44836 88.30348 115.3330 72.10060 156.3750 86.56925       NA
#> 6       NA 109.57126       NA 111.8527       NA 119.7508 99.30241 109.79633
```

- `study`: An integer vector with study labels.
- `group`: An integer vector with group labels.
- `t`: An integer vector with time points of outcome measurement.
- `n`: An integer vector with numbers of observations in groups at time points.
- `mean`: A vector of means of the outcome in groups at time points.
- `m`: A vector of medians of the outcome in groups at time points.
- `s`: A vector of standard deviations of the outcome in groups at time points.
- `se`: A vector of standard errors of the mean of the outcome in groups at time points.
- `ll95`: A vector of the lower limits of the 95 percent confidence interval of the mean outcome in groups at time points.
- `ul95`: A vector of the upper limits of the 95 percent confidence interval of the mean outcome in groups at time points.
- `ll95`: A vector of the lower limits of the 90 percent confidence interval of the mean outcome in groups at time points.
- `ul95`: A vector of the upper limits of the 90 percent confidence interval of the mean outcome in groups at time points.
- `ll99`: A vector of the lower limits of the 99 percent confidence interval of the mean outcome in groups at time points.
- `ul95`: A vector of the upper limits of the 99 percent confidence interval of the mean outcome in groups at time points.
- `a`: A vector of minima of the outcome in groups at time points.
- `b`: A vector of maxima of the outcome in groups at time points.
- `lq`: A vector of lower quartils of the outcome in groups at time points.
- `uq`: A vector of upper quartils of the outcome in groups at time points.

Unification

`co.co`

The `co.co` function uniforms a differently or incompletely reported continuous outcome across studies included in a meta-analysis by converting the provided results for each group to mean and standard deviation where possible. For this purpose, conversion and imputation methods previously introduced in the literature are used:

- Obtaining Mean and Standard Deviation from Mean and Standard Error by Algebraic Conversion

- Obtaining Mean and Standard Deviation from Mean and Confidence Interval by Algebraic Conversion
- Obtaining Mean and Standard Deviation from Mean and Range
  - The "Range"-Method (Higgins, Li, and Deeks 2019)
  - Method by Walter and Yao (Walter and Yao 2007)
- Obtaining Mean and Standard Deviation form Median and Range
  - Method by Hozo et al. (Hozo, Djulbegovic, and Hozo 2005)
  - Method by Wan et al. (Wan et al. 2014)
- Obtaining Mean and Standard Deviation from Median and Interquartile Range
  - Method by Wan et al. (Wan et al. 2014)
  - Cochrane Method (Higgins, Li, and Deeks 2019)
- Obtaining Mean and Standard Deviation form Median, Range, and Interquartile Range
  - Method by Bland (Bland 2015)
  - Method by Wan et al. (Wan et al. 2014)
- Obtaining Mean and Standard Deviation from Mean and Missing Standard Deviation by Imputation
  - Method by Furukawa et al. (Furukawa et al. 2006)
  - Method by Marinho et al. (Marinho et al. 2003)

If the data is structured similar to the example data set with the same names for the variables, the function can be used by defining the data set in a single argument:

```
dataCoUniform <- co.co(data=dataCoRaw)
```

If the data set is structured differently with different names for the variables, the function can be used by defining vectors of study labels, group labels, time points of outcome measurement, numbers of observations, and measures of central tendency and dispersion separately:

```
dataCoUniform <- co.co(study=dataCoRaw$study,group=dataCoRaw$group,t=dataCoRaw$t,
                       n=dataCoRaw$n,mean=dataCoRaw$mean,m=dataCoRaw$m,
                       s=dataCoRaw$s,se=dataCoRaw$se,ll95=dataCoRaw$ll95,
                       ul95=dataCoRaw$ul95,ll90=dataCoRaw$ll90,ul90=dataCoRaw$ul90,
                       ll99=dataCoRaw$ll99,ul99=dataCoRaw$ul99,a=dataCoRaw$a,
                       b=dataCoRaw$b,lq=dataCoRaw$lq,uq=dataCoRaw$uq)
```

The following optional arguments can be used to choose between different methods of conversion and imputation:

- `meanSd.meanAB`: An optional character string indicating which method is used for obtaining mean and standard deviation from mean and range. Either "range" for the range method (default), or "walterYao" for the method by Walter and Yao.
