ENDPOINT-ORIENTED CLUSTERING OF INDIVIDUAL PATIENT DATA

A new approach for the health economic evaluation of medical interventions

Marie Hinnenthal

Vollständiger Abdruck der von der Fakultät für Wirtschafts- und Organisationswissenschaften der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)

genehmigte Dissertation.

Gutachter:

1. Univ.-Prof. Dr. Andreas Brieden
2. Univ.-Prof. Dr. Claudius Steinhardt

Die Dissertation wurde am 22.12.2016 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Wirtschafts- und Organisationswissenschaften am 24.06.2017 angenommen. Die mündliche Prüfung fand am 11.07.2017 statt.

der Bundeswehr
Universität München

ENDPOINT-ORIENTED CLUSTERING OF INDIVIDUAL PATIENT DATA

A new approach for the health economic evaluation of medical interventions

Marie Hinnenthal

DISSERTATION
zur Erlangung des akademischen Grades eines
Doktor der Wirtschafts- und
Sozialwissenschaften (Dr. rer. pol.)

Dezember 2016

ENDPOINT-ORIENTED CLUSTERING OF INDIVIDUAL PATIENT DATA

A new approach for the health economic evaluation of medical interventions

Marie Hinnenthal

Abstract

In the age of big data, the collection of individual patient information, with the help of clinical trials, is becoming an increasingly important area of healthcare. These large amounts of data have the potential to provide an improved medical care for patients. Especially with regard to the health economic evaluation of medical interventions, the analysis of this data can lead to a better patient-oriented medication in terms of evidence-based medicine. Meta-analytic approaches for the evaluation of clinical drug studies only estimate a weighted mean value of the measured endpoints. The whole potential of the big amounts of individual patient data is therefore not nearly exploited. The collected patients' information, e.g. the differences in the socio-economic parameters of different patient groups and the associated heterogeneity in the efficacy of a drug, are not sufficiently considered in the common analysis. For approaches like the subgroup analysis, which uses such information, the considered groups are often too small to provide statistically well-founded results. Therefore, new methods are needed for the analysis of the extensive patient data. The new approaches presented in this thesis are all based on an innovative endpoint-oriented clustering, developed by Brieden and Gritzmann. The algorithm identifies hidden multidimensional structures and forms sufficiently large clusters in which patients are grouped with similar combinations of their characteristic values. The new invented methods, which are applied on the identified patient collectives, deal with the evaluation and prediction of the efficacy of medical interventions and the identification of clinical heterogeneity in the treatment effects.

Zusammenfassung

Im Zeitalter von 'Big Data' nimmt das Sammeln von individuellen Patientendaten mit Hilfe von klinischen Studien einen immer größer werdenden Stellenwert im Gesundheitswesen ein. Dabei haben diese großen Mengen an Daten das Potential, eine verbesserte medizinische Versorgung für Patienten zu bieten. Insbesondere im Hinblick auf die gesundheitsökonomische Bewertung von medizinischen Interventionen, kann die Auswertungen dieser Daten zu einer besseren patientenorientierten Medikation im Sinne der evidenzbasierten Medizin führen. Mit den herkömmlichen metaanalytischen Methoden für die Auswertung von klinischen Medikamentenstudien werden lediglich gewichtete Mittelwerte der gemessenen Endpunkte bestimmt. Das ganze Potential, das die großen Mengen an Daten mit sich bringt, wird somit nicht annähernd ausgeschöpft. Die gesammelten Patienteninformationen, wie z.B. die Unterschiede in den sozio-ökonomischen Parametern von unterschiedlichen Patientengruppen und die damit einhergehende Heterogenität in der Wirksamkeit eines Medikaments, werden nicht ausreichend in die Analyse mit einbezogen. Bei Ansätzen wie z.B. der Subgruppenanalyse, die solche Informationen verwertet, sind oft die betrachteten Gruppen zu klein um statistisch fundierte Aussagen treffen zu können. Daher ist es notwendig neue Methoden für die Auswertung der großen Datenmengen zu entwickeln. Die in dieser Arbeit vorgestellten neuen Ansätze basieren alle auf einem innovativen endpunkt-orientierten Clustering-Algorithmus, der von Brieden und Gritzmann entwickelt wurde. Dieser identifiziert versteckte multidimensionale Strukturen und bildet genügend große Kollektive in denen Patienten gruppiert werden, die im Hinblick auf die Kombination ihrer charakteristischen Ausprägungen ähnlich sind. Die darauf aufbauenden Methoden dienen der Bewertung und Vorhersage der Effizienz medizinischer Interventionen und der Identifizierung der Heterogenität in der Wirksamkeit eines Medikaments in den identifizierten Patientenkollektiven.

Acknowledgment

It was a long, stony path, rich in detours, until the completion of my PhD thesis. During this time a lot of people assisted me with good advice and encouragement and made this thesis possible. To express my gratefulness, I want to thank all these people for their overwhelming support.

First of all, I want to thank my doctoral advisor Prof. Dr. Andreas Brieden for all his unlimited engagement in all aspects. I'm especially grateful for the possibility to deal with such an interesting and innovative topic. Furthermore, I want to thank him for creating such a friendly and pleasant working atmosphere.

Then, I want to send a big thank-you to my colleagues and friends Michael, Bernhard, Falk, Sabine, Saskia and Wolfgang and of course to all my other colleagues at the university. We had a very great time not only during the professional discussions. Therefore, the lunch and the coffee breaks offered great opportunities to talk about all the important things in this world.

A further very big thank-you goes to my big family. To my parents Rita and Hans, my brothers Boris, Jascha and Michel with their wifes Sabine, Miriam and Kerstin, my nephews Rasmus, Piet and Theo, my nieces Lotte and Ida and last but not least Juli and Karo... They were and are always there for me and supported me in all situations. I love you all!

Last but not least, I want to thank my friends Raita and Giannina and all the others for giving me the encouragement to finish this thesis and for the necessary distractions.

Contents

List of Figures

List of Tables

List of Abbreviations

AMNOG Reform of the Market for Medicinal Products

CBA cost-benefit analysis

CEA cost-efficiency analysis

CMA cost-minimization analysis

CUA cost-utility analysis

EBM evidence-based medicine

G-BA Federal Joint Committee

IQWiG Institute for Quality and Efficiency in Health Care

MLE maximum-likelihood estimation

NNT Number needed to treat

OR Odds ratio

QALY quality adjusted life year

QoL quality of life

RCT randomized controlled trial

RR Risk ratio

RD Risk difference

SGB V Social Code Book V

SHI Statutory Health Insurance

SR systematic review

1. Introduction

In the last few years, the collection of big data has been revolutionary for the healthcare industry [51]. Large amounts of data have been generated by medical care and medically relevant researches [54]. Driven by the potential to improve the quality of healthcare delivery, these massive quantities of data serve to support a wide range of medical and healthcare functions [54]. Among others, this includes the health economic evaluation, in terms of evidence-based medicine, to provide customized, personalized information for all patients. Especially healthcare institutions, like the Institute for Quality and Efficiency in Health Care (IQWiG) or pharmaceuticals companies, use this data for the evaluation of the benefit of new invented drugs. Thereby, the emphasis is on collecting data on the health of patients with the help of clinical trials. The problem is that this data is becoming bigger and more complex, especially with the development of new storage technologies, so that they can not be easily managed with traditional or common data management tools and methods [67]. Big data in healthcare is overwhelming not only because of its volume but also because of the diversity of data types [67]. It covers clinical data, like physician's written notes and prescriptions, patient data in electronic patient records, machine generated data, like monitored vital signs, social media posts and less patient-specific information, including emergency care data, news feeds and articles in medical journals [54]. Statistical, mathematical, visualization and computational approaches from a wide range of disciplines are needed to keep apace of the complexity in big data and to advance medicine [51]. The common way for handling big data, in terms of health economic evaluation, is the use of meta-analytic approaches. The main assumption of most of these methods is that there is only one true effect of a medical intervention for all treated patients. Therefore, these approaches deal with the summary and synthesis of patient data derived by endpoint-oriented clinical studies to provide evidence about the efficacy of a medical intervention. With the consideration of several clinical

studies, it is possible to achieve more statistical power and a higher accuracy of the treatment effect estimates. The aggregation of the results of the single trials to one summary treatment effect is then, roughly speaking, carried out by considering mean outcomes of the different studies. This implies the assumption that the benefit of a drug is independent of factors such as gender, age or the body mass index (BMI). In suitable, multivariate subgroup analyses, such as the analysis of the efficacy of a drug among women with a certain BMI in a specific age group, in general the respective collectives are too small to derive significant results [21]. This clinical variety of characteristic values combinations might lead to the so-called clinical heterogeneity in the treatment effects. Unfortunately, it is difficult to consider this heterogeneity with the common meta-analytic approaches.

In this thesis, new theories for the handling of big patient data and the analysis of the benefit of medical interventions, in terms of health economic evaluation, are introduced. The theories are based on the results of the application of an innovative endpoint-oriented clustering technology that just aims to identify sufficiently large and homogeneous patient collectives by means of similar characteristic values combinations. With the use of this technology, clinical heterogeneity is taken into account in the analysis of the treatment's efficacy. The main assumption is that there is a treatment effect for each patient collective which has been identified by the endpoint-oriented clustering algorithm. The new invented cluster-based methods, which are applied on the identified patient collectives, deal with the evaluation and prediction of the efficacy of medical interventions and the identification of clinical heterogeneity in the treatment effects.

To get an understanding of the relevance and necessity of these new invented cluster-based analyses, in Chapter 2 of this thesis, we will give an introduction to the health economic evaluation and its methods for the assessment of the benefit of medical interventions in terms of evidenced-based medicine. Thereby, the role of the scientific IQWiG is condensed. Since the health economic evaluation methods are based on significant data on the efficacy of the analyzed drug, in Chapter 3 we will give an overview of meta-analysis, which is currently the main supplier of summaries and syntheses of patient data derived by endpoint-oriented

clinical trials. Thereby, we will present the standard treatment effect estimates for binary and cardinal outcome or the so-called endpoint of a patient. Then, we will give an overview of the two established models, the fixed-effects and the random-effects model, for the aggregation of the treatment effect estimates in the single studies. In this context, in Chapter 4 of this thesis, we will highlight the causes and the assessment of the resulting heterogeneity in the treatment effects of the single trials with the help of the Q-statistic and its related indices. Furthermore, we will present methods which take heterogeneity into account in terms of meta-analysis. Unfortunately, with the use of these methods, like the subgroup analysis and the meta-regression, several problems arise and we will give a recommendation how these existent problems can be handled with the help of new invented cluster-based analysis methods. Therefore, in Chapter 5, the innovative endpoint-oriented clustering approach is introduced. We will present the general definitions which are necessary for the formulation of the clustering algorithm. In this context, we distinguish between supervised and unsupervised learning approaches which are commonly used in the field of machine learning. Unsupervised techniques discover hidden structures in high-dimensional patient data sets without prior information. In contrast, supervised learning approaches train a mapping based on a given set of characteristic values combinations of patients and their endpoints. For more information, please refer to [45]. On the patient collectives, identified by the unsupervised or supervised learning clustering approaches, new statistical cluster-based analysis methods have been developed for the evaluation of medical interventions, the identification of heterogeneity and the prediction of the efficacy of the analyzed drugs. In Chapter 6, the empirical results of the new invented cluster-based analyses and the common meta-analytic techniques are presented and compared. All methods had been applied on a patient data set derived by clinical studies on a new invented and two standard antidepressants. We will see that the existent problem of heterogeneity could be handled with the help of the new cluster-based approaches. In this context, we will have a closer look at the cluster-based identification of heterogeneity within and across clusters. Furthermore, the results of the cluster-based prediction of the efficacy of all antidepressants and the assessment of the prediction's reliability is demonstrated. In Chapter 7, we will summarize the main empirical results and we will discuss possible future work.

2. Health economic evaluation

Health economics is an interdisciplinary science which deals with the production, distribution and consumption of health goods in medicare and combines health science and economics [75]. Health economist are engaged in analyzing the supply and demand of health services under consideration of existent information asymmetries [75]. The health economic efficiency states that the recoverable services of the statutory health insurances and the services of the medical service provisioner have to be sufficient, appropriate and economical [11]. Services, which are not necessary or uneconomical, may not be used by the insured or may not be produced by the service providers. Also the health insurance funds may not authorize them [11][75]. Therefore, health economists' task is to find a balance between the medical possibilities, its financing and quality [75]. Methods and models for the assessment and evaluation of medical interventions and new health technologies, by means of medication, labor technologies or diagnostics [11], have been developed. For this so-called health economic evaluation, clinical, economic and epidemiological data is applied [58]. All used data sources have to be described accurately, their choice has to be motivated and their suitability and validity has to be evaluated [58]. In Chapter 3, we want to give an overview of the standard method, the meta-analysis, for providing clinical data for the assessment and evaluation of health services. But since this method is critical questioned due to suitability and validity of data, we will introduce new innovative cluster-based methods in Chapter 5, the cluster-based meta-analysis and further approaches, for the supply of suitable and valid data for the health economic evaluation.

To give an understanding of the relevance and necessity of the new invented cluster-based analyses, we want to give an introduction to heath economic evaluation. In Section 2.1 of this chapter, we will give an overview of the general health

economic evaluation methods for which the cluster-based analysis could provide its suitable and valid data. We will only give a short introduction, a detailed description of each method can be found in [58]. In Section 2.2, which is mainly based on [36], we will highlight the health economic evaluation in Germany as a task of the scientific IQWiG and give an overview of its legal responsibilities, its health economic evaluation methods and the role of evidence-based medicine in the evaluation of medical interventions in Germany.

2.1. General evaluation methods

In health economic evaluation of medical interventions, the costs and the patients' outcome, e.g. the influence on the expectancy of life or the health status, are compared [11]. Since health economic analyses serve as decision support, it is necessary to provide a minimum of transparency [11]. Therefore, specific standards for the implementation and publication of the methods in health economic analysis have been developed [11]. When analyzing the costs and the benefit of a health economic service or medical intervention, different evaluation methods can be applied depending on the unit of the measured outcome [11]. A summary of these methods is shown in Table 2.1. There are three different possibilities for the measurement of the outcome. If it is in natural one-dimensional units, like clinical parameters, e.g. the remission or response rate or gained life years, we choose the cost-efficiency analysis (CEA). In case of outcome measured in utility values, the common approach is the cost-utility analysis (CUA). If we have monetary units, we take the cost-benefit analysis (CBA) [11]. For each method, the costs are expressed in a monetary unit. However, the most simplest way for the evaluation of medical interventions is the cost-minimization analysis (CMA). The descriptions of the different methods in the following subsections are mainly based on [11] and [58]. We will start with the CMA.

Cost-minimization analysis

In the CMA, two or more different medical interventions are compared due to their net costs to identify the most cost-efficient alternative. The evaluation method is only feasible under the assumptions that the treatment effects are equal. The equality has to be presented transparently and comprehensible, e.g.

Method	Cost evaluation	Patient' outcome Evaluation	Comparison
CMA	monetary	none	none
CEA	monetary	natural units	costs per outcome unit
CUA	monetary	utility values	costs per QALY
CBA	monetary	monetary	net costs

Table 2.1.: Methods for health economic evaluation

with the help of meta-analysis [76].

The types of interventions, which can be evaluated with this method, are limited. A common example of a CMA is the comparison of generic equivalents of medical interventions. If a manufacturer want to launch a generic medication, he has to demonstrate that its product is biologically equivalent to the already established medication. Thus, in case the compared medications have the same chemical composition, dose and pharmaceutical properties, only the cost of the medication itself needs to be considered due to the same assumed outcome. Another example of the conduction of a CMA, is the comparison of costs for a medication which is administered in different settings. A common example is the administration of an intravenous antibiotics in a hospital, compared to the medication of the same antibiotics at home via a home health care service [46].

Cost-efficiency analysis

The next evaluation method we want to discuss is the CEA. In this approach, the costs and the efficacy of medical interventions are compared. Thereby, the costs are represented in monetary units and the patients' outcome is expressed in non-monetary units, e.g. gained life years or a clinical parameters. The requirements for the conduction of the CEA is that the examined interventions have the same qualitative clinical endpoints, by means of the same effect measure for the evaluation of the treatment effect. It is used especially for the comparison of two mutually exclusive interventions. The comparative criterion t_{CEA} is carried out in costs per outcome unit, where the intervention is chosen with the lower t_{CEA} value. For the example of measuring the outcome in gained life years, the

comparative criterion is

$$t_{CEA} = \frac{\text{costs [monetary unit]}}{\text{life-years-gained benefit}}.$$

Using this method, it is not possible to evaluate interventions with more than one effect. For example, if safety-related services not only avoid cases of death but also assaults, then CEA fails because it does not take into account the aggregation of multi-dimensional effects. Another disadvantage of this method is that there is indeed a ranking of mutually exclusive interventions but there is no recommendation [11] for which value a conduction of an intervention is no longer reasonable. The only feasible application is the evaluation of medical interventions when apportioning a fixed budget among those. In this case, this intervention should be taken with the lowest t_{CEA} value and continue until the budget is used up. Thereby, a problem which occurs, is the determination of the budgets' amount.

Cost-utility analysis

The next evaluation method is the CUA. Like it is stated above, this method is used if the outcome is measured in utility values, by means of values of a cardinal utility function. The costs are also represented in a monetary unit. The CUA is taken for the comparison of two mutually exclusive interventions. The big advantage of this method is that all effects of a medical intervention, e.g. the prolongation of life and the change of the health status, can be considered in the evaluation by a suitable weighting scheme. For the conduction of a CUA, different approaches have been developed. The interested reader is referred to [58] for more information. The most common utility measure is the quality adjusted life year (QALY). Using this approach, all possible health conditions are evaluated on a scale where death is rated by 0 and the status of perfectly health is rated by 1. The utility function is then defined in a way that a representative individual would be indifferent e.g. between the scenario 'survive one year with health status 0.5' and scenario 'survive half a year with health status 1'. With this approach, all effects of a medical intervention can be compared and aggregated to one summary index. This index can be interpreted as growth in the QALYs.

For the comparison of interventions, the comparative criterion

$$t_{CUA} = \frac{\text{costs [monetary unit]}}{\text{benefit [utility unit]}}$$

is used and the intervention with the lower value is taken. In contrast to the CEA, with this method medical interventions with effects on different clinical levels can be compared due to the calculated utility values.

But this method also has mentionable disadvantages. For the determination of the utility function, it has to be fixed who rates the different states of health, because different persons may have different opinions on the scaling. Also the CUA only provides a ranking of mutually exclusive interventions but does not recommend a limit for the comparative criterion t_{CUA} for the conduction of the intervention. Here, also the splitting of a fixed budget is a feasible application, with the problem of the optimal determination of the budgets' amount.

Cost-benefit analysis

The last evaluation method we want to discuss is the CBA. For this method, e.g. each prolongation of life and each change in the state of health is expressed in a monetary value. Using this method, each considered intervention can be evaluated separately. The evaluation criterion is defined by

$$t_{CBA} = \frac{\text{costs [monetary unit]}}{\text{utility [monetary unit]}}.$$

An intervention is recommended as long as $t_{CBA} < 1$ or $t'_{CBA} > 0$, if

$$t'_{CBA} = \text{utility [monetary unit]} - \text{costs [monetary unit]}.$$

In comparison to the CEA and CUA, this method answers the question which amount should be generally spent for medical interventions to prolong the life expectancy and to improve the quality of life.

Lemma 2.1.1. *The CEA is suitable for the comparison of two mutually exclusive medical interventions with one-dimensional effect. On the other hand the*

CUA is also used for the evaluation of interventions with more than one effect. However, without the specification of a specific budget there is no conclusion if the intervention should be conducted. The CBA is a monetary evaluation of life and health and therefore enables the evaluation of each single intervention.

The big advantage of the CBA is that there is a recommendation for the conduction of a medical intervention. Certainly, it is based on another concept for the measurement of the effects of interventions. Generally, it emanates from the assumption that there is a subjective utility and uses the willingness to pay of the affected patients as a measure for the monetary value. The CBA is also used by IQWiG for the health economic evaluation in Germany.

2.2. Health economic evaluation in Germany

In Germany the scientific IQWiG is responsible for the evaluation of the benefit and harm of medical interventions, as well as their economic implications to contribute to an continuous improvement of the quality and efficiency of the German population's medicare [36]. IQWiG was established as part of the German Health Care Reform in 2004 by the Federal Joint Committee (G-BA) and is largely funded by contributions of Statutory Health Insurance (SHI) members. For this purpose, a levy is determined by the G-BA and is paid by all German medical practices and hospitals treating SHI-insured patients. Its legal foundations and responsibilities can be found in the Social Code Book V (SGB V) and have been adapted and extended within the framework of further German Health Care Reforms in the last few years [36]. The Institute's main focus in the evaluation of medical interventions is the benefit for patients. Therefore, questions like 'Does the intervention prolong life or reduce symptoms? Does it improve the quality of life of the patient?' need to be answered. Based on this neutral health information on current medical knowledge and the patient-oriented work, the population is able to make informed decisions on health care interventions [36].

