Aus dem Institut für Biometrie und klinische Epidemiologie
der Medizinischen Fakultät Charité - Universitätsmedizin Berlin

DISSERTATION

Nonparametric procedures for the analysis of complex data
Nichtparametrische Verfahren für die Analyse von komplexen Datensätzen

zur Erlangung des akademischen Grades
Doctor of Philosophy (PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Kerstin Rubarth
aus Kirchheim unter Teck

Erstbetreuung: Prof. Dr. Frank Konietschke

Datum der Promotion: 30.11.2023

Contents

List of Figures

List of Tables

List of abbreviations

ANOVA Analysis of Variance

ATS ANOVA-Type-Statistic

CCA Complete Case Analysis

GEE Generalized Estimating Equations

FDA Food and Drug Administration

FDI First Dorsal Interrosseus

LH Leading Hemisphere

LOCF Last Observation Carried Forward

MAR Missing At Random

MCAR Missing Completely At Random

MCTP Multiple Contrast Test Procedure

MI Multiple Imputation

MMV Moyamoya Vasculopathy

MNAR Missing Not At Random

MSE Mean Squared Error

NH Non-Leading Hemisphere

nTMS navigated Transcranial Magnetic Stimulation

RM Repeated Measures

RMT Resting Motor Threshold

SICI Short Interval Cortical Inhibition

WTS Wald-Type-Statistic

Abstract

English An often encountered problem in pre-clinical, early clinical or translational studies is the analysis of complex data structures. In such studies, sample sizes are typically quite small; outcomes might not be normally distributed; are highly skewed or are not even on a metric scale. In these situations, nonparametric inference methods should be preferred over parametric procedures. Furthermore, an issue that often arises in practical applications is the occurrence of missing data. We develop nonparametric methods for the analysis of repeated measures designs that are based on all-available information instead of using completely observed subjects only. Neither any specific data distribution nor equal covariance matrices across the (treatment) groups are required. The methods can be applied to metric, highly skewed, ordinal, ordered categorical and even binary data in a unified way. No adjustment for ties in the data is necessary as opposed to classical nonparametric methods. We further generalize the framework to allow for possibly dependent replicates or clustered data. One typical example where clustered data frequently arise are animal experiments, where several animals share the same cage. The assumption of independence between animals from the same cage is likely to be violated since it can be assumed that these animals are more similar than animals from other cages, for example in terms of their behaviour. In this dissertation, statistical hypotheses are formulated in terms of the nonparametric relative effect, which is easy to understand and to interpret. We present quadratic-type as well as multiple contrast test-type procedures including simultaneous confidence intervals for the analysis of such designs. Extensive simulation studies evaluate the precision of the proposed estimators as well as type-I error rates and the power in various settings. It turns out that the methods are applicable in many different situations. Real world data sets exemplify the application of the newly developed procedures.

German Ein häufig auftretendes Problem in präklinischen, frühen klinischen oder translationalen Studien ist die Analyse komplexer Datenstrukturen. In solchen Studien ist der Stichprobenumfang in der Regel recht klein; die Parameter sind möglicherweise nicht normalverteilt, schief verteilt oder liegen nicht auf einer metrischen Skala. In solchen Situationen sollten nichtparametrische Inferenzmethoden gegenüber parametrischen Modellen bevorzugt werden. Ein weiteres häufiges Problem sind fehlende Werte. Wir entwickeln nichtparametrische Methoden für die Analyse von Modellen mit wiederholten Messungen, die auf allen verfügbaren Informationen beruhen, anstatt nur die Information von vollständig beobachteten Subjekten zu verwenden. Es sind weder eine bestimmte Datenverteilung noch gleiche Kovarianzmatrizen der Messwiederholungen der (Behandlungs-)Gruppen erforderlich. Die Methoden können auf metrische, sehr schiefe, ordinale, geordnete kategoriale und sogar binäre Daten in einheitlicher Weise angewendet werden. Im Gegensatz zu den klassischen nichtparametrischen Methoden ist keine Anpassung für Bindungen in den Daten erforderlich.
Wir verallgemeinern das Modell sowie die Prozeduren, um mögliche abhängige Wiederholungen oder geclusterte Daten zu berücksichtigen. Ein typisches Beispiel für geclusterte Daten sind Daten aus Tierexperimenten, in denen meist mehrere Tiere in einem Käfig gehalten werden. Hierbei sollte von der Annahme der Unabhängigkeit der Tiere in einem Käfig abgesehen werden, da davon ausgegangen werden kann, dass sich Tiere aus demselben Käfig ähnlicher sind als Tiere aus anderen Käfigen. In dieser Dissertation werden Hypothesen in nichtparametrischen relativen Effekten formuliert, welche leicht verständlich und einfach zu interpretieren sind. Für die Analyse solcher Modelle werden sowohl quadratische als auch multiple Kontrasttestverfahren einschließlich simultaner Konfidenzintervalle vorgestellt. Umfangreiche Simulationsstudien evaluieren die Präzision der vorgeschlagenen Schätzer sowie die Typ-I-Fehlerraten und die Power in verschiedenen Settings. Es zeigt sich, dass die Methoden in vielen verschiedenen Situationen anwendbar sind. Reale Datensätze veranschaulichen die Anwendung der neu entwickelten Verfahren.

1 Introduction

In many scientific applications, especially in medicine, subjects are observed repeatedly under different conditions or time points. Such repeated measures (RM) designs are a cost saving alternative to general factorial designs where only independent units are used. The simplest example of such RM designs is a paired design involving one homogeneous group only. In the analysis of such designs, it is important to consider a possible dependence between the repeated measures. If such a dependency is ignored, the variance of differences or sums will be under- or overestimated and thus, the analysis could lead to wrong decisions. Furthermore, one common assumption when modeling such designs is the assumption of sphericity, which means that the variances of any differences between two repeated measures are equal [4]. However, this assumption is often violated in practical applications and is not easy to verify. Hence, statistical procedures which allow for any positive definite covariance structure should be preferred in order to obtain valid results. In most cases, the structure of the covariance matrix is unknown a priori and methods which make no assumption in terms of any covariance structure represent a robust alternative to traditional approaches. Another issue arises whenever sample sizes are quite small, outcomes are not normally distributed, highly skewed or ordinal. In such scenarios, many parametric methods for the analysis of RM data, e.g. RM-Analysis of Variance (ANOVA), Mixed Models or Generalized Estimating Equations (GEE) should not be considered since they rely on assumptions such as normality and homoscedastic covariance matrices.
Besides specifying a correct model, it is important to consider a suitable effect measure to describe any treatment effect or time differences in repeated measure models. The choice of the appropriate effect measure relies often on the scale of the variable. One commonly used effect size for the analysis of metric data in two groups is simply the difference between two means. However, analyses by using the mean can be distorted if data is highly skewed or if many outliers are present. In case of ordinal outcomes, means are not even defined - however, this fact is often ignored in practical applications.
In contrast to mean-based approaches, nonparametric rank-based methods are a suitable option for the analysis of metric, ordinal, ordered categorical and even binary data in a unified way.
So called Wilcoxon-Mann-Whitney-type effects, also known as relative (treatment) effects or probabilistic index defined as $p = P(X < Y) + \frac{1}{2}P(X = Y)$, ($X, Y$ random variables coming from different distributions F_X, F_Y) are intuitive nonparametric effects which can be used in these general situations. In this dissertation, we present generalizations to the work of Konietschke et al. [5], who developed inference methods in a simple repeated measures design, in more complex scenarios.
One issue that naturally arises in practical applications due to the nature of RM designs is the occurrence of missing data. There are many reasons for missing values in a data set, e.g. measurement errors, subjects forget to answer an item on a questionnaire, drop out of the study or die before the study ends. However, it is important to reflect on the missing value mechanism. Missing data can be missing completely at random (MCAR), where the missingness is completely independent of predictors and responses. The completely observed cases of such a data set represent an unbiased sample of the population but using only complete observations results in a loss of power. Additionally, data is rarely MCAR. Missing values can also be missing at random (MAR), which means that the probability of missingness depends on other observed information, e.g. predictors. To obtain correct results, an adjustment on this information must be made. However, if data is missing not at random (MNAR), the missingness depends on the variable itself and analyses are biased. The problem is that there is no way to be sure about the missing value mechanism in an actual data set. The easiest way to handle missing values is to constrain the analysis simply on the completely observed cases (complete case analysis, CCA). As mentioned before, this should only be done in scenarios where strong arguments for MCAR exist. Further, this approach reduces the sample size and thus, results in a loss of power. Another strategy is to impute the missing values to maintain the sample size and the power. There are several ways to impute missing data; the most naïve approach would be to simply impute the missings with the mean or median of the corresponding observed values, which typically leads to biased results. Mean or median imputation, as well as regression based or stochastic regression imputation are examples for so called "Single Imputation Methods". More advanced methods, so called "Model-based Methods" include the Expectation Maximization approach and Multiple Imputation (MI). The latter is structured in three consecutive steps. First, several imputed data sets are generated where the missing values are replaced with "plausible" values, generated by using an appropriate imputation algorithm. Second, each imputed data set is analyzed separately with an appropriate statistical technique. In the final third step, parameters and standard errors from the separate analyses are combined according

to Rubins's Rules [6]. The advantage of using MI methods instead of single imputations is that the uncertainty due to the missing values is directly modeled. However, there are many choices to be made when conducting MI analyses, all leading to possibly different results, higher computation times and no final complete data set. Additionally, as pointed out by Ramosaj et al. [7], a MI approach (Multiple Imputation by Chained Equation) and a nonparametric imputation approach yield extremely inflated type-I error rates, i.e. extremely liberal test decision. Thus, the approach within this dissertation will be to not impute the missing data but to use all-available information from the data sets.
Another common scenario in (pre-) clinical and lab experiments is the presence of dependent replicates or clustered data, not necessary equally sized. One typical example for clustered data are animal experiments where several animals share the same cage. It can be assumed that those animals are more similar than animals from different cages. Another example are lab experiments, where, for example, the number of bacteria grown in the same petri dish is investigated. Ignoring the possibly present correlated data structure and assuming independence of the subjects within a cluster (e.g. cage or petri dish) introduces bias by inflating type-I error rates due to an underestimation of standard errors. Furthermore, estimation of treatment effects is not straightforward due to the presence of intra-cluster correlations and unequally sized clusters. A common approach is shrinking the observations from one cluster to a single value by using a summary measure, e.g. the mean or median. Then, a statistical method for independent observations is applied to obtain estimators, p-values and confidence intervals. However, this approach is quite unsatisfactory since a lot of existing information is disregarded and thus, the power of the experiment is decreased. Therefore, special procedures for clustered data have been developed that take all observations into account. For the parametric framework, several procedures of the analysis of such clustered data exists, e.g., Linear Mixed Models or Generalized Estimating Equations (GEEs). However, these methods are based on assumptions such as multivariate normality or linear relationships, which are hard to verify in practical applications, especially in studies with small sample sizes. Several achievements in the nonparametric framework for special clustered data designs have already been accomplished by Roy et al. [8], Gao et al. [9], Akritas and Brunner [10] and Brunner et al. [11], see more in chapter 2.5. Within this dissertation, a general framework for the analysis of factorial longitudinal data with a clustered data structure is developed. The new methods improve already existing methods upon these following aspects:

- Traditional nonparametric methods for repeated measures data are formulated in terms of distribution functions ($H_0^F : F_1 = ... = F_d$, where d is the number of time points or conditions and F_s is the distribution function at time or condition s), which implies that variance homogeneity under the null hypothesis is needed, which is rarely present in practical applications. Thus, allowing for different distributions under the null hypothesis makes the methods more robust. Therefore, the hypotheses are not formulated in terms of distribution functions but in terms of so called *nonparametric relative effects* p ($H_0^p : p_1 = ... = p_d$), which will be explained later in chapter 2.3.

- Existing methods can only be used for testing the global hypothesis of no time- or intervention/treatment effect, e.g. if any time point or intervention group differs significantly from the other time points or groups. However in most practical applications, the main research question is not answered by global testing procedures since these procedures do not report which time point or group is significantly different from the others. Thus, pairwise testing procedures have to be performed after obtaining a significant result of a global testing procedure. However, multiple testing is an issue when applying several pairwise testing procedures, and adjustment techniques such as the famous Bonferroni-correction tend to be quite conservative since the correlation of the test statistics is not taken into account. Further, decisions from global testing procedures and subsequent post-hoc tests may not be compatible, e.g. the global test could reject the null hypothesis but no local test rejects any null hypothesis. Therefore, so called Multiple Contrast Test Procedures (MCTP), which test all hypotheses simultaneously will be generalized to complex data settings, such as repeated measures with missing values and/or clustered data.

- Another disadvantage of procedures formulated in terms of distribution functions is that they cannot be converted into confidence intervals. According to international regulatory authorities such as the ICH E9 [12] for clinical trials, the computation of confidence intervals is requested, since they display variability of the data. Since the newly proposed methods are formulated

in terms of relative effects, confidence intervals which are compatible to test decisions can be obtained.

- Some procedures, e.g. Domhof et al. [13], are not even computable in some scenarios, since the proposed estimator of the covariance matrix might not be positive semidefinite. In contrast, the covariance matrix estimators developed within this dissertation are always positive semidefinite.

The thesis is structured as follows: After the introduction, a general model for repeated measures data with missing values and a clustered data structure is presented. In a next step, nonparametric relative effects and hypotheses for such designs are introduced. Furthermore, an overview of issues that arise in scenarios with missing or clustered data is presented as well as a short overview of traditional procedures for these situations. Nonparametric point estimators and test procedures will be presented afterwards. In the results section, the theoretical properties of the procedures are shortly presented, as well as results from Monte-Carlo simulation studies. Furthermore, the results of the analysis of the study by Acker et al. [1], in which the new methods are applied, are also included in the results section.

Aim of this thesis

The Ph.D. position of Kerstin Rubarth was funded by the German Research Foundation by a joint project of Prof. Dr. Frank Konietschke and Prof. Dr. Markus Pauly called "Resampling-based inference methods for the evaluation of complex models in biometrics" (grant KO 4680/3-2). Thus, Prof. Pauly and one postdoc, Dr. Paavo Sattler, of his team coauthored the two methodological papers.
The global aim of this thesis was to generalize the rank-based nonparametric procedures for simple longitudinal data by Konietschke et al. [5] to scenarios with missing and clustered data. These generalisations are motivated by statistical consultations and joint projects with clinical partners at the Charité, as data sets regularly appear there for which there is not yet an adequate methodology available. For example, the data set on Moyamoya disease by Acker et al. [1] contains a large amount of missing values and at the time point of the requested analysis, no adequate nonparametric method was available to analyze the data. Therefore, the statistical inference was based only on the completely observed subjects. This data set is re-analyzed within this thesis using the newly proposed methods. These aspects are addressed in the following *Thesis articles*:

1. Kerstin Rubarth, Markus Pauly, and Frank Konietschke. Ranking procedures for repeated measures designs with missing data: Estimation, testing and asymptotic theory. *Statistical Methods in Medical Research*, 31(1):105–118, 2022. PMID: 34841991

2. Kerstin Rubarth, Paavo Sattler, Hanna Zimmermann, and Frank Konietschke. Estimation and testing of wilcoxon–mann–whitney effects in factorial clustered data designs. *Symmetry*, 14:244, 01 2022

3. Güliz Acker, Davide Giampiccolo, Kerstin Rubarth, Robert Mertens, Anna Zdunczyk, Juliane Hardt, Daniel Jussen, Heike Schneider, Tizian Rosenstock, Vera Mueller, Thomas Picht, and Peter Vajkoczy. Motor excitability in bilateral moyamoya vasculopathy and the impact of revascularization. *Neurosurgical Focus*, 51:E7, 09 2021

2 Methods

As already described in the introduction, specifying a suitable model for RM data is of high importance and should ideally be done before collecting any data. Therefore, a general nonparametric model for factorial repeated measures with a clustered structure will be described in the following section. No distributional assumptions such as normality is required. Further, the model can be used to represent and analyze metric, ordinal, and ordered categorical data as well as binary data in a unified way.