- `meanSd.mIqrRange`: An optional character string indicating which method is used for obtaining mean and standard deviation from median, interquartile range, and range. Either "Bland" for the method by bland, or "wanEtal" for the method by Wan et al. (default).
- `meanSd.mIqr`: An optional character string indicating which method is used for obtaining mean and standard deviation from median and interquartile range. Either "wanEtal" for the method by Wan et al. (default), or "cochrane" for Cochrane method.
- `meanSd.mRange`: An optional character string indicating which method is used for obtaining mean and standard deviation from median and range. Either "hozoEtal" for the method by Hozo et al., or "wanEtal" for the method by Wan et al. (default).

- `meanSd.mean`: An optional character string indicating which method is used for obtaining mean and standard deviation from mean and missing standard deviation by imputation. Either "furukawaEtal" for the method by Furukawa et al. (default), or "marinhoEtal" for the method by Marinho et al..

The function returns a data frame object that contains study labels, group labels, time points of outcome measurement, numbers of observations in groups at time points, means of outcome in groups at time points, standard deviations of outcome in groups at time points:

```
head(dataCoUniform)
#>   study group t  n      mean         s
#> 1     1     1 1 26  98.25075 16.414850
#> 2     2     1 1 26        NA 27.255960
#> 3     3     1 1 12  89.21373 21.325936
#> 4     4     1 1 25  97.29694 24.909225
#> 5     5     1 1 21 101.81822 24.043639
#> 6     6     1 1 11 105.54045  8.127579
```

Effect Measure Calculation

Depending on the study design of the studies, different effect measures can be calculated and pooled in a meta-analysis. Version 0.0.1 of the *uniform* package allows to calculate effect measures for continuous outcome data given by interventional controlled studies.

`em.co_interventionalControlled`

The data frame returned by the `co.co` function can be included in the `em.co_interventionalControlled` function. The function calculates the absolute difference in means between the intervention and control group or the standardized difference in means Hedges' g, either adjusted for small sample bias or unadjusted (Hedges and Olkin 1985). Before, the method by Follmann et al. for obtaining the mean change from baseline with standard deviation from mean baseline with standard deviation and mean postintervention with standard deviation is used to calculate the mean change from baseline with standard deviation in case of missed reporting (Follmann et al. 1992).

In addition to the data frame object returned by the `co.co` function, the following information must be passed into the `em.co_interventionalControlled` function:

- `groupIntervention`: A character string that specifies how the elements of the group vector that define the intervention groups are labeled (e.g., "1" for intervention groups).
- `groupControl`: A character string that specifies how the elements of the group vector that define the control groups are labeled (e.g., "2" for control groups).
- `tBaseline`: A character string that specifies how the elements of the time point vector that define baseline measurements are labeled (e.g., "1" for baseline measurements).
- `tPost`: A character string that specifies how the elements of the time point vector that define postintervention measurements are labeled (e.g., "2" for postintervention measurements).
- `tChange`: A character string that specifies how the elements of the time point vector that define change from baseline measurements are labeled (e.g., "3" for change from baseline measurements).

The function can then be executed by:

```
dataEm<-em.co_interventionalControlled(data=dataCoUniform,
                                       groupIntervention="1",groupControl="2",tBaseline="1",
                                       tPost="2",tChange="3")
```

The following optional arguments can be used to change the type of effect measure and specify the calculation:

- `em`: An optional character string indicating which summary measure is calculated. Either "md" for mean difference (default), or "smd" for Hedges g' as standardized mean difference.
- `smdMethod`: An optional character string indicating if Hedges g' is adjusted to account for a positive bias for small sample sizes. Either "hedgesAdjusted" for Hedges adjusted g', or "hedgesUnadjusted" for Hedges unadjusted g'.
- `combineChangePost`: An optional logical indicating whether postintervention measurements should be used instead of change from baseline measurements for studies for which change from baseline measurements cannot be obtained and when the absolute mean difference is used as effect measure. Default is FALSE.

With usage of optional arguments, the function may be executed by:

```
dataEm<-em.co_interventionalControlled(data=dataCoUniform,groupIntervention="1",
                                       groupControl="2",tBaseline="1",tPost="2",tChange="3",
                                       em="smd",smdMethod="hedgesAdjusted",
                                       combineChangePost=TRUE)
```

The function returns a data frame object that contains study labels, numbers of observations in the intervention groups, numbers of observations in the control groups, effect measures, and standard errors for the studies included in the meta-analysis.