2.2.1. Legal responsibility

The Institute takes action in issues on quality and efficiency of services in the context of statutory health insurance, like the examination of advantages and disadvantages of medical interventions for patients. IQWiG operates for example in areas like research, representation and assessment of the current scientific evidence on diagnostic and therapeutic procedures for selected diseases. Furthermore, it is responsible for the preparation of scientific reports, expert opinions, and comments on quality and efficiency issues of the statutory health insurance service. During the preparation of its reports, the Institute ensures high transparency of their procedures and appropriate involvement of third parties. In all important phases of report preparation, the law obliges the Institute to provide the opportunity of comment to experts, manufacturers and relevant organizations representing the interests of patients and self-help groups of chronically ill and disabled persons, as well as to the Federal Government Commissioner for Patients Affairs. Then, the Institute considers these comments in its assessments. Another task is the appraisal of evidence-based clinical practice guideline on the most relevant diseases from an epidemiological point of view. But one of the most important tasks is the assessment of the benefit and costs of drugs and the associated provision of easily understandable information for all patients and consumers on the quality and efficiency of health care services, as well as on the diagnosis and treatment of diseases of substantial epidemiological relevance. In order to perform the mentioned tasks, IQWiG has to award scientific contracts to external experts. Just like the Institute's scientific staff, also the external experts have to disclose all connections to associations and contract organizations, particularly to pharmaceutical and medical devices industries, including details on the type and amount of any remuneration received. This measure ensures the professional independence of the Institute.

Like it is stated in the SGB V, only the G-BA commissions IQWiG. Furthermore, the institutions which form the G-BA, the Federal Ministry of Health and the organizations relevant for the fulfillment of the interests of patients and self-help groups of chronically ill and disabled persons, as well as the Federal Government Commissioner for the needs of the patient, can apply for commissioning the Institute at the G-BA. After the fulfillment of the tasks the commission's

results of the Institute are submitted to the G-BA as recommendations. The G-BA has to consider these recommendations under its terms of reference.

2.2.2. Economic evaluation

Due to the resolutions made in the context of the Reform of the Market for Medicinal Products (AMNOG), at the beginning of 2011, the G-BA can commission the Institute to assess the benefit of drugs with new active ingredients shortly after market entry. Therefore, dossiers summarizing the results of studies have to be submitted by the manufacturer. In connection with this benefit analysis, the G-BA can also commission the Institute to conduct a health economic evaluation. Therefore, cost-benefit relationships of medical interventions are opposed to provide information which are considered for cost absorption by the insurance community. In this context, the cost-benefit analysis of IQWiG, described in [37], is mainly used if price negotiation between the SHI head association and the pharmaceutical company fails and there is no compliance in following arbitration proceedings. Thereby, due to law, the suitability and reasonableness of the cost absorption by the insurance community has to be considered by the G-BA. To do so, the G-BA receives information about the suitability on the basis of the CBA conducted by IQWiG. Information about the reasonableness are provided by economic impact analysis. The evaluation of suitability and reasonableness, with regard to cost absorption, is conducted under consideration if there is a justifiable relation between costs and benefit of the medical intervention which has to arise from the CBA. With respect to this evaluation, the Institute must ensure that the assessment of the benefit is conducted according to internationally approved standards of evidence-based medicine and health economics. The criteria to determine the benefit for patients are named in law like the increase in life expectancy, the improvement in health status and quality of life (QoL) and the reduction in disease duration and adverse effects.

As part of the health care structure, law amendments were made in the year 2012 which enables the G-BA to initiate clinical trials for new interventions unless the benefits hasn't been sufficiently proved yet and, however, the method reveals the potential of a necessary treatment alternative. Even external applicants, e.g.

medical device manufacturers, can apply for a clinical trial by presenting relevant documents of the method's potential to the G-BA which has established the criteria for the evaluation of this potential. The G-BA normally contracts the Institute to check the documents, whether there is a potential for the application.

2.2.3. Evidence-based medicine

An important basis of IQWiG's work is evidence-based medicine (EBM). According to [57], evidence-based medicine has been defined as 'integrating individual clinical expertise with the best available external clinical evidence from systematic research'. In particular, it means health care for patients that is not based solely on personal opinions and conventions, but on evidence which has been surveyed by scientific methods [57]. Therefore, EBM includes approaches which should protect from wrong decisions and expectations. In this context, a wrong decision might be that a beneficial medical intervention is not included in the provision of health care or that an useless or even harmful medication is widely spread. The Institute submits certain evidence, e.g. that medication A is more efficient than medication B for the treatment of patients with a specific disease. Thus, the Institute's application is not the treatment of single patients but the identification of evidence for the benefit of patients.

The characteristic standard element of EBM is the structured and systematic way to find answers to medical questions. Therefore, the approach is to formulate the scientific question precisely. In medicine there is almost always the decision between at least two interventions. Furthermore, it has to be defined how the benefit of a therapy should be measured and which consequences can be expected. Hereby, the question to be answered is if life can be prolonged and if disorders and life quality can be improved. The EBM formulates explicitly that in medicine, for the description of the benefit of medical interventions, only probabilities or predictions for patient groups are possible. Therefore, to give evidence about the benefit, studies with a sufficiently large number of participants are required. For the planning, conduction and evaluation of these clinical trials, rules and methods have been developed which fulfill international standards. Thereby, a very important strategy of EBM is the identification of appropriate clinical

trials due to their design and conduction for the evaluation of interventions by aggregating the results of the single trials. This statistical method is called meta-analysis and is introduced in the next chapter. If there are big differences between the results of the single trials, this heterogeneity has to be considered. To deal with this arising problem, IQWiG uses subgroup analysis which is also explained in Chapter 3. The new invented cluster-based technologies introduced in Chapter 5 are based on the idea of this subgroup analysis.

3. Meta-analysis

Meta-analysis has been defined by Glass in 1976 to be the 'statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings' [25]. The term meta-analysis has been adopted within other disciplines and has proven particularly popular in clinical research, although Glass was involved in social science [25]. In medicine, the increase in the number of meta-analyses being conducted began in the 1980s, due to the greater importance of evidence-based medicine and the requirement of reliable summaries of the vast and expanding volume of clinical trials [55][77]. In this thesis, we will focus on meta-analyses of randomized controlled trials (RCTs). Besides fixed controlled trials and further study forms, especially the RCT design is a very popular and approved standard for conducting clinical studies to evaluate medical interventions or to compare two different medications [55]. This study design gets its name from the mode of carrying out the study. Compared to fixed controlled trials, the participants of RCTs are randomly grouped into control and treatment group. The treatment group receives the medical intervention and the control group receives a placebo, an already approved standard therapy or no intervention at all [25]. For all participants of a study, an outcome is measured, e.g. a binary outcome which states if the participant has responded to the administered medication respectively placebo or not. The evaluation of a clinical trial is done by estimating and making inferences about the treatment effect, depending on the current medical research question. But if one conducted clinical trial is analyzed separately, it might be too small or too limited in scope to get unequivocal or generalizable conclusions about the effect of a medical intervention. To strengthen the evidence about the efficacy, statistical methods, like the meta-analysis, are needed [24] for the summary of the collected data generated by RCTs.

In this chapter, we will give an introduction to meta-analysis of RCTs. We will start in Section 3.1 with the presentation of the idea of meta-analysis. Then, on this basis, we will give a detailed description in Section 3.2 how the treatment effect in the single clinical trials can be determined, depending on the scale of the measured outcome of the participating patients. Therefore, we will have a look on effect measures of binary and cardinal patient outcome data. Since in meta-analysis it is assumed that there is one true treatment effect for all patients, in Section 3.3 we will give an introduction how the treatment effect estimates of the single clinical trials can be aggregated to one estimated summary treatment effect. Therefore, we will present the fixed-effects and the random-effects model and we will discuss the differences between these two approaches. In the end of this chapter, in Section 3.4, it is described how meta-analyses are published to enable evidence-based medicine. The following sections are mainly based on [5] and [77].

3.1. Overview

Meta-analysis is classified to secondary research and is conducted similar to a primary study where it is common to analyze the mean value and the standard deviation of the outcome data in the treatment group of a study [5]. In the first step, all appropriate studies, including patients with the same administered medical intervention, are collected. In the second step, the treatment effect of each single study is estimated on the basis of an appropriate effect measure, e.g. the Risk Ratio for binary outcome. Possible effect measures under consideration of the type of outcome will be discussed later in this chapter. Since in meta-analysis it is assumed that there is one true treatment effect for all patients, it is common to estimate a summary effect by calculating the weighted average of the single study estimates. Therefore, the so-called fixed-effects and the random-effects model are usually conducted. We will have a detailed look at these approaches later in this chapter. There are different ways in which the results of a clinical trial are available. Most frequently, the treatment effect of a clinical trial is published e.g. in a scientific paper or a clinical trial report as a summarizing statistic. The other way, and this is most preferable way, individual patient data, like the outcome and specific attributes are available for each patient of the included studies.

In this thesis, we will focus on clinical trials with individual patient data. In this case, it is also conventional to undertake meta-regression or subgroup analysis. These approaches will be discussed in Chapter 4, when dealing with heterogeneity.

3.2. Treatment effect estimates

As already mentioned in the introductory paragraph, in meta-analysis we assume that there is one true treatment effect of a medical intervention for all patients. This treatment effect can be determined by the weighted aggregation of the single estimated effects of the included studies which deal with the same research question concerning the evaluation of a medical intervention. For this synthesis, an effect measure based on the patients' outcome is needed. In this section, we will give a short overview of the different types of effect measures, depending on the scale of the measured outcome of the patients participating in a study. As outlined in [30] and shown in Table 3.1, there are different effect measures for dichotomous, cardinal and ordinal outcome data. The outcome data of each participant is called dichotomous or binary if it has one of only two possible occurrences, e.g. an event happens or no event happens. The outcome data is called cardinal if the measurement is numerical and the use of the arithmetic mean as the measure of average can be justified, e.g. the MADRS Score for the evaluation of ataractics. Another type of effect measure for cardinal outcome data is the time-to-event data discussed in [30]. In this case, for each individual of a study, the time until a defined event occurs is measured. If the outcome is one of several ordered categories, we talk about ordinal data, e.g. counts or rates which can be got from counting the number of events each individual has experienced. In the following, we will give a detailed introduction to the effect measures for dichotomous or binary outcome data and for cardinal outcome data. For the interested reader a detailed introduction to the effect measures of the other types can be found in [77].

3.2.1. Binary data

Before we get to the definition of the effect measures for dichotomous or binary outcome data, with the two possible occurrences success and failure, we need to

Type of outcome data	Effect measures
Binary data	Risk ratio (RR) Odds ratio (OR) Risk difference (RD) Number needed to treat (NNT)
Cardinal data	Mean difference Standardized mean difference Time-to-event
Ordinal data	Rates Counts

Table 3.1.: Effect measures depending on individual outcome data

summarize the individual binary outcomes of study j, $j = 1, \ldots, n_{st}$, where n_{st} is the number of studies considered for the meta-analysis. For the determination of the effect measures, we assume that for each study the outcome of an individual in the treatment group is seen as a Bernoulli distributed random variable

$$T_j \sim \mathcal{B}(p_{T_j}),$$

with unknown probability of success p_{T_j}. The outcome of an individual in the control group is also assumed to be Bernoulli distributed

$$C_j \sim \mathcal{B}(p_{C_j}),$$

with unknown probability of success p_{C_j}. We consider the individual outcome in the treatment group as in the control group of all studies as independent distributed random variables. Then the number of success S_{T_j} in the treatment group of study j is binomial distributed,

$$S_{T_j} \sim \mathcal{B}(n_{T_j}, p_{T_j}, s_{T_j}),$$

with known number of individuals n_{T_j} in the treatment group and realization s_{T_j} of S_{T_j} of study j. Analogously, the number of success S_{C_j} in the control group of

	Success	Failure	
Treatment group	s_{T_j}	f_{T_j}	n_{T_j}
Control group	s_{C_j}	f_{C_j}	n_{C_j}
	s_j	f_j	n_j

Table 3.2.: 2×2-table of binary patients' outcome of study j

study j is also binomial distributed,

$$S_{C_j} \sim \mathcal{B}(n_{C_j}, p_{C_j}, s_{C_j}),$$

with known number of individuals in the control group n_{C_j} and realization s_{C_j} of S_{C_j}. These results can be summarized in the so-called classic 2×2-table [5][77]. In this table all individuals in the treatment group and the control group of a study are subdivided into a success and a failure group, as the outcome of an individual is only one of these two classifications. In Table 3.2 this 2×2-table is presented. The number of individuals in the treatment group who experienced a success is described by s_{T_j} and the number of individuals who experienced a failure is described by f_{T_j}. We have an analogue notation for the control group. The number of all individuals of study j is denoted by n_j. To get an estimate for the unknown probabilities p_{T_j} and p_{C_j}, it is common to use the maximum-likelihood estimation (MLE) [5].

Theorem 3.2.1. *Let Table 3.2 be the 2×2-table of the binary patients' outcome of study j. Then*

$$\tilde{P}_{T_j} := \frac{S_{T_j}}{n_{T_j}}$$

is the maximum likelihood estimator for the probability of success in the treatment group of study j and

$$\tilde{p}_{T_j} := \frac{s_{T_j}}{n_{T_j}}$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$.

Proof. To apply the MLE we use the presented 2×2-table [5]. The likelihood

function for the estimator $\tilde{P}_{T_j}$ of p_{T_j} is

$$L(s_{T_j}, p_{T_j}) = \binom{n_{T_j}}{s_{T_j}} p_{T_j}(1 - p_{T_j})^{(n_{T_j} - s_{T_j})}.$$

With the maximization of this function we get the maximum-likelihood estimate

$$\tilde{p}_{T_j} = \frac{s_{T_j}}{n_{T_j}}.$$

□

Theorem 3.2.2. *Let Table 3.2 be the* 2×2*-table of the binary patients' outcome of study* j*. Then*

$$\tilde{P}_{C_j} := \frac{S_{C_j}}{n_{C_j}}$$

is the maximum likelihood estimator for the probability of success in the control group of study j *and*

$$\tilde{p}_{C_j} := \frac{s_{C_j}}{n_{C_j}}$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$*.*

Proof. The likelihood function for the estimator $\tilde{P}_{C_j}$ of p_{C_j} is analogously

$$L(s_{C_j}, p_{C_j}) = \binom{n_{C_j}}{s_{C_j}} p_{C_j}(1 - p_{C_j})^{n_{C_j} - s_{C_j}}$$

and with the maximization of this function we get

$$\tilde{p}_{C_j} = \frac{s_{C_j}}{n_{C_j}}$$

for the maximum-likelihood estimate.

□

After calculating and before analyzing and aggregating the treatment effects of the single trials to one summary treatment effect, the values of the discussed ratio effect measures below, the Risk Ratio and the Odds Ratio, are usually transformed by applying the natural logarithm to the original value [5][30]. The untransformed effect measures have three common characteristics:

1. 0 is the lowest value that occurs

2. Value 1 is referred to no treatment effect

3. The highest value is infinity

This means that the number scale is not symmetric. An example of this asymmetric scale is given in [30]: Whilst an Odds Ratio of 0.5 (a halving) and an Odds Ratio of 2 (a doubling) are opposites, such that they should average to no effect, the average of 0.5 and 2 is not an Odds Ratio of 1 but an Odds Ratio of 1.25. With the transformation of the original value, by applying the natural logarithm, the scale becomes symmetric. The natural logarithm of 0 is not defined but

$$\lim_{x \to 0} \ln(x) = -\infty.$$

Furthermore, it applies that

$$\ln(1) = 0$$

and

$$\lim_{x \to \infty} \ln x = \infty.$$

In the example for the Odds Ratio of 0.5 the natural logarithm is $\ln(0.5) = -0.69$ and for the Odds Ratio of 2 holds that $\ln(2) = 0.69$. The average of -0.69 and 0.69 is 0 which is the transformed value of an Odds Ratio of 1 ($\exp(0) = 1$). This correctly implies no average treatment effect. The variance and the standard error are also calculated for the transformed estimated treatment effect to yield not only a summary treatment effect but also confidence limits and so on in logarithmic units. This has the effect of making e.g. the confidence intervals appear symmetric.

Risk Ratio

The first treatment effect estimate we want to discuss is the Risk Ratio. As stated in [5] the Risk Ratio is simply the ratio of two risks. To be more specific it is the ratio of the estimated probability of success in the treatment group $\tilde{p}_{T_j}$ and the estimated probability of success in the control group $\tilde{p}_{C_j}$. The Risk Ratio is therefore defined as follows [26][77].

Definition 3.2.3. *Let Table 3.2 be the* 2×2*-table of the binary patients' outcome of study* j*. Then*

$$\tilde{\Theta}_j := \ln\left(\frac{\tilde{P}_{T_j}}{\tilde{P}_{C_j}}\right)$$

is the estimator for the treatment effect of study j*. The estimation*

$$\mathrm{RR}_j := \frac{\tilde{p}_{T_j}}{\tilde{p}_{C_j}}$$

is called the Risk Ratio and

$$\tilde{\theta}_j := \ln\left(\mathrm{RR}_j\right)$$

the transformed Risk Ratio of study j*, for* $j = 1, \ldots, n_{st}$*.*

As stated in 3.2.1, before aggregating the values of the effect measure of all studies, they have to be transformed. Therefore, we calculate the estimated transformed treatment effect $\tilde{\theta}_j$ for each study. Since we assume that there is one true treatment effect for all patients, $\tilde{\theta}_j$ is the realization of the approximately normally distributed random variable

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \mathrm{var}(\tilde{\Theta}_j)),$$

with unknown expected value θ, the true treatment effect, and variance $\mathrm{var}(\tilde{\Theta}_j)$, which can be asymptotically estimated due to the delta method explained in [41][77]. A detailed derivation for the distribution of the estimated treatment effect can be found in Section 3.3.

Theorem 3.2.4. *Let Table 3.2 be the* 2×2*-table of the binary patients' outcome of study* j*. Then*

$$\tilde{\Sigma}_j^2 = \frac{1}{S_{T_j}} - \frac{1}{n_{T_j}} + \frac{1}{S_{C_j}} - \frac{1}{n_{C_j}}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ *of study* j*, with the corresponding estimation*

$$\tilde{\sigma}_j^2 = \frac{1}{s_{T_j}} - \frac{1}{n_{T_j}} + \frac{1}{s_{C_j}} - \frac{1}{n_{C_j}}$$

and $\mathrm{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}_j^2$*, for* $j = 1, \ldots, n_{st}$*, derived by the delta method.*

Odds Ratio

As the Risk Ratio is the ratio of two risks, the Odds Ratio is logically the ratio of two odds [5]. It is the ratio of the odds of success in the treatment group and the odds of success in the control group. It is defined as follows [26][77].

Definition 3.2.5. *Let Table 3.2 be the 2×2-table of the binary patients' outcome of study j. Then*

$$\tilde{\Theta}_j := \ln\left(\frac{\tilde{P}_{T_j}(1-\tilde{P}_{C_j})}{\tilde{P}_{C_j}(1-\tilde{P}_{T_j})}\right)$$

is the estimator for the treatment effect of study j. The estimation

$$\mathrm{OR}_j := \frac{\tilde{p}_{T_j}(1-\tilde{p}_{C_j})}{\tilde{p}_{C_j}(1-\tilde{p}_{T_j})}$$

is called the Odds Ratio and

$$\tilde{\theta}_j := \ln(\mathrm{OR}_j)$$

the transformed Odds Ratio of study j, for $j = 1, \ldots, n_{st}$.

Like the Risk Ratio, the Odds Ratio is calculated for each study and has to be transformed to $\tilde{\theta}_j$ before interpreting the results. Since $\tilde{\theta}_j$ is the realization of the approximately normally distributed variable

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \mathrm{var}(\tilde{\Theta}_j)),$$

the variance $\mathrm{var}(\tilde{\Theta}_j)$ of the treatment effect $\tilde{\Theta}_j$ can be estimated due the delta method [41]. This is shown in the next theorem [77][26].

Theorem 3.2.6. *Let Table 3.2 be the 2×2-table of the binary patients' outcome of study j. Then*

$$\tilde{\Sigma}_j^2 = \frac{1}{S_{T_j}} + \frac{1}{f_{T_j}} + \frac{1}{S_{C_j}} + \frac{1}{f_{C_j}}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ of study j, with the corresponding estimation

$$\tilde{\sigma}_j^2 = \frac{1}{s_{T_j}} + \frac{1}{f_{T_j}} + \frac{1}{s_{C_j}} + \frac{1}{f_{C_j}}$$

and $\text{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}_j^2$*, for* $j = 1, \ldots, n_{st}$*, derived by the delta method.*

Risk Difference

Last but not least, we want to introduce the Risk Difference which is also an effect measure for binary outcome data. Unlike the ratio effect measures, the Risk Difference is, like the name implies, the difference between two risks [5]. In detail it is the difference between the estimated probability of success in the treatment group and the estimated probability of success in the control group. The Risk Difference is defined as follows [77].