2.1 General model for clustered factorial repeated measures

The subjects within a RM model with possibly dependent replications from the i-th group can be described with random vectors

$$\boldsymbol{X}_{ik} = (\boldsymbol{X}_{i1k}, .., \boldsymbol{X}_{idk}), i = 1, ..., a \text{ and } k = 1, ..., n_i.$$

Here, a and d denote the number of groups and time points or conditions respectively and n_i denotes the sample size of group i. The number of groups a and time points d will be considered as fixed and additionally, d is assumed to be smaller than the number of subjects in any group i (i.e. $d < n_i$). Then, indicators λ_{isk} are defined which specify subject k in group i at time s as observed or missing:

$$\lambda_{isk} \quad = \quad \begin{cases} 1, & X_{isk} \text{ is observed} \\ 0, & X_{isk} \text{ is missing.} \end{cases}$$

The (possibly) dependent replicates of subject k at time s in group i are denoted as

$$\boldsymbol{X}_{isk} = (\lambda_{isk}, (X_{isk1}, \ldots, X_{iskm_{isk}})), \tag{1}$$

with m_{isk} being the number of dependent replicates of subject k at time s in group i. Furthermore, $\lambda_{is} = \sum_{k=1}^{n_i} \lambda_{isk}$ denotes the number of observed subjects in group i at time s and $m_{is} = \sum_{k=1}^{n_i} m_{isk}$ denotes the number of dependent replicates in group i at time s. Finally, $N = \sum_{i=1}^{a} n_i$ denotes the total sample size.
In order to account for metric, ordinal, ordered categorical or binary data and ties in the data, we use the *normalized version* of the distribution function, introduced by Ruymgaart [16] as

$$F_{is}(x) = P(X_{isku} < x) + \frac{1}{2} P(X_{isku} = x),$$

which is the mean of the left- and right-continuous distribution functions $F^-(x) = P(X_{isku} < x)$ and $F^+(x) = P(X_{isku} \leq x)$, respectively.
Many scenarios can be described by this general model. The nonparametric formulation allows for a unified analysis of both metric, discrete, ordinal and even binary data. Furthermore, no linear treatment or time effects and no specific covariance structure are assumed. The variance of the observations can be completely different and any correlation structure between the observations can be present. In the following section, a detailed introduction of a nonparametric effect measure, the so called *relative effect* is presented.

2.2 Nonparametric relative effects

Model (1) does not contain any parameters to describe any differences between the distributions of time points and groups. Further, describing differences between groups or time points in terms of the mean is not appropriate for highly skewed metric data or metric data with many outliers. Furthermore, in case of ordinal data, means are even not defined. One could think of categories in a survey such as "strongly agree", "agree", "disagree" and "strongly disagree". The sum of e.g., "strongly agree" and "strongly disagree" is not defined and does not make any sense. Therefore, to allow for robust analyses, marginal distribution functions will be used for a unified analysis of e.g., non-normally distributed or ordinal data. Thus, the *unweighted mean distribution function* over all groups and time points is defined as

$$G(x) = \frac{1}{ad} \sum_{i=1}^{a} \sum_{s=1}^{d} F_{is}(x). \tag{2}$$

Here, *unweighted* means that the group sizes $n_1, ..., n_a$ do not have an impact on $G(x)$. Basically, the mean distribution $G(x)$ describes the distribution of a randomly sampled observation from an experiment. Then, the so called *nonparametric relative effect*

$$p_{is} = \int G dF_{is} = P(Z < X_{is11}) + \frac{1}{2} P(Z = X_{is11}), i = 1, \ldots, a; s = 1, \ldots, d, \tag{3}$$

relates the marginal distribution F_{is} to the mean distribution G. Here, Z denotes a random variable from the mean distribution G, which is stochastically independent from X_{is11}. These relative effects

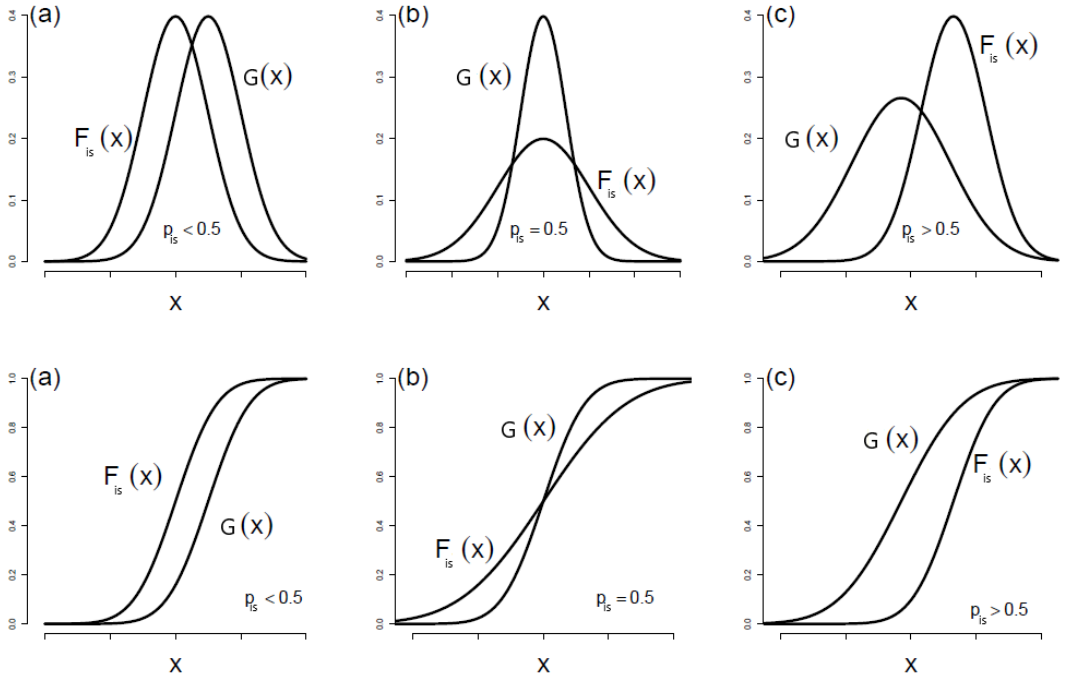

Figure 1: Interpretation of nonparametric relative effects, first row: density functions of F_{is} and G, second row: distribution functions of F_{is} and G, adapted from Konietschke et al. [2].

are also called *relative marginal effects* by Brunner, Domhof and Langer [17] as well as *probabilistic index* by Acion et al. [18]. Their interpretation is very intuitive, they simply describe the probability, that a randomly sampled observation Z from the whole experiment is smaller than a randomly sampled observation from group i at time s.
As opposed to the definition of the mean distribution function in equation (2), Kruskal [19] introduced a *weighted mean distribution function* which accounts for the group sizes by defining the mean distribution

$$H(x) = \frac{1}{Nd} \sum_{i=1}^{a} \sum_{s=1}^{d} n_i F_{is}(x). \tag{4}$$

However, relative effects based on equation (4) with $r_{is} = \int H dF_{is}$ depend on the sample sizes $n_1, ..., n_a$, are therefore no model constants and thus, should not be used to formulate hypotheses. Therefore this definition of relative effects by using the *weighted mean distribution function* will not be regarded in this thesis.
It directly follows from equation (3) that data in group i at time s tend to be larger as a randomly sampled observation from the whole experiment (i.e. from the mean distribution Z) if $p_{is} > 0.5$ and that there is no tendency to larger nor smaller values if $p_{is} = 0.5$. One can also directly compare two groups i and j at two time points s and t by comparing the relative effects p_{is} and p_{jt}. If $p_{is} > p_{jt}$ then it follows that data in group i at time s tend to be larger than data in group j at time t and if $p_{is} = p_{jt}$ then there is no tendency to larger nor smaller values between both groups i and j at both time points s and t. The relationship between two normal distributions F_{is} and G and the corresponding relative effect is depicted in figure 1. Until now, the distribution of the observations was not specified and as already mentioned, nonparametric relative effects represent a robust alternative to mean based approaches in case of highly skewed or heavy tailed distributions. However, if the actual data generating distribution is known, one can use the parameters of the distribution in order to describe effects. Assuming independent and normally distributed data with $G \sim \mathcal{N}(\mu, \sigma^2)$ and

$X_{is1} \sim \mathcal{N}(\mu_{is}, \sigma_{is}^2)$, then one can calculate the relative effect in group i at time s as

$$p_{is} = \Phi\left(\frac{\mu_{is} - \mu}{\sqrt{\sigma^2 + \sigma_{is}^2}}\right),$$

where Φ denotes the distribution function of a standard normal distribution. Further, the following relationship between the mean difference and the relative effect exists:

$$\mu_{is} - \mu = \Phi^{-1}(p_{is})\sqrt{\sigma^2 + \sigma_{is}^2}.$$

Another point to consider is that effect sizes should not be dependent on the choice of the scale. Often, the scale of ordinal data is chosen in an arbitrary way and the advantage of using such nonparametric relative effects is that these relative effects are invariant under order preserving (strongly monotone) transformations of the data.
To conclude, the nonparametric relative effects can be used to quantify treatment or time effects in many biomedical experiments as well as in many other applications. However, only reporting effects is not sufficient in many practical applications. In most cases, researchers want to test hypotheses and present confidence intervals. In the next section, different ways of formulating nonparametric hypotheses will be discussed.

2.3 Nonparametric global and multiple hypotheses

Akritas und Brunner [20] as well as Brunner and Puri [21] first presented nonparametric ranking procedures for testing hypotheses formulated in terms of distribution functions. Within the framework of this dissertation, these hypotheses could be formulated as follows:

$$H_0^F : F_{11} = ... = F_{ad} \text{ or } H_0^F : \boldsymbol{CF} = \boldsymbol{0},$$

Here $\boldsymbol{F} = (F_{11}, ..., F_{ad})'$ denotes the vector of distribution functions and $\boldsymbol{C}$ is a contrast matrix. Hence, H_0^F implies that all distributions in each group at each time point are equal. As already mentioned, the procedures for testing H_0^F have the disadvantage that they do not allow for variance heteroscedasticity under the null hypotheses, i.e. they assume a specific covariance structure under the null, which is unrealistic in many practical applications. Another issue is that no confidence intervals can be derived from these procedures. However, they display variability in the data and uncertainty of estimation and are therefore required in many biomedical applications by regulatory authorities, such as stated by the ICH E9 [12]. Furthermore, the interpretability of the method and their implications are difficult to understand for users without a statistical background. Therefore, Konietschke et al. [5] formulated hypotheses in RM designs in terms of nonparametric relative effects instead of distribution functions. As already explained before, if $p_{is} = p_{jt}$ holds, than data from group i and j at time s and t do not tend to larger nor smaller values. Hence, the null hypothesis of no treatment or time effect can be formulated as

$$H_0^p : p_{11} = ... = p_{ad} \text{ or } H_0^p : \boldsymbol{Cp} = \boldsymbol{0}.$$

Here, $\boldsymbol{p} = (p_{11}, ..., p_{ad})'$ denotes the vector of nonparametric relative effects. In case of normally distributed data, the null hypotheses $H_0^\mu : \mu_{is} - \mu = 0$ and $H_0^p : p_{is} = \frac{1}{2}$ are equivalent. Furthermore, it holds that $p_{is} = \frac{1}{2}$ even if the variances are unequal. Therefore, Brunner and Munzel [11] suggested to call procedures for testing $H_0^p : p_{is} = \frac{1}{2}$ the "Nonparametric Behrens-Fisher Problem". Here, data can have different distribution and thus, different variances or shape parameters even under the null hypothesis of no treatment or time effect which is the most realistic case in many practical applications. Hence, H_0^F can be considered as the more 'strict' hypothesis in comparison with H_0^p and it directly implies H_0^p, i.e. $H_0^F : \boldsymbol{CF} = \boldsymbol{0} \Longrightarrow H_0^p : \boldsymbol{Cp} = \boldsymbol{0}$.
However, testing only the global hypothesis H_0^p or H_0^F usually does not answer the primary research question. Usually, it is not only of interest whether a difference between several treatment groups or time points exists, but which specific treatment group or time point differs (from the others) is key. To specify multiple hypotheses, so called *contrast matrices* will be used, which will be briefly explained. A formal property of a contrast matrix is that each of its rows sum up to zero, i.e. $\boldsymbol{C}\boldsymbol{1}_{ad} = \boldsymbol{0}$, with

$\mathbf{1}_{ad}$ being a column vector of 1s of length ad and

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_q \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1ad} \\ \vdots & \dots & \vdots \\ c_{q1} & \dots & c_{qad} \end{pmatrix},$$

with q being the number of formulated hypotheses. These matrices are used to translate practical research questions into formal statistical hypotheses. Depending on the research question of interest, different contrast matrices can be used. The most commonly used contrast matrices are *Tukey*-type [22] and *Dunnett*-type [23] contrast matrices. An example of such contrast matrices would be given in a scenario with three treatment groups measured at one time point. Then, the Tukey matrix is defined as

$$\mathbf{C}^{\text{Tukey}} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

and the Dunnett matrix is defined as

$$\mathbf{C}^{\text{Dunnett}} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

Finally, global and multiple nonparametric hypotheses can be described in this scenario by using, for example, a Tukey-type contrast matrix:

$$H_0^p : \begin{cases} p_{11} = p_{21} \\ p_{11} = p_{31} \\ p_{21} = p_{31} \end{cases} \iff H_0^p : \boldsymbol{C}^{\text{Tukey}}\boldsymbol{p} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} p_{11} \\ p_{21} \\ p_{31} \end{pmatrix} = \mathbf{0}.$$

Generally, the *Tukey* contrast matrix compares all treatment groups or repeated measures with each other and is therefore labelled as the *all-pairwise* contrast, whereas the *Dunnett* contrast matrix is applied when all groups or time points should be compared with one *control* group or time point, such as

$$H_0^p : \begin{cases} p_{11} = p_{21} \\ p_{11} = p_{31} \\ p_{21} = p_{31} \end{cases} \iff H_0^p : \boldsymbol{C}^{\text{Dunnett}}\boldsymbol{p} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_{11} \\ p_{21} \\ p_{31} \end{pmatrix} = \mathbf{0}.$$

In this example, the first treatment group represents the control group.
Bretz et al. [24] present several other versions of contrast matrices, e.g. those that detect change points or trends. Furthermore, user-specified contrast matrices for special hypotheses can also be defined, as long as they fulfill the property of a contrast matrix.
So far, the relative effects have been described as hypothetical quantities. Section 2.6 describes how to estimate them from the data. However, before introducing estimators and test procedures, a short overview of the problems of missing data and clustered data are presented in the following two sections.

2.4 Missing Data

A famous quote by David Hand is "We should be suspicious of any data set (large or small) which appears perfect" [25]. The occurrence of missing data is a commonly encountered problem in biomedical sciences and the quality of studies is often assessed by whether or how many data is missing. Reviews on reporting practices by Wood et al. 2004 [26], Powney et al. [27], Klebanoff and Cole 2008 [28] and many others indicate that the amount of missing values is often not stated in scientific papers, default methods such as list-wise deletion, which will be explained below, are often not even mentioned in case of their application and that it is often unclear how many subjects were available for tables, plots and models. Helpful and well-known reporting guidelines which include guidance on how to report studies with missing data are the CONSORT statement [29] for clinical trials as well as the STROBE statement [30] for observational studies.
Missing values can occur due to various reasons and Rubin [6] first presented three classes for missing data, also called missing mechanisms:

- *Missing Completely At Random (MCAR)* If data is MCAR, then the cause of the missing data is independent from the data. An example for MCAR is a survey study, where the respondents simply forgot to answer an item. If data is MCAR, the deletion of subjects with missing data (list-wise deletion) only reduces the power to detect effects, however it does not introduce any statistical bias. Hence, many researchers argue that their data is MCAR - however, this assumption is hard to justify and is often unrealistic.

- *Missing At Random (MAR)* If data is MAR, than the missingness depends on other observed data. In a survey study, an item would be MAR if women would be more prone to not answer questions regarding their weight than men. Thus, in the group of women, the data would be MCAR. Many modern methods such as multiple imputation [25] assume the data to be MAR.