```
head(dataEm)
#>   study nIntervention nControl         smd        se
#> 1     1            26       26  0.39786965 0.2800807
#> 2     2            26       26 -0.07910397 0.2774585
#> 3     3            12       12  0.53719940 0.4155464
#> 4     4            25       25  0.27115222 0.2841395
#> 5     5            21       21  0.79729749 0.3206334
#> 6     6            11       11  0.89229024 0.4471166
```

Afterwards, the effect measures can easily be pooled by using functions of other packages, e.g. the `metagen` function of the meta package (Balduzzi, Rücker, and Schwarzer 2019).

References

Balduzzi, Sara, Gerta Rücker, and Guido Schwarzer. 2019. "How to Perform a Meta-Analysis with r: A Practical Tutorial." *Evidence-Based Mental Health* 22 (4): 153–60.

Bland, Martin. 2015. "Estimating Mean and Standard Deviation from the Sample Size, Three Quartiles, Minimum, and Maximum." *International Journal of Statistics in Medical Research* 4 (1): 57–64.

Follmann, Dean, Paul Elliott, IL Suh, and Jeffrey Cutler. 1992. "Variance Imputation for Overviews of Clinical Trials with Continuous Response." *Journal of Clinical Epidemiology* 45 (7): 769–73.

Furukawa, Toshi A, Corrado Barbui, Andrea Cipriani, Paolo Brambilla, and Norio Watanabe. 2006. "Imputing Missing Standard Deviations in Meta-Analyses Can Provide Accurate Results." *Journal of Clinical Epidemiology* 59 (1): 7–10.

Hedges, LV, and I Olkin. 1985. "Statistical Methods for Meta-Analysis." Academic Press.

Higgins, Julian PT, Tianjing Li, and Jonathan J Deeks. 2019. "Choosing Effect Measures and Computing Estimates of Effect." *Cochrane Handbook for Systematic Reviews of Interventions*, 143–76.

Hozo, Stela Pudar, Benjamin Djulbegovic, and Iztok Hozo. 2005. "Estimating the Mean and Variance from the Median, Range, and the Size of a Sample." *BMC Medical Research Methodology* 5 (1): 1–10.

Marinho, Valeria CC, Julian Higgins, Stuart Logan, and Aubrey Sheiham. 2003. "Fluoride Toothpastes for Preventing Dental Caries in Children and Adolescents." *Cochrane Database of Systematic Reviews*, no. 1.

Tripepi, G, KJ Jager, FW Dekker, C Wanner, and C Zoccali. 2007. "Measures of Effect: Relative Risks, Odds Ratios, Risk Difference, and 'Number Needed to Treat'." *Kidney International* 72 (7): 789–91.

Walter, SD, and X Yao. 2007. "Effect Sizes Can Be Calculated for Studies Reporting Ranges for Outcome Variables in Systematic Reviews." *Journal of Clinical Epidemiology* 60 (8): 849–52.

Wan, Xiang, Wenqian Wang, Jiming Liu, and Tiejun Tong. 2014. "Estimating the Sample Mean and Standard Deviation from the Sample Size, Median, Range and/or Interquartile Range." *BMC Medical Research Methodology* 14 (1): 1–13.

Chapter 9

Discussion

9.1 Summary

This master's thesis was conducted with the goal of comparing different methods for the conversion and imputation of differently or incompletely reported continuous outcomes, setting up a guidance on which methods should be used, and creating an R package for the conversion and imputation of continuous outcomes and calculation of uniform effect measures.

Based on the results of simulation studies that compared conversion and imputation methods in terms of precision when applied for computing uniform effect measures for interventional controlled studies that are pooled in a meta-analysis, the conversion and imputation methods written in bold in the following should be used:

- Obtaining Mean and Standard Deviation from Mean and Standard Error by **Algebraic Conversion**
- Obtaining Mean and Standard Deviation from Mean and Confidence Interval by **Algebraic Conversion**
- Obtaining Mean and Standard Deviation from Mean and Range

  - **The "Range"-Method** [5]
  - Method by Walter and Yao [7]

- Obtaining Mean and Standard Deviation form Median and Range
  - Method by Hozo et al. [10]
  - **Method by Wan et al.** [11]

- Obtaining Mean and Standard Deviation from Median and Interquartile Range
  - **Method by Wan et al.** [11]
  - Cochrane Method [5]

- Obtaining Mean and Standard Deviation form Median, Range, and Interquartile Range
  - Method by Bland [12]
  - **Method by Wan et al.** [11]

- Obtaining Mean and Standard Deviation from Mean and Missing Standard Deviation by Imputation
  - **Method by Furukawa et al.** [13]
  - Method by Marinho et al. [**?**, 14]

- Obtaining the Standardized Mean Difference and Standard Error from a 2 x 2 Contingency Table for a Dichotomized Continuous Outcome
  - **Method by Cox** [16]
  - Method by Glas et al. [17]
  - Method by Hedges and Olkin [19]

- Obtaining Mean Change from Baseline with Standard Deviation from Mean Baseline, Mean Postintervention, Mean Baseline Standard Deviation, and Mean Postintervention Standard Deviation with a **Method by Follmann et al.**

The R package *uniform* was created. It can be downloaded from GitHub and allows the unification of a differently or incompletely reported continuous outcome across studies included in a meta-analysis and the calculation of consistent effect measures for interventional controlled studies.