Definition 3.2.7. *Let Table 3.2 be the* 2×2*-table of the binary patients' outcome of study* j*. Then*

$$\tilde{\Theta}_j := \tilde{P}_{T_j} - \tilde{P}_{C_j}$$

is the estimator for the treatment effect of study j*. The corresponding estimation*

$$\tilde{\theta}_j = \text{RD}_j := \tilde{p}_{T_j} - \tilde{p}_{C_j}$$

is called the Risk Difference of study j*, for* $j = 1, \ldots, n_{st}$*.*

With $\tilde{\theta}_j$ being the realization of the approximately normally distributed variable

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \text{var}(\tilde{\Theta}_j)),$$

the variance $\text{var}(\tilde{\Theta}_j)$ can be estimated as follows [41][77].

Theorem 3.2.8. *Let Table 3.2 be the* 2×2*-table of the binary patients' outcome of study* j*. Then*

$$\tilde{\Sigma}_j^2 = \frac{s_{T_j} f_{T_j}}{(n_{T_j})^3} + \frac{s_{C_j} f_{C_j}}{(n_{C_j})^3}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ *of study* j*, with the corresponding estimation*

$$\tilde{\sigma}_j^2 = \frac{s_{T_j} f_{T_j}}{(n_{T_j})^3} + \frac{s_{C_j} f_{C_j}}{(n_{C_j})^3}$$

and $\text{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}_j^2$*, for* $j = 1, \ldots, n_{st}$*, derived by the delta method.*

After calculating the realization $\tilde{\theta}_j$ and the estimate of $\text{var}(\tilde{\Theta}_j)$ of the treatment effect $\tilde{\Theta}_j$, a summary treatment effect is determined by aggregating the single transformed effect measure values of all studies. The different models for this aggregation will be discussed later in Section 3.3. In the end, the aggregated treatment effect can be re-transformed [5], when using the Risk Ratio or the Odds Ratio.

Choice of an effect measure

Like it is discussed in [5] and in [77], the selection of one of the introduced effect measures for meta-analysis depends on different factors. The transformed Odds Ratio is the preferred effect measure as the adherence of corresponding test statistics to their asymptotic normal or χ^2 distribution is closest [63]. A further advantage of the use of the transformed Odds Ratio over the use of the transformed Risk Ratio is that if the probability of failure $(1 - \tilde{p}_{T_j})$ in the treatment group respectively $(1-\tilde{p}_{C_j})$ in the control group is used instead of the probability of success $\tilde{p}_{T_j}$ respectively $\tilde{p}_{C_j}$, the resulting transformed Odds Ratio will be of opposite sign and equal magnitude. The resulting transformed Risk Ratio will also be of opposite sign but not of equal magnitude [77]. The big disadvantage of the use of the Risk Difference is that the resulting values are restricted to the interval $[-1, 1]$. Therefore, the on asymptotic theory based confidence interval can include points outside these limits [77]. Nevertheless an appropriate effect measure should be chosen for the selected trials depending on the underlying available data. To get a reliable result after aggregation, for each trial the same effect measure has to be taken.

3.2.2. Cardinal data

For the description of the effect measures for cardinal data, we assume that for study j, $j = 1, \ldots, n_{st}$, the observed outcome Y_{iT_j} of individual i, $i = 1, \ldots, n_{T_j}$, in the treatment group is approximately normally distributed,

$$Y_{iT_j} \sim \mathcal{N}(y_{T_j}, \sigma^2_{T_j}),$$

with realization y_{iT_j}, unknown expected value y_{T_j} and variance $\sigma^2_{T_j}$. The outcome Y_{iC_j} of individual i, $i = 1, \ldots, n_{C_j}$, in the control group is also assumed to be at

least approximately normally distributed,

$$Y_{iC_j} \sim \mathcal{N}(y_{C_j}, \sigma^2_{C_j}),$$

with realization y_{iC_j}, unknown expected value y_{C_j} and variance $\sigma^2_{C_j}$. Thereby, n_{T_j} is the number of patients in the treatment group and n_{C_j} the number of patients in the control group of study j. The effect measures for cardinal data we want to discuss are based on the mean patients' outcome. Therefore, we use the arithmetic mean outcome as estimator for the expected outcome in the treatment group and the control group of study j. Before we get to the definition of this unbiased estimator, we summarize the patients' outcome in the treatment group and the control group to the following sets.

Definition 3.2.9. *Let Y_{iT_j} be the outcome of patient i in the treatment group of study j with realization y_{iT_j} and let Y_{iC_j} be the outcome of patient i in the control group of study j with realization y_{iC_j}, then*

$$\mathcal{Y}_{T_j} := \{y_{iT_j}\}_{i=1}^{n_{T_j}}$$

is the set of patients in the treatment group and

$$\mathcal{Y}_{C_j} := \{y_{iC_j}\}_{i=1}^{n_{C_j}}$$

is the set of patients in the control group of study j, for $j = 1, \ldots, n_{st}$.

With this definition, we can specify the unbiased estimator for the expected outcome in the treatment and control group [5][77].

Theorem 3.2.10. *Let $\mathcal{Y}_{T_j}$ be the set of patients in the treatment group of study j, then*

$$\tilde{Y}_{T_j} := \frac{1}{n_{T_j}} \sum_{i=1}^{n_{T_j}} Y_{iT_j}$$

is the unbiased estimator for the expected outcome in the treatment group of study

j and

$$\tilde{y}_{T_j} := \frac{1}{n_{T_j}} \sum_{i=1}^{n_{T_j}} y_{iT_j}$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$.

Theorem 3.2.11. *Let* $\mathcal{Y}_{C_j}$ *be the set of patients in the control group of study* j*, then*

$$\tilde{Y}_{C_j} := \frac{1}{n_{C_j}} \sum_{i=1}^{n_{C_j}} Y_{iC_j}$$

is the unbiased estimator for the expected outcome in the control group of study j *and*

$$\tilde{y}_{C_j} := \frac{1}{n_{C_j}} \sum_{i=1}^{n_{C_j}} y_{iC_j}$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$.

The estimators of variance $\sigma^2_{T_j}$ and $\sigma^2_{C_j}$ of the patients' outcome in the treatment and control group of study j are given in the following theorems [5][77].

Theorem 3.2.12. *Let* $\mathcal{Y}_{T_j}$ *be the set of patients in the treatment group of study* j *and let* $\tilde{Y}_{T_j}$ *be the estimator for the expected outcome with estimation* $\tilde{y}_{T_j}$*, then*

$$\tilde{\Sigma}^2_{T_j} := \frac{1}{n_{T_j} - 1} \sum_{i=1}^{n_{T_j}} (Y_{iT_j} - \tilde{Y}_{T_j})^2$$

is the unbiased estimator for the variance of the outcome in the treatment group of study j *and*

$$\tilde{\sigma}^2_{T_j} := \frac{1}{n_{T_j} - 1} \sum_{i=1}^{n_{T_j}} (y_{iT_j} - \tilde{y}_{T_j})^2$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$.

Theorem 3.2.13. *Let* $\mathcal{Y}_{C_j}$ *be the set of patients in the control group of study* j *and let* $\tilde{Y}_{C_j}$ *be the estimator for the expected outcome with estimation* $\tilde{y}_{C_j}$*, then*

$$\tilde{\Sigma}^2_{C_j} := \frac{1}{n_{C_j} - 1} \sum_{i=1}^{n_{C_j}} (Y_{iC_j} - \tilde{Y}_{C_j})^2$$

is the unbiased estimator for the variance of the outcome in the control group of study j and

$$\tilde{\sigma}^2_{C_j} := \frac{1}{n_{C_j} - 1} \sum_{i=1}^{n_{C_j}} (y_{iC_j} - \tilde{y}_{C_j})^2$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$.

Absolute difference between means

The first treatment effect for cardinal data we want to discuss, is the absolute difference between the mean outcome of the treatment and the control group of study j, $j = 1, \ldots, n_{st}$. This difference is specified in the following definition [77].

Definition 3.2.14. *Let $\tilde{Y}_{T_j}$ be the estimator for the expected outcome in the treatment group with estimation $\tilde{y}_{T_j}$ and let $\tilde{Y}_{C_j}$ be the estimator for the expected outcome in the control group with estimation $\tilde{y}_{C_j}$ of study j. Then*

$$\tilde{\Theta}_j := \tilde{Y}_{T_j} - \tilde{Y}_{C_j}$$

is the estimator for the treatment effect of study j. The corresponding estimation

$$\tilde{\theta}_j := \tilde{y}_{T_j} - \tilde{y}_{C_j}$$

is called the absolute difference between the mean outcomes in the treatment and control group of study j, for $j = 1, \ldots, n_{st}$.

The effect measure value $\tilde{\theta}_j$ is the realization of the approximately normally distributed random variable

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \mathrm{var}(\tilde{\Theta}_j)),$$

with unknown true treatment effect θ and variance $\mathrm{var}(\tilde{\Theta}_j)$. The estimation of $\mathrm{var}(\tilde{\Theta}_j)$ is given in the next remark [77].

Remark 3.2.15. *Let $\tilde{\Sigma}^2_{T_j}$ be the estimator for the variance of the outcome in the treatment group with realizations $\tilde{\sigma}^2_{T_j}$. Furthermore let $\tilde{\Sigma}^2_{C_j}$ be the estimator for*

the variance of the outcome in the control group with realizations $\tilde{\sigma}^2_{C_j}$ of study j. Then

$$\tilde{\Sigma}^2_j := \frac{\tilde{\Sigma}^2_{T_j}}{n_{T_j}} + \frac{\tilde{\Sigma}^2_{C_j}}{n_{C_j}}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ of study j,

$$\tilde{\sigma}^2_j := \frac{\tilde{\sigma}^2_{T_j}}{n_{T_j}} + \frac{\tilde{\sigma}^2_{C_j}}{n_{C_j}}$$

is the corresponding estimation and $\mathrm{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}^2_j$, for $j = 1, \ldots, n_{st}$.

If we assume, that the variance $\sigma^2_{T_j}$ of the outcome Y_{iT_j} of individuals in the treatment group and the variance $\sigma^2_{C_j}$ of the outcome of individuals in the control group are the same,

$$\sigma^2_{T_j} = \sigma^2_{C_j} =: \sigma^2_{P_j},$$

like it is assumed in most of the parametric data analysis techniques, then for the variance of the estimated treatment effect $\tilde{\Theta}_j$ holds

$$\mathrm{var}(\tilde{\Theta}_j) = \sigma^2_{P_j} \left(\frac{n_{T_j} + n_{C_j}}{n_{T_j} n_{C_j}} \right)$$

[5]. But before we get to this definition, we need an estimate for the variance $\sigma^2_{P_j}$ which is specified in the following remark [77].

Remark 3.2.16. *Let $\tilde{\Sigma}^2_{T_j}$ be the estimator for the variance of the outcome in the treatment group with realization $\tilde{\sigma}^2_{T_j}$. Furthermore let $\tilde{\Sigma}^2_{C_j}$ be the estimator for the variance of the outcome in the control group with realization $\tilde{\sigma}^2_{C_j}$ of study j. Then*

$$\tilde{\Sigma}^2_{P_j} := \frac{(n_{T_j} - 1)\tilde{\Sigma}^2_{T_j} + (n_{C_j} - 1)\tilde{\Sigma}^2_{C_j}}{n_{T_j} + n_{C_j} - 2},$$

is the estimator for the variance of the pooled outcome in the treatment and control group and

$$\tilde{\sigma}^2_{P_j} := \frac{(n_{T_j} - 1)\tilde{\sigma}^2_{T_j} + (n_{C_j} - 1)\tilde{\sigma}^2_{C_j}}{n_{T_j} + n_{C_j} - 2},$$

the corresponding estimation, for $j = 1, \ldots, n_{st}$.

Now the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ can

be formulated [77].

Remark 3.2.17. *Let $\tilde{\Sigma}^2_{P_j}$ be the estimator for the variance of the pooled outcome in the treatment and control group with realization $\tilde{\sigma}^2_{P_j}$. Then*

$$\tilde{\Sigma}^2_j := \tilde{\Sigma}^2_{P_j}\left(\frac{n_{T_j}+n_{C_j}}{n_{T_j}n_{C_j}}\right)$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ of study j,

$$\tilde{\sigma}^2_j := \tilde{\sigma}^2_{P_j}\left(\frac{n_{T_j}+n_{C_j}}{n_{T_j}n_{C_j}}\right)$$

is the corresponding estimation and $\mathrm{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}^2_j$, for $j = 1, \ldots, n_{st}$.

Standardized difference between means

The second effect measure for cardinal data we want to introduce, is the standardized difference between the mean outcome of the treatment and the control group of study j, $j = 1, \ldots, n_{st}$. It is used in meta-analysis when the studies all assess the equal outcome but measure it in a variety of ways, e.g. all studies measure depression but they use different psychometric scales. In this circumstance, it is necessary to standardize the results of the single studies to a uniform scale before they can be aggregated [30]. This standardized difference is defined as follows [26][77].

Definition 3.2.18. *Let $\tilde{Y}_{T_j}$ be the estimator for the expected outcome in the treatment group with estimation $\tilde{y}_{T_j}$ and let $\tilde{Y}_{C_j}$ be the estimator for the expected outcome in the control group with estimation $\tilde{y}_{C_j}$ of study j. Furthermore, let $\tilde{\Sigma}^2_{P_j}$ be the estimated variance of the pooled outcome with estimation $\tilde{\sigma}^2_{P_j}$ of study j. Then*

$$\tilde{\Theta}_j := \frac{\tilde{Y}_{T_j} - \tilde{Y}_{C_j}}{\tilde{\Sigma}_{P_j}}$$

is the estimator for the treatment effect of study j. The corresponding estimation

$$\tilde{\theta}_j := \frac{\tilde{y}_{T_j} - \tilde{y}_{C_j}}{\tilde{\sigma}_{P_j}}$$

is called the standardized difference between the mean outcomes in the treatment and control group of study j, for $j = 1, \ldots, n_{st}$.

As for the absolute difference, the estimated treatment effect $\tilde{\theta}_j$ is the realization of the approximately normally distributed random variable

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \mathrm{var}(\tilde{\Theta}_j)),$$

with unknown true treatment effect θ and variance $\mathrm{var}(\tilde{\Theta}_j)$. The variance of $\tilde{\Theta}_j$ can be approximated as follows [77].

Remark 3.2.19. *Let n_j be the number of patients of study j. Furthermore let n_{T_j} be the number of patients in the treatment group and n_{C_j} the number of patients in the control group of study j. Then*

$$\tilde{\Sigma}_j^2 := \frac{n_j}{n_{T_j} n_{C_j}}$$

is the approximated variance of the estimated treatment effect $\tilde{\Theta}_j$,

$$\tilde{\sigma}_j^2 := \frac{n_j}{n_{T_j} n_{C_j}}$$

is the corresponding estimation and $\mathrm{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}_j^2$, for $j = 1, \ldots, n_{st}$.

For each discussed effect measure holds that the estimated treatment effect $\tilde{\theta}_j$ is the realization of the approximately normally distributed random variable

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \tilde{\sigma}_j^2),$$

with unknown true treatment effect θ and variance $\mathrm{var}(\tilde{\Theta}_j) \approx \tilde{\sigma}_j^2$. Since

$$\frac{\tilde{\Theta}_j - \theta}{\tilde{\sigma}_j}$$

is then standardized normally distributed, for the $(1-\alpha)$ confidence interval $\mathcal{I}_j$ of the true treatment effect θ of study j, for $j = 1, \ldots, n_{st}$, follows

$$\mathcal{I}_j = [\tilde{\theta}_j - z[1 - \frac{\alpha}{2}]\tilde{\sigma}_j, \tilde{\theta}_j + z[1 - \frac{\alpha}{2}]\tilde{\sigma}_j],$$

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

3.3. Summary treatment effect estimate

In this section, we will give an introduction to the most common models for calculating the estimated summary treatment effect, the fixed-effects model and the random-effects model. Those two model will be introduced like it is done in [77] and [5]. In both models the estimation of the true treatment effect of an intervention is based on a weighted aggregation of the estimated treatment effects of the studies considered for the meta-analysis. In this thesis, we also call the true treatment effect of an intervention the summary treatment effect due to the method of obtainment. But before we get to the introduction of those two models, we have to make some assumptions. First of all, the selected clinical trials have the same study design, comparing a treatment group with a control group. We also assume that the same individual outcome is recorded in each selected clinical trial. And last but not least, we assume that the same parametrization of the treatment effect and the same method of estimating this treatment effect is used in each study [77]. As stated above, these assumptions might be violated and the results of the meta-analysis might be unreliable, especially if the used treatment effect estimates of the studies are taken from summary statistics. Such unreliable results might lead to wrong decisions when evaluating medical interventions. Beside the violation of the general assumptions there might be differences in the compilation of each study. This leads to methodological heterogeneity across trials. This problem occurs especially if a meta-analysis is undertaken retrospectively. In comparison to prospectively planned multi-center studies, retrospectively conducted meta-analysis haven't followed a common protocol for e.g. the selection of the population of a study. Nevertheless, if heterogeneity is derived by study estimates or if there is a strong presumption that heterogeneity is existent, the reasons have to be analyzed. Due to the importance of heterogeneity in meta-analysis, we will discuss this topic more precisely in Chapter 4.

3.3.1. Fixed-effects model

The first model we want to discuss is the fixed-effects model. At first, we want to take a look at the estimation of the true fixed treatment effect θ. We assume that for the conduction of a meta-analysis, n_{st} relevant studies have been collected with the same study design. In each study the treatment effect of an intervention is measured by comparing the treatment group with a control group and for these groups the same individual outcome is reported and the same effect measure is estimated, e.g. the transformed Odds Ratio for binary individual outcome data.

Let $\tilde{\Theta}_j$ be an independent approximately normally distributed estimated treatment effect of study j,

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \sigma_j^2),$$

for $j = 1, \ldots, n_{st}$, with correspondent realization $\tilde{\theta}_j$, unknown expected value θ, the true fixed treatment effect, and variance σ_j^2. The fixed-effect model is given by

$$\tilde{\theta}_j := \theta + \epsilon_j,$$

where ϵ_j is the estimation error of study j. The estimation error is a realization of a normally distributed random variable

$$\mathcal{E}_j \sim \mathcal{N}(0, \sigma_j^2),$$

with expected value 0. The variance is denoted by σ_j^2 and is also called intra-study variance. In the fixed-effects model the estimated variance of the estimated treatment effect $\tilde{\Theta}_j$ of study j is treated as if it were the true variance σ_j^2 [77]. From this it follows that the estimated treatment effect $\tilde{\Theta}_j$, for $j = 1, \ldots, n_{st}$, is normally distributed

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \tilde{\sigma}_j^2),$$

with expected value θ, the true treatment effect, and variance $\tilde{\sigma}_j^2$ [77].

To get an estimate of the true fixed treatment effect θ, we need to use an estimator $\hat{\Theta}$ determined by the weighted mean of the existent treatment effects of all included studies, $\tilde{\Theta}_1, \ldots, \tilde{\Theta}_n$. To do so it is common to apply the inverse variance method. The idea behind this method is, in comparison to the calculation of

a simple mean value, that the weight w_j given to study j is the inverse of the variance of the estimated treatment effect $\tilde{\Theta}_j$,

$$w_j = \frac{1}{\tilde{\sigma}_j^2},$$

for $j = 1, \ldots, n_{st}$. For the calculation of the estimated summary treatment effect $\hat{\theta}$ follows

$$\hat{\theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\theta}_j w_j}{\sum_{i=j}^{n_{st}} w_j},$$

as a realization of the unbiased estimator

$$\hat{\Theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\Theta}_j w_j}{\sum_{j=1}^{n_{st}} w_j}.$$

Since $\tilde{\sigma}_j^2$ is treated as the true variance, $\hat{\Theta}$ is the maximum likelihood estimator of θ. With the use of this inverse variance method, larger studies, which have smaller variance by means of a more precise treatment effect, are given more weight than smaller studies which have larger standard errors [30]. With this method of weighting, the variance of the estimated summary treatment effect

$$\mathrm{var}(\hat{\Theta}) = \mathrm{var}\left(\frac{\sum_{j=1}^{n_{st}} \tilde{\Theta}_j w_j}{\sum_{j=1}^{n_{st}} w_j}\right) = \frac{1}{\sum_{j=1}^{n_{st}} w_j}$$

is minimized. Since

$$\frac{\hat{\Theta} - \theta}{\sqrt{\mathrm{var}(\hat{\Theta})}}$$

is standard normally distributed, for the $(1-\alpha)$ confidence interval $\mathcal{I}$ of the true treatment effect θ follows

$$\mathcal{I} = [\hat{\theta} - z[1 - \frac{\alpha}{2}]\sqrt{\mathrm{var}(\hat{\Theta})}, \hat{\theta} + z[1 - \frac{\alpha}{2}]\sqrt{\mathrm{var}(\hat{\Theta})}],$$

where $z[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution. From this it follows that $P(\theta \in \mathcal{I}) = 1 - \alpha$.