- *Missing Not At Random (MNAR)* If missing data is not MCAR nor MAR, then the missing data is considered to be missing not at random. In a survey study, richer people would not report their salary. Scenarios with MNAR data are very complex and not easy to handle. The most important task would be to detect data that cause the missingness and to perform sensitivity analyses.

Ideally after determining the missing mechanism, one must decide how to handle the missing data problem. A standard approach is *list-wise deletion* which is also called *complete-case analysis*. Here, all subjects which have at least one missing value are discarded from the analysis. This procedure is very easy to implement, however it is only valid under the MCAR assumption and can be extremely inefficient in terms of the statistical power if many missing values occur. Therefore, another "naive" procedure is often applied: *mean* or *median imputation*, where the missing values are substituted with the respective mean or median of the observed data. This approach is again only valid under MCAR and distorts the distributions, underestimates variances and therefore lead to biased estimates. For longitudinal data exists a similar procedure, called *Last Observation Carried Forward (LOCF)*, where the missing value is imputed with the last previously observed data point. This approach has been formerly suggested by the FDA for the analysis of clinical trials [25]. However, the Panel on Handling Missing Data advises against it [31] and Molenberghs and Kenward [32] showed that estimates obtained after using LOCF can be biased even under MCAR. A more advanced technique is *regression imputation*, where other variables are used to predict the missing value by first building a regression model with the variable which has missing values as the dependent variable. In a second step, the missing values are replaced by predictions from the regression model. Again, the approach is valid in MCAR scenarios and also in MAR scenarios, if the variables which determine the missingness of the dependent variable are included in the regression model as independent predictors. However, this procedure often underestimates variances and overestimates correlations. Therefore, a refinement of this method is made through *stochastic regression imputation* where a noise term is added to the prediction of the regression model in order to account for the uncertainty of the imputation. However, Rubin [33] argued, that handling a missing data problem by imputing only one value cannot be correct in general since even for a given model, imputed values cannot be calculated with certainty. Therefore, his idea was to create multiple imputed data sets in order to reflect the uncertainty of the imputation. In a first step, the approach produces m different imputed data sets, where several imputation algorithms are available. In a second step, for each of the m imputed data sets, the analysis is run separately, producing m different results. Finally, the results are then pooled by using Rubin's rules [6]. So called Multiple Imputation (MI) is nowadays considered to be the state of the art imputation technique and is explained in detail by van Buuren [25], among others. The benefits of using MI are that the procedure incorporates residual and model uncertainty as well as a correct estimation of variances and thus leading to correct inferences. The imputed data sets can be analyzed by using standard methods for completely observed data sets. However, the application of multiple imputation is more complex than using standard methods and its computation time can be quite long. Furthermore, many imputation methods exist which all lead to slightly different results and reporting is not straightforward since there is no final complete data set. Especially when sample sizes are quite small, no or few covariates were observed such as in preclinical experiments and one would prefer a simple but still sophisticated analysis, any imputation method might not be an appropriate choice.
Another option is an *all-available case analysis*, where all calculations are performed on observed data. The difference to a *complete-case analysis* is, that the observed information from a subject is still used even if some values from the subject are missing. However, also this method has some

drawbacks: the estimates can be biased if the data are not MCAR and covariance matrices may be not positive semidefinite which is a requirement for most statistical procedures. However, Domhof et al. [13] developed an all-available case procedure for the nonparametric analysis of longitudinal data with missing values. The hypotheses were formulated in terms of distribution functions (H_0^F), see Section 2.3, and the resulting covariance matrices may be not positive semidefinite in some scenarios with many missing values. Therefore, the approach of this dissertation is to develop methods for analyzing hypotheses formulated in terms of relative effects (H_0^p) in general RM designs with missing data, where non-positive semidefinite covariance matrices do not occur, even if the amount of missing values is high.

2.5 Clustered Data

Another commonly encountered problem in many (bio-) medical applications is the occurrence of clustered data. A cluster is considered to be a group of possibly dependent subjects (*possibly dependent replicates*), such as students in a class, animals sharing the same cage or bacteria cultivated in the same petri dish. Several methods have been proposed on how to deal with clustered data. A brief summary of these methods will be outlined. First of all, many researchers simply *ignore* clustered data structures and treat the possibly dependent replicates as if they were independent and apply standard procedures for independent data. This approach typically results in too small standard errors if a correlation between the clustered subjects is present and thus, results in too small p-values leading to possibly false positive findings. Another commonly used approach is summarizing the information from a cluster by using a *summary measure* such as the mean, median or a weighted mixture of both. By using this procedure, the information from a cluster is condensed to one numeric value and standard procedures for independent data can be applied on the transformed values. However, this approach goes in hand with a loss of information from individual subjects, a loss of statistical power and a reduced precision of point estimators. Furthermore, the number of dependent subjects (cluster sizes) are neglected. Therefore, also this approach should be avoided and whenever a clustered structure is suspected, special procedures for clustered data should be considered.
First, parametric methods which account for the clustered structure were developed such as *Linear Mixed Models* by Laird and Ware [34] and *Generalized Estimating Equations* by Liang and Zeger [35]. Many parametric methods for several designs such as longitudinal data and differently scaled outcome variables have been proposed by many other authors. In contrast to the parametric framework, the development of nonparametric methods for clustered data started later. Rosner and Groove [36] first proposed a generalization of the classic Mann-Whitney test for clustered observations. Furthermore, Dutta and Datta [37], Rosner et al [38] and Datta and Satten [39] also introduced methods for testing $H_0^F : F_1 = F_2$ in case of two independent groups with dependent replicates. The work by Larocque et al. [40] was the first solution for testing hypotheses formulated in terms of relative effects in a two sample design with clustered data, which was then refined by the work of Roy et al. [8], who proposed different weighting schemes for the clusters. Roy et al. [8] developed two versions of estimators, an *unweighted* and a *weighted version*. By using the unweighted version, each cluster adds the same weight to the estimation of point estimators, irrespectively of its size, whereas by using the weighted version, larger clusters add more weight to the estimation. As Roy et al. [8] developed the methodology for paired data, this dissertation generalizes the weighting scheme to clustered data embedded in general factorial designs with repeated measures and allows for the occurrence of missing data. Missing data here means, that no dependent replicates could be observed for one subject at a specific time point. Furthermore, the theory and notation within this dissertation allow not only for weighted and unweighted estimation but for other weighting schemes, e.g. a mixture of weighted and unweighted estimation.

2.6 Point estimators

The distribution functions in Section 2.1 and thus, the relative effects in Section 2.2 are unknown and must be estimated in order to be able to test hypotheses, calculate confidence intervals and draw inferential conclusions. Brunner and Puri [21] as well as Brunner, Domhof and Langer [17] first presented estimators for nonparametric effects and these results were later used by Konietschke and Brunner [41], Konietschke et al. [5], Konietschke et al. [42], Konietschke and Pauly [43] and others in longitudinal and factorial designs.

The main idea is to first estimate the unknown distribution functions F_{is} and G and then to plug in the empirical distribution functions $\widehat{F}_{is}$ and $\widehat{G}$ into the integral representation of $p_{is} = \int G dF_{is}$.
In a model without missing and clustered data such as

$$\boldsymbol{X}_{ik} = (X_{i1k}, ..., X_{idk})', i = 1, ..., a; k = 1, ..., n_i,$$

an estimator of F_{is} would be given by

$$\widehat{F}_{is}(x) = \frac{1}{n_i} \sum_{k=1}^{n_i} c(x - X_{isk}),$$

with

$$c(u) = \begin{cases} 0, u < 0 \\ \frac{1}{2}, u = 0 \\ 1, u > 0. \end{cases}$$

Thus, the estimator $\widehat{F}_{is}(x)$ simply counts how many observations X_{isk} in group i at time s are smaller than a fixed scale point x and divides this sum by the number of subjects in group i. Additionally, the case $u = 0$ accounts for ties in the data and therefore, no special adjustment for ties needs to be made at a later stage as opposed to other nonparametric methods.
Thus, an estimator of $G(x)$ is then given by the average of all estimated marginal distribution functions:

$$\widehat{G}(x) = \frac{1}{ad} \sum_{i=1}^{a} \sum_{s=1}^{d} \widehat{F}_{is}(x). \tag{5}$$

Finally, the estimators $\widehat{G}(x)$ and $\widehat{F}_{is}(x)$ can be plugged into equation (3) in order to obtain

$$\widehat{p}_{is} = \int \widehat{G} \widehat{F}_{is} = \frac{1}{ad} \sum_{j=1}^{a} \sum_{t=1}^{d} \frac{1}{n_i} \sum_{k=1}^{n_i} \widehat{F}_{jt}(X_{isk}) = \frac{1}{ad} \sum_{j=1}^{a} \sum_{t=1}^{d} \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{1}{n_j} \sum_{\ell=1}^{n_j} c(X_{isk} - X_{jt\ell}). \tag{6}$$

Thus, the vector $\widehat{\boldsymbol{p}} = (\widehat{p}_{11}, ..., \widehat{p}_{ad})'$ contains all relative effects for each group at each time point.
Generalizing these methods to the case of repeated measures with missing or clustered data is not straightforward and approaches to do so will be discussed in the following sections.

2.6.1 Point estimators in scenarios with missing data

First, an estimator for scenarios with missing data is presented. Recall model definition in Equation (1), where indicators λ_{isk} were introduced to mark an observation as missing. The distribution functions are then estimated by using all observed data and by disregarding only the missing values (all-available cases):

$$\widehat{F}_{is}(x) = \frac{1}{\lambda_{is}} \sum_{k=1}^{n_i} \lambda_{isk} c(x - X_{isk}).$$

This approach is reasonable if data is MCAR, since the distribution function would be over- or under-estimated if the missing values were potentially smaller or larger than the observed values (MAR or MNAR scenarios). However, simulation studies, see section 3.2, show that the procedures work quite well even in MAR scenarios.
Then, an estimator of p_{is} is given by plugging in the estimator of the distribution function which accounts for missing values in equation (3):

$$\begin{aligned} \widehat{p}_{is} &= \int \widehat{G} d\widehat{F}_{is} = \frac{1}{ad} \sum_{j=1}^{a} \sum_{s=1}^{d} \frac{1}{\lambda_{is}} \sum_{i=1}^{n_i} \widehat{F}_{jt}(X_{isk}) \lambda_{isk} \\ &= \frac{1}{ad} \sum_{k=1}^{n_i} \frac{\lambda_{isk}}{\lambda_{is}} \sum_{j=1}^{a} \sum_{t=1}^{d} \sum_{\ell=1}^{n_j} \frac{\lambda_{jt\ell}}{\lambda_{jt}} c(X_{isk} - X_{jt\ell}). \end{aligned}$$

2.6.2 Point estimators in scenarios with clustered data

Roy et al. [8] proposed two approaches for estimating the distribution functions in a two sample setting with dependent replicates by developing *unweighted* and *weighted* estimators for distribution functions and relative effects. In case of *unweighted* estimation, all clusters add the same weight to the estimation of the distribution function, disregarding the size of the clusters, whereas in case of *weighted* estimation, larger clusters add more weight to the estimation of the distribution function than smaller clusters. This idea is generalized to factorial repeated measures designs with missing data in this dissertation. The unweighted and weighted distribution functions are then defined as

$$\widehat{F}_{is}^{(\upsilon_1)}(x) = \frac{1}{\lambda_{is}} \sum_{k=1}^{n_i} \frac{1}{m_{isk}} \sum_{u=1}^{m_{isk}} c(x - X_{isku})\lambda_{isk} \text{ (unweighted estimator)}$$

and

$$\widehat{F}_{is}^{(\upsilon_2)}(x) = \frac{1}{m_{is}} \sum_{k=1}^{n_i} \sum_{u=1}^{m_{isk}} c(x - X_{isku})\lambda_{isk} \text{ (weighted estimator).}$$

In case of the unweighted estimator $\widehat{F}_{is}^{(\upsilon_1)}(x)$, the count function is averaged separately for each cluster and the average of these averages is calculated and therefore, each cluster adds the same weight to the estimation of F_{is}. In contrast to the unweighted estimation approach, in case of the weighted estimator $\widehat{F}_{is}^{(\upsilon_2)}(x)$, the counts are averaged over all clusters, which means that larger clusters add more weight to the estimation than smaller clusters.
In order to define both estimators (unweighted and weighted versions) and other possible versions in a unified way, general weights are defined as:

$$w_{isk}^{\upsilon_1} = \frac{1}{\lambda_{is} m_{isk}} \quad \text{and} \quad w_{isk}^{\upsilon_2} = \frac{1}{m_{is}}.$$

Estimators of F_{is} and G are then given by

$$\widehat{F}_{is}^{*}(x) = \sum_{k=1}^{n_i} \sum_{u=1}^{m_{isk}} w_{isk}^{*} c(x - X_{isku})\lambda_{isk}, * \in \{\upsilon_1, \upsilon_2\}$$

and

$$\widehat{G}^{*} = \frac{1}{ad} \sum_{i=1}^{a} \sum_{s=1}^{d} \widehat{F}_{is}^{*} = \frac{1}{ad} \sum_{i=1}^{a} \sum_{s=1}^{d} \sum_{k=1}^{n_i} \lambda_{isk} \sum_{u=1}^{m_{isk}} w_{isk}^{*} c(x - X_{isku}).$$

An estimator of p_{is} can then be written in a unified way as

$$\begin{aligned}\widehat{p}_{is}^{*} &= \int \widehat{G}^{*} d\widehat{F}_{is}^{*} = \frac{1}{ad} \sum_{j=1}^{a} \sum_{t=1}^{d} \sum_{k=1}^{n_i} \sum_{u=1}^{m_{isk}} \lambda_{isk} w_{isk}^{*} \widehat{F}_{jt}^{*}(X_{isku}) \\ &= \frac{1}{ad} \sum_{k=1}^{n_i} \sum_{j=1}^{a} \sum_{t=1}^{d} \sum_{\ell=1}^{n_j} \sum_{u=1}^{m_{isk}} \sum_{v=1}^{m_{jt\ell}} \lambda_{isk} \lambda_{jt\ell} w_{isk}^{*} w_{jt\ell}^{*} c(X_{isku} - X_{jt\ell v}).\end{aligned}$$

Finally, the proposed estimators can be used to test hypotheses in terms of distribution functions (H_0^F) as well as relative effects (H_0^p). The focus of the next section lies on statistical test procedures for H_0^p.

2.7 Test Procedures

In order to be able to conduct statistical inference in general factorial repeated measures designs allowing for missing and clustered data, test statistics for global and multiple local hypotheses are derived. All test procedures within this dissertation are formulated in terms of relative effects $\boldsymbol{p}$ by using an appropriate contrast matrix $\boldsymbol{C} \in \mathbb{R}^{ad}$ such as

$$H_0^p : \boldsymbol{C}\boldsymbol{p} = \boldsymbol{0},$$

which means that asymptotically valid test procedures for testing a family of hypotheses

$$\boldsymbol{\Omega} = \{H_0^\ell : \boldsymbol{c}_\ell' \boldsymbol{p} = 0, \ell = 1, ..., q\}$$

will be derived.

2.7.1 Global Test Procedures

First, global test procedures are presented. These procedures are a commonly used approach in many scientific fields whenever several groups or time points are compared. The global null hypothesis formulated in relative effects is always defined as $H_0^{\text{global}} : p_{11} = p_{12} = ... = p_{ad}$. If a test procedure rejects this global hypothesis, it can be concluded that at least one of the relative effects differs from the others. This statement is rarely of practical importance since no information is given about which time point or group led to the rejection of the null hypothesis. Therefore, subsequent post-hoc tests, ideally adjusted for multiple testing, are applied to detect which relative effect is statistically different from the others. Here, a local null hypothesis is defined as $H_0^{\text{local}} : p_{is} = p_{jt}$. However, this procedure has some issues; the research question answered by the global hypothesis is mostly never answered with the result of global testing procedures. Furthermore, incompatible results can occur: for example, the global testing procedure rejects the global hypothesis but no local hypothesis is rejected which is difficult to interpret and report in practical applications. However, quadratic test procedures are commonly used. Within this dissertation, two global testing procedures are derived, the *Wald*-type statistic (WTS) and the *ANOVA*-type statistic (ATS). These were first introduced by Brunner et al. [11] as well as Domhof et al. [13] for the repeated measures design with missing data and are also known as *quadratic test procedures*.