The *co.co* function of the *uniform* package uniforms a differently of incompletely reported continuous outcome across studies. The above mentioned conversion and imputation methods were integrated in the function. The methods in bold are the default methods. There are two exceptions. Methods for obtaining the standardized mean difference and standard error from a 2x2 contingency table for a dichotomized continuous outcome and the method by Follmann et al. for obtaining the mean change from baseline with standard deviation from mean baseline with standard deviation and mean postintervention with standard deviation were not integrated since these methods conceptually differ from the other methods, which is explained in Chapter 3. In short, these methods do not result in the mean and standard deviation for a single group of observations of a continuous outcome.

The *em.co_interventionalControlled* function calculates effect measures for continuous outcome data given by interventional controlled studies and uniformed by the co.co function. The method by Follmann et al. for obtaining the mean change from baseline with standard deviation from mean baseline with standard deviation and mean postintervention with standard deviation is integrated in this function.

9.2 Previous Research

Some of the works that introduced the included conversion and imputation methods also performed simulation studies. In most of these studies, measures of central tendency and/or dispersion (e.g., mean and/or standard deviation) of an outcome in a single group of observations were used as estimand, while effect measures (absolute difference in means and standardized difference in means) comparing an outcome between two groups of observations were used as estimands in the simulation studies of this thesis.

The previous simulation studies by Hozo et al. [10], Wan et al. [11], and Bland [12] are presented in more detail in Chapter 3. Shortly summarized, Wan et al. compared their newly developed methods for obtaining mean and standard deviation from median and range, for obtaining mean and standard deviation from median and interquartile range, and for obtaining mean and standard deviation from median, range, and interquartile range to existing methods and found an improvement in precision. Bland aimed at improving the method by Hozo et al. for obtaining mean and standard deviation from median and range by incoroprating the interquartile range and found an improvement in precision in simulation studies. The findings of our simulation studies do not contradict these findings. Walter and Yao [7] suggested a method for obtaining mean and standard deviation from mean and range that is more sophisticated than the "range"-method by the incorporation of a conversion factor that decreases with sample size. Walter and Yao did not back this method up by simulation studies, and interestingly, the simulation studies of this thesis do not suggest higher precision.

9.3 Limitations

First, the simulation studies of this thesis aimed at comparing conversion and imputation methods for differently or incompletely reported continuous outcomes in terms of precision when applied for computing uniform effect mea-

sures for interventional controlled studies that are pooled in a meta-analysis. Study designs different from the interventional controlled setting could lead to other results. Second, the simulation studies of this thesis included several data generation mechanisms with different assumptions. Other scenarios are possible and may lead to different results. Third, data from 1000 meta-analyses per data generation scenario were simulated. Each meta-analysis included between 10 and 30 studies that were also simulated. To keep the time expense reasonable, the number of simulated meta-analyses per data generation scenario had to be kept at the relatively low number of 1000, which affected the accuracy of the results. Fourth, the conversion and imputation methods included in this thesis were found in a literature search that was not systematic with a previously defined search string and inclusion of several data bases. Consequently, the likelihood that relevant conversion and imputation methods may have been missed is increased.

9.4 Outlook

The current version 0.0.1 of the *uniform* R package allows to uniform a differently or incompletely reported continuous outcome across studies included in a meta-analysis by using the *co.co* function. The function can be broadly applied on different study designs. The *em.co_interventionalControlled* function allows to calculate consistent effect measures for continuous outcome data given by interventional controlled studies and uniformed by the *co.co* function. Calculation of effect measures for other study designs shall be included in a subsequent version of the *uniform* package. Afterwards, the package shall be uploaded to The Comprehensive R Archive Network (CRAN). The usage of the package shall be published in a journal article.