3.3.2. Random-effects model

If heterogeneity across studies can not be explained or if data for the analysis of heterogeneity is not available, however, it is possible to consider the variation in the study estimates by using the random-effects model. As in the fixed-effects model, we have to assume that the n_{st} relevant studies considered for the meta-analysis have the same study design, that the treatment effect is measured by comparing the treatment group with a control group and that in each study the same individual outcome is reported and the same treatment effect is estimated.

In the fixed-effects model we assume that there is one true fixed treatment effect θ for all studies and that the variation of the estimated treatment effect $\tilde{\theta}_j$ across studies is due to an estimation error ϵ_j. In contrast to this model, the random-effects model allows the variation of the true treatment effect across studies. For example, the treatment effect might be higher (or lower) in studies where the participants are older, more educated, or healthier than in others [5]. Thus, it is assumed that the treatment effect θ_j of study j, for $j = 1, \ldots, n_{st}$, is a realization of the normally distributed random variable

$$\Theta_j \sim \mathcal{N}(\theta, \tau^2),$$

with expected treatment effect θ and variance τ^2. It holds

$$\theta_j = \theta + \nu_j,$$

with the study-specific random-effect ν_j of study j. ν_j is the realization of a normally distributed random variable

$$N \sim \mathcal{N}(0, \tau^2),$$

with expected random-effect 0 and the so-called between-study or inter-study variance τ^2 [5]. Like for the fixed-effects approach, let $\tilde{\Theta}_j$ be an independent normally distributed treatment effect observation

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta_j, \sigma_j^2),$$

with correspondent unknown expected value θ_j and variance σ_j^2, for $j = 1, \ldots, n_{st}$. Then, the observed treatment effect of study j can be described by

$$\tilde{\theta}_j = \theta_j + \epsilon_j,$$

where ϵ_j is the estimation error and a realization of a normally distributed random variable

$$\mathcal{E}_j \sim \mathcal{N}(0, \sigma_j^2),$$

with expected estimation error 0 and variance σ_j^2. The random-effects model is then given by

$$\tilde{\theta}_j := \theta_j + \epsilon_j = \theta + \nu_j + \epsilon_j,$$

for $j = 1, \ldots, n_{st}$, where ϵ_j is the estimation error and ν_j is the random-effect of study j. With $\mathcal{E}_j$ and N being independent distributed it follows that

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, \sigma_j^2 + \tau^2),$$

where τ^2 is unknown and has to be estimated from the underlying data. Thus, the distributional assumption is that

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, w_j^{-1} + \hat{\tau}^2),$$

where $w_j^{-1} + \hat{\tau}^2$ is treated as the true variance of $\tilde{\Theta}_j$. $\hat{\tau}^2$ is the estimate of τ^2 and $w_j^{-1} = \tilde{\sigma}_j^2$ is the observed variance. The estimation of τ^2 is described in the next section. With

$$w_j^* = \frac{1}{w_j^{-1} + \hat{\tau}^2}$$

it follows that

$$\tilde{\Theta}_j \sim \mathcal{N}(\theta, (w_j^*)^{-1}).$$

If $(w_j^*)^{-1}$ is treated to be the true variance of $\tilde{\Theta}_j$, then the asymptotically unbiased maximum likelihood estimator of the true treatment effect is given by

$$\hat{\Theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\Theta}_j w_j^*}{\sum_{j=1}^{n_{st}} w_j^*},$$

with realization

$$\hat{\theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\theta}_j w_j^*}{\sum_{j=1}^{n_{st}} w_j^*}.$$

With this approach of estimating the expected true treatment effect θ, the approximated variance of the estimated summary treatment effect is

$$\text{var}(\hat{\Theta}) = \text{var}\left(\frac{\sum_{j=1}^{n_{st}} \tilde{\Theta}_j w_j^*}{\sum_{j=1}^{n_{st}} w_j^*}\right) = \frac{1}{\sum_{j=1}^{n_{st}} w_j^*}.$$

Since

$$\frac{\hat{\Theta} - \theta}{\sqrt{\text{var}(\hat{\Theta})}}$$

is standardized normally distributed, for the $(1-\alpha)$ confidence interval $\mathcal{I}$ of the expected true treatment effect θ follows

$$\mathcal{I} = [\hat{\theta} - z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\hat{\Theta})}, \hat{\theta} + z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\hat{\Theta})}],$$

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

The weight w_j^* will be close to the fixed-effects weight w_j if τ^2 is small. If so, it follows that the estimate of the true effect of the random-effects model is similar to the estimate of the fixed-effects model and thus, the standard errors and the confidence intervals are similar in both models. In the other case, if τ is large, the estimate of θ of the random-effects model gets closer to the mean value and the standard error and the confidence interval will be larger.

Estimation of τ^2

As stated in [77], one possible approach of estimating τ^2 is based on the method of moments. The interested reader might refer to [77] for a detailed description of this method. In the random-effects model, the fixed-effect estimate of the true treatment effect is given by

$$\hat{\theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\theta}_j w_j}{\sum_{j=1}^{n_{st}} w_j}.$$

This is an unbiased estimate, but its variance is calculated as

$$\mathrm{var}(\hat{\Theta}) = \mathrm{var}\left(\frac{\sum_{j=1}^{n_{st}} \tilde{\Theta}_j w_j}{\sum_{j=1}^{n_{st}} w_j}\right) = \frac{1}{\sum_{j=1}^{n_{st}} w_j} + \frac{\tau^2 \sum_{j=1}^{n_{st}} (w_j)^2}{(\sum_{j=1}^{n_{st}} w_j)^2}.$$

For the estimation of τ^2, we will apply the so-called Q-statistic which will be introduced in the next chapter. The realization is defined by

$$q := \sum_{j=1}^{n_{st}} w_j (\tilde{\theta}_j - \hat{\theta})^2 = \sum_{j=1}^{n_{st}} w_j (\tilde{\theta}_j - \theta)^2 - \left(\sum_{j=1}^{n_{st}} w_j\right) (\hat{\theta} - \theta)^2.$$

For now we will take this definition without further explanation. With the calculated $\mathrm{var}(\hat{\Theta})$, for the expected value of the Q-statistic follows

$$\begin{aligned} \mathrm{E}(Q) &= \sum_{j=1}^{n_{st}} w_j \,\mathrm{var}(\tilde{\Theta}_j) - \left(\sum_{j=1}^{n_{st}} w_j\right) \mathrm{var}(\hat{\Theta}) \\ &= (1 - n_{st}) + \tau^2 \left(\sum_{j=1}^{n_{st}} w_j - \frac{\sum_{j=1}^{n_{st}} w_j^2}{\sum_{j=1}^{n_{st}} w_j}\right). \end{aligned}$$

As described in [23], [24] and [59], with the application of the method of moments, we get the estimation $\hat{\tau}^2$ for the inter-study variance τ^2,

$$\hat{\tau}^2 = \max\left\{0; \frac{q - (n_{st} - 1)}{\sum_{j=1}^{n_{st}} w_j - \frac{\sum_{j=1}^{n_{st}} w_j^2}{\sum_{j=1}^{n_{st}} w_j}}\right\}.$$

Since the deviated quotient could be a negative value for the estimation of τ^2, we need to get the maximum of zero and the given value. $\hat{\tau}^2 = 0$ leads to the fixed-effects model [77].

The Q-statistic, as well as the inter-study variance, are very important measures for dealing with heterogeneity across trials and will be discussed in detail in the framework of heterogeneity in Chapter 4.

3.3.3. Comparison of both approaches

When conducting the fixed-effects model approach, we assume that there is one true fixed treatment effect θ in all included studies, thus, for all participating patients. The only reason for the variation of the treatment effect is a sampling error in the corresponding clinical trial. Then the summary treatment effect is our estimate of this common true effect. Therefore, we choose the fixed-effects model if we believe that all the considered studies for the meta-analysis are expected to share a common treatment effect.

In the random-effects model, we assume that there is a true treatment effect for each study as a realization of the normally distributed random variable

$$\Theta_j \sim \mathcal{N}(\theta, \mathrm{var}(\Theta_j)),$$

with expected true effect θ which varies across the included studies. The summary treatment effect is our estimate of the expected value θ. The study weights used for the determination of the summary treatment effect are more balanced under the random-effects model than under the fixed-effects model. Large studies are assigned less and small studies more relative weight. Another consequence is, that the standard error of the summary treatment effect and the confidence intervals are wider under the random-effects model than under the fixed-effects model. Thus, if we assume that there is no common treatment effect due to some random-effect across the single trials or if the single treatment effects are gathered from published literature, it is more justified to use the random-effects model.

3.4. Systematic reviews

Finally, meta-analyses are published as a key component in systematic reviews (SRs). A SR is commonly prepared to answer a defined research question and to reduce bias by identifying, collecting and summarizing all relevant studies that fits the pre-defined criteria. SRs themselves are not only publicized in academic forums but are also promoted and disseminated by organizations and databases specifically developed for this purpose. E.g. the Cochrane Collabora-

tion (www.cochrane.org) is a widely recognized and respected international and not-for-profit organization that promotes, supports, and disseminates SRs and meta-analyses on the efficacy of interventions in the health care field [74]. And as pointed out in Chapter 2, meta-analyses and therefore SRs play an important role in the health economic evaluation of medical interventions conducted by the IQWiG.

4. Heterogeneity

Like it is discussed in Section 3.3, it is common to distinguish between the variability within studies (intra-study variance) and the variability across studies (inter-study variance). In general, these variabilities or heterogeneity might be due to differences in the study design and the conduction, as well as in the participants, the interventions or the reported outcomes [31][72]. Like it is stated in [30], it can be helpful to distinguish between different types of heterogeneity. The variability among the participants, the interventions and the reported outcome may be described as clinical heterogeneity and the variability in the study design and conduction as well as the risk of bias may be described as methodological heterogeneity [27]. Variability in the treatment effects, being estimated in the different studies, is known as statistical heterogeneity, which is a consequence of clinical or methodological heterogeneity, or both [27][44]. Statistical heterogeneity manifests itself in the observed treatment effects of the considered studies, being more different from each other than one would expect due to random error alone [27]. We will follow conventions and refer to statistical heterogeneity simply as heterogeneity. In general, for the conduction of meta-analysis, it is desirable to have clinical homogeneity, by means of participating patients with similar combinations of their characteristic values which have an influence on the efficacy of the analyzed medical intervention [60]. Because if there is clinical homogeneity, the assumption that there is only one true treatment effect in all studies is intuitively plausible and the use of e.g. the fixed-effects model is reliable [27]. But unfortunately, the knowledge of the influencing characteristic values combinations is often not existent [53]. And since a certain number of clinical trials and a certain number of participants in the considered studies are needed to get high statistical power, the chance of clinical homogeneity is low and heterogeneity has to be assumed. Statistical power refers to the likelihood of detecting, within a sample, an effect or relationship that exists within the population. More formally stated,

'the power of a statistical test of a null hypothesis is the probability that it will lead to the rejection of the null hypothesis, i.e., the probability that it will result in the conclusion that the phenomenon exists' [22]. Finally, it can be stated that it is important to analyze, to quantify and to interpret the existing heterogeneity.

In Section 4.1, we will give an overview of the general causes of heterogeneity. Thereby, we will have a closer look at methodological heterogeneity, random heterogeneity and regional heterogeneity, which might be caused by sociodemographic, biographical and clinical parameters. Then, in Section 4.2, we will analyze how heterogeneity can be assessed. Thereby, we will discuss the Q-statistic and further related indices for the quantification of heterogeneity. On this basis, we will discuss the limitations of the discussed assessment methods. In Section 4.3, we will have a look on how heterogeneity can be considered in meta-analysis. Amongst others, in this context, we will give a short overview of subgroup analysis and meta-regression. A recommendation for the consideration of heterogeneity on the basis of the innovative geometric clustering approach is given in the last Section 4.4.

4.1. Causes of heterogeneity

Like it is stated in [44], there are different reasons for heterogeneity. Clinical heterogeneity, and especially the associated regional heterogeneity, can occur due to intrinsic and extrinsic factors, like a variety of regional influencing factors [21]. Due to scientific literature, examples of these factors have been presented which can influence results of clinical trials culturally specific or across regions [43][53], like the traditions of diagnosis and access to the health system, the therapy according to national guidelines, conventional medical pretreatment and inpatient treatments in different countries, effects in the placebo response to the respective treatment results and clinical data for the inclusion into studies [35][39][49][50][65][66][70]. Especially within the evaluation of European and US-American studies, parameters of Table 4.1 have been identified as influencing factors. E.g. there might be different realizations of relevant characteristics of the patient population in different studies, e.g. in one study there might be patients with a high BMI, in another study there might be more patients with a

Conduct clinical studies in general	
Intrinsic	Socio-demographic data, especially body height, weight and body mass index (BMI)
Extrinsic	Access to health care Facilities for inpatient / outpatient care Pretreatment, guidelines compliance Recruitment (advertising vs. allocation) Medical education
Conduct psychiatric studies	
Intrinsic	Psychiatric anamnesis and psychometric Duration of depressive episode Severity of disease Somatization (for depression) Co-morbidity
Extrinsic	Access to mental health care Stigmatization of mental disorders Assessment of co-morbidity

Table 4.1.: Intrinsic and extrinsic influencing factors of regional heterogeneity in treatment studies

low BMI. We expect different results in the estimated treatment effect of the two studies if the efficacy of the tested medication is dependent on the body mass index. Consequently, these factors have to be considered and may have an effect on the composition of the study population, the conduction of studies in the respective centers and the evaluation of endpoints [49][39][35].

Methodological heterogeneity can occur as a result of methodological factors. If for example the administration or dosage of a medication has an influence on the treatment effect and differs between two considered studies, we also expect different results in the estimated treatment effects [30]. Another reason for methodological heterogeneity might be the usage of different effect measures in the included studies [27]. Significant statistical heterogeneity arising from methodological diversity or differences in outcome assessments suggests that the studies are not all estimating the same quantity, but does not necessarily suggest that the true intervention effect varies [30].

A third reason for heterogeneity is the so-called random heterogeneity which is also assigned to methodological heterogeneity. This type of heterogeneity occurs by dividing the participants of a study into control group and treatment group randomly. Even if the distributions of two studies are identical, the same result can not be expected if there is a random classification into treatment group and control group, especially if the number of participants in the study is small [44].

4.2. Assessment of heterogeneity

Consequently, if there are distinctive country effects in multinational studies or if there is evidence of increased heterogeneity in meta-analyses, this has to be explained. There are different way for the assessment, like the evaluation of the Q-statistic with the help of the p-value [31][59]. This statistic has been already introduced in Section 3.3.2 and will be discussed in detail in the following. Related indices for the quantification of heterogeneity, like the l^2 index and the H^2-statistic, are presented in the ensuing section. For further assessment methods, please refer to [27][72] or [52]. In the end we will address the limitations of the discussed methods.

4.2.1. Statistical test for the identification of heterogeneity

In this section, we want to discuss how heterogeneity across trials can be identified. The most common ways for the identification is the use of the so-called Cochran's χ^2 test [68][69]. Since this test was also discussed by DerSimonian and Laird, it is also known as the DerSimonian and Laird Q-test [24] or only Q-test. The procedure is shown in Figure 4.1. Although this test is an instrument to check if the data are homogeneous, it is usually called test of heterogeneity [68]. Actually it tests the null hypothesis

$$H_0 : \theta_1 = \cdots = \theta_{n_{st}},$$

where θ_j is the true treatment effect of study j, for $j = 1, \ldots, n_{st}$ [68]. This implies that we assume the same treatment effect in all included studies and $\tau^2 = 0$ [38].

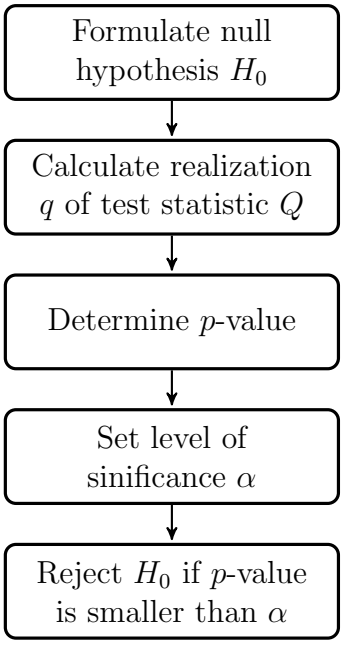

Figure 4.1.: Hypothesis test procedure for the identification of heterogeneity across trials

The alternative hypothesis, that at least one of the estimated treatment effects of the n_{st} studies differ from the others, is mathematically denoted by

$$H_1 : \exists k \in \{1, \ldots, n_{st}\} : \theta_k \neq \theta_j, \forall j \in \{1, \ldots, n_{st}\}, j \neq k.$$

[68] For the conduction of the Q-test we need the so-called Q-statistic, which has already been applied in Section 3.3.2 for the estimation of the inter-study variance. Thus, for the identification of heterogeneity in the treatment effects across studies, the realization

$$q = \sum_{j=1}^{n_{st}} w_j (\tilde{\theta}_j - \hat{\theta})^2$$

of the Q-statistic

$$Q = \sum_{j=1}^{n_{st}} w_j (\tilde{\Theta}_j - \hat{\Theta})^2$$

has to be calculated [31][59]. $\hat{\Theta}$ is the estimated summary treatment effect with realization $\hat{\theta}$, calculated on the basis of the fixed-effects model discussed in Section

3.3, and $\tilde{\Theta}_j$ with realization $\tilde{\theta}_j$ is the estimated treatment effect of study j. The weight is denoted by w_j and depends on the relied model [30]. Since

$$\hat{\theta} = \frac{\sum_{j=1}^{n_{st}} w_j \tilde{\theta}_j}{\sum_{j=1}^{n_{st}} w_j},$$

it is also conventional to formulate q as

$$q = \sum_{j=1}^{n_{st}} w_j \tilde{\theta}_j^2 - \frac{(\sum_{j=1}^{n_{st}} w_j \tilde{\theta}_j)^2}{\sum_{j=1}^{n_{st}} w_j}.$$

As it is stated in [68], Q approximately follows a χ^2 distribution with $(n_{st} - 1)$ degrees of freedom. The realization of the Q-statistic is commonly evaluated with the help of the p-value,

$$pv = P(Q \geq q|H_0) = 1 - F_{\chi^2_{(n_{st}-1)}}(q),$$

where $F_{\chi^2_{(n_{st}-1)}}$ is the cumulative function of the χ^2 distribution with $(n_{st} - 1)$ degrees of freedom. It represents the probability to get a result like the calculated q or a higher value due to chance under the assumption that all treatment effects are equal in the single studies [38]. Generally, a level of significance α is pre-defined. If the p-value is greater than α, one would suggest that the observed inter-study variance is plausibly due to chance and therefore, the null hypothesis cannot be safely rejected. Conversely, a p-value smaller than α indicates a small possibility that the observed inter-study variance is due to chance and therefore this indicates statistically significant heterogeneity across studies [62] and the null hypothesis can be rejected.

In fact, the Q-test only gives us the information about the presence or the absence of heterogeneity but it does not measure the extent of it. The reason therefore is the missing standardization and the resultant invariance with respect to the number n_{st} and the size of studies included in the meta-analysis. More specifically, a higher number of studies leads to a higher Q-statistic value [31].

4.2.2. Indices for the quantification of heterogeneity

In this section, we want to give an introduction how heterogeneity across trials can be quantified. With the following indices, the assessment of the extent of heterogeneity of a meta-analysis and the comparison of heterogeneity across meta-analyses are enabled.

An obvious instrument for the quantification of heterogeneity is the inter-study variance τ^2 [31] of the estimated treatment effect across trials. The estimation of the inter-study variance is done as a part of the random-effects model and has been discussed in Section 3.3.2. It quantifies the difference in the estimated treatment effects that cannot be explained by the intra-study variance alone [38][72]. The advantage of the inter-study variance is that its estimates are invariant due to either the number or size of studies in a meta-analysis. E.g. the estimates do not systematically increase with either the number or size of studies [31].

Another way to assess the extent of heterogeneity is the calculation of the Q-statistic-based l^2 index which has been invented by Higgins and Thompson [30][33]. The l^2 index is defined by

$$l^2 := \max\left\{0, \frac{q-(n_{st}-1)}{q}\right\} \cdot 100\%,$$

where q is the realization of the Q-statistic and $(n_{st}-1)$ is the degree of freedom [30][42]. This index is obviously a standardization of the Q-statistic and therefore, in contrast to Q, independent of the number n_{st} of studies in the meta-analysis. l^2 is interpreted as the percentage of variability in the treatment effect estimates which is attributable to heterogeneity between studies rather than to sampling errors [30]. According to [30], this index can be interpreted as it is listed in Table 4.2.