2.7.2 Multiple Contrast Test Procedure

The Multiple Contrast Test Procedure (MCTP), first introduced by Bretz et al. [24] and generalized to the nonparametric framework by Konietschke et al. [44] is a multiple test procedure and thus a test for local null hypotheses. The idea of this procedure is to reverse the order of classic test procedures. In a first step, for each local hypothesis $H_0^{(\ell)} : \boldsymbol{c}_\ell' \boldsymbol{p} = 0$ one test statistic $T_\ell^p = \sqrt{N} \frac{\boldsymbol{c}_\ell' (\widehat{\boldsymbol{p}} - \boldsymbol{p})}{\widehat{\sigma}_\ell}$, $\ell = 1, ..., q$ is calculated, where $\widehat{\sigma}_\ell^2$ is a consistent variance estimator of $\sqrt{N}(\boldsymbol{c}_\ell' (\widehat{\boldsymbol{p}} - \boldsymbol{p}))$. In a next step, all q test statistics are collected in a vector $\boldsymbol{T} = \left(T_1^p, ..., T_q^p\right)'$. Konietschke et al. [5] showed that the vector $\boldsymbol{T}$ is asymptotically multivariate normally distributed with expectation $\boldsymbol{0}$ and unknown correlation matrix $\boldsymbol{R}$. The local hypothesis $H_0^\ell : \boldsymbol{c}_\ell' \boldsymbol{p} = 0$ will be rejected at multiple α level if $|T_\ell^p| \geq z_{1-\alpha}(\boldsymbol{R})$. Here, $z_{1-\alpha}(\boldsymbol{R})$ denotes the two-sided $(1-\alpha)$-equi-coordinate quantile of the multivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$, which means that

$$P\left(\bigcap_{\ell=1}^{q} \{-z(1-\alpha, \boldsymbol{R}) < X_\ell < z(1-\alpha, \boldsymbol{R})\}\right) = 1 - \alpha$$

holds for $(X_1, ..., X_q)' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$. Figure 2 illustrates different equi-coordinate quantiles for different bivariate normal distributions with different correlations. Further information on equi-coordinate quantiles can be found in Bretz et al. [24]. An advantage of this procedure is that the critical value takes the correlation of the test statistics $\boldsymbol{T}$ into account. Thus, further adjustments for multiple testing are not necessary. Furthermore, the approach directly allows the computation of asymptotic simultaneous $(1 - \alpha)$-confidence intervals with

$$CI_\ell = \left[\boldsymbol{c}_\ell' \widehat{\boldsymbol{p}} \mp \frac{z_{1-\alpha}(\boldsymbol{R})}{\sqrt{N}} \widehat{\sigma}_\ell\right].$$

These confidence intervals are always compatible with the corresponding local test decisions, which means that it cannot happen that the ℓ-th confidence interval CI_ℓ contains zero but the ℓ-th null hypothesis is rejected.
In a second step, the global null hypothesis $H_0^p : \boldsymbol{Cp} = \boldsymbol{0} = \frac{1}{2}$ will be rejected if at least one local hypothesis is rejected, i.e. if $T_0 = max\left\{|T_1^p|, ..., |T_q^p|\right\} \geq z_{1-\alpha}(\boldsymbol{R})$.
The correlation matrix $\boldsymbol{R}$ is almost always unknown in practical applications and must be therefore estimated. Estimating the correlation and covariance matrices is not trivial, especially in scenarios with missing data and a clustered data structure. Detailed derivations and approximations for small sample sizes can be found in Rubarth et al. [14] and Rubarth et al. [15].

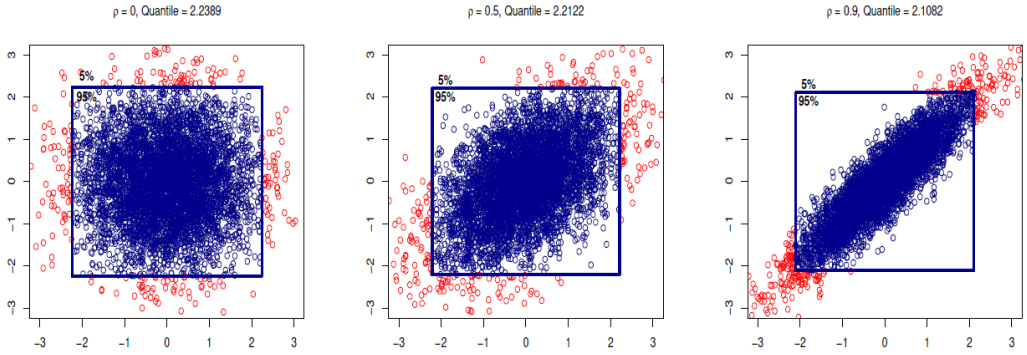

Figure 2: Two-sided equi-coordinate 95 % quantiles of different bivariate distributions $N(\mathbf{0}, \boldsymbol{I}_2 + \rho(\boldsymbol{J}_2 - \boldsymbol{I}_2))$, adapted from Gunawardana [3]

3 Results

In this section, several mathematical results as well as results from simulation studies will be summarized. Furthermore, the data from the third thesis article is again analyzed with the newly proposed methods.

3.1 Mathematical results

Within this dissertation, the following mathematical results in the nonparametric framework of factorial repeated measures with clustered and missing data could be achieved.

- In a first step, *unbiased* and *strongly consistent* estimators $\widehat{\boldsymbol{p}}^{(1)}$ (unweighted) and $\widehat{\boldsymbol{p}}^{(2)}$ (weighted) of the vector of relative effects $\boldsymbol{p}$ were constructed.
- In a second step, the asymptotic distributions of $\sqrt{N}(\widehat{\boldsymbol{p}}^{(1)} - \boldsymbol{p})$ and $\sqrt{N}(\widehat{\boldsymbol{p}}^{(2)} - \boldsymbol{p})$ were derived.
- It was further shown that the statistics $\sqrt{N}(\widehat{\boldsymbol{p}}^{(1)} - \boldsymbol{p})$ and $\sqrt{N}(\widehat{\boldsymbol{p}}^{(2)} - \boldsymbol{p})$ follow a multivariate normal distribution with expectation $\mathbf{0}$ and covariance matrices $\boldsymbol{V}_N^1$ (unweighted) and $\boldsymbol{V}_N^{(2)}$ (weighted), respectively.
- Finally, positive semidefinite covariance estimators $\widehat{\boldsymbol{V}}_N^{(1)}$ and $\widehat{\boldsymbol{V}}_N^{(2)}$ of $\boldsymbol{V}_N^{(1)}$ and $\boldsymbol{V}_N^{(2)}$ could be derived.
- Furthermore, by using all these results, quadratic- as well as multiple contrast-type test procedures could be derived.
- In the paper of Rubarth et al. [14], the Greenhouse-Gaisser method, introduced by Box [45], was applied in order to improve the performance of the ATS, since previous simulation studies indicated that the procedure tends to be liberal in some scenarios, with many missing values, 'extreme' heteroscedasticity or small sample sizes.

All test procedures rely on asymptotics and therefore, their performance in terms of their type-I error control as well as their power was evaluated in Monte-Carlo simulation studies in the first and second thesis articles. Furthermore, the precision of the unweighted and weighted estimator in terms of the bias and mean squared error (MSE) was investigated in thesis article 2.

3.2 Simulation results

All results of extensive simulation studies can be found in the first thesis article [14] and the second thesis article [15] as well as an detailed overview of all simulated scenarios. Here, a brief summary of these findings is presented. First, some statements about the procedures' ability to maintain the *type-I error* level are given.

- First, the procedures for testing H_0^p are robust in terms of their type-I error if data is generated under H_0^F instead of H_0^p, which is intuitive, since equality in distribution directly implies equality of relative effects.
- The WTS procedure is in every scenario extremely liberal, as already shown by Konietschke et al. [5]. Therefore, the procedure should only be used in very large trials or experiments, e.g. $n_i > 100$, approximately. Thus, the subsequent findings are centered around the ATS and the MCTP.
- In case of heteroscedastic covariance matrices, it turns out that in most scenarios, the procedures hold the type-I error quite accurately. However, in case of smaller sample sizes, e.g. $n_i = 15$ and missing rates of 30%, the procedures over-reject the null hypothesis.
- In most scenarios, the newly proposed modification of the ATS by using the Greenhouse-Gaisser method holds the type-I error better than the MCTP, while being slightly conservative.
- Due to the nonparametric nature of the procedures, no differences between the procedures' performances in scenarios with different data generating distributions could be detected.
- The ATS and MCTP exhibit a conservative behaviour if strong correlations and a high amount of missing data is present, as already noted by Konietschke et al. [42], Friedrich et al. [46], Munzel et al. [47], Harrar et al. [48] as well as Amro et al. [49].
- The simulation studies suggest that the ATS and MCTP can be used in scenarios with $n_i \geq 15$.
- A comparison between MCAR and MAR scenarios was conducted with the conclusion that the performance of the procedure is robust against violations of the MCAR assumption, although the theory was grounded on MCAR.
- Furthermore, the ATS performs better in equally sized scenarios, whereas the MCTP performs better in unbalanced settings.
- Regarding scenarios with clustered data, the procedures become more accurate in terms of their type-I error control if more dependent replicates are present.
- Furthermore, in case of increasing intra-cluster correlations, the type-I error of the ATS decreases, whereas the type-I error of the MCTP increases.
- The type-I error rates of the ATS using either the weighted or unweighted version of the estimator are comparable, whereas the type-I error rates of the MCTP by using the weighted version tend to be slightly larger than by using the unweighted version of the estimator.

Next, the performance of the ATS and MCTP in terms of their *power* to detect alternatives is briefly outlined.

- The power of the ATS and MCTP are comparable in MCAR and MAR scenarios, none exhibits a superior performance.
- As expected, analyses using completely observed units only is inferior to the newly proposed procedures of this dissertation, which use all-available information, i.e. do not disregard units which could not be completely observed.
- Again, the power of the procedures was not affected by the data generating distribution.

To conclude, the precision of the weighted and the unweighted estimators in terms of their MSE and bias is investigated. The bias and MSE are defined here as

$$\text{bias} = \frac{1}{n_{\text{sim}}} \sum_{i_{\text{sim}}=1}^{n_{\text{sim}}} \frac{1}{ad} \sum_{i=1}^{a} \sum_{s=1}^{d} \left(\widehat{p}_{is}^{*} - \frac{1}{2} \right)$$

$$\text{MSE} = \frac{1}{n_{\text{sim}}} \sum_{i_{\text{sim}}=1}^{n_{\text{sim}}} \frac{1}{ad} \sum_{i=1}^{a} \sum_{s=1}^{d} \left(\widehat{p}_{is}^{*} - \frac{1}{2} \right)^2,$$

where n_{sim} denotes the number of simulation runs and $n_{\text{sim}} = 10,000$ for each scenario.

- The behaviour of the unweighted and weighted version of the estimator in terms of their MSE is quite comparable in most scenarios. Roy et al. [8] concluded from their simulation study, that the precision of both estimators depend on the correlation.
- Scenarios with small sample sizes exhibit larger MSEs and negative biases whereas scenarios with larger sample sizes exhibit smaller MSES and positive biases, especially in the weighted version of the estimator.
- MSEs increase with increasing missing rates. No specific behaviour of the biases can be detected in terms of increasing missing rates.
- Furthermore, MSEs decrease with an increasing number of dependent replicates. Again, no specific behaviour of the biases can be detected.
- Scenarios with higher intra-cluster correlations exhibit larger MSEs, the same can be observed for the bias of the weighted estimator. However, scenarios with higher intra-cluster correlations exhibit smaller biases in case of unweighted estimation.

3.3 New analysis of study by Acker et al. [1]

In this subsection, the study of Acker et al [1] of patients with Moyamoya vasculopathy (MMV) is reconsidered.

3.3.1 Introduction

Moyamoya is a rare cerebrovascular condition and is one of the leading causes of stroke in children and young adults [50]. One of its consequences is motor cortical dysfunction which has been shown to be reversible after revascularization due to the brain's adaptive properties [51]. So called transcranial magnetic stimulation (TMS) which induces a electrical depolarization with a focal, rapid magnetic field induction, allows to interrogate the brain in a noninvasive manner [52]. The aim of the study by Acker et al. [1] was to analyze the corticospinal excitability and the role of bypass surgery in restoring cortical motor function in MMV patients by using navigated TMS (nTMS). Intra- and interhemispheric differences were analyzed before and after surgery, where the clinically more affected hemisphere, also called leading hemisphere (LH), was operated first.

3.3.2 Statistical Analysis

In the prospective trial, a total of $n = 30$ patients with bilateral MVV were identified. An extensive description of the data can be found in Acker et al. [1] and is therefore omitted here. The relevant outcome variables were

- *Resting motor treshold* (RMT), which is defined as the amount of TMS intensity which produces a motor-evoked potential that exceeds a defined peak to peak amplitude in 50 % of the time in a set number of trials [53].
- *Cortical representation area of the first dorsal interosseus (FDI) muscle* (AREA). Dorsal interossei are four muscles in the back of the hand. Cortical representations recreates features of the outside world, here for the first dorsal interossesus, which can be interpreted and evaluated in the brain [54].
- *Short interval cortical inhibition* (SICI) is defined as the relative amplitude reduction of motor evoked potentials by subtreshold conditioning stimuli [55].

Due to the medium sample size and non-normally distributed variables, the analyses were conducted by using nonparametric procedures such as Brunner-Munzel tests [11] and nonparametric ANOVA using ranks [17]. An issue of the study was the occurrence of missing data in some patients which was was assumed to be MCAR. Due to the fact that at the time of analysis, the newly proposed methods of this dissertation were not available, the analyses using nonparametric ANOVA were restricted to completely observed patients only. Since quadratic test procedures such as an ANOVA procedure can only indicate whether there is any significant time effect, the analysis is now re-done by using the

MCTP and by incorporating all-available information from the patients.
The variables RMT and AREA can be described by using the following model:

$$\boldsymbol{X} = \left((\lambda_1, X_1), (\lambda_2, X_2), (\lambda_3, X_3), (\lambda_4, X_4) \right)',$$

where

- X_1 denotes the pre-operative measurement of RMT or AREA of the LH,
- X_2 denotes the pre-operative measurement of RMT or AREA of the non-leading hemisphere (NH),
- X_3 denotes the post-operative measurement of RMT or AREA of the LH,
- X_4 denotes the post-operative measurement of RMT or AREA of the NH and
- $\lambda_1, ..., \lambda_4$ indicate, whether $X_1, ..., X_4$ could be observed.

The vector of relative effects is defined as follows:

$$\boldsymbol{p} = (p_1, p_2, p_3, p_4)',$$

where p_1 is the relative effect regarding the pre-operative values from the LH, p_2 is the relative effect regarding the pre-operative values from the NH, p_3 is the relative effect regarding the post-operative values from the LH and p_4 is the relative effect regarding the post-operative values from the NH.
Note that no factorial or clustered structure is present. The research question whether there are intra- or interhemispheric differences between the hemispheres in terms of RMT and AREA can be answered by using the following contrast matrix

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

Hence, the leading hemisphere is compared with the non leading hemisphere pre-operatively as well as post-operatively (interhemispheric differences) and a comparison between the leading hemisphere as well as the non leading hemisphere in pre- and post-operative conditions is conducted (intrahemispheric differences).
Due to the fact that for the variable SICI a time series including measurements at Baseline, 3 ms, 5 ms, 7 ms, 10 ms and 17 ms, the variables are described by using the following model:

$$\boldsymbol{X} = \left((\lambda_1, X_1), (\lambda_2, X_2), ..., (\lambda_{24}, X_{24}) \right)',$$

where

- $X_1, ..., X_6$ denote the pre-operative measurements of SICI of the leading hemisphere at Baseline, 3 ms, 5 ms, 7 ms, 10 ms and 17 ms,
- $X_7, ..., X_{12}$ denote the pre-operative measurements of SICI of the non-leading hemisphere at Baseline, 3 ms, 5 ms, 7 ms, 10 ms and 17 ms,
- $X_{13}, ..., X_{18}$ denote the post-operative measurements of SICI of the leading hemisphere at Baseline, 3 ms, 5 ms, 7 ms, 10 ms and 17 ms,
- $X_{19}, ..., X_{24}$ denote the post-operative measurements of SICI of the non-leading hemisphere at Baseline, 3 ms, 5 ms, 7 ms, 10 ms and 17 ms and
- $\lambda_1, ..., \lambda_{24}$ denote the respective indicators.