9.5 Conclusion

In this master thesis, conversion and imputation methods for a differently or incompletely reported continuous outcome were compared in terms of bias and included in the newly created *uniform* R package. Bias was compared using simulation studies with different data generation scenarios. Methods resulting in the highest precision were set as default. Version 0.0.1 of the *uniform* R package allows to uniform a differently or incompletely reported continuous outcome across studies and to calculate effect measures for interventional controlled studies.

Bibliography

[1] I. Masic, M. Miokovic, and B. Muhamedagic, "Evidence based medicine–new approaches and challenges," *Acta Informatica Medica*, vol. 16, no. 4, p. 219, 2008.

[2] A. Jinha, "Article 50 million: An estimate of the number of scholarly articles in existence," *Learned Publishing*, vol. 23, pp. 258–263, 07 2010.

[3] N. Mikolajewicz and S. V. Komarova, "Meta-analytic methodology for basic research: a practical guide," *Frontiers in physiology*, vol. 10, p. 203, 2019.

[4] G. Tripepi, K. Jager, F. Dekker, C. Wanner, and C. Zoccali, "Measures of effect: relative risks, odds ratios, risk difference, and 'number needed to treat'," *Kidney international*, vol. 72, no. 7, pp. 789–791, 2007.

[5] J. P. Higgins, T. Li, and J. J. Deeks, "Choosing effect measures and computing estimates of effect," *Cochrane handbook for systematic reviews of interventions*, pp. 143–176, 2019.

[6] M. Eid, *Statistik und Forschungsmethoden mit Online-Materialien*, 5th ed., M. . Gollwitzer, Ed. Beltz, 2017.

[7] S. Walter and X. Yao, "Effect sizes can be calculated for studies reporting ranges for outcome variables in systematic reviews," *Journal of clinical epidemiology*, vol. 60, no. 8, pp. 849–852, 2007.

[8] L. H. C. Tippett, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, vol. 17, no. 3/4, pp. 364–387, 1925. [Online]. Available: http://www.jstor.org/stable/2332087

[9] H. Ahrens, "Pearson, es, ho. o. hartley: Biometrika tables for statisticians, vol. 2. cambridge university press, london 1972. xviii, 385 s.,£ 5.00," 1974.

[10] S. P. Hozo, B. Djulbegovic, and I. Hozo, "Estimating the mean and variance from the median, range, and the size of a sample," *BMC medical research methodology*, vol. 5, no. 1, pp. 1–10, 2005.

[11] X. Wan, W. Wang, J. Liu, and T. Tong, "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range," *BMC medical research methodology*, vol. 14, no. 1, pp. 1–13, 2014.

[12] M. Bland, "Estimating mean and standard deviation from the sample size, three quartiles, minimum, and maximum," *International Journal of Statistics in Medical Research*, vol. 4, no. 1, pp. 57–64, 2015.

[13] T. A. Furukawa, C. Barbui, A. Cipriani, P. Brambilla, and N. Watanabe, "Imputing missing standard deviations in meta-analyses can provide accurate results," *Journal of clinical epidemiology*, vol. 59, no. 1, pp. 7–10, 2006.

[14] V. C. Marinho, J. Higgins, S. Logan, and A. Sheiham, "Fluoride toothpastes for preventing dental caries in children and adolescents," *Cochrane database of systematic reviews*, no. 1, 2003.

[15] J. Sánchez-Meca, F. Marín-Martínez, and S. Chacón-Moscoso, "Effect-size indices for dichotomized outcomes in meta-analysis." *Psychological methods*, vol. 8, no. 4, p. 448, 2003.

[16] D. R. Cox and E. J. Snell, *Analysis of binary data*. Routledge, 2018.

[17] G. Glas, B. McGraw, and M. Smith, "Meta-analysis in social research. beverly hills," 1981.

[18] R. L. Thorndike, "Personnel selection: test and measurement technique. new york: Johh wiley and sons," 1949.

[19] L. Hedges and I. Olkin, "Statistical methods for meta-analysis," 1985.

[20] D. Follmann, P. Elliott, I. Suh, and J. Cutler, "Variance imputation for overviews of clinical trials with continuous response," *Journal of clinical epidemiology*, vol. 45, no. 7, pp. 769–773, 1992.