A further index for the quantification of heterogeneity is the H^2-statistic. It is derived by the Q-statistic and the realization is defined as

$$h^2 := \frac{q}{(n_{st}-1)},$$

Result	Interpretation heterogeneity
0% to 40%	might not be important
30% to 60%	may represent moderate heterogeneity
50% to 90%	may represent substantial heterogeneity
75% to 100%	considerable heterogeneity

Table 4.2.: Interpretation of the l^2 index

with $(n_{st} - 1)$ degrees of freedom [42]. Like it is stated in [31], H^2 describes the relative excess in Q over its degrees of freedom. The ratio of Q to its degrees of freedom has been suggested previously as a measure of the extent of heterogeneity in [31]. Since Q is χ^2 distributed it holds $E(Q) = n_{st} - 1$. In the absence of heterogeneity, $h^2 = 1$ indicates homogeneity of the treatment effects [31][56]. Study simulations showed that the value of H^2 does not intrinsically depend on the number of studies (unlike Q) and increases appropriately as τ^2 increases [31].

4.2.3. Limitations of tests and indices

Addressing heterogeneity represents one of the most troublesome aspects of meta-analyses [31]. Especially methods for testing statistical heterogeneity, like the Q-test have to be considered critically due to their low statistical power in case of a low number of studies included [42]. Unfortunately, this is the most common situations in meta-analysis. And if we do not have an appropriate number of studies, heterogeneity might exists even if the Q-statistic is not statistically significant. Due to this problem, the choice of an appropriate α for the evaluation of the p-value of the χ^2 test is very difficult. It is recommended to use a significance level of $\alpha = 0.10$ instead of the usually taken significance level of $\alpha = 0.05$ [68]. In fact, it is not considered that a not significant test of heterogeneity or a p-value higher than 0.10 in a meta-analysis with a little number of studies included, is either due to less statistical power of the Q-test or due to homogeneity of the result of the studies. Nevertheless, as pointed out in [62], it may be argued that heterogeneity should be analyzed no matter what p-value is observed. In addition the Q-test has excessive power to detect clinically unimportant heterogeneity if there is a high number of clinical trials involved [42]. Due to the common little number of studies included, this constellation is seldom.

Another factor which has an negative influence on the statistical power of the Q-test is the weight assigned to the studies. Simulation studies showed that the statistical power decreases if the sum of the study weights decreases or if the weights are unequally distributed, especially if one study has a high share in the sum of the weights [28]. Further simulation studies discovered that the statistical power of the Q-test is proportional to the inter-study variance and inversely proportional to the intra-study variance [42]. These results can also be mathematically proven and are discussed in [29][38].

But not only statistical tests for the identification of heterogeneity derive their limit. Also the indices for the quantification of the extent of heterogeneity show restrictions in their usage. The inter-study variance τ^2, for the quantification of heterogeneity across trials, does not enable the comparison of heterogeneity across meta-analyses of different types of outcomes, such as dichotomous and continuous outcome. Furthermore, as the inter-study variance depends on the chosen effect measure, the isolated interpretation of this estimation can be difficult [31].

With regard to H^2, there is a slight average bias for small numbers of studies (< 8), as it is stated in [31]. Variability in H^2 is large for small numbers of studies, so it will be difficult in practice to distinguish moderate heterogeneity from chance. The variability can be reduced by increasing the number of studies [31]. In fact, there is a mathematical relationship between the statistical test for heterogeneity (based on Q) and the value of H over varying numbers of studies. This relationship is shown in [31]. With a small number of studies, statistically significant heterogeneity would be evident only when the impact of heterogeneity, as measured by the H^2-statistic, is high. This explicitly highlights the poor properties of the test when there are few studies [31].

4.3. Consideration of heterogeneity in meta-analysis

We should always keep in mind that there are insuperable problems with regard to heterogeneity like it is discussed in Section 4.2.3. But there are several ways to consider heterogeneity in meta-analysis. In the following we want to present the most common approaches how to deal with heterogeneity in case of individual patient data [72].

4.3.1. Choice of an appropriate treatment effect measure

The first way to deal with heterogeneity is the choice of an appropriate measure for the estimation of the treatment effect in each study. Practical applications of meta-analysis showed that heterogeneity may be a consequence of an inappropriate choice of treatment effect measures. E.g. as described in [30], when studies collect continuous outcome data using different scales or different units, extreme heterogeneity may be apparent when using the mean difference. This problem could be solved when using the more appropriate standardized mean difference. Furthermore, the choice of an effect measure for dichotomous outcomes (Odds Ratio, Risk Ratio or Risk Difference) may affect the degree of heterogeneity among results. In particular, when control group risks vary, homogeneous Odds Ratios or Risk Ratios will necessarily lead to heterogeneous Risk Differences and vice versa. However, it remains unclear whether homogeneity of treatment effects in a particular meta-analysis is a suitable criterion for choosing between these measures.

4.3.2. Fixed-effects model vs random-effects model

Like it is discussed in Section 3.3.3, we choose the fixed-effects model if we believe that all the studies included in the meta-analysis are expected to share a common treatment effect and that there is no heterogeneity across the included studies ($\tau^2 = 0$). This could imply that the participants of all included studies are chosen due to similarity of their characteristic values combinations which might have an influence on the efficacy of the administered medical intervention. However, this is a relatively rare situation. If we assume that there is no common treatment

effect, it is more justified to use the random-effects model.

The Cochrane Eyes and Vision Group suggests to conduct no meta-analysis if the p-value of the Q-test for heterogeneity is less than 0.05. If the p-value is between 0.05 and 0.10 it is recommended to take the fixed-effects model and the random-effects model and if the p-value is higher than 0.10 they suggest to use only the fixed-effects model [32]. But the choice between a fixed-effects and a random-effects model should not be made only on the p-value of the Q-test. The number of trials and the distribution of the estimated treatment effects need to be considered additionally [77].

4.3.3. Subgroup analysis

A way to deal with clinical heterogeneity in RCTs, in case of individual patient data, is the use of subgroup analysis [72]. Also IQWiG uses this method for the evaluation of medical interventions. Its aim is the identification of either consistency of, or differences in the treatment effects among different characteristics of patients [78]. For the conduction of a subgroup analysis, the participating patients are slitted into subgroups within studies, with the intention to compare the treatment effects of those subgroups [78]. E.g. for the evaluation of the efficacy of an intervention it might be reasonable to separate the patients by gender if it is known that the investigated intervention has a different effect on men and women. Subgroup analyses may be also done for subsets of studies, such as different geographical locations or to answer specific questions about particular patient groups, types of interventions or types of studies [30]. In case of existent individual data, such subsets of participants are easily analyzed. It is possible to conduct meta-analyses within those subgroups or to conduct meta-analyses that combine several subgroups. The comparison of the treatment effects in the subgroups then enables the investigation of whether a medical intervention has different effects in different subgroups. If there are only two subgroups, it is possible to calculate the summary treatment effects and their confidence intervals in the two groups. Then the overlap of both confidence intervals has to be investigated. If there is no overlap, this indicates that the difference is statistical significance. It has to be considered that the difference may still be statistically significant if the confidence intervals only overlap to a small degree [30]. Another

simple approach is recommended by [5]. Here a standard test for heterogeneity is undertaken by considering the treatment effects of each subgroup instead of the treatment effects of the studies. For the description of the differences in the treatment effects in the different subgroups an l^2-statistic is calculated. For the use of this approach it has to be assumed that the subgroups are statistically independent. In particular this means that a participant should only be assigned to one subgroup [30].

A problem which arises when presenting the results of a subgroup analysis is that there might be patients not fitting into one of the defined subsets. This could lead to a hindrance of an effective intervention or a treatment with an ineffective or even harmful intervention [60]. Furthermore, in suitable, multivariate subgroup analyses, in general the respective collectives are too small to derive significant results [21]. The new invented endpoint-oriented clustering approach, introduced in Chapter 5, is based on the idea of the subgroup analysis. It considers this problem by identifying patient collectives with the help of an innovative clustering algorithm so that all patients can be assigned to one specific cluster.

4.3.4. Meta-regression

Another very common approach for addressing heterogeneity across the treatment effects of studies is the so-called meta-regression [72]. The conduction of a meta-regression is similar to a classical multi-linear regression where the relationship between a scalar outcome variable (dependent variable) and one or more explanatory variables (independent variables) is modeled. In meta-regression the normally distributed outcome variable is the estimated treatment effect of a study, e.g. the transformed Odds Ratio or the transformed Risk Ratio. Characteristics of studies that might influence the treatment effect define the explanatory variables. These must be at least ordered (binary, ordinal or cardinal) but not necessary normally distributed. In literature those characteristics are often called 'potential effect modifiers' or 'covariates' [30]. Meta-regression can be conducted with one or more than one covariates. The more common situation is the conduction with only one covariate due to rarely published data. Meta-regression can be used for the analysis of the impact of covariates on the treatment effect,

as well as for the prediction of treatment effects for studies with patients with specific characteristic values. The interested reader is referred to [64] for more information.

Even if appropriate statistical methods have been used for meta-regression, there are a number of limitations to the interpretation of the results. The most important problem is that there might be characteristics with a nominal level of scale, e.g. reason for withdrawal or the cultural background. Then the common meta-regression can not be applied. Due to a transformation technique used in the context of the new invented clustering algorithm, also nominal scaled random variables can be considered for the evaluation of medical interventions on patient collectives identified by the new invented cluster-based approach.

4.4. Recommendation for the consideration of heterogeneity

As stated in [71], meta-analysis is critically discussed regarding the heterogeneity between studies and the enthusiasm for meta-analysis can not be shared by the broader medical community [40][73]. Meta-analysis is not seen as an exact statistical science that give simple answers to complex clinical problems. It is rather seen as a valuable objective descriptive technique which often furnishes clear qualitative conclusions about broad treatment policies, but whose quantitative results have to be interpreted cautiously [71]. There are solid arguments that treatment effects, that are mildly beneficial to the average, may have differing effects on individuals [40]. However, not only due to this critical perception on meta-analysis it is important to find new approaches to solve the prevalent problem of heterogeneity between clinical trials. Therefore, we recommend an approach which is slightly based on the idea of subgroup analysis: the innovative endpoint-oriented clustering technology introduced in Chapter 5, that aims to identify sufficiently large and homogeneous patient collectives due to similar combination of the characteristic values, for the evaluation of medical interventions. The technique and the mathematical optimization model was originally developed for an agricultural economics application. A more detailed descrip-

tion of the application can be found for example in [14]. However, the method can be applied not only for the problem just mentioned, indeed it is a general and innovative technique, developed by Brieden and Gritzmann, to detect hidden structures in high-dimensional data. Studies on the structure, complexities or on efficient approximation of the problem can be found in [7][10][15][16][12][17][18]. For a summary have a look at [8] or [9].

5. Geometric clustering of patient data

In the health care sector, one of the most important tasks is to analyze the efficacy and safety of medical interventions with the help of the large amounts of patient data derived by clinical trials. Since the number of participating patients in a single trial is too low for a reliable evaluation of the efficacy of a medication, the common way is the application of meta-analytic techniques to evaluate larger collectives built out of the available data. The aggregation of the results of the single trials is then, roughly speaking, carried out by only calculating mean values out of the estimated treatment effects of the different studies, like it is explained in Chapter 3. Thus, only the patients' outcome is used for the analysis and all other patient data, collected in the framework of the conducted trials, which could provide important information for improved medical care, is not considered. Furthermore, this disregard could also lead to false decisions in terms of health economic evaluation of medical interventions. E.g. it might be dramatic if meta-analysis states that a new drug is not efficient in general and has to be taken off the market but for a patient collective, with specific characteristic values combinations, the drug achieves great progress in the healing process. However, the use of meta-analysis implies the assumption that the efficacy of a drug is independent of factors such as gender, age, BMI (body mass index), etc. As we have discussed in Chapter 4, clinical heterogeneity, which has to be considered when evaluating medical interventions, exists between the outcome of patients, especially due to different combinations of the patients' characteristic values. In a suitable, multivariate subgroup analysis, which considers the patients' characteristics and the related heterogeneity, in general the respective collectives are too small to derive significant results. Another problem which occurs is that patients might not fit to the analyzed subgroups. The innovative

endpoint-oriented geometric clustering approach and the new invented cluster based analyses, which will be presented in this chapter, are based on the idea of subgroup analysis and take the existent problems into account. With the assumption that the patients' characteristics do have an influence on the outcome, it is reasonable to recap the individual patient data of all trials and to classify homogeneous collectives among the included trials by means of similar combinations of the patients' characteristic values. Then on these sufficiently large and homogeneous patient collectives, identified by the endpoint-oriented clustering algorithm, the new introduced cluster-based approaches can be applied for the evaluation and prediction of the efficacy of a medical intervention and the identification of heterogeneity in the treatment effects to enable improved medical care for patients in terms of evidence-based medicine.

In Section 5.1 to Section 5.6 of this chapter, we introduce the theoretical principles of the endpoint-oriented geometric clustering approach. They represent the main ideas of the classification of heterogeneous patient data into homogeneous patient collectives and are based on the work of Brieden and Gritzmann, see for example [13], [14] and [15], and the work of Borgwardt [6] and Öllinger [48]. Basic definitions are given on the mathematical interpretation of a clustering and how it is related to polytopes. We show the connection between a vertex of a polytope and a feasible clustering which leads to a partition of the Euclidean space into convex polyhedral cells. This enables the application of the endpoint-oriented geometric clustering approach for the classification of patient data in homogeneous collectives. Additionally, the introduced data transformation technique allows the computation of a clustering for non-metric data by using the conditional probabilities. In Section 5.7, the geometric clustering approach is used as an unsupervised learning approach and composes the basis for the conduction of the new invented cluster-based meta-analysis. Therefore, we adopted the definitions and terms of the classic meta-analysis to the terms of cluster-based meta-analysis. In Section 5.8, we will have a look at a the unsupervised clustering algorithm as the basis for the cluster-based identification of heterogeneity of the patients' outcome with different characteristic values combinations. In the last Section 5.9 of this chapter, we examine the new clustering approach as a supervised learning approach for the cluster-based prediction of the efficacy of an

administered medication.

5.1. General terminology

We start by providing the necessary terminology for the conduction of an endpoint-oriented geometric clustering for the identification of sufficiently large and homogeneous patient collectives. Therefore, we need to consider the available individual patient data of all relevant trials as geometric objects. Each patient has to be expressed by the position of a point in the geometric space and the distance is measured by the Euclidean norm. We start by the description of a patient as a geometric object.

Definition 5.1.1. *Let $d \in \mathbb{N}$ be the number of relevant patient characteristics or attributes and for $l = 1 \ldots, d$ let A_l be a discrete random variable, the representative of the ith characteristic and let Ω_i be the correspondent sample spaces of the characteristic A_i with its characteristic values $\{a_i\}_{a_i \in \Omega_i}$.*

When thinking of patients' attributes, there might be random variables with non-metric sample spaces which makes it hard to consider a patient as a point in the geometric space $\mathbb{R}^d$. To handle this problem, we use the transformation technique explained in Section 5.5. With this transformation of the data, the patients with their d characteristic values can be treated as the mentioned data points and the number of the patient attributes d then represents the dimension of the underlying geometric space in which the geometric clustering can be conducted.

Definition 5.1.2. *The number of patients in all considered trials is denoted by $N \in \mathbb{N}$. The patient data set of characteristics is then defined by*

$$X^{all} := \{x_j\}_{j=1}^{N} \subset \mathbb{R}^d$$

with $x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$, for $j = 1, \ldots, N$ and $a_l \in \Omega_l$ for $l = 1, \ldots, d$.

For the estimation of the efficacy of a drug administered to a patient with specific characteristic values combinations, we also need to define the patient's outcome measured in the corresponding trial. Therefore, let Y be a normally or Bernoulli distributed random variable, which represents the outcome of patient j, and let Ω be the correspondent sample space of Y. With the consideration of Y, the d-dimensional space enlarges to the $(d+1)$-dimensional space $\mathbb{R}^d \times \Omega$ and the patient data set is defined as follows.

Definition 5.1.3. *Let X^{all} be the patient data set of characteristics and Y the correspondent outcome with its sample space Ω. Then the patient data set is defined by*

$$S^{all} := \{(x_j, y_j)\}_{j=1}^{N} \subset \mathbb{R}^d \times \Omega$$

with $x_j \in X^{all}$ and $y_j \in \Omega$

In RCTs the study participants are divided into control and treatment group. For the identification, we use the treatment group assignment for patient j, for $j = 1, \ldots, N$.

Definition 5.1.4. *Let $S^{all} = \{(x_j, y_j)\}_{j=1}^{N} \subset \mathbb{R}^d \times \Omega$ be the patient data set. Then*

$$t_j := \begin{cases} 1 & \text{if patient } j \text{ is in the treatment group} \\ 0 & \text{if patient } j \text{ is in the control group} \end{cases}$$

is the treatment group assignment for patient j, for $j = 1, \ldots, N$, and defines whether a patient is in the treatment or in the control group.

Furthermore, we need to know in which study a patient has participated. Therefore, we define the study assignment vector for patient j, for $j = 1, \ldots, N$. The following definition gives the required information.

Definition 5.1.5. *Let $S^{all} = \{(x_j, y_j)\}_{j=1}^{N} \subset \mathbb{R}^d \times \Omega$ be the patient data set and n_{st} the number of studies included. Then $st_j = (st_{j1}, \ldots, st_{jn_{st}}) \in \{0,1\}^{n_{st}}$ with*

$$st_{jm} := \begin{cases} 1 & \text{if patient } j \text{ is in study } m \\ 0 & \text{else} \end{cases},$$

for $m = 1, \ldots, n_{st}$, is the study assignment vector for patient j, $j = 1, \ldots, N$.

If we use the supervised learning clustering approach, e.g. for the prediction of the efficacy of a certain medication, the patient data set is divided into training and testing data set. The clustering approach is then conducted on the training data set. For the comparison with other supervised learning approaches it is common to take 80% of all data points as training data set and the remaining 20% as testing data.

Definition 5.1.6. *Let X^{all} be the patient data set of characteristics and Y the correspondent outcome with its sample space Ω. $X = \{x_j\}_{j=1}^{n} \subset X^{all}$ be the training patient data set of characteristics for the conduction of the clustering approach and $X^{te} := X^{all} \backslash X$ be the correspondent testing patient data set of characteristics. Then the patient training data set is defined by*

$$S := \{(x_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega,$$

with $x_j \in X$ and $y_j \in \Omega$. The correspondent testing patient data set is defined by

$$S^{te} := \{(x_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega,$$

with $x_j^{te} \in X^{te}$ and $y_j^{te} \in \Omega$. n denotes the number of patients in the training data set and n^{te} the number of patients in the testing data set with $n^{te} \leq n \leq N$ and $n^{te} + n = N$.

With these patient specific definitions, the general terms for the conduction of the geometric clustering method can be specified.

Definition 5.1.7 (Cluster size)**.** *Let $k \in \mathbb{N}$ represent the number of clusters. Then κ_i with $\sum_{i=1}^{k} \kappa_i = n$ is the number of patients in cluster Cl_i, for $i = 1, \ldots, k$. We call κ_i the cluster size of cluster Cl_i.*

With this basic determination, it is possible to define the partition of a data set X into k preferably homogeneous clusters as k-clustering.

Definition 5.1.8 (k-clustering)**.** *A k-clustering $Cl = (Cl_1, \ldots, Cl_k)$ is a partition of a set $X \to \mathbb{R}^d$ into k nonempty sets $Cl_1, \ldots, Cl_k$. The ith entry of Cl is called the cluster Cl_i of the clustering Cl for $i = 1, \ldots, k$.*

For high statistical power when conducting statistical tests on a cluster, the size is of high importance. Therefore, we define the shape of a k-clustering, which is based on the sizes $\kappa_1, \ldots, \kappa_k$ of a clustering.

Definition 5.1.9 (Shape of a k-clustering)**.** *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a k-clustering of a data set X and $\kappa_i = |Cl_i|$ the size of cluster Cl_i, for $i = 1, \ldots, k$, then*

$$|Cl| := (|Cl_1|, \ldots, |Cl_k|)$$

is called the shape of a k-clustering.

Statistical power in a cluster can be seen as the power to detect true differences in the patients' outcome or treatment effects across clusters. To get this high statistical power within a cluster with regard to statistical significance, it is desirable to have a certain number of patients in each cluster. Therefore we define the (k, l, u)-clustering where the cluster size is restricted by an upper and especially a lower boundary.

Definition 5.1.10 ((k, l, u)-clustering)**.** *A k-clustering $Cl = (Cl_1, \ldots Cl_k)$ with $l = (l_1, \ldots, l_k) \in \mathbb{N}^k$, $u = (u_1, \ldots, u_k) \in \mathbb{N}^k$ and $l_i \leq \kappa_i \leq u_i$, for $i = 1, \ldots, k$, is called (k, l, u)-clustering.*

If the number of clusters k and the boundaries $l, u \in \mathbb{N}^k$ are clear from the context and fixed, we will call the (k, l, u)-clustering also bounded-shape clustering (BSC). In this case, we use the notation $BSC(X) = BSC(k, l, u) := \mathcal{C}(X, k, l, u)$.

It is also possible to fix the shape of the clustering by means of fixing the cluster sizes κ_i, $i = 1, \ldots, k$. The resulting clustering is called $(k, (\kappa_1, \ldots, \kappa_k))$-clustering and is defined as follows.