Here, analogously to the case of RMT or AREA, each time point is compared in an interhemispheric (differences between LH and NH) and intrahemispheric (pre- and postoperative differences between the same hemisphere) manner separately.

Comparison ($\widehat{p}_j$ vs $\widehat{p}_i$)	$\widehat{p}_i - \widehat{p}_j$	95%-Confidence Interval	p-value
LH vs. NH: pre-op.	0.058	[-0.069, 0.185]	0.581
LH: pre-op. vs. post-op	0.130	[0.026, 0.236]	0.011
NH: pre-op. vs. post-op	0.001	[-0.088, 0.089]	1.000
LH vs. NH: post-op.	- 0.072	[-0.225, 0.080]	0.545

Table 1: Nonparametric analysis of RMT by using the MCTP

3.3.3 Results

First, the analysis of RMT using all-available information (Table 1) confirms the findings of analysis using completely observed data only. RMT in the LH tended to lower values than in the NH pre-operatively, which was reversed post-operatively, without reaching statistical significance. Again, it could be shown that the RMT increased in the LH after revascularization ($\widehat{p}_i - \widehat{p}_j = 0.130$, pval = 0.011), whereas the RMT in the NH remained unchanged ($\widehat{p}_i - \widehat{p}_j = 0.001$, pval = 1.000).

Furthermore, the new analysis of AREA (Table 2) did not reveal any new findings compared to

Comparison ($\widehat{p}_j$ vs $\widehat{p}_i$)	$\widehat{p}_i - \widehat{p}_j$	95%-Confidence Interval	p-value
LH vs. NH: pre-op.	-0.096	[-0.241, 0.049]	0.277
LH: pre-op. vs. post-op	-0.099	[-0.292, 0.094]	0.491
NH: pre-op. vs. post-op	0.087	[-0.196, 0.370]	0.821
LH vs. NH: post-op.	0.090	[-0.079, 0.258]	0.460

Table 2: Nonparametric analysis of AREA by using the MCTP

the complete-case analysis. The former analysis of SICI only compared inter- and intrahemispheric

Comparison ($\widehat{p}_j$ vs $\widehat{p}_i$)	$\widehat{p}_i - \widehat{p}_j$	95%-Confidence Interval	p-value
LH vs. NH: pre-op. (baseline)	0.014	[-0.162, 0.190]	1.000
LH vs. NH: pre-op. (3 ms)	-0.091	[-0.284, 0.103]	0.802
LH vs. NH: pre-op. (5 ms)	0.072	[-0.104, 0.248]	0.899
LH vs. NH: pre-op. (7 ms)	0.008	[-0.196, 0.213]	1.000
LH vs. NH: pre-op. (10 ms)	0.025	[-0.192, 0.242]	1.000
LH vs. NH: pre-op. (17 ms)	0.092	[-0.066, 0.251]	0.570
LH: pre-op. vs. post- op (baseline)	0.041	[-0.202, 0.284]	1.000
LH: pre-op. vs. post- op (3 ms)	-0.060	[-0.273, 0.155]	0.992
LH: pre-op. vs. post- op (5 ms)	-0.046	[-0.372, 0.281]	1.000
LH: pre-op. vs. post- op (7 ms)	-0.005	[-0.363, 0.354]	1.000
LH: pre-op. vs. post- op (10 ms)	-0.017	[-0.367, 0.335]	1.000
LH: pre-op. vs. post- op (17 ms)	0.001	[-0.326, 0.328]	1.000
NH: pre-op. vs. post- op (baseline)	-0.046	[-0.234, 0.142]	0.997
NH: pre-op. vs. post- op (3 ms)	-0.041	[-0.202, 0.121]	0.996
NH: pre-op. vs. post- op (5 ms)	0.024	[-0.151, 0.200]	1.000
NH: pre-op. vs. post- op (7 ms)	-0.041	[-0.275, 0.193]	1.000
NH: pre-op. vs. post- op (10 ms)	-0.022	[-0.205, 0.162]	1.000
NH: pre-op. vs. post- op (17 ms)	0.010	[-0.193, 0.213]	1.000
LH vs. NH: post-op. (baseline)	-0.019	[-0.188, 0.150]	1.000
LH vs. NH: post-op. (3 ms)	-0.009	[-0.140, 0.122]	1.000
LH vs. NH: post-op. (5 ms)	-0.093	[-0.301, 0.115]	0.836
LH vs. NH: post-op. (7 ms)	-0.055	[-0.344, 0.236]	1.000
LH vs. NH: post-op. (10 ms)	-0.063	[-0.337, 0.211]	0.998
LH vs. NH: post-op. (17 ms)	-0.082	[-0.300, 0.137]	0.938

Table 3: Nonparametric analysis of SICI by using the MCTP

differences regarding baseline values. Thus, this analysis (3) investigated also differences at all other measured time points. However, all estimators $\widehat{p}_i - \widehat{p}_j$ are very close to zero and p-values are very close to one, indicating no effect of the nTMS on SICI.

To conclude, the new analysis using all-available information did not reveal any new findings. However, it was possible to test all relevant hypotheses simultaneously and to construct simultaneous confidence intervals for the differences of relative effects.

4 Discussion and Outlook

Within this dissertation, a novel approach for the analysis of factorial longitudinal data with missing data and a possible clustered data structure was presented. Therefore, the proposed methods can be applied to a large variety of different data sets. Especially in pre- and early clinical trials, laboratory and animal experiments, sample sizes are rather small and outcomes are often not metric but ordinal, binary or highly skewed. The newly proposed procedures can be used in a unified way for the analysis of such data. Classic nonparametric procedures test hypotheses formulated in terms of distribution functions. Formulating such hypotheses is rather strict and mostly not relevant in applied research, since expectations as well as variances are assumed to be equal under this formulation of the null hypothesis. Therefore, the hypotheses within this dissertation are formulated in terms of relative effects. These relative effects simply describe if data from one group or time point tend to be larger or smaller than data from another group or time point. Further, the proposed methods allow the calculation of confidence intervals, which display uncertainty in the estimation and are typically required by regulatory authorities, e.g. ICH E9 [12], when reporting the results from a trial.
However, the proposed methods have some limitations. As already presented in the results section, the type-I error rate depends on the sample size. It was shown that in very small sample sizes, e.g. $n_i = 10$, high correlations, 'extreme' heteroscedastic covariances or a high amount of missing data, the procedures do not hold the type-I error rate accurately. By applying so-called resampling procedures, this problem could be tackled. Further, estimation of relative effects and the calculation of their corresponding confidence intervals becomes an issue when the estimated relative effect is close to 0 or 1. A further practical limitation is the interpretation and dissemination of results of hypotheses formulated in terms of relative effects to clinicians and applied researchers without a statistical or methodological background. Furthermore, a common misconception of applied researchers is that the well-known Wilcoxon-Mann-Whitney-test compares the median of two groups [56], instead of evaluating two samples in terms of the relative effect described in this thesis. Therefore, the statistical literacy of non-statisticians should be fostered, with a special focus on nonparametric procedures.
Additionally, no sample size formulas for such designs are available. Right now, Monte Carlo simulations can be performed in order to obtain the power of the proposed procedures for a given sample size and given effects in the data. However, this approach is quite unpractical or even impossible to conduct for applied researchers without programming skills. Therefore, closed sample size formulas and approximations via simulations, implemented in *R* [57] functions, will be derived and provided in the near future. Besides providing sample size formulas, the proposed methods will be also added to an already existing *R*-package called *nparLD* [58], which already contains nonparametric methods for the analysis of factorial longitudinal data and provides procedures for testing hypotheses formulated in terms of relative effects. However, the implemented procedures do not allow for missing data nor clustered data. If missing data occur, the analysis is conducted on only completely observed data. Therefore, this work closes an important methodological gap which was motivated by consulting cases at the *Institute of Biometry and Clinical Epidemiology* at the *Charité Universitätsmedizin Berlin*, where no adequate analysis method was available.

References

[1] Güliz Acker, Davide Giampiccolo, Kerstin Rubarth, Robert Mertens, Anna Zdunczyk, Juliane Hardt, Daniel Jussen, Heike Schneider, Tizian Rosenstock, Vera Mueller, Thomas Picht, and Peter Vajkoczy. Motor excitability in bilateral moyamoya vasculopathy and the impact of revascularization. *Neurosurgical Focus*, 51:E7, 09 2021.

[2] Frank Konietschke. *Analyse nichtparametrischer Therapie-Effekte in praklinischen Experimenten und klinischen Studien mit Messwiederholungen*. Habilitation, Georg-August-Universität Göttingen, 2012.

[3] Kalupahana L. D. Asanka M. Gunawardana. *Nonparametric multiple comparisons and simultaneous confidence intervals for general multivariate factorial designs*. PhD thesis, University of Texas at Dallas, 2020.

[4] Geert Verbeke and Geert Molenberghs. *Linear Mixed Models For Longitudinal Data*. 01 2005.

[5] Frank Konietschke, Arne Bathke, Ludwig Hothorn, and Edgar Brunner. Testing and estimation of purely nonparametric effects in repeated measures designs. *Computational Statistics & Data Analysis*, 54(8):1895 – 1905, 2010.

[6] Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 12 1976.

[7] Burim Ramosaj, Lubna Amro, and Markus Pauly. A cautionary tale on using imputation methods for inference in matched pairs design. *Bioinformatics (Oxford, England)*, 36, 02 2020.

[8] Akash Roy, Solomon W. Harrar, and Frank Konietschke. The nonparametric behrens-fisher problem with dependent replicates. *Statistics in Medicine*, 38(25):4939–4962, 2019.

[9] Xin Gao. A nonparametric procedure for the two-factor mixed model with missing data. *Biometrical journal. Biometrische Zeitschrift*, 49:774–88, 10 2007.

[10] Michael Akritas and Edgar Brunner. A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61(2):249 – 277, 1997.

[11] Edgar Brunner, Ulrich Munzel, and Madan L Puri. Rank-score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70(2):286 – 317, 1999.

[12] EMA. Ich e9 statistical principles for clinical trials, June 2020.

[13] Sebastian Domhof, Edgar Brunner, and Wayne Osgood. Rank procedures for repeated measures with missing values. *Sociological Methods & Research*, 30(3):367–393, 2002.

[14] Kerstin Rubarth, Markus Pauly, and Frank Konietschke. Ranking procedures for repeated measures designs with missing data: Estimation, testing and asymptotic theory. *Statistical Methods in Medical Research*, 31(1):105–118, 2022. PMID: 34841991.

[15] Kerstin Rubarth, Paavo Sattler, Hanna Zimmermann, and Frank Konietschke. Estimation and testing of wilcoxon–mann–whitney effects in factorial clustered data designs. *Symmetry*, 14:244, 01 2022.

[16] Frits Ruymgaart. *A unified approach to the asymptotic distribution theory of certain midrank statistics*, volume 821, pages 1–18. 11 2006.

[17] Edgar Brunner, Sebastian Domhof, and Frank Langer. *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*. 01 2002.

[18] Laura Acion, John J Peterson, Scott Temple, and Stephan Arndt. Probabilistic index: An intuitive non-parametric approach to measuring the size of treatment effects. *Statistics in medicine*, 25(4):591–602, 2006.

[19] William Kruskal. A nonparametric test for the several sample problem. *Ann. Math. Statist.*, 23(4):525–540, 12 1952.

[20] Michael Akritas, Steven Arnold, and Edgar Brunner. Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, 92:258–265, 03 1997.

[21] Edgar Brunner and Madan Puri. Nonparametric methods in factorial designs. *Stat. Papers*, 42:1–52, 06 2001.

[22] John W Tukey. The problem of multiple comparisons. *Multiple comparisons*, 1953.

[23] Charles W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 50(272):1096–1121, 1955.

[24] Frank Bretz, Alan Genz, and Ludwig Hothorn. On the numerical availability of multiple comparison procedures. *Biometrical Journal*, 43(5):645–656, 2001.

[25] Stef Van Buuren. *Flexible imputation of missing data*. CRC press, 2018.

[26] Angela M. Wood, Ian R. White, and Simon G. Thompson. Are missing outcome data adequately handled? a review of published randomized controlled trials in major medical journals. *Clinical Trials*, 1(4):368–376, 2004. PMID: 16279275.

[27] Matthew Powney, Paula Williamson, Jamie Kirkham, and Ruwanthi Kolamunnage-Donà. A review of the handling of missing longitudinal outcome data in clinical trials. *Trials*, 15:237, 06 2014.

[28] Mark A. Klebanoff and Stephen R. Cole. Use of multiple imputation in the epidemiologic literature. *American journal of epidemiology*, 168 4:355–7, 2008.

[29] Kenneth Schulz, Douglas Altman, and David Moher. Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine*, 8:18, 03 2010.

[30] Jan Vandenbroucke, Erik Elm, Douglas Altman, Peter Gøtzsche, Cynthia Mulrow, Stuart Pocock, Charles Poole, James Schlesselman, and Matthias Egger. Strengthening the reporting of observational studies in epidemiology (strobe): Explanation and elaboration. *Epidemiology (Cambridge, Mass.)*, 18:805–35, 12 2007.

[31] Roderick J. Little, Ralph D'Agostino, Michael L. Cohen, Kay Dickersin, Scott S. Emerson, John T. Farrar, Constantine Frangakis, Joseph W. Hogan, Geert Molenberghs, Susan A. Murphy, James D. Neaton, Andrea Rotnitzky, Daniel Scharfstein, Weichung J. Shih, Jay P. Siegel, and Hal Stern. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 367(14):1355–1360, 2012. PMID: 23034025.

[32] Geert Molenberghs and Michael Kenward. Missing data in clinical studies. 05 2007.

[33] Donald Rubin. The design of a general and flexible system for handling nonresponse in sample surveys. *The American Statistician*, 58:298–302, 02 2004.

[34] Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

[35] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 04 1986.

[36] Bernard A. Rosner and Douglas Grove. Use of the mann-whitney u-test for clustered data. *Statistics in medicine*, 18 11:1387–400, 1999.

[37] Sandipan Dutta and Somnath Datta. A rank-sum test for clustered data when the number of subjects in a group within a cluster is informative. *Biometrics*, 72, 11 2015.

[38] Bernard Rosner, Robert Glynn, and Mei-Ling Lee. The wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*, 62:185–92, 03 2006.

[39] Somnath Datta and Glen Satten. Rank-sum tests for clustered data. *Journal of the American Statistical Association*, 100:908–915, 02 2005.

[40] Denis Larocque, Riina Haataja, Jaakko Nevalainen, and Hannu Oja. Two sample tests for the nonparametric behrens–fisher problem with clustered data. *Journal of Nonparametric Statistics*, 22(6):755–771, 2010.

[41] Frank Konietschke and Edgar Brunner. Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes. *Computational Statistics Data Analysis*, 53:730–741, 01 2009.

[42] Frank Konietschke, Solomon W. Harrar, Katharina Lange, and Edgar Brunner. Ranking procedures for matched pairs with missing data — asymptotic theory and a small sample approximation. *Computational Statistics & Data Analysis*, 56(5):1090 – 1102, 2012. Second Issue for COMPUTATIONAL STATISTICS FOR CLINICAL RESEARCH.

[43] Frank Konietschke and Markus Pauly. A studentized permutation test for the nonparametric behrens-fisher problem in paired data. *Electronic Journal of Statistics*, 6, 01 2012.

[44] Frank Konietschke, Ludwig Hothorn, and Edgar Brunner. Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics*, 6, 01 2012.