[21] M. Borenstein, L. V. Hedges, J. P. Higgins, and H. R. Rothstein, "Effect sizes based on means," *Introduction to meta-analysis*, pp. 21–32, 2009.

[22] ——, "A basic introduction to fixed-effect and random-effects models for meta-analysis," *Research synthesis methods*, vol. 1, no. 2, pp. 97–111, 2010.

[23] R. DerSimonian and N. Laird, "Meta-analysis in clinical trials," *Controlled clinical trials*, vol. 7, no. 3, pp. 177–188, 1986.

[24] T. P. Morris, I. R. White, and M. J. Crowther, "Using simulation studies to evaluate statistical methods," *Statistics in medicine*, vol. 38, no. 11, pp. 2074–2102, 2019.

[25] F. M. Giorgi, C. Ceraolo, and D. Mercatelli, "The r language: An engine for bioinformatics and data science," *Life*, vol. 12, no. 5, p. 648, 2022.

[26] S. Balduzzi, G. Rücker, and G. Schwarzer, "How to perform a meta-analysis with r: a practical tutorial," *Evidence-based mental health*, vol. 22, no. 4, pp. 153–160, 2019.

[27] W. Viechtbauer, "Conducting meta-analyses in r with the metafor package," *Journal of statistical software*, vol. 36, no. 3, pp. 1–48, 2010.

[28] H. Wickham, J. Hester, W. Chang, and J. Bryan, *devtools: Tools to Make Developing R Packages Easier*, 2022, https://devtools.r-lib.org/, https://github.com/r-lib/devtools.

[29] github, "Github," 2023. [Online]. Available: https://github.com/

Appendix

Appendix 1: Possibilities for Reporting Outcome Data

- Dichotomous data
  - Number of participants with each outcome in each participant group
  - Proportion of participants with each outcome in each participant group
- Continuous data
  - Preinterventional measurement, postinterventional measurement, and/or change from baseline in each participant group
    - Mean with standard deviation
    - Mean with standard error
    - Mean with confidence interval
    - Mean with range
    - Mean with missing dispersion measure
    - Median with range
    - Median with interquartile range
    - Median with range and interquartile range

        * Median with missing dispersion measure
        * Dichotomized with the number of participants inside an outcome range
        * Dichotomized with the proportion of participants inside an outcome range

- Ordinal data
    - Number of participants in each category of the ordinal scale in each group
    - Proportion of participants with each category of the ordinal scale in each group

- Counts and rates
    - Number of events in each group and person-time at risk in each group
    - Number of events in each group and time at risk in each group

- Time-to-event data
    - Number of patients who are followed up in each period and each group and number of events in each period and each group

Appendix 2: List of Effect Measures with Corresponding Outcome Data Needed

- Ratio measures
    - Odds ratio: dichotomous data
    - Risk ratio: dichotomous data
    - Proportional odds ratio: ordinal data

  - Rate ratio: counts and rates
  - Hazard ratio: time-to-event data
  - Ratio of means: continuous data
  - Number needed to treat: dichotomous data
- Difference measures
  - Absolute difference: continuous data
  - Standardized difference
    - Hedges g': continuous data
    - Glass' delta: continuous data
    - Standardized difference in means in terms of the minimal important difference: continuous data
    - Mean prevented fraction: continuous data
    - Difference in the mean percentage change from baseline: continuous data
  - Risk difference: dichotomous data
  - Rate difference: dichotomous data

Acknowledgement

My special thanks go to my supervisor, Dr. Lorenz Uhlmann, for his statistical expertise and organizational support in carrying out this work. In my eyes, Dr. Uhlmann has great didactic skills and I very much hope that he will continue to serve the university in addition to his career in business.

I would like to thank Prof. Peter Schlattmann for the co-assessment of this master's thesis and, after co-supervising my doctoral thesis, for his support over the years.

I would like to thank my staff physician, Prof. Daniela Hartmann, for making it possible for me to attend classes. I would like to thank deputy chief physician Prof. Michael Flaig and chief physician Prof. Lars E. French for the financial support.

My thanks go to my good friend Jan Simmler for the idea.

I would like to thank my colleague Melia-Evelina Fleischmann for her encouragement during the writing of this master thesis.

To my parents, Marlis and Wolfgang Kendziora, I express my unlimited gratitude for their lifelong support.