Definition 5.1.11 ($(k,(\kappa_1,\ldots,\kappa_k))$-clustering)**.** *A k-clustering of X,*

$$Cl = (Cl_1,\ldots,Cl_k),\ \textit{with}\ |Cl| = (\kappa_1,\ldots,\kappa_k)$$

is called $(k,(\kappa_1,\ldots,\kappa_k))$-clustering.

Since we do not need fixed cluster sizes but at least lower boundaries for the improvement of the statistical power, in the following we will have a closer look at the (k,l,u)-clustering. With the specification of the upper and lower boundaries for the cluster sizes, it is possible to get a fixed number of feasible (k,l,u)-clusterings.

Remark 5.1.12. *For a given set X of size n, lower and upper boundaries $l,u \in \mathbb{N}^k$ and the number of cluster $k \in \mathbb{N}$, the number of feasible (k,l,u)-clustering is*

$$\sum_{\substack{l_i \le \kappa_i \le u_i, i\in\{1,\ldots,k\} \\ \sum_i \kappa_i = n}} \frac{n!}{\prod_{i=1}^k \kappa_i! \prod_{i=1}^k m_i!}$$

with $m_i := |\{\kappa_i : \kappa_i \in \{\kappa_1,\ldots,\kappa_k\}|$.

These number of feasible (k,l,u)-clusterings can be summarized to a set of feasible (k,l,u)-clusterings, like it is done in the next definition.

Definition 5.1.13 (Set of feasible (k,l,u)-clusterings)**.** *Let X be a data set of size n and k be the number of clusters then*

$$\mathcal{C}(X,k,l,u) := \{Cl : Cl\ \textit{is a (k,l,u)-clustering of}\ X\}$$

is the set of (k,l,u)-clusterings of X.

We can also define the set of feasible k-clusterings without the predefined upper and lower boundaries for the cluster sizes.

Definition 5.1.14 (Set of feasible k-clusterings)**.** *Let X be a set of size n and k be the number of clusters then*

$$\mathcal{C}(X,k) := \{Cl : Cl\ \textit{is a k-clustering of}\ X\}$$

is the set of k-clusterings of X.

To get a feasible (k, l, u)-clustering the upper and lower boundaries for the cluster sizes have to fulfill a general standard.

Remark 5.1.15. *Let X be a set of size n and $l, u \in \mathbb{N}^k$ lower and upper boundaries. A clustering $Cl = (Cl_1, \dots Cl_k)$ is feasible with respect to l, u if*

$$\sum_{i=1}^{k} l_i \leq n \leq \sum_{i=1}^{k} u_i$$

holds.

In the following, we will always assume that the boundaries fulfill this requirement and allow feasible clusterings.

In the next step, we will have a closer look at the connection between the vertexes of a special polytope, the so called bounded-shape partition polytope, and the determination of a clustering. We will see that each vertex of this bounded-shape partition polytope represents a clustering, like it is discussed by Brieden and Gritzmann in [15] and [16]. To get there, we need some more definitions. A clustering can be identified by the center of gravities of each cluster, which may be intuitively, or by the sum of data points in a cluster [16]. In the latter case, upper and lower boundaries can be defined.

Definition 5.1.16 (Cluster sum)**.** *Let $Cl = (Cl_1, \dots Cl_k)$ be a clustering of a set X. Then the cluster sum s_i of a cluster Cl_i is defined as*

$$s_i := \sum_{x \in Cl_i} x,$$

for $i = 1, \dots, k$. The vector $v(Cl) := (s_1^T, \dots, s_k^T)^T \in \mathbb{R}^{k \cdot d}$ is called the cluster sum vector.

Definition 5.1.17 (Center of gravity)**.** *Let $Cl = (Cl_1, \dots Cl_k)$ be a clustering of*

a set X. Then the center of gravity c_i of a cluster Cl_i is defined as

$$c_i := \frac{s_i}{\kappa_i},$$

for $i = 1, \ldots, k$.

As it is mentioned above, a clustering can be identified as a vertex of the so-called bounded-shape partition polytope. Now we want to have a look how this polytope is defined. To do so, we need the set of the cluster sum vectors of all feasible (k, l, u)-clusterings of data set X.

Definition 5.1.18 (Set of Cluster Sum Vectors)**.** *Let X be a set of size n. Then $V := V(X; k, l, u) := \{v(Cl) : Cl \in \mathcal{C}(X, k, l, u)\}$ is the set of all cluster sum vectors.*

With this definition, we are able to define the bounded-shape partition polytope as the convex hull of the cluster sum vectors of all feasible (k, l, u)-clusterings.

Definition 5.1.19 (Bounded-Shape Partition Polytope (BSPP))**.** *Let X be a subset of $\mathbb{R}^d$. The bounded-shape partition polytope is defined as the convex hull of all cluster sum vectors*

$$BSPP = BSPP(k, l, u) = BSPP(k, l_1, \ldots, l_k, u_1, \ldots, u_k) := convV(X; k, l, u)$$

As it is mentioned above, each vertex of the bounded-shape partition polytope represents a feasible (k, l, u)-clusterings by its cluster sum. This fact can be formulated as follows.

Lemma 5.1.20. *Let v^* be a vertex of a $BSPP$. Then there is exactly one (k, l, u)-clustering $Cl = (Cl_1, \ldots Cl_k)$ with $v(Cl) = v^*$. We call this the clustering of v^*.*

Proof. The proof is given in [16] as the bounded-shape partition polytopes are constrained in the subspace of the described gravity bodies. □

Lemma 5.1.20 allows the conclusion that (k, l, u)-clusterings can be identified by vertices of polytopes, which has already been shown by Barnes, Hoffman and

Rothblum in [3]. As a result of this conclusion, it can be shown according to [34] that the bounded-shape partition polytope can also be explained by linear constraints which form the basis for the linear program defined in Section 5.2. The solution of this linear program then represents a feasible clustering [3]. To get there, we need one further definition of the decision variable which indicates whether a cluster contains a data point or not.

Definition 5.1.21. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a clustering of a set $X = \{x_j\}_{j=1}^n$ and for all $i \in \{1, \ldots, k\}$ and all $j \in \{1, \ldots, n\}$ let*

$$\xi_{ij} = \begin{cases} 1 & \text{if } x_j \in Cl_i \\ 0 & \text{if } x_j \neq Cl_i \end{cases}$$

be a decision variable indicating whether cluster Cl_i contains x_j, $\xi_{ij} = 1$, or not, $\xi_{ij} = 0$.

Now we have all necessary components for the definition of the bounded-shape partition polytope with linear constraints.

Definition 5.1.22. *Let $k, l_i, u_i \in \mathbb{N}$, for $i = 1, \ldots, k$, with $\sum_{i=1}^k l_i \leq n \leq \sum_{i=1}^k u_i$. We call the polytope defined by the constraints*

$$\begin{aligned} \sum_{j=1}^{n} \xi_{ij} &\leq u_i \quad (i \leq k) \\ \sum_{j=1}^{n} \xi_{ij} &\geq l_i \quad (i \leq k) \\ \sum_{i=1}^{k} \xi_{ij} &= 1 \quad (j \leq n) \\ \xi_{ij} &\geq 0 \quad (i \leq k, j \leq n) \end{aligned}$$

the bounded-shape partition polytope $BSPP(k, l, u)$.

With this definition, we can formulate the geometric clustering approach as maximization problem in the following section. Since this resulting linear pro-

gram is based on the bounded-shape partition polytope, the optimization problem is also known as the bounded-shape partition problem.

5.2. Clustering as an optimization problem

Our goal is to classify the patients of all available trial into collectives as homogeneous as possible with the help of the geometric clustering approach to enable a more precise evaluation of the underlying medical intervention. Therefore, under all possible clusterings, we want to find a specific clustering with an upper and especially a lower boundary for the number of patients in the resulting collectives which maximizes the target function f. The specific target function is defined in Section 5.4. As it is stated in [34], this specific clustering, also called bounded-shape clustering, can be obtained by computing the solution $\xi = (\xi_{ij}) \in \{0,1\}^{k \times n}$ of the following linear program with constraints based on the bounded-shape partition polytope.

$$
\begin{aligned}
& \max \quad f(\xi) \\
\sum_{j=1}^{n} \xi_{ij} \quad & \leq \quad u_i \quad (i \leq k) \\
\sum_{j=1}^{n} \xi_{ij} \quad & \geq \quad l_i \quad (i \leq k) \\
\sum_{i=1}^{k} \xi_{ij} \quad & = \quad 1 \quad (j \leq n) \\
\xi_{ij} \quad & \in \quad \{0,1\} \quad (i \leq k, j \leq n)
\end{aligned}
$$

In [34] it is also shown that this bounded-shape partition problem can be solved in polynomial time due to the total unimodularity of the underlying matrix derived from the constraints of the problem formulation. This is also the reason why the solution of the relaxation is integral and therefore it is sufficient to solve the relaxation of

$$
\begin{aligned}
&\max \quad f(\xi) \\
\sum_{j=1}^{n} \xi_{ij} \quad &\leq \quad u_i \quad (i \leq k) \\
\sum_{j=1}^{n} \xi_{ij} \quad &\geq \quad l_i \quad (i \leq k) \\
\sum_{i=1}^{k} \xi_{ij} \quad &= \quad 1 \quad (j \leq n) \\
\xi_{ij} \quad &\geq \quad 0 \quad (i \leq k, j \leq n).
\end{aligned}
$$

We can also show according to [3] that the solution of this formulated problem, as a vertex of the bounded-shape partition polytope, is a separable clustering.

5.3. Separability

To show that the solution of the linear program provides a separable clustering, we need to know what separability means and what it implies. Thus, we need to define linear separability first.

Definition 5.3.1 (Weakly linearly separable)**.** *Let $A, B \subset \mathbb{R}^d$. A and B are weakly linearly separable, if there is a hyperplane $H_{a,\beta} \subset \mathbb{R}^d$ with $a \in \mathbb{R}^d \backslash \{0\}$ and $\beta \in \mathbb{R}$ such that $A \subset H^{\geq}_{a,\beta}$ and $B \subset H^{\leq}_{a,\beta}$.*

Definition 5.3.2 (Strictly linearly separable)**.** *Let $A, B \subset \mathbb{R}^d$. A and B are strictly linearly separable, if there is a hyperplane $H_{a,\beta} \subset \mathbb{R}^d$ with $a \in \mathbb{R}^d \backslash \{0\}$ and $\beta \in \mathbb{R}$ such that $A \subset H^{>}_{a,\beta}$ and $B \subset H^{<}_{a,\beta}$.*

The separability of a clustering is given, if all pairs of its clusters are separable.

Definition 5.3.3 (Separability of clusterings)**.** *Let $Cl = (Cl_1, \ldots Cl_k)$ be a clustering of a set X. Cl allows (weak, strict) linear separation (or is weakly, strictly linear separable) if Cl_i and Cl_j are (weakly, strictly) linearly separable for any $i \neq j$, $i, j \in \{1, \ldots, k\}$.*

As it is stated by Barnes, Hoffman and Rothblum in [3], with these terms we have everything to explain the relationship between the vertices of the bounded-shape partition polytope and the separability of the associated clustering.

Theorem 5.3.4. *Let v^* be a vertex of the bounded-shape partition polytope. Then the bounded-shape clustering Cl^* associated with $v^* = v(Cl^*)$ allows strict linear separation.*

Proof. The proof is given in [3]. □

The separability of a clustering is one desirable property for a uniquely separation of a patient data set into patient collectives. As already discussed, another desirable property would be the separation into patient collectives as homogeneous as possible. Therefore, we need a target function for the measurement of homogeneity which can be maximized under the bounded-shape partition polytope.

5.4. Criterion for homogeneous collectives

Homogeneous collectives out of a heterogeneous patient data set X can be obtained by grouping patients with similar characteristic values combinations to one collective. For the division of patients into these homogeneous collectives, a clustering should therefore be separable and the homogeneity in a cluster should be maximized. This homogeneity can be obtained by maximizing the distances between the centers of each cluster in the geometric space. Additionally the predefined upper or lower bounds for the number of patients in one collective must not be exceeded or undershot. Thus, we need a feasible target function for the optimization problem defined in Section 5.2.

The target function for the maximization of the pairwise distances between the centers of gravity of a cluster can be formulated as

$$\max_{Cl=(Cl_1,\ldots,Cl_k)\in\mathcal{C}} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \|c_i - c_j\|_2^2,$$

with the center of gravity

$$c_i = \frac{1}{\kappa_i} \sum_{j=1}^{n} x_j \xi_{ij},$$

of cluster Cl_i, for $i = 1, \ldots, k$. Here, for the determination of the pairwise distance, we use the Euclidean norm. Unfortunately, this target function leads to a nonlinear optimization problem which is known to be $\mathbb{NP}$-hard [4].

To handle this, we piecewise linearly approximate the nonlinear maximization problem. In [16] it is shown, that in case of $\sum_{x \in X} x = 0$, the maximization of the pairwise distances between the centers of gravity is equivalent to the maximization of the total linear cluster distance

$$\max_{Cl=(Cl_1,\ldots,Cl_k)\in\mathcal{C}} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \sum_{x_i \in Cl_i} \sum_{x_j \in Cl_j} (v_i - v_j)^T (s_i - s_j),$$

with $v = (v_1^T, \ldots, v_k^T)^T \in \mathbb{R}^{d \cdot k}$ and s_i being the cluster sum of cluster Cl_i, $i = 1, \ldots, k$. In [48] it is shown, that this is related to the so-called least-square assignment (LSA).

Definition 5.4.1 (Least-Square Assignment (LSA))**.** *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of a set X and $v = (v_1^T, \ldots, v_k^T)^T \in \mathbb{R}^{d \cdot k}$, Cl is called a least square assignment (LSA) of X to v, if and only if it minimizes*

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \xi_{ij} \|x_j - v_i\|_2^2$$

over the $BSPP(k, l, u)$.

As the least-square assignment minimizes the variance of the points assigned to a cluster, outliers could be allocated to a cluster with only few points. This property of the LSA isn't always desirable for the identification of homogeneous patient collectives. These found collectives should contain enough patients for the conduction of cluster-based meta-analysis, for a more precise estimate of a drug's efficacy or for the conduction of correspondent tests of hypothesis. Only few patients in the collectives could falsify the results of meta-analysis or the prediction of a drug's efficacy. It is possible to set appropriate strict lower bounds

or to conduct an approach which is an approximation of the LSA. This approach leads to proper filled cluster without setting strict lower bounds which is desirable for the evaluation of drug intervention on homogeneous patient collectives.

Definition 5.4.2 (Cluster Sum Assignment (CSA)). *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of a set* X *and* $v = (v_1^T, \ldots, v_k^T)^T \in \mathbb{R}^{d \cdot k}$*,* Cl *is called a cluster sum assignment (CSA) of* X *to* v*, if and only if it maximizes*

$$\sum_{i=1}^{k} \sum_{j=1}^{n} \xi_{ij} v_i^T x_j$$

over the $BSPP(k, l, u)$*.*

In [48] it is shown that the LSA and the CSA are equivalent and the resulting clusterings of the correspondent approaches are the same in case of standardized sites a_i. In this case and if data set X is replaced by the centered set

$$(X)^c = X - \frac{\sum_{i=1}^{n} x_i}{n},$$

so that $\sum_{x \in X^c} x = 0$, the CSA can be interpreted as a clustering with the centers of gravity c_i being pushed away from the unit ball with respect to directions a_i, for $i = 1, \ldots, k$, since a norm maximization can be interpreted as scaling up a unit ball to fit the feasible region.

As the main result of this section, we show that the solution of the maximization of the cluster sum assignment over the bounded-shape partition polytope,

$$\begin{aligned}
\max_{\xi=(\xi_{ij})\in\{0,1\}^{k\times n}} & \sum_{i=1}^{k}\sum_{j=1}^{n}\xi_{ij}v_i^T x_j \\
\sum_{j=1}^{n}\xi_{ij} &\leq u_i \quad (i\leq k) \\
\sum_{j=1}^{n}\xi_{ij} &\geq l_i \quad (i\leq k) \\
\sum_{i=1}^{k}\xi_{ij} &= 1 \quad (j\leq n) \\
\xi_{ij} &\geq 0 \quad (i\leq k, j\leq n),
\end{aligned}$$

is a linearly separable clustering.

Theorem 5.4.3. *Let $X \subset \mathbb{R}^d$ be a data set and $k, l_i, u_i \in \mathbb{N}$ for all $i \in \{1,\ldots,k\}$ with $\sum_{i=1}^{k} l_i \leq n \leq \sum_{i=1}^{k} u_i$ the parameter set. Let $BSPP(k,l,u)$ be the corresponding bounded-shape partition polytope and*

$$v = (v_1^T,\ldots,v_k^T)^T \in \mathbb{R}^{d\cdot k}.$$

Then we can find a vertex ve^* of the $BSPP$ with $v^T ve^* \geq v^T ve$ for any $ve \in BSPP$ by solving the linear program.

This theorem represents the basis for the division of the patient data set X into homogeneous patient collectives. The corresponding clustering is optimal regarding the cluster sum and represents a vertex of the bounded-shape partition polytope. Furthermore, it allows strict linear separability like it is shown in [34]. It can also be shown that this separability leads to a partition of $\mathbb{R}^d$ into convex polyhedral cells which are known as power diagrams. In [48] it is discussed that a vertex of a bounded-shape partition polytope induces such power diagram. For more details about power diagrams please refer to [1].

5.5. Transformation of data

The geometric clustering approach is conducted in a geometric space, therefore, for each characteristic of a patient, we need quantitative characteristic values. But since the characteristics of a patient can be of any level of scale, they have to be quantified first. This transformation then enables the use of the geometric clustering approach. In the following, we will give an introduction to this transformation technique. It is based on the work of Öllinger [48]. If the clustering approach is unsupervised, then the training data set is the whole data set and we set $S = S^{all}$. If the clustering approach is supervised and we have a division of the patient data S^{all}, we use the notations S for the training and S^{te} for the testing data set. The idea of the transformation technique is to replace the characteristic value $A_i = a_i$ by the conditional expected value $E(Y|A_i = a_i)$, for $i = 1, \ldots, d$, $a_i \in \Omega_i$, of the patients' outcome or endpoint given the characteristic value $A_i = a_i$. This is why we call the clustering algorithm endpoint-oriented.

Since the number of characteristic values combinations increases proportionally with the increase in the number of characteristic values, which leads to a extremely small ratio between the actual and possible combinations, we need to classify characteristics with a high number of characteristic values. To do so, the most intuitive way is to define classes of equal class width. A second approach is the division into classes of equal class density. But we can also set the class boundaries intuitively.

For the transformation of the classified patient data set S, we assume that the patients' outcome Y is a discrete random variable with sample space $\Omega = \{y_1, \ldots, y_n\}$. The elements of Ω represent the outcome of all patients in set S. For the motivation of the idea to replace the characteristic values by the conditional expected outcome, we take a similar approach to the Naive Bayes classification. In this approach, the conditional probability for an outcome $Y = y$ of a new patient given specific characteristic values $A_1 = a_1, \ldots, A_d = a_d$ can be predicted. For the calculation of the conditional probability, the Bayes' rule

$$P(Y = y|A_1 = a_1, \ldots, A_d = a_d) = \frac{P(A_1 = a_1, \ldots, A_d = a_d|Y = y)P(Y = y)}{\sum\limits_{z \in \Omega} P(A_1 = a_1, \ldots, A_d = a_d|Y = z)P(Y = z)}$$

can be applied. For this approach, we need the conditional probabilities

$$P(A_1 = a_1, \ldots, A_d = a_d | Y = y)$$

which are not easy to estimate, given the patient data set S. With the implication of the following definition, these conditional probabilities can be written as the product of conditional probabilities of the single characteristics A_i, $i = 1, \ldots, d$, given Y.

Theorem 5.5.1 (Conditional Independence)**.** *Let X, Y and Z be random variables. Then X and Y are conditionally independent given Z if and only if the probability distribution of X is independent of the value of Y given Z. This is equivalent to*

$$P(X = x_i, Y = y_j | Z = z_k) = P(X = x_i | Z = z_k) P(Y = y_j | Z = z_k),\ \forall i, j, k.$$

For the characteristics $A_1, \ldots, A_d$, this leads to

$$P(A_1 = a_1, \ldots A_d = a_d | Y = y) = \prod_{i=1}^{d} P(A_i = a_i | Y = y).$$

We assume that the characteristics of the patients are conditional independent. With this assumption we get the new formula

$$P(Y = y | A_1 = a_1, \ldots, A_d = a_d) = \frac{P(Y = y) \prod_{i=1}^{d} P(A_i = a_i | Y = y)}{\sum_{z \in \Omega} P(Y = z) \prod_{i=1}^{d} P(A_i = a_i | Y = z)}$$

for the calculation of the probability of a specific outcome $Y = y$ under given characteristic values $A_1 = a_1, \ldots, A_d = a_d$, which is much easier to calculate with the given set of patient data points.