[45] George E. P. Box. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *The Annals of Mathematical Statistics*, 25(2):290 – 302, 1954.

[46] Sarah Friedrich, Frank Konietschke, and Markus Pauly. A wild bootstrap approach for nonparametric repeated measurements. *Computational Statistics & Data Analysis*, 113:38 – 52, 2017.

[47] Ulrich Munzel. Nonparametric methods for paired samples. *Statistica Neerlandica*, 53(3):277–286, 1999.

[48] Solomon W. Harrar, Merga B. Feyasa, and Eshetu Wencheko. Nonparametric procedures for partially paired data in two groups. *Computational Statistics & Data Analysis*, 144:106903, 2020.

[49] Lubna Amro, Frank Konietschke, and Markus Pauly. Multiplication-combination tests for incomplete paired data. *Statistics in Medicine*, 38, 01 2018.

[50] Güliz Acker, S Goerdes, Ulf Schneider, P Schmiedek, Marcus Czabanka, and Peter Vajkoczy. Distinct clinical and radiographic characteristics of moyamoya disease amongst european caucasians. *European journal of neurology : the official journal of the European Federation of Neurological Societies*, 22, 04 2015.

[51] Farsin Hamzei, Joachim Liepert, Christian Dettmers, Cornelius Weiller, and Michel Rijntjes. Two different reorganization patterns after rehabilitative therapy: An exploratory study with fmri and tms. *NeuroImage*, 31:710–20, 07 2006.

[52] B.L. Day, Dirk Dressler, C.D. Marsden, K. Nakashima, John Rothwell, and P. Thompson. Electric and magnetic stimulation of human motor cortex: Surface emg and single motor unit responses. *The Journal of physiology*, 412:449–73, 06 1989.

[53] Jeffrey Borckardt, Ziad Nahas, Jejo Koola, and Mark George. Estimating resting motor thresholds in transcranial magnetic stimulation research and practice: A computer simulation evaluation of best methods. *The journal of ECT*, 22:169–75, 10 2006.

[54] Edvard Moser, Yasser Roudi, Menno Witter, Clifford Kentros, Tobias Bonhoeffer, and May-Britt Moser. Grid cells and cortical representation. *Nature reviews. Neuroscience*, 15, 06 2014.

[55] Gintaute Samusyte, Hugh Bostock, John Rothwell, and Martin Koltzenburg. Short-interval intracortical inhibition: Comparison between conventional and threshold-tracking techniques. *Brain Stimulation*, 11(4):806–817, 2018.

[56] Michael P. Fay and Michael A. Proschan. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, 4(none):1 – 39, 2010.

[57] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[58] Kimihiro Noguchi, Yulia R. Gel, Edgar Brunner, and Frank Konietschke. nparld: an r software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical software*, 50(12), 2012.

Appendix A Statutory Declaration

I, Kerstin Rubarth, affirm in lieu of oath by my handwritten signature that I have written the submitted dissertation with the topic: "Nonparametric procedures for the analysis of complex data" (German title: "Nichtparametrische Verfahren für die Analyse von komplexen Datensätzen") independently and without the undisclosed help of third parties and have not used any sources or aids other than those indicated.
All passages that are based literally or in spirit on publications or lectures of other authors are marked as such in correct citation. I am responsible for the sections on methodology and results (in particular figures, graphs and tables).
I further assure that I have correctly identified the data, data evaluations and conclusions generated in collaboration with other persons and that I have correctly identified my own contribution as well as the contributions of other persons (see share declaration). I have correctly identified texts or parts of texts that were created or used jointly with others.
My shares in any publications on this dissertation correspond to those stated in the joint declaration with the first supervisor below. The ICMJE (International Committee of Medical Journal Editors; www.icmje.og) guidelines on authorship have been followed for all publications arising from this dissertation. I further declare that I am committed to comply with the statutes of the Charité - Universitätsmedizin Berlin to ensure good scientific practice.
Furthermore, I affirm that I have not already submitted this dissertation in the same or similar form to another faculty.
I am aware of the significance of this affidavit and the criminal consequences of an untrue affidavit (§§156, 161 of the Criminal Code).

Date and signature of doctoral candidate

Appendix B Declaration of own contributions to the publications

Kerstin Rubarth contributed to the following to the below listed publications:

Thesis Article 1: Kerstin Rubarth, Markus Pauly, and Frank Konietschke. Ranking procedures for repeated measures designs with missing data: Estimation, testing and asymptotic theory. *Statistical Methods in Medical Research*, 31(1):105–118, 2022. PMID: 34841991

Kerstin Rubarth developed the methodology, conducted the related simulation studies and analyzed the data set independently. Moreover, she generated all figures and tables within the manuscript. She wrote more than 90% of the manuscript independently. Her first supervisor, Frank Konietschke, was always supporting in advisory capacity. The manuscript was finalized after helpful discussions with the second supervisor (Markus Pauly).

Thesis Article 2: Kerstin Rubarth, Paavo Sattler, Hanna Zimmermann, and Frank Konietschke. Estimation and testing of wilcoxon–mann–whitney effects in factorial clustered data designs. *Symmetry*, 14:244, 01 2022

Kerstin Rubarth developed jointly with Paavo Sattler the methodology, conducted the related simulation studies and analyzed the data set independently. Moreover, she generated all figures and tables within the manuscript. She wrote more than 90% of the manuscript independently. Her first supervisor, Frank Konietschke, was always supporting in advisory capacity.

Thesis Article 3: Güliz Acker, Davide Giampiccolo, Kerstin Rubarth, Robert Mertens, Anna Zdunczyk, Juliane Hardt, Daniel Jussen, Heike Schneider, Tizian Rosenstock, Vera Mueller, Thomas Picht, and Peter Vajkoczy. Motor excitability in bilateral moyamoya vasculopathy and the impact of revascularization. *Neurosurgical Focus*, 51:E7, 09 2021

Kerstin Rubarth was essentially involved in conducting the statistical analyses of the paper. She evaluated the three outcome variables by using elaborate nonparametric procedures and discussed the results and implications with the clinicians. Furthermore, she generated figure 2 as well as the supplementary figure and tables. She approved the final version of the article for publication and is responsible for the content together with the other authors.

Date, signature and stamp of first supervising university professor

Date and signature of doctoral candidate

Appendix C Reprints of articles

In the electronic version of my work, DOI-routed reference links are included instead of my three thesis articles.

- **Thesis Article 1** Kerstin Rubarth, Markus Pauly, and Frank Konietschke. Ranking procedures for repeated measures designs with missing data: Estimation, testing and asymptotic theory. *Statistical Methods in Medical Research*, 31(1):105–118, 2022. PMID: 34841991
  https://doi.org/10.1177/09622802211046389
- **Thesis Article 2** Kerstin Rubarth, Paavo Sattler, Hanna Zimmermann, and Frank Konietschke. Estimation and testing of wilcoxon–mann–whitney effects in factorial clustered data designs. *Symmetry*, 14:244, 01 2022
  https://doi.org/10.3390/sym14020244
- **Thesis Article 3** Güliz Acker, Davide Giampiccolo, Kerstin Rubarth, Robert Mertens, Anna Zdunczyk, Juliane Hardt, Daniel Jussen, Heike Schneider, Tizian Rosenstock, Vera Mueller, Thomas Picht, and Peter Vajkoczy. Motor excitability in bilateral moyamoya vasculopathy and the impact of revascularization. *Neurosurgical Focus*, 51:E7, 09 2021
  https://doi.org/10.3171/2021.6.FOCUS21280

C.1 Article 1

Journal Data Filtered By: **Selected JCR Year: 2020** Selected Editions: SCIE,SSCI
Selected Categories: **"HEALTH CARE SCIENCES and SERVICES"**
Selected Category Scheme: WoS
Gesamtanzahl: 108 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	npj Digital Medicine	2,406	11.653	0.007450
2	Implementation Science	14,750	7.327	0.020010
3	BMJ Quality & Safety	7,335	7.035	0.014360
4	ACADEMIC MEDICINE	22,828	6.893	0.027600
5	JOURNAL OF CLINICAL EPIDEMIOLOGY	36,224	6.437	0.028360
6	HEALTH AFFAIRS	22,442	6.301	0.047250
7	MEDICAL EDUCATION	14,132	6.251	0.011580
8	JOURNAL OF TELEMEDICINE AND TELECARE	5,052	6.184	0.005600
9	VALUE IN HEALTH	12,642	5.725	0.017860
10	JOURNAL OF MEDICAL INTERNET RESEARCH	26,102	5.428	0.039100
11	JOURNAL OF GENERAL INTERNAL MEDICINE	26,727	5.128	0.028950
12	International Journal of Integrated Care	2,034	5.120	0.002800
13	International Journal of Health Policy and Management	2,118	5.007	0.005600
14	PHARMACOECONOMICS	6,543	4.981	0.009170
15	Journal of Personalized Medicine	1,071	4.945	0.002290
16	MILBANK QUARTERLY	4,632	4.911	0.005270
17	JMIR mHealth and uHealth	7,694	4.773	0.015520
18	PALLIATIVE MEDICINE	7,332	4.762	0.009100
19	BMC Medical Research Methodology	16,557	4.615	0.017550

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
20	JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION	12,078	4.497	0.016910
21	JOURNAL OF MEDICAL SYSTEMS	8,017	4.460	0.009500
22	Internet Interventions-The Application of Information Technology in Mental and Behavioural Health	1,658	4.333	0.003310
22	JOURNAL OF RURAL HEALTH	2,866	4.333	0.004050
24	QUALITY OF LIFE RESEARCH	19,584	4.147	0.017860
25	JMIR Serious Games	641	4.143	0.000970
26	INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	7,651	4.046	0.010440
27	HEALTH TECHNOLOGY ASSESSMENT	6,820	4.014	0.009010
28	MEDICAL CARE RESEARCH AND REVIEW	2,973	3.929	0.003190
29	JOURNAL OF HEALTH ECONOMICS	9,140	3.883	0.014840
29	Patient-Patient Centered Outcomes Research	1,738	3.883	0.003240
31	ADVANCES IN HEALTH SCIENCES EDUCATION	3,798	3.853	0.004540
32	MEDICAL TEACHER	12,121	3.650	0.011270
33	JOURNAL OF PAIN AND SYMPTOM MANAGEMENT	15,063	3.612	0.015920
34	SUPPORTIVE CARE IN CANCER	18,289	3.603	0.023280
35	BMJ Supportive & Palliative Care	2,039	3.568	0.004180
36	Telemedicine and e-Health	6,307	3.536	0.007090
37	Digital Health	676	3.495	0.001640
38	HEALTH SERVICES RESEARCH	9,969	3.402	0.013340
39	HEALTH EXPECTATIONS	4,873	3.377	0.008540
40	HEALTH POLICY AND PLANNING	6,543	3.344	0.010630

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
41	BMC Palliative Care	2,800	3.234	0.005230
42	Risk Management and Healthcare Policy	945	3.200	0.001770
43	Health and Quality of Life Outcomes	12,375	3.186	0.011700
44	HEALTH ECONOMICS	7,932	3.046	0.011310
45	STATISTICAL METHODS IN MEDICAL RESEARCH	6,654	3.021	0.015730
46	MEDICAL CARE	23,770	2.983	0.016890
47	HEALTH POLICY	9,058	2.980	0.010010
48	JOURNAL OF PALLIATIVE MEDICINE	8,370	2.947	0.010440
48	Perspectives on Medical Education	1,193	2.947	0.003210
50	Journal of Managed Care & Specialty Pharmacy	2,384	2.903	0.006700
51	Journal of Patient Safety	1,525	2.844	0.002620
52	HASTINGS CENTER REPORT	2,050	2.683	0.002620
53	Journal of Healthcare Engineering	1,992	2.682	0.003760
54	Health Informatics Journal	1,497	2.681	0.002300
55	BMC HEALTH SERVICES RESEARCH	25,956	2.655	0.041730
56	EVALUATION & THE HEALTH PROFESSIONS	1,622	2.651	0.000920
57	Healthcare	2,175	2.645	0.004300
58	Annals of Palliative Medicine	1,097	2.595	0.002230
59	MEDICAL DECISION MAKING	6,391	2.583	0.007240
60	SCANDINAVIAN JOURNAL OF PRIMARY HEALTH CARE	1,842	2.581	0.001840
61	Applied Health Economics and Health Policy	1,726	2.561	0.003070
62	Disability and Health Journal	2,139	2.554	0.004030

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
63	International Journal of Evidence-Based Healthcare	1,721	2.548	0.003700
64	EUROPEAN JOURNAL OF CANCER CARE	4,547	2.520	0.006590
65	American Journal of Hospice & Palliative Medicine	3,368	2.500	0.004780
66	Population Health Management	1,333	2.459	0.002930
67	JOURNAL OF MEDICAL ECONOMICS	2,936	2.448	0.004760
68	Informatics for Health & Social Care	511	2.439	0.000830
69	JOURNAL OF EVALUATION IN CLINICAL PRACTICE	5,408	2.431	0.005340
70	Therapeutics and Clinical Risk Management	4,454	2.423	0.006680
71	TEACHING AND LEARNING IN MEDICINE	1,914	2.414	0.002730
72	Chronic Illness	898	2.409	0.000830
73	Journal of Multidisciplinary Healthcare	1,737	2.404	0.003190
74	Journal of Interprofessional Care	4,227	2.338	0.003940
75	Current Opinion in Supportive and Palliative Care	1,552	2.302	0.001960
76	JOURNAL OF HEALTH POLITICS POLICY AND LAW	1,586	2.265	0.002870
77	JOURNAL OF PALLIATIVE CARE	1,241	2.250	0.000730
78	AMERICAN JOURNAL OF MANAGED CARE	5,296	2.229	0.008030
79	JOURNAL OF PUBLIC HEALTH POLICY	1,228	2.222	0.001240
80	Expert Review of Pharmacoeconomics & Outcomes Research	2,176	2.217	0.002560
81	INTERNATIONAL JOURNAL OF TECHNOLOGY ASSESSMENT IN HEALTH CARE	2,522	2.188	0.001760
82	METHODS OF INFORMATION IN MEDICINE	1,601	2.176	0.001340

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
83	Gaceta Sanitaria	2,417	2.139	0.002190
84	JOURNAL OF SCHOOL HEALTH	5,051	2.118	0.004120
85	JOURNAL OF THE HISTORY OF MEDICINE AND ALLIED SCIENCES	614	2.088	0.000570
86	American Health and Drug Benefits	757	2.070	0.001490
87	INTERNATIONAL JOURNAL FOR QUALITY IN HEALTH CARE	7,191	2.038	0.005210
88	Australian Health Review	2,253	1.990	0.002530
89	Families Systems & Health	1,285	1.950	0.001440
90	Simulation in Healthcare-Journal of the Society for Simulation in Healthcare	2,043	1.929	0.002900
91	AMERICAN JOURNAL OF MEDICAL QUALITY	1,737	1.852	0.002700
92	Journal of Comparative Effectiveness Research	907	1.744	0.002170
93	INQUIRY-THE JOURNAL OF HEALTH CARE ORGANIZATION PROVISION AND FINANCING	975	1.730	0.001880
94	INTERNATIONAL JOURNAL OF HEALTH SERVICES	2,012	1.663	0.001770
95	EASTERN MEDITERRANEAN HEALTH JOURNAL	3,247	1.628	0.002060
96	JOURNAL OF BEHAVIORAL HEALTH SERVICES & RESEARCH	1,596	1.505	0.001910
97	JOURNAL OF MANIPULATIVE AND PHYSIOLOGICAL THERAPEUTICS	3,239	1.437	0.001830
98	MEDICAL HISTORY	854	1.419	0.001210
99	JOURNAL OF CONTINUING EDUCATION IN THE HEALTH PROFESSIONS	1,661	1.355	0.001360
100	BULLETIN OF THE HISTORY OF MEDICINE	997	1.314	0.001030
101	Australian Journal of Primary Health	1,185	1.307	0.001590

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
102	TECHNOLOGY AND HEALTH CARE	1,506	1.285	0.001810
103	CAMBRIDGE QUARTERLY OF HEALTHCARE ETHICS	907	1.284	0.001030
104	Geospatial Health	862	1.212	0.001030
105	Journal of the American Association of Nurse Practitioners	1,043	1.165	0.002030
106	Journal for Healthcare Quality	769	1.095	0.001260
107	Quality Management in Health Care	561	0.926	0.000510
108	JBI Evidence Implementation	0	Not Available	0.000000

C.2 Article 2

Journal Data Filtered By: **Selected JCR Year: 2020** Selected Editions: SCIE,SSCI Selected Categories: **"MULTIDISCIPLINARY SCIENCES"** Selected Category Scheme: WoS

Gesamtanzahl: 73 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	NATURE	915,925	49.962	1.089400
2	SCIENCE	814,971	47.728	0.895760
3	National Science Review	5,889	17.275	0.011400
4	Nature Communications	453,215	14.919	1.238540
5	Science Advances	65,205	14.136	0.218640
6	Nature Human Behaviour	5,549	13.663	0.023120
7	Science Bulletin	8,832	11.780	0.016400
8	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	799,058	11.205	0.806620
9	Journal of Advanced Research	5,927	10.479	0.006800
10	GigaScience	5,876	6.524	0.018630
11	Scientific Data	10,617	6.444	0.034470
12	Frontiers in Bioengineering and Biotechnology	7,470	5.890	0.011340
13	ANNALS OF THE NEW YORK ACADEMY OF SCIENCES	52,619	5.691	0.021430
14	iScience	5,235	5.458	0.012300
15	Research Synthesis Methods	3,926	5.273	0.007520
16	NPJ Microgravity	594	4.415	0.001790
17	Scientific Reports	541,615	4.379	1.232500
18	PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES	24,950	4.226	0.025400

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
19	Journal of the Royal Society Interface	16,834	4.118	0.022010
20	JOURNAL OF KING SAUD UNIVERSITY SCIENCE	3,276	4.011	0.002870
21	Advanced Theory and Simulations	1,201	4.004	0.002780
22	GLOBAL CHALLENGES	1,047	3.847	0.002860
23	FRACTALS-COMPLEX GEOMETRY PATTERNS AND SCALING IN NATURE AND SOCIETY	2,667	3.665	0.002570
24	SCIENCE AND ENGINEERING ETHICS	2,796	3.525	0.003700
25	PROCEEDINGS OF THE JAPAN ACADEMY SERIES B-PHYSICAL AND BIOLOGICAL SCIENCES	2,218	3.493	0.001940
26	PLoS One	857,723	3.240	1.081150
27	PeerJ	29,906	2.984	0.069540
28	Royal Society Open Science	11,155	2.963	0.030990
29	INTERNATIONAL JOURNAL OF BIFURCATION AND CHAOS	8,572	2.836	0.006590
30	COMPLEXITY	7,133	2.833	0.009620
31	SCIENCE PROGRESS	689	2.774	0.000380
32	JOURNAL OF THE ROYAL SOCIETY OF NEW ZEALAND	944	2.750	0.000680
33	Symmetry-Basel	9,999	2.713	0.011650
34	PROCEEDINGS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES	22,295	2.704	0.015330
35	Journal of Taibah University for Science	2,141	2.688	0.002210
36	MIT Technology Review	1,156	2.563	0.002680
37	Facets	488	2.535	0.001170

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
38	ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING	8,349	2.334	0.008350
39	SOUTH AFRICAN JOURNAL OF SCIENCE	3,313	2.197	0.001680
40	SCIENTIFIC AMERICAN	7,590	2.142	0.003070
41	Frontiers in Life Science	520	2.000	0.000800
42	Science of Nature	954	1.954	0.002580
43	Journal of Radiation Research and Applied Sciences	1,611	1.770	0.001870
44	ANAIS DA ACADEMIA BRASILEIRA DE CIENCIAS	4,708	1.753	0.004490
45	JOURNAL OF THE INDIAN INSTITUTE OF SCIENCE	561	1.742	0.000510
46	RENDICONTI LINCEI-SCIENZE FISICHE E NATURALI	1,217	1.627	0.001380
47	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES INDIA SECTION A-PHYSICAL SCIENCES	658	1.544	0.000850
48	Proceedings of the Romanian Academy Series A-Mathematics Physics Technical Sciences Information Science	485	1.523	0.000680
49	Jove-Journal of Visualized Experiments	20,722	1.355	0.031050
50	DISCRETE DYNAMICS IN NATURE AND SOCIETY	2,504	1.348	0.002530
51	ISSUES IN SCIENCE AND TECHNOLOGY	607	1.255	0.001080
52	ADVANCES IN COMPLEX SYSTEMS	677	1.226	0.000540
53	Iranian Journal of Science and Technology Transaction A-Science	1,237	1.194	0.001790
54	CURRENT SCIENCE	13,179	1.102	0.006070
55	Proceedings of the Estonian Academy of Sciences	638	1.045	0.000350
56	Sains Malaysiana	2,386	1.009	0.001890
57	INTERDISCIPLINARY SCIENCE REVIEWS	391	1.000	0.000250

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
58	TRANSACTIONS OF THE ROYAL SOCIETY OF SOUTH AUSTRALIA	644	0.968	0.000200
59	Kuwait Journal of Science	341	0.948	0.000490
60	SCIENTIST	336	0.853	0.000360
61	NATIONAL ACADEMY SCIENCE LETTERS-INDIA	739	0.788	0.000590
62	DEFENCE SCIENCE JOURNAL	1,181	0.707	0.000500
63	Maejo International Journal of Science and Technology	260	0.636	0.000140
64	SCIENCEASIA	923	0.615	0.000440
65	HERALD OF THE RUSSIAN ACADEMY OF SCIENCES	415	0.560	0.000750
66	ACTA SCIENTIARUM-TECHNOLOGY	509	0.550	0.000420
67	AMERICAN SCIENTIST	2,848	0.548	0.000600
68	Chiang Mai Journal of Science	732	0.523	0.000590
69	JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA	465	0.515	0.000240
70	ENDEAVOUR	668	0.444	0.000480
71	COMPTES RENDUS DE L ACADEMIE BULGARE DES SCIENCES	743	0.378	0.000720
72	NEW SCIENTIST	1,124	0.319	0.001000
73	All Life	16	Not Available	0.000000

C.3 Article 3

Journal Data Filtered By: **Selected JCR Year: 2020** Selected Editions: SCIE,SSCI
Selected Categories: **"SURGERY"** Selected Category Scheme: WoS
Gesamtanzahl: 211 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	JAMA Surgery	12,793	14.766	0.037320
2	ANNALS OF SURGERY	64,045	12.969	0.062660
3	JOURNAL OF HEART AND LUNG TRANSPLANTATION	15,107	10.247	0.021660
4	JOURNAL OF NEUROLOGY NEUROSURGERY AND PSYCHIATRY	37,094	10.154	0.026380
5	ENDOSCOPY	14,018	10.093	0.017740
6	AMERICAN JOURNAL OF TRANSPLANTATION	32,841	8.086	0.037980
7	Digestive Endoscopy	4,707	7.559	0.006120
8	Hepatobiliary Surgery and Nutrition	1,292	7.293	0.002280
9	EUROPEAN JOURNAL OF VASCULAR AND ENDOVASCULAR SURGERY	12,166	7.069	0.013270
10	Journal of Hepato-Biliary-Pancreatic Sciences	4,958	7.027	0.004550
11	BRITISH JOURNAL OF SURGERY	29,311	6.939	0.024180
12	AMERICAN JOURNAL OF SURGICAL PATHOLOGY	26,272	6.394	0.020620
13	JAMA Otolaryngology-Head & Neck Surgery	5,965	6.223	0.012380
14	JOURNAL OF THE AMERICAN COLLEGE OF SURGEONS	20,819	6.113	0.023470
15	International Journal of Surgery	16,011	6.071	0.018760
16	Journal of NeuroInterventional Surgery	7,426	5.836	0.016070
17	LIVER TRANSPLANTATION	11,872	5.799	0.011700
18	World Journal of Emergency Surgery	2,562	5.469	0.004460

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
19	ANNALS OF SURGICAL ONCOLOGY	37,490	5.344	0.043690
20	JOURNAL OF BONE AND JOINT SURGERY-AMERICAN VOLUME	53,702	5.284	0.033030
21	JOURNAL OF THORACIC AND CARDIOVASCULAR SURGERY	33,066	5.209	0.026550
22	JOURNAL OF NEUROSURGERY	43,275	5.115	0.027680
23	Burns & Trauma	937	5.099	0.001610
24	Bone & Joint Journal	9,587	5.082	0.020810
25	TRANSPLANTATION	27,214	4.939	0.024800
26	ARTHROSCOPY-THE JOURNAL OF ARTHROSCOPIC AND RELATED SURGERY	20,208	4.772	0.020680
27	Hernia	5,350	4.739	0.005050
28	Surgery for Obesity and Related Diseases	10,541	4.734	0.016650
29	PLASTIC AND RECONSTRUCTIVE SURGERY	45,656	4.730	0.030360
30	NEUROSURGERY	34,635	4.654	0.022250
31	JAMA Facial Plastic Surgery	1,662	4.611	0.003460
32	DISEASES OF THE COLON & RECTUM	16,654	4.585	0.011010
33	SURGICAL ENDOSCOPY AND OTHER INTERVENTIONAL TECHNIQUES	31,681	4.584	0.034550
34	EJSO	12,510	4.424	0.016820
35	KNEE SURGERY SPORTS TRAUMATOLOGY ARTHROSCOPY	21,052	4.342	0.025830
36	ANNALS OF THORACIC SURGERY	41,620	4.330	0.035100

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
37	Aesthetic Surgery Journal	5,829	4.283	0.007010
38	JOURNAL OF VASCULAR SURGERY	31,991	4.268	0.024580
39	EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY	18,818	4.191	0.020200
40	CLINICAL ORTHOPAEDICS AND RELATED RESEARCH	44,823	4.176	0.023370
41	OBESITY SURGERY	19,974	4.129	0.022980
42	Annals of Cardiothoracic Surgery	2,434	4.101	0.004750
43	Neurosurgical Focus	9,818	4.047	0.011120
44	LASERS IN SURGERY AND MEDICINE	7,332	4.025	0.004040
45	SURGERY	25,223	3.982	0.024930
46	JOURNAL OF NEUROSURGICAL ANESTHESIOLOGY	1,988	3.956	0.001470
47	Colorectal Disease	8,971	3.788	0.008860
48	TRANSPLANT INTERNATIONAL	5,770	3.782	0.006680
49	Techniques in Coloproctology	3,480	3.781	0.004150
50	HPB	6,923	3.647	0.009900
51	WOUND REPAIR AND REGENERATION	7,277	3.617	0.003950
52	JOURNAL OF NEUROSURGERY-SPINE	10,175	3.602	0.011700
53	JOURNAL OF REFRACTIVE SURGERY	5,617	3.573	0.005710
54	Perioperative Medicine	497	3.535	0.001450
55	OTOLARYNGOLOGY-HEAD AND NECK SURGERY	19,487	3.497	0.015940
56	Surgical Oncology Clinics of North America	1,726	3.495	0.002490

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
57	Neurospine	498	3.492	0.000920
58	JOURNAL OF ENDOVASCULAR THERAPY	4,318	3.487	0.004840
59	JOURNAL OF SURGICAL ONCOLOGY	14,157	3.454	0.015960
59	SHOCK	9,489	3.454	0.009910
61	JOURNAL OF GASTROINTESTINAL SURGERY	13,867	3.452	0.014020
62	LANGENBECKS ARCHIVES OF SURGERY	5,093	3.445	0.004690
63	DERMATOLOGIC SURGERY	10,750	3.398	0.006330
64	BJS Open	683	3.396	0.001990
65	WORLD JOURNAL OF SURGERY	21,622	3.352	0.020430
66	JOURNAL OF CATARACT AND REFRACTIVE SURGERY	16,061	3.351	0.011300
67	International Wound Journal	4,946	3.315	0.006060
68	Journal of Trauma and Acute Care Surgery	11,459	3.313	0.018910
69	SURGICAL ONCOLOGY-OXFORD	3,000	3.279	0.003930
70	LASERS IN MEDICAL SCIENCE	7,562	3.161	0.006890
71	HEAD AND NECK-JOURNAL FOR THE SCIENCES AND SPECIALTIES OF THE HEAD AND NECK	16,925	3.147	0.017350
72	ARCHIVES OF ORTHOPAEDIC AND TRAUMA SURGERY	8,816	3.067	0.008580
73	NEUROSURGICAL REVIEW	3,616	3.042	0.003650
74	JOURNAL OF THE AMERICAN ACADEMY OF ORTHOPAEDIC SURGEONS	8,269	3.020	0.009880

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
75	JOURNAL OF SHOULDER AND ELBOW SURGERY	16,721	3.019	0.016110
76	Gland Surgery	1,734	2.953	0.003350
77	International Journal of Computer Assisted Radiology and Surgery	3,780	2.924	0.006620
78	Journal of Surgical Education	4,853	2.891	0.009740
79	JOURNAL OF RECONSTRUCTIVE MICROSURGERY	3,242	2.873	0.003150
80	CLINICAL TRANSPLANTATION	6,281	2.863	0.009180
81	Journal of Vascular Surgery-Venous and Lymphatic Disorders	1,769	2.859	0.002710
82	Updates in Surgery	1,663	2.797	0.002760
83	PHOTOMEDICINE AND LASER SURGERY	3,542	2.796	0.002140
84	INTERNATIONAL JOURNAL OF ORAL AND MAXILLOFACIAL SURGERY	10,324	2.789	0.009490
85	Asian Journal of Surgery	1,603	2.767	0.002180
86	Seminars in Pediatric Surgery	2,233	2.754	0.002820
86	World Journal of Surgical Oncology	6,910	2.754	0.008440
88	BURNS	10,973	2.744	0.007660
89	SURGICAL CLINICS OF NORTH AMERICA	4,502	2.741	0.003520
90	Journal of Plastic Reconstructive and Aesthetic Surgery	8,470	2.740	0.008970
91	Frontiers in Surgery	1,238	2.718	0.002650
92	Operative Neurosurgery	2,150	2.703	0.003930
93	Journal of Hand Surgery-European Volume	6,037	2.688	0.004290
94	Ostomy Wound Management	1,451	2.629	0.000810
95	DIGESTIVE SURGERY	2,652	2.588	0.002300

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
96	INJURY-INTERNATIONAL JOURNAL OF THE CARE OF THE INJURED	19,692	2.586	0.020790
97	World Journal of Gastrointestinal Surgery	1,299	2.582	0.002120
98	INTERNATIONAL JOURNAL OF COLORECTAL DISEASE	8,097	2.571	0.008470
99	AMERICAN JOURNAL OF SURGERY	21,583	2.565	0.016660
100	SURGERY TODAY	5,872	2.549	0.005150
101	International Journal of Medical Robotics and Computer Assisted Surgery	2,222	2.547	0.002610
102	JOURNAL OF PEDIATRIC SURGERY	20,364	2.545	0.014870
103	JOURNAL OF INVESTIGATIVE SURGERY	1,586	2.533	0.001420
104	NEUROSURGERY CLINICS OF NORTH AMERICA	2,414	2.509	0.002500
105	MINIMALLY INVASIVE THERAPY & ALLIED TECHNOLOGIES	1,319	2.442	0.001060
106	MICROSURGERY	4,125	2.425	0.003260
107	SURGEON-JOURNAL OF THE ROYAL COLLEGES OF SURGEONS OF EDINBURGH AND IRELAND	1,702	2.392	0.001960
108	Journal of Neurosurgery-Pediatrics	5,911	2.375	0.007520
109	Clinics in Colon and Rectal Surgery	1,666	2.373	0.001870
110	Scandinavian Journal of Surgery	1,361	2.360	0.001640
111	Advances in Skin & Wound Care	1,782	2.347	0.001400
112	AESTHETIC PLASTIC SURGERY	5,410	2.326	0.004130
113	Seminars in Plastic Surgery	1,256	2.314	0.001140