The approach for the transformation of the characteristic values of a patient is similar to the Naive Bayes approach and is based on the indirect estimation of

the conditional expected value

$$E(Y|A_1 = a_1, \ldots, A_d = a_d) = \sum_{y \in \Omega} yP(Y = y|A_1 = a_1, \ldots, A_d = a_d).$$

This conditional expected value also contains the conditional probabilities of $Y = y$, $y \in \Omega$, given $A_1 = a_1, \ldots, A_d = a_d$. These probabilities can be written as

$$P(Y = y|A_1 = a_1, \ldots, A_d = a_d) = \sum_{i=1}^{d} \beta_i P(Y = y|A_i = a_i),$$

with

$$\beta_i = \hat{\beta}_i \frac{P(Y = y|A_1 = a_1, \ldots, A_d = a_d)}{P(Y = y|A_i = a_i)}, \ \sum_{i=1}^{d} \hat{\beta}_i = 1.$$

The conditional expected value can then be calculated as the convex combination of the one dimensional conditional expected values $E(Y|A_i = a_1)$, $i = 1, \ldots, d$

$$\begin{aligned} E(Y|A_1 = a_1, \ldots, A_d = a_d) &= \sum_{y \in \Omega} yP(Y = y|A_1 = a_1, \ldots, A_d = a_d) \\ &= \sum_{y \in \Omega} y(\sum_{i=1}^{d} \beta_i P(Y = y|A_i = a_i)) \\ &= \sum_{i=1}^{d} \beta_i E(Y|A_i = a_i). \end{aligned}$$

Similar to the Naive Bayes approach we reduce the multidimensional conditional expected value to the product of the one dimensional expected values, which are much easier to estimate, as we will see in the following passages.

Since Y and $A_1, \ldots, A_d$ are discrete random variables, the one dimensional conditional expected value $E(Y|A_i = a_i)$ can be calculated by the Bayes' rule

$$\begin{aligned} E(Y|A_i = a_i) &= \sum_{y \in \Omega} yP(Y = y|A_i = a_i) \\ &= \sum_{y \in \Omega} y \frac{P(Y = y, A_i = a_i)}{P(A_i = a_i)}, \end{aligned}$$

for $i = 1, \ldots, d$.

In the case of a Bernoulli distributed random variable Y, with the two occurrences $\Omega = \{0, 1\}$, where 1 stands e.g. for response and 0 for no response to a the administered medication, the conditional expected value is equivalent to the conditional probabilities.

Remark 5.5.2. *If Y is a Bernoulli distributed binary random variable with probability p, $Y \sim Be(p)$, for the conditional expected value holds*

$$E(Y|X = x) = P(Y = 1|X = x) =: p_{|x}.$$

The conditional expected value

$$E(Y|A_i = a_i),\ i \in \{1, \ldots, d\},\ a_i \in \Omega_i$$

or the equivalent conditional probability

$$P(Y = 1|A_i = a_i),\ i \in \{1, \ldots, d\},\ a_i \in \Omega_i,$$

if Y is a binary random variable, is the new quantified value replacing the original patient characteristic value a_i of characteristic A_i, for $i = 1, \ldots, d$. Each patient data point $x_j \in X = \{x_1, \ldots x_n\} \subset \mathbb{R}^d$, $x_j = (a_1, \ldots, a_d)$, is then represented by the vector of their conditional expected values

$$(a_1, \ldots, a_d) \to (E(Y|A_1 = a_1) \ldots, E(Y|A_d = a_d))$$

or the equivalent conditional probabilities

$$(a_1, \ldots, a_d) \to (P(Y = 1|A_1 = a_1) \ldots, P(Y = 1|A_d = a_d))$$

in the binary case. With this transformation, the geometric clustering approach of the following section is applicable.

Since the probability $P(A_i = a_i)$ and the joint probability $P(Y = y, A_i = a_i)$, which are needed for the calculation of the conditional expected value, are generally unknown, they have to be estimated by the corresponding conditional means

of the given patient data set. First of all, we calculate the estimation for the training data set.

Theorem 5.5.3. *Let $S = \{(x_j, y_j)\}_{j=1}^n \subset \mathbb{R}^d \times \Omega$, with $x_j = (x_{j1}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$, be a sample of the d characteristics, represented by A_i, $i = 1, \ldots, d$, and the discrete outcome of a patient is denoted by Y. Then*

$$\hat{a}_i = \hat{\Theta}(Y|A_i = a_i) := \frac{\sum_{j=1}^n y_j \mathbb{1}_{\{x_{ji}\}}(a_i)}{\sum_{j=1}^n \mathbb{1}_{\{x_{ji}\}}(a_i)}, \text{ for } i = 1, \ldots, d$$

is an unbiased estimation for the conditional expected value $E(Y|A_i = a_i)$.

Proof. As it is stated above, the conditional expected value is defined by

$$E(Y|A_i = a_i) = \sum_{y \in \Omega} y \frac{P(Y = y, A_i = a_i)}{P(A_i = a_i)},$$

for $i = 1, \ldots, d$, based on the Bayes' rule. The estimator for the joint probability

$$\hat{P}(Y = y, A_i = a_i) = \frac{\sum_{j=1}^n \mathbb{1}_{\{y_j\}}(y) \mathbb{1}_{\{x_{ji}\}}(a_i)}{n}$$

and the estimator of the single probability

$$\hat{P}(A_i = a_i) = \frac{\sum_{j=1}^n \mathbb{1}_{\{x_{ji}\}}(a_i)}{n}$$

result from the conditional frequencies and lead to a mean estimation of the conditional expected value $E(Y|A_i = a_i)$. □

If the patients' outcome is Bernoulli distributed, the estimation is given by the following theorem.

Theorem 5.5.4. *Let $S = \{(x_j, y_j)\}_{j=1}^n \subset \mathbb{R}^d \times \Omega$, with $x_j = (x_{j1}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$, be a sample of the d characteristics, represented by A_i, $i = 1, \ldots, d$, and the binary outcome of a patient is denoted by Y. Then*

$$\hat{a}_i = \hat{\Theta}(Y = 1|A_i = a_i) := \frac{\sum_{j=1}^n \mathbb{1}_{\{1\}}(y_j) \mathbb{1}_{\{x_{ji}\}}(a_i)}{\sum_{j=1}^n \mathbb{1}_{\{x_{ji}\}}(a_i)}, \text{ for } i = 1, \ldots, d$$

is an unbiased estimator for the conditional expected value $P(Y = 1|A_i = a_i)$.

With these theorems, we have everything for the transformation of the training data set $S = \{(x_j, y_j)\}_{j=1}^n$, with $x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$, into quantitative values to conduct the geometric clustering approach. In the first step, for $i = 1, \ldots d$, the values in the sample space Ω_i of the characteristic A_i of a patient have to be transformed into the corresponding estimated conditional expected values,

$$a_i \to \hat{a}_i, \, \forall a_i \in \Omega_i.$$

This results in a transformed random variable $\hat{A}_i$ with its transformed sample space $\hat{\Omega}_i = \{\hat{a}_i | a_i \in \Omega_i\}$. In a second step all values of the data set S have to be replaced by the corresponding estimated conditional expected values,

$$x_{ji} = a_i \to \hat{x}_{ji} := \hat{a}_i, \text{ for } j = 1, \ldots n, \text{ and } i = 1 \ldots d.$$

As a result we get a transformed data set

$$S = \{(x_j, y_j)\}_{j=1}^n \subset \mathbb{R}^d \times \Omega \to \hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^n \subset \mathbb{R}^d \times \Omega$$

with

$$\hat{x}_j = (\hat{x}_{j1}, \hat{x}_{j2}, \ldots, \hat{x}_{jd}) = (\hat{a}_1, \hat{a}_2 \ldots, \hat{a}_d)$$

and $\hat{a}_i \in \hat{\Omega}_i$, for $i = 1, \ldots, d$.

While the training data set is transformed based on the conditional expected values of the training data set, the testing data set has to be transformed differently. Due to the fact that the outcome of a patient is unknown, the characteristic values of the patients have to be replaced by the estimations for the conditional expected values of the training data set. The testing data set

$$S^{te} = \{(x_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega,$$

with $x_j^{te} = (x_{j1}^{te}, x_{j2}^{te}, \ldots, x_{jd}^{te}) = (a_1, \ldots, a_d)$ is transformed to

$$\hat{S}^{te} = \{(\hat{x}_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega$$

by its conditional expected values based on S,

$$x_{ji}^{te} = a_i \rightarrow \hat{x}_{ji}^{te} := \hat{a}_i, \text{ for } j = 1, \dots n^{te}, \text{ and } i = 1 \dots d.$$

As a result we get a transformed testing patient data set

$$S^{te} = \{(x_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega \rightarrow \hat{S}^{te} = \{(\hat{x}_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega$$

with

$$\hat{x}_j^{te} = (\hat{x}_{j1}^{te}, \hat{x}_{j2}^{te}, \dots, \hat{x}_{jd}^{te}) = (\hat{a}_1, \hat{a}_2 \dots, \hat{a}_d)$$

and $\hat{a}_i \in \hat{\Omega}_i$, for $i = 1, \dots, d$.

In the next step, we can conduct a geometric clustering on the classified and transformed training data set $\hat{S}$. Additionally we can assign the transformed testing data set to a cluster by evaluating its position with respect to the separating hyperplanes of the power diagram induced by the clustering.

.

5.6. Geometric clustering algorithm

Now, we have all components for the formulation of the geometric clustering algorithm according to [48], for the division of a patient data set S into homogeneous patients collectives. In a first step, all relevant characteristics $A_1, \dots, A_d$ of a patient which could have an impact on the efficacy of a medical intervention have to be identified. Furthermore, the patients' outcome Y due to the administered medication has to be defined. Thereby, the patient data set S is determined. In a second step, based on the identified characteristics and the patients' outcome, set S has to classified and transformed to $\hat{S}$ by applying the transformation approach discussed in Section 5.5. With this preparation of the individual patient data, the geometric clustering approach described in Section 5.4 can be conducted on the transformed patient data $\hat{S}$. This approach is embedded in an iterative sequence, in which an optimal clustering is computed in each step with respect to a given site vector $v = (v_1^T, \dots, v_k^T)^T \in \mathbb{R}^{d \cdot k}$ as a solution of the linear program. The output by means of an optimal clustering of each step is taken as input for a next

step by applying the standardized sum of the clustering solution as new sites

$$(v_1, \ldots, v_k) \to (\frac{c_1}{\|c_1\|}, \ldots, \frac{c_k}{\|c_k\|}) = (\frac{s_1}{\|s_1\|}, \ldots, \frac{s_k}{\|s_k\|}).$$

In the following, this procedure for the computation of an optimal clustering is described in each step. We assume that all relevant characteristics $A_1, \ldots, A_d$ of the participating patients have been identified and the patients' outcome Y due to a medical intervention has been defined. Furthermore, the data set has been divided into a training and a testing data set in case of a supervised learning approach. In case of an unsupervised learning approach all patient data serve as the training data set. The first algorithm deals with the transformation of the patient data set S.

Algorithm 1: Transformation of the training data set S

Data: $S = \{(x_j, y_j)\}_{j=1}^n \subset \mathbb{R}^d \times \Omega$ with
$x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$ and $a_i \in \Omega_i$;

Result: $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^n \subset \mathbb{R}^d \times \Omega$ with
$\hat{x}_j = (\hat{x}_{j1}, \hat{x}_{j2}, \ldots, \hat{x}_{jd}) = (\hat{a}_1, \ldots, \hat{a}_d)$ and $\hat{a}_i \in \hat{\Omega}_i = \{\hat{a}_i | a_i \in \Omega_i\}$,
$i = 1, \ldots, d$;

```
/* Transformation of characteristic values                    */
```

for $i = 1$ **to** d **do**
  for $a_i \in \Omega_i$ **do**
    $a_i \to \hat{a}_i$

```
/* Generation of transformed training data set                */
```

for $i = 1$ **to** d **do**
  for $j = 1$ **to** n **do**
    $x_{ij} = a_i \to \hat{x}_{ij} = \hat{a}_i$

While the training data set is transformed based on the estimations of the conditional expected outcome of the training patient data set, the characteristic values of the testing data set S^{te} are also replaced by the estimations for the conditional expected values of the training patient data set.

Algorithm 2: Transformation of the testing data set S^{te}

Data: $S^{te} = \{(x_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega$ with
$x_j^{te} = (x_{j1}^{te}, x_{j2}^{te}, \ldots, x_{jd}^{te}) = (a_1, \ldots, a_d)$ and $a_i \in \Omega_i$, $\{\hat{a}_i | a_i \in \Omega_i\}_{i=1}^{d}$;

Result: $\hat{S}^{te} = \{\hat{x}_j^{te}, y_j^{te}\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega$ with
$\hat{x}_j^{te} = (\hat{x}_{j1}^{te}, \hat{x}_{j2}^{te}, \ldots, \hat{x}_{jd}^{te}) = (\hat{a}_1, \ldots, \hat{a}_d)$;

```
/* Generation of transformed testing data set                    */
```

for $i = 1$ **to** d **do**
  for $j = 1$ **to** n^{te} **do**
    $x_{ij}^{te} = a_i \rightarrow \hat{x}_{ij}^{te} = \hat{a}_i$

With the transformation of the data set to quantitative values, the geometric clustering approach can be conducted.

Algorithm 3: Calculation of a clustering Cl

Data: $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega$ with
$\hat{x}_j = (\hat{x}_{j1}, \hat{x}_{j2}, \ldots, \hat{x}_{jd}) = (\hat{a}_1, \ldots, \hat{a}_d)$; boundaries $l_i, u_i \in \mathbb{N}$;
$\sum_{i=1}^{k} l_i \leq n \leq \sum_{i=1}^{k} u_i$; number of cluster k; initial site vector
$v = (v_1^T, \ldots, v_k^T)^T \in \mathbb{R}^{d \cdot k}$;

Result: (k, l, u)-clustering $Cl = (Cl_1, \ldots, Cl_k)$; $\xi = (\xi_{ij}) \in \{0,1\}^{k \times n}$;

```
/* Solve the linear program and return a feasible solution       */
```

$$
\begin{aligned}
\max_{\xi=(\xi_{ij})\in\{0,1\}^{k\times n}} & \sum_{i=1}^{k}\sum_{j=1}^{n} \xi_{ij} v_i^T x_j \\
\sum_{j=1}^{n} \xi_{ij} &\leq u_i \quad (i \leq k) \\
\sum_{j=1}^{n} \xi_{ij} &\geq l_i \quad (i \leq k) \\
\sum_{i=1}^{k} \xi_{ij} &= 1 \quad (j \leq n) \\
\xi_{ij} &\geq 0 \quad (i \leq k, j \leq n),
\end{aligned}
$$

Lemma 5.6.1. *Algorithm 3 computes a (k, l, u)-clustering by linear programming with $k \cdot n$ variables and $(k+1) \cdot n + 2 \cdot k$ constraints.*

Algorithm 3 is the basic module of the following final algorithm for the geometric clustering approach for the division of heterogeneous individual patient data into homogeneous patient collectives.

Algorithm 4: The iterative calculation of the clustering Cl

Data: $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega$ with
$\hat{x}_j = (\hat{x}_{j1}, \hat{x}_{j2}, \ldots, \hat{x}_{jd}) = (\hat{a}_1, \ldots, \hat{a}_d)$; boundaries $l_i, u_i \in \mathbb{N}$
$\sum_{i=1}^{k} l_i \leq n \leq \sum_{i=1}^{k} u_i$; number of cluster k; initial site vector
$v = (v_1^T, \ldots, v_k^T)^T \in \mathbb{R}^{d \cdot k}$;

Result: (k, l, u)-clustering $Cl = (Cl_1, \ldots, Cl_k)$;

1. To get a (k, l, u)-clustering with the correspondent assignment $(\xi_{ij}) \in \{0, 1\}^{k \times n}$, apply algorithm 3 with site vector $v = (v_1^T, \ldots, v_k^T)^T$;

2. Update
$$(v_1, \ldots, v_k) = (\frac{s_1}{\|s_1\|}, \ldots, \frac{s_k}{\|s_k\|}),$$
with $s_i := \sum_{j=1}^{n} \xi_{ij} \hat{x}_j$;

3. If the objective function value
$$\sum_{i=1}^{k} \sum_{j=1}^{n} \xi_{ij} \frac{(\hat{x}_j)^T v_i}{\|v_i\|}$$
increases during the last iteration go to 1., else return the current assignment;

Theorem 5.6.2. *Algorithm 4 terminates with a feasible (k, l, u)-clustering that is a cluster sum assignment.*

The proof for Theorem 5.6.2 can be found in [48]. It can also be shown that the found (k, l, u)-clustering, as a vertex of a bounded-shape partition polytope, induces a power diagram, like it is shorty discussed in Section 5.4. After the computation of the (k, l, u)-clustering $Cl = (Cl_1, \ldots, Cl_k)$, each data point of the transformed training data set $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega$ will be assigned to the corresponding cluster by using the cluster assignment vector.

Definition 5.6.3 (Cluster assignment vector)**.** *Let $Cl = (Cl_1, \ldots, Cl_k)$ be the (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and $\xi = (\xi_{ij}) \in \{0,1\}^{k \times n}$ the corresponding assignment of Cl. Then*

$$Cl(\hat{x}_j) = (Cl_1(\hat{x}_j), \ldots, Cl_k(\hat{x}_j)) \in \{0,1\}^k$$

with

$$Cl_i(\hat{x}_j) := \xi_{ij},\ i = 1, \ldots, k$$

is called the cluster assignment vector of $\hat{x}_j$.

For the evaluation of the performance of the clustering with the help of the testing data set $\hat{S}^{te} = \{(\hat{x}_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega$ and for the prediction of a treatment effect of a patient data point $\{x, y\} \in \mathbb{R}^d \times \Omega$, the assignment is defined as follows.

Definition 5.6.4. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the patient data set $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega$ and $\xi = (\xi_{ij}) \in \{0,1\}^{k \times n}$ the corresponding assignment of Cl. With $c_1, \ldots, c_k$, the centers of gravity, let*

$$d_i(x) := \|c_i - x\|_2,\ i = 1, \ldots, k$$

be the distance of data point x to the center of gravity c_i. Then

$$Cl(x) = (Cl_1(x), \ldots, Cl_k(x)) \in \{0,1\}^k$$

with

$$Cl_i(x) := \begin{cases} 1 & \text{if } d_i \leq d_j,\ \forall j = 1, \ldots, k \text{ and } i \neq j \\ 0 & \text{else} \end{cases}$$

is called the cluster assignment vector of x.

In this definition, the assignment of a data point to a cluster is not necessarily unique. It might hold $\sum_{i=1}^{k} Cl_i(x) \neq 1$. In this case, where a data point is assigned to more than one cluster, we will take the cluster with the highest cluster number.

Definition 5.6.5. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the patient*

data set $\hat{S}$ and let

$$Cl(x) = (Cl_1(x), \ldots, Cl_k(x))$$

be the cluster assignment vector of x then

$$Cl^*(x) = (Cl_1^*(x), \ldots, Cl_k^*(x)) \in \{0,1\}^k$$

with

$$Cl_i^*(x) := \begin{cases} 1 & \text{if } C_k(x) = 0,\ \forall k > i \\ 0 & \text{else} \end{cases}$$

is called the corrected cluster assignment vector of x.

As it is stated in Section 5.3, a bounded-shape clustering, as a vertex of the bounded-shape partition polytope, allows strict linear separation. With this definition, the assignment of a data point x to a cluster is unique and it holds $\sum_{i=1}^{k} Cl_{v_i}^*(x) = 1$.

Definition 5.6.6 (Cluster value)**.** *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and*

$$Cl(x) = (Cl_1(x), \ldots, Cl_k(x))$$

the cluster assignment vector of x. Then the function

$$f(x|Cl) = \sum_{i=1}^{k} f_i(Cl_i) Cl_i(x)$$

is called the cluster value of the clustering Cl for x. $f_i : Cl \to \mathbb{R}; Cl_i \to f_i(Cl_i)$ is called the cluster value of cluster Cl_i.

The cluster value in Definition 5.6.6 is not specified yet. The choice of the cluster value depends on the scientific question. These possible cluster values will be discussed in the following sections.

5.7. Cluster-based meta-analysis

In meta-analysis, we assume that there is one true treatment effect θ of a medication for all patients. There are two different models, the fixed-effects model and the random-effects model, for the determination of an estimated summary treatment effect by the weighted aggregation of the single estimated treatment effects of all considered trials. In comparison to that approach, in the new invented cluster-based meta-analysis, we assume that there is one true treatment effect θ^i for each patient collective Cl_i, $i = 1, \ldots, k$. In each cluster we find individuals with similar characteristic values combinations, which might have a crucial influence on the efficacy of a medical intervention. E.g. elderly male patients might respond differently to a medical intervention than younger female patients. Like it is also discussed in Chapter 4, the assumption of different true treatment effects in clusters of similar patients, by means of similar characteristic values combinations, is more intuitive than one general true treatment effect for all patients. The different estimates for the treatment effect in the patient collectives, depending on the individual outcome, is described in Section 5.7.2. For the determination of the estimation $\tilde{\theta}^i$ of the true treatment effect for each patient collective Cl_i, we present two different new invented model approaches, the cluster-based fixed-effects and the cluster-based random-effects model which will be discussed in Section 5.7.3 and Section 5.7.4. For the justification of this assumption, we use the new invented Q_{Cl}-statistic which will be introduced in Section 5.7.5. In Section 5.7.6, we will describe how heterogeneity within clusters can be assessed. The following theory is based on the joint working paper of Brieden and Hinnenthal [19].