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
114	Journal of Neurosurgical Sciences	1,160	2.279	0.001460
115	Orthopaedics & Traumatology-Surgery & Research	5,600	2.256	0.008540
116	Journal of Cosmetic and Laser Therapy	1,570	2.247	0.001480
117	JOURNAL OF HAND SURGERY-AMERICAN VOLUME	14,875	2.230	0.009410
118	Photobiomodulation Photomedicine and Laser Surgery	319	2.222	0.000320
119	ACTA NEUROCHIRURGICA	11,865	2.216	0.008690
120	KNEE	6,011	2.199	0.006850
121	JOURNAL OF SURGICAL RESEARCH	17,062	2.192	0.018750
122	EUROPEAN JOURNAL OF PEDIATRIC SURGERY	2,105	2.191	0.001900
123	JSLS-Journal of the Society of Laparoendoscopic Surgeons	2,396	2.172	0.001670
124	Surgical Infections	3,093	2.150	0.004170
125	World Neurosurgery	23,506	2.104	0.043750
126	BMC Surgery	2,772	2.102	0.004400
127	CANADIAN JOURNAL OF SURGERY	3,251	2.089	0.002590
128	JOURNAL OF CRANIO-MAXILLOFACIAL SURGERY	7,808	2.078	0.010200
129	Surgical Innovation	1,669	2.058	0.002490
130	International Journal of Lower Extremity Wounds	1,152	2.057	0.000940
131	Journal of Visceral Surgery	1,265	2.043	0.001700
132	CLINICS IN PLASTIC SURGERY	2,943	2.017	0.002210
133	Visceral Medicine	579	1.960	0.001290

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
134	Journal of Hand Therapy	2,077	1.950	0.001580
135	Facial Plastic Surgery Clinics of North America	1,180	1.918	0.001320
136	CURRENT PROBLEMS IN SURGERY	629	1.909	0.000400
137	Interactive Cardiovascular and Thoracic Surgery	6,701	1.905	0.008390
138	ANNALS OF THE ROYAL COLLEGE OF SURGEONS OF ENGLAND	5,057	1.891	0.003910
139	JOURNAL OF CARDIOVASCULAR SURGERY	2,091	1.888	0.002010
140	Journal of Laparoendoscopic & Advanced Surgical Techniques	4,506	1.878	0.005580
141	CLINICAL NEUROLOGY AND NEUROSURGERY	7,916	1.876	0.008290
142	STEREOTACTIC AND FUNCTIONAL NEUROSURGERY	2,298	1.875	0.001840
143	ANZ JOURNAL OF SURGERY	5,909	1.872	0.005790
144	Geriatric Orthopaedic Surgery & Rehabilitation	817	1.870	0.001540
145	Annals of Surgical Treatment and Research	1,110	1.859	0.002120
146	Journal of Burn Care & Research	4,287	1.845	0.004050
147	PEDIATRIC SURGERY INTERNATIONAL	5,254	1.827	0.004730
147	THORACIC AND CARDIOVASCULAR SURGEON	2,225	1.827	0.002190
149	Journal of Neurological Surgery Part B-Skull Base	1,330	1.826	0.002420
150	Computer Assisted Surgery	213	1.787	0.000470
151	Thoracic Surgery Clinics	1,127	1.750	0.001480
152	OPHTHALMIC PLASTIC AND RECONSTRUCTIVE SURGERY	4,117	1.746	0.003270

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
153	EUROPEAN SURGICAL RESEARCH	1,325	1.745	0.000770
154	NEUROLOGIA MEDICO-CHIRURGICA	3,545	1.742	0.002530
155	Journal of Korean Neurosurgical Society	2,692	1.729	0.002520
156	SURGICAL LAPAROSCOPY ENDOSCOPY & PERCUTANEOUS TECHNIQUES	3,343	1.719	0.002720
157	CIRUGIA ESPANOLA	999	1.653	0.001140
158	BRITISH JOURNAL OF ORAL & MAXILLOFACIAL SURGERY	6,230	1.651	0.004820
159	Journal of Cardiothoracic Surgery	2,757	1.637	0.003630
160	JOURNAL OF CARDIAC SURGERY	3,035	1.620	0.002960
161	BRITISH JOURNAL OF NEUROSURGERY	3,840	1.596	0.003290
162	NEUROCHIRURGIE	1,151	1.553	0.001180
163	WOUNDS-A COMPENDIUM OF CLINICAL RESEARCH AND PRACTICE	1,501	1.546	0.001520
164	ANNALS OF PLASTIC SURGERY	10,891	1.539	0.008480
165	Annals of Transplantation	1,328	1.530	0.001840
166	Annals of Thoracic and Cardiovascular Surgery	1,322	1.520	0.001120
167	General Thoracic and Cardiovascular Surgery	1,831	1.517	0.002490
168	CHILDS NERVOUS SYSTEM	7,901	1.475	0.006050
169	ANNALS OF VASCULAR SURGERY	7,121	1.466	0.009160
170	Journal of Plastic Surgery and Hand Surgery	998	1.462	0.001600
171	FACIAL PLASTIC SURGERY	1,675	1.446	0.001580

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
172	CLEFT PALATE-CRANIOFACIAL JOURNAL	5,115	1.433	0.002930
173	Journal of Minimal Access Surgery	874	1.407	0.001330
174	Acta Cirurgica Brasileira	1,731	1.388	0.001450
175	Brazilian Journal of Cardiovascular Surgery	1,136	1.312	0.001240
176	Ophthalmic Surgery Lasers & Imaging Retina	3,424	1.300	0.003930
177	Journal of Foot & Ankle Surgery	4,150	1.286	0.005080
178	INTERNATIONAL JOURNAL OF SURGICAL PATHOLOGY	1,930	1.271	0.001840
179	Journal of Neurological Surgery Part A-Central European Neurosurgery	788	1.268	0.001310
180	SURGICAL AND RADIOLOGIC ANATOMY	4,351	1.246	0.002900
181	Videosurgery and Other Miniinvasive Techniques	700	1.195	0.000920
182	Progress in Transplantation	1,005	1.187	0.001180
183	PEDIATRIC NEUROSURGERY	2,512	1.162	0.000910
184	Journal of Orthopaedic Surgery	2,058	1.118	0.003050
185	ACTA CHIRURGICA BELGICA	1,149	1.090	0.000810
186	Vascular and Endovascular Surgery	1,619	1.089	0.001590
187	TRANSPLANTATION PROCEEDINGS	13,635	1.066	0.008260
188	JOURNAL OF CRANIOFACIAL SURGERY	10,436	1.046	0.011300
189	Handchirurgie Mikrochirurgie Plastische Chirurgie	743	1.018	0.000430
190	Turkish Neurosurgery	1,673	1.003	0.001810
191	MINERVA CHIRURGICA	655	1.000	0.000560

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
191	Seminars in Vascular Surgery	734	1.000	0.000650
191	UNFALLCHIRURG	1,710	1.000	0.001230
194	Hand Surgery & Rehabilitation	383	0.969	0.000990
195	CHIRURG	1,451	0.955	0.000990
196	European Surgery-Acta Chirurgica Austriaca	330	0.953	0.000470
197	Plastic Surgery	300	0.947	0.000670
198	ZENTRALBLATT FUR CHIRURGIE	832	0.942	0.000490
199	Annali Italiani di Chirurgia	812	0.766	0.000650
200	AMERICAN SURGEON	6,964	0.688	0.005930
201	HEART SURGERY FORUM	696	0.676	0.000700
202	Annales de Chirurgie Plastique Esthetique	824	0.660	0.000630
203	Indian Journal of Surgery	1,514	0.656	0.001690
204	Bariatric Surgical Practice and Patient Care	166	0.607	0.000280
205	NEUROCIRUGIA	405	0.553	0.000300
206	SOUTH AFRICAN JOURNAL OF SURGERY	419	0.375	0.000360
207	Cirugia y Cirujanos	502	0.361	0.000540
208	CESKA A SLOVENSKA NEUROLOGIE A NEUROCHIRURGIE	242	0.350	0.000160
209	Turk Gogus Kalp Damar Cerrahisi Dergisi-Turkish Journal of Thoracic and Cardiovascular Surgery	332	0.332	0.000230
210	OPERATIVE TECHNIQUES IN SPORTS MEDICINE	398	0.280	0.000290
211	Facial Plastic Surgery & Aesthetic Medicine	143	Not Available	0.000000

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
211	Joint Diseases and Related Surgery	0	Not Available	0.000000

Appendix D Curriculum Vitae

My curriculum vitae does not appear in the electronic version of my work for reasons of data protection.

Appendix E List of publications

Number of first authorships: 2
Number of co-authorships: 13

The following journal impact factors are based on https://jcr.clarivate.com/ (accessed October 10 2022, 5:30 pm)

1. A Recurrent Neural Network Model for Predicting Activated Partial Thromboplastin Time After Treatment With Heparin: Retrospective StudyBoie SD, Engelhardt LJ, Coenen N, Giesa N, Rubarth K, Menk M, Balzer F., JMIR Med Inform. 2022 Oct 13;10(10):e39187.
   Impact factor: 3.228

2. Body computed tomography in sepsis: predictors of CT findings and patient outcomes in a retrospective medical ICU cohort study Body computed tomography in sepsis: predictors of CT findings and patient outcomes in a retrospective medical ICU cohort study., Emerg Radiol. 2022 Aug 4
   Impact factor: 0.55

3. Prognostic value of lymph node involvement in oral squamous cell carcinoma Voss JO, Freund L, Neumann F, Mrosk F, Rubarth K, Kreutzer K, Doll C, Heiland M, Koerdt S., Clin Oral Investig. 2022 Jul 27.
   Impact factor: 3.607

4. Cerebrovascular Events in Suspected Sepsis: Retrospective Prevalence Study in Critically Ill Patients Undergoing Full-Body Computed Tomography.Pohlan J, Nawabi J, Witham D, Schroth L, Krause F, Schulze J, Gelen S, Ahlborn R, Rubarth K, Dewey M..Front Neurol. 2022 May 9;13:811022.
   Impact factor: 4.086

5. A frozen decade: Ten years outcome of atrial fibrillation ablation using a single shot device for pulmonary vein isolation. Bergau L, Sciacca V, Nesapiragasan V, Rubarth K, Konietschke F, Fink T, El Hamriti M, Imnadze G, Dagher L, Braun M, Khalaph M, Guckel D, Heintze J, Noelker G, Vogt J, Sommer P, Sohns C., J Cardiovasc Electr. 2022
   Impact factor: 2.871

6. Pattern of cervical lymph node metastases in squamous cell carcinoma of the upper oral cavity – How to manage the neck. Doll, C., Mrosk, F., Wuester, J., Runge, A.-S., Neumann, F., Rubarth, K., Heiland, M., Kreutzer, K., Voss, J., Raguse, J.-D., Koerdt, S., Oral Oncology. 2022;130
   Impact factor: 5.337

7. Obesity is Associated With Myelin Oligodendrocyte Glycoprotein Antibody Associated Disease in Acute Optic Neuritis. (Preprint) Stiebel-Kalish H, Rubarth K, Blum K, Tiosano A, Lotan I, Hellmann MA, Wilf-Yarkoni A, Bialer O, Flanagan EP, Pittock SJ, Bhatti MT, Schmitz-Hübsch T, Friedemann P, Asseyer S, Chen, J.
   Impact factor: NA

8. Estimation and Testing of Wilcoxon–Mann–Whitney Effects in Factorial Clustered Data Designs. Rubarth, K., Sattler, P., Zimmermann, H.G., Konietschke F., Symmetry. 2022;14(2).
   Impact factor: 2.713

9. Magnetic field-induced interactions between phones containing magnets and cardiovascular implantable electronic devices: Flip it to be safe? Lacour, P., Dang, P.L., Heinzel, F.R., Parwani, A.S., Bähr, F., Kucher, A., Hohendanner, F., Niendorf, T., Rahimi, F., Saha, N., Han, H., Rubarth, K., Sherif, M., Boldt, L.H., Pieske, B., Blaschke, F., Heart Rhythm. 2022;19(3):372-80.
   Impact factor: 6.343

10. Ranking procedures for repeated measures designs with missing data: Estimation, testing and asymptotic theory. Rubarth, K., Pauly, M., Konietschke, F., Stat Methods Med Res. 2022;31(1):105-18.
    Impact factor: 3.021

11. Relevance of CT for the detection of septic foci: diagnostic performance in a retrospective cohort of medical intensive care patients. Pohlan, J., Witham, D., Opper Hernando, M.I., Muench, G., Anhamm, M., Schnorr, A., Farkic, L., Breiling, K., Ahlborn, R., Rubarth, K., Praeger, D., Dewey, M., Clin Radiol. 2022;77(3):203-9.
    Impact factor: 2.350

12. Catheter ablation for atrial fibrillation in patients with end-stage heart failure and eligibility for heart transplantation. Sohns, C., Marrouche, N.F., Costard-Jäckle, A., Sossalla, S., Bergau, L., Schramm, R., Fuchs, U., Omran, H., Rubarth, K., Dumitrescu, D., Konietschke, F., Rudolph, V., Gummert, J., Sommer, P., Fox, H., ESC Heart Fail. 2021;8(2):1666-74.
    Impact factor: 4.411

13. Motor excitability in bilateral moyamoya vasculopathy and the impact of revascularization. Acker, G., Giampiccolo, D., Rubarth, K., Mertens, R., Zdunczyk, A., Hardt, J., Jussen, D., Schneider, H., Rosenstock, T., Mueller, V., Picht, T., Vajkoczy, P., Neurosurg Focus. 2021;51(3).
    Impact factor: 4.047

14. Tumor-Associated Microglia/Macrophages as a Predictor for Survival in Glioblastoma and Temozolomide-Induced Changes in CXCR2 Signaling with New Resistance Overcoming Strategy by Combination Therapy. Urbantat, R.M., Jelgersma, C., Brandenburg, S., Nieminen-Kelhä, M., Kremenetskaia, I., Zollfrank, J., Mueller, S., Rubarth, K., Koch, A., Vajkoczy, P., Acker, G., Int J Mol Sci. 2021;22(20).
    Impact factor: 5.924

15. What statistics instructors need to know about concept acquisition to make statistics stick. Kruppa, J., Rohmann, J., Herrmann, C., Sieg, M., Rubarth, K., Piper, S., J Univ Teach Learn Pract. 2021;18(2).
    Impact factor: 0.55

16. Cool enough? Lessons learned from cryoballoon-guided catheter ablation for atrial fibrillation in young adults.Bergau, L., El Hamriti, M., Rubarth, K., Dagher, L., Molatta, S., Braun, M., Khalaph, M., Imnadze, G., Nölker, G., Nowak, C.P., Fox, H., Sommer, P., Sohns, C.,J Cardiovasc Electr. 2020;31(11):2857-64.
    Impact factor: 2.871

Appendix F Acknowledgements

First of all, I would like to thank my excellent supervisor team, Prof. Dr. Frank Konietschke, Prof. Dr. Markus Pauly and Prof. Dr. Geraldine Rauch. In particular, I am extremely thankful to Frank Konietschke, who introduced me into the world of nonparametric statistics. Frank Konietschke was always available and supported me whenever I had a question regarding my research or project related issues. I learned a lot while working with Frank Konietschke, not only how to apply my theoretical knowledge in practice but also how to communicate with scientists from other fields.
Furthermore, I would like to thank Dr. Paavo Sattler for very fruitful discussions and collaborations.

Moreover, I would like to thank all current and former colleagues at the Institute of Biometry and Clinical Epidemiology at the Charité - Universitätsmedizin Berlin. I very much enjoy the work climate there. Especially, I would like to thank Dr. Asanka Gunawardana and Dr. Claus Peter Nowak, who supported and encouraged me during my first days and months at the Charité. Furthermore, I would like to thank Dr. Sophie Piper, Dr. Uwe Schöneberg and Annett Kaphahn for their support and advice at everyday work.
Furthermore, I would like to thank the PhD studies "Health Data Sciences" team, especially Prof. Dr. Tobias Kurth and Jessica Rohmann for offering a great selection of courses and support during my PhD journey.
Finally, I would like to express my deep gratitude to my friends, Lorenz and my parents for being there whenever I needed them.