5.7.1. Clustering

For applying the two new invented models, we use the endpoint-oriented geometric clustering method as unsupervised learning approach. Since we want to analyze the efficacy of a treatment in the patient collectives, it is important to include all available characteristics, which might have an influence on the treatment effect, except the administered medication. Then, the patient data set

$S = S^{all}$, with all relevant characteristic values and the outcome of a patient, is transformed like it is discussed in Section 5.5. The clustering Algorithm 4 is then applied on the transformed patient data set

$$\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega.$$

We then obtain a (k, l, u)-clustering $Cl = (Cl_1, \ldots, C_k)$ as a partition of the transformed patient data set. Each cluster now represents a collective with patients of similar combinations of their characteristic values. The required cluster value according to Definition 5.6.6 for each cluster then results from the new invented fixed-effects or random-effects model, as summary treatment effect of the corresponding cluster. But before we get to the introduction of those models, we need some basic terms which are defined in the following. We begin with the definition of the study index set of cluster Cl_i.

Definition 5.7.1. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$. Then*

$$\mathcal{I}_{Cl_i} := \{j |\ \textit{study } j \textit{ is included in cluster } Cl_i\} \subseteq \{1, \ldots, n_{st}\}$$

is the study index set of cluster Cl_i and includes the indices of all studies included in cluster Cl_i, for $i = 1, \ldots, k$. Thereby, n_{st} denotes the number of all studies.

For the definition of the treatment effect for cardinal outcome data, we need to specify the patient data set of the control and the treatment group of study j in cluster Cl_i.

Definition 5.7.2. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S} = \{(\hat{x}_l, y_l)\}_{l=1}^{n}$, $\xi = (\xi_{il}) \in \{0, 1\}^{k \times n}$ the corresponding assignment of Cl, st_l the study assignment vector and t_l the treatment group assignment for patient l. Then*

$$\hat{S}_{T_j}^{i} := \{(\hat{x}_l, y_l) | \xi_{il} = 1 \wedge st_{lj} = 1 \wedge t_l = 1\}_{l=1}^{n} \subseteq \hat{S}$$

is the patient data set of the treatment group and

$$\hat{S}^i_{C_j} := \{(\hat{x}_l, y_l) | \xi_{il} = 1 \wedge st_{lj} = 1 \wedge t_l = 0\}_{l=1}^n \subseteq \hat{S}$$

is the patient data set of the control group of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

If there is a low number of patients included in the single trials considered for the cluster-based meta-analysis, the cardinality of the sets $\hat{S}^i_{T_j}$ and $\hat{S}^i_{C_j}$ might be too low to guarantee high statistical power. In this case, it is possible to include the characteristic variable 'study' as influencing variable. The treatment effect of a patient collective is then not calculated by aggregating the treatment effects of the singe trials but by calculating a treatment effect estimate, like the Risk Ratio, out of the outcome of all patients included in the corresponding cluster without differentiating between the single trials. Therefore, we only have to summarize all trials to one single trial and set $n_{st} = 1$. This approach would then increase the statistical power.

5.7.2. Treatment effect estimates

In Section 3.2 we discussed treatment effect estimates for binary and cardinal data in meta-analysis. Those treatment effect estimates are calculated for each study included in the meta-analysis. In the cluster-based meta-analysis, the treatment effect estimates are determined for each study included in a cluster. In case of binary outcome data of the participating patients, several estimations are defined, like the transformed Risk Ratio or the transformed Odds Ratio of a study in the corresponding cluster. In case of cardinal data, the treatment effect is represented by the mean difference or the standardized mean difference of the outcome of the treatment and control group in the identified patient collectives.

Binary data

Before we define the cluster-based treatment effect estimates based on the binary outcome of the study participants, we need to have a look at the marginal and joint distribution of the patients of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. We assume that the outcome of an individual in the treatment group

	success	failure	
Treatment group	$s^i_{T_j}$	$f^i_{T_j}$	$n^i_{T_j}$
Control group	$s^i_{C_j}$	$f^i_{C_j}$	$n^i_{C_j}$
	s^i_j	f^i_j	n^i_j

Table 5.1.: 2×2-table of binary outcome of study j in cluster Cl_i

of study j in cluster Cl_i is seen as a Bernoulli distributed random variable

$$T^i_j \sim Be(p^i_{T_j}),$$

with unknown probability of success $p^i_{T_j}$. The outcome of an individual in the control group of study j in cluster Cl_i is also assumed to be Bernoulli distributed,

$$C^i_j \sim Be(p^i_{C_j}),$$

with the unknown probability of success $p^i_{C_j}$. The occurrences success, represented by 1, and failure, represented by 0, of study j in cluster Cl_i, for the treatment and the control group can be summarized in Table 5.1.

For binary data, we define the Risk Ratio RR^i_j, the Odds Ratio OR^i_j and the Risk Difference RD^i_j of study j in cluster Cl_i based on Table 5.1. Thereby, $n^i_{T_j}$ denotes the absolute number of patients in the treatment group and $n^i_{C_j}$ the absolute number of patients in the control group of study j in cluster Cl_i. The number of patients with success of study j in cluster Cl_i is denoted by s^i_j and the patients with failure is denoted by f^i_j. For the joint distribution, the number of patients with success in the treatment group of study j in cluster Cl_i is defined by $s^i_{T_j}$ and the number of those in the control group is consequently denoted by $s^i_{C_j}$. Analogously, the number of patients with failure in the treatment group $f^i_{T_j}$ and in the control group $f^i_{C_j}$ are defined.

The number of patients with success in the treatment group of study j in cluster Cl_i is binomial distributed,

$$S^i_{T_j} \sim B(n^i_{T_j}, p^i_{T_j}, s^i_{T_j}),$$

just as the number of patients with success in the control group,

$$S^i_{C_j} \sim B(n^i_{C_j}, p^i_{C_j}, s^i_{C_j}).$$

Since the probabilities of success of the treatment and the control group of a study in the corresponding cluster are also unknown, we need an estimation for these probabilities.

Theorem 5.7.3. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let Table 5.1 be the 2×2-table of the binary patients' outcome of study j in cluster Cl_i. Then*

$$\tilde{P}^i_{T_j} = \frac{S^i_{T_j}}{n^i_{T_j}} \text{ and } \tilde{P}^i_{C_j} = \frac{S^i_{C_j}}{n^i_{C_j}}$$

are the estimators for the probability of success in the treatment and the control group of study j in cluster Cl_i and

$$\tilde{p}^i_{T_j} = \frac{s^i_{T_j}}{n^i_{T_j}} \text{ and } \tilde{p}^i_{C_j} = \frac{s^i_{C_j}}{n^i_{C_j}},$$

the corresponding estimations, for $i = 1, \dots, k$ and $j \in \mathcal{I}_{Cl_i}$.

Proof. The likelihood function for the estimator $\tilde{P}^i_{T_j}$ of $p^i_{T_j}$ is

$$L(s^i_{T_j}, p^i_{T_j}) = \binom{n^i_{T_j}}{s^i_{T_j}} p^i_{T_j} (1 - p^i_{T_j})^{n^i_{T_j} - s^i_{T_j}}.$$

With the maximization of this function we get

$$\tilde{p}^i_{T_j} = \frac{s^i_{T_j}}{n^i_{T_j}},$$

for $i = 1, \dots, k$ and $j \in \mathcal{I}_{Cl_i}$. The same holds for the estimation $\tilde{p}^i_{C_j}$. □

Based on Table 5.1, we are able to define the cluster-based treatment effect estimations, the Risk Ratio RR^i_j, the Odds Ratio OR^i_j and the Risk Difference RD^i_j of study j in cluster Cl_i. We start by defining the Risk Ratio estimation.

Definition 5.7.4. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let Table 5.1 be the 2×2-table of the binary patients' outcome of study j in cluster Cl_i. Then*

$$\tilde{\Theta}_j^i := \ln\left(\frac{\tilde{P}_{T_j}^i}{\tilde{P}_{C_j}^i}\right) = \ln\left(\frac{S_{T_j}^i n_{C_j}^i}{S_{C_j}^i n_{T_j}^i}\right)$$

is the estimator for the treatment effect of study j in cluster Cl_i. The estimation

$$\mathrm{RR}_j^i := \frac{\tilde{p}_{T_j}^i}{\tilde{p}_{C_j}^i} = \frac{s_{T_j}^i n_{C_j}^i}{s_{C_j}^i n_{T_j}^i}$$

is called the Risk Ratio and

$$\tilde{\theta}_j^i := \ln(\mathrm{RR}_j^i)$$

the transformed Risk Ratio of study j in cluster Cl_i, for $i = 1, \dots, k$ and $j \in \mathcal{I}_{Cl_i}$.

Like it is discussed in the next section, we assume that the estimated treatment effect $\tilde{\theta}_j^i$ of study j in cluster Cl_i is a realization of the approximately normally distributed random variable

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta^i, \mathrm{var}(\tilde{\Theta}_j^i),$$

with unknown expected value θ^i and variance $\mathrm{var}(\tilde{\Theta}_j^i)$. The variance $\mathrm{var}(\tilde{\Theta}_j^i)$ of the transformed Risk Ratio of study j in cluster Cl_i can be approximated as follows by using the delta method.

Theorem 5.7.5. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let Table 5.1 be the 2×2-table of the binary patients' outcome of study j in cluster Cl_i. Then*

$$(\tilde{\Sigma}_j^i)^2 = \frac{1}{S_{T_j}^i} - \frac{1}{n_{T_j}^i} + \frac{1}{S_{C_j}^i} - \frac{1}{n_{C_j}^i}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j^i$ of study j in cluster Cl_i, with the corresponding estimation

$$(\tilde{\sigma}_j^i)^2 = \frac{1}{s_{T_j}^i} - \frac{1}{n_{T_j}^i} + \frac{1}{s_{C_j}^i} - \frac{1}{n_{C_j}^i},$$

and $\text{var}(\tilde{\Theta}_j^i) \approx (\tilde{\sigma}_j^i)^2$, *for* $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

The next cluster-based treatment effect estimate is the Odds Ratio of study j in cluster Cl_i.

Definition 5.7.6. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$ *and let Table 5.1 be the* 2×2*-table of the binary patients' outcome of study* j *in cluster* Cl_i. *Then*

$$\tilde{\Theta}_j^i := ln\left(\frac{\tilde{P}_{T_j}^i(1-\tilde{P}_{C_j}^i)}{\tilde{P}_{C_j}^i(1-\tilde{P}_{T_j}^i)}\right)$$

is the estimator for the treatment effect of study j *in cluster* Cl_i. *The estimation*

$$\text{OR}_j^i := \frac{\tilde{p}_{T_j}^i(1-\tilde{p}_{C_j}^i)}{\tilde{p}_{C_j}^i(1-\tilde{p}_{T_j}^i)}$$

is called the Odds Ratio and

$$\tilde{\theta}_j^i := \ln(\text{OR}_j^i)$$

the transformed Odds Ratio of study j *in cluster* Cl_i, *for* $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

The approximation of the variance of the transformed Odds Ratio of study j in cluster Cl_i is given in the next theorem due by using the delta method.

Theorem 5.7.7. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$ *and let Table 5.1 be the* 2×2*-table of the binary patients' outcome of study* j *in cluster* Cl_i. *Then*

$$(\tilde{\Sigma}_j^i)^2 := \frac{1}{S_{T_j}^i} + \frac{1}{f_{T_j}^i} + \frac{1}{S_{C_j}^i} + \frac{1}{f_{C_j}^i}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j^i$ *of study* j *in cluster* Cl_i, *with corresponding estimation*

$$(\tilde{\sigma}_j^i)^2 := \frac{1}{s_{T_j}^i} + \frac{1}{f_{T_j}^i} + \frac{1}{s_{C_j}^i} + \frac{1}{f_{C_j}^i},$$

and $\text{var}(\tilde{\Theta}_j^i) \approx (\tilde{\sigma}_j^i)^2$, *for* $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

The last cluster-based treatment effect estimate is the Risk Difference of study j in cluster Cl_i. We close this section with the definition of the Risk Difference and the determination of the approximated variance by using the delta method.

Definition 5.7.8. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let Table 5.1 be the 2×2-table of the binary patients' outcome of study j in cluster Cl_i. Then*

$$\tilde{\Theta}^i_j := \tilde{P}^i_{T_j} - \tilde{P}^i_{C_j}$$

is the estimator for the treatment effect of study j in cluster Cl_i. The corresponding estimation

$$\tilde{\theta}^i_j = \mathrm{RD}^i_j := \tilde{p}^i_{T_j} - \tilde{p}^i_{C_j}$$

is called the Risk Difference of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

Theorem 5.7.9. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let Table 5.1 be the 2×2-table of the binary patients' outcome of study j in cluster Cl_i. Then*

$$(\tilde{\Sigma}^i_j)^2 := \frac{S^i_{T_j} f^i_{T_j}}{(n^i_{T_j})^3} + \frac{S^i_{C_j} f^i_{C_j}}{(n^i_{C_j})^3}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}^i_j$ of study j in cluster Cl_i, with corresponding estimation

$$(\tilde{\sigma}^i_j)^2 := \frac{s^i_{T_j} f^i_{T_j}}{(n^i_{T_j})^3} + \frac{s^i_{C_j} f^i_{C_j}}{(n^i_{C_j})^3},$$

and $\mathrm{var}(\tilde{\Theta}^i_j) \approx (\tilde{\sigma}^i_j)^2$, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

Cardinal data

Like it is described in Section 3.2.2, we assume that for study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$, the outcome $Y^i_{lT_j}$ of individual l, $l = 1, \ldots, n^i_{T_j}$, in the treatment group is approximately normally distributed,

$$Y^i_{lT_j} \sim \mathcal{N}(y^i_{T_j}, (\sigma^i_{T_j})^2),$$

with realization $y^i_{lT_j}$, unknown expected value $y^i_{T_j}$ and variance $(\sigma^i_{T_j})^2$. The outcome $Y^i_{lC_j}$ of individual l, $l = 1, \ldots, n^i_{C_j}$, in the control group is also assumed to be at least approximately normally distributed,

$$Y^i_{lC_j} \sim \mathcal{N}(y^i_{C_j}, (\sigma^i_{C_j})^2),$$

with realization $y^i_{lC_j}$, unknown expected value $y^i_{C_j}$ and variance $(\sigma^i_{C_j})^2$. We use the arithmetic mean value as estimator for the expected outcome in the treatment and control group of study j in cluster Cl_i.

Theorem 5.7.10. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let $\hat{S}^i_{T_j} = \{(\hat{x}^i_{lT_j}, y^i_{lT_j})\}_{l=1}^{n^i_{T_j}}$ be the patient data set in the treatment group of study j in cluster Cl_i. Then*

$$\tilde{Y}^i_{T_j} := \frac{1}{n^i_{T_j}} \sum_{i=1}^{n^i_{T_j}} Y^i_{lT_j}$$

is the unbiased estimator for the expected outcome in the treatment group of study j in cluster Cl_i and

$$\tilde{y}^i_{T_j} := \frac{1}{n^i_{T_j}} \sum_{l=1}^{n^i_{T_j}} y^i_{lT_j}$$

the corresponding estimation, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

Theorem 5.7.11. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let $\hat{S}^i_{C_j} = \{(\hat{x}^i_{lC_j}, y^i_{lC_j})\}_{l=1}^{n^i_{C_j}}$ be the patient data set in the control group of study j in cluster Cl_i. Then*

$$\tilde{Y}^i_{C_j} := \frac{1}{n^i_{C_j}} \sum_{i=1}^{n^i_{C_j}} Y^i_{lC_j}$$

is the unbiased estimator for the expected outcome in the control group of study j and

$$\tilde{y}^i_{C_j} := \frac{1}{n^i_{C_j}} \sum_{i=1}^{n^i_{C_j}} y^i_{lC_j}$$

the corresponding estimation, for $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

The estimators of variance $(\sigma^i_{T_j})^2$ and $(\sigma^i_{C_j})^2$ of the patients' outcome in the treatment and control group of study j in cluster Cl_i are given in the following theorems.

Theorem 5.7.12. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$, $\hat{S}^i_{T_j} = \{(\hat{x}^i_{lT_j}, y^i_{lT_j})\}_{l=1}^{n^i_{T_j}}$ *be the treatment group and* $\tilde{Y}^i_{T_j}$ *be the estimator for the expected outcome in the treatment group, with estimation* $\tilde{y}^i_{T_j}$*, of study* j *in cluster* Cl_i*. Then*

$$(\tilde{\Sigma}^i_{T_j})^2 := \frac{1}{n^i_{T_j} - 1} \sum_{l=1}^{n^i_{T_j}} (Y^i_{lT_j} - \tilde{Y}^i_{T_j})^2$$

is the unbiased estimator for the variance of the outcome in the treatment group of study j *in cluster* Cl_i *and*

$$(\tilde{\sigma}^i_{T_j})^2 := \frac{1}{n^i_{T_j} - 1} \sum_{l=1}^{n^i_{T_j}} (y^i_{lT_j} - \tilde{y}^i_{T_j})^2$$

the corresponding estimation, for $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

Theorem 5.7.13. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$, $\hat{S}^i_{C_j} = \{(\hat{x}^i_{lC_j}, y^i_{lC_j})\}_{l=1}^{n^i_{C_j}}$ *be the control group and* $\tilde{Y}^i_{C_j}$ *be the estimator for the expected outcome in the control group, with estimation* $\tilde{y}^i_{C_j}$*, of study* j *in cluster* Cl_i*. Then*

$$(\tilde{\Sigma}^i_{C_j})^2 := \frac{1}{n^i_{C_j} - 1} \sum_{l=1}^{n^i_{C_j}} (Y^i_{lC_j} - \tilde{Y}^i_{C_j})^2$$

is the unbiased estimator for the variance of the outcome in the control group of study j *and*

$$(\tilde{\sigma}^i_{C_j})^2 := \frac{1}{n^i_{C_j} - 1} \sum_{l=1}^{n^i_{C_j}} (y^i_{lC_j} - \tilde{y}^i_{C_j})^2$$

the corresponding estimation, for $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

The absolute difference between the mean outcome of the treatment and the control group of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$ is then specified in the following definition.

Definition 5.7.14. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$, let $\tilde{Y}^i_{T_j}$ be the estimator for the expected outcome in the treatment group, with estimation $\tilde{y}^i_{T_j}$, and let $\tilde{Y}^i_{C_j}$ be the estimator for the expected outcome in the control group, with estimation $\tilde{y}^i_{C_j}$, of study j in cluster Cl_i. Then*

$$\tilde{\Theta}^i_j := \tilde{Y}^i_{T_j} - \tilde{Y}^i_{C_j}$$

is the estimator for the treatment effect of study j. The corresponding estimation

$$\tilde{\theta}^i_j := \tilde{y}^i_{T_j} - \tilde{y}^i_{C_j}$$

is called the absolute difference between the mean outcomes in the treatment and control group of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

The effect measure value $\tilde{\theta}^i_j$ is the realization of the approximately normally distributed random variable

$$\tilde{\Theta}^i_j \sim \mathcal{N}(\theta^i, \mathrm{var}(\tilde{\Theta}^i_j)),$$

with unknown true treatment effect θ^i and variance $\mathrm{var}(\tilde{\Theta}^i_j)$ which can be approximated as follows.

Remark 5.7.15. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$, let $(\tilde{\Sigma}^i_{T_j})^2$ be the estimator for the variance of the outcome in the treatment group, with realizations $(\tilde{\sigma}^i_{T_j})^2$. Furthermore let $(\tilde{\Sigma}^i_{C_j})^2$ the estimator for the variance in the control group, with realizations $(\tilde{\sigma}^i_{C_j})^2$, of study j in cluster Cl_i. Then*

$$(\tilde{\Sigma}^i_j)^2 := \frac{(\tilde{\Sigma}^i_{T_j})^2}{n^i_{T_j}} + \frac{(\tilde{\Sigma}^i_{C_j})^2}{n^i_{C_j}}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}^i_j$ of study j

in cluster Cl_i,

$$(\tilde{\sigma}_j^i)^2 := \frac{(\tilde{\sigma}_{T_j}^i)^2}{n_{T_j}^i} + \frac{(\tilde{\sigma}_{C_j}^i)^2}{n_{C_j}^i}$$

is the corresponding estimation and $\mathrm{var}(\tilde{\Theta}_j^i) \approx (\tilde{\sigma}_j^i)^2$, *for* $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

If we assume that the variance $(\sigma_{T_j}^i)^2$ of the outcome of individuals in the treatment group and the variance $(\sigma_{C_j}^i)^2$ of the outcome of individuals in the control group of study j in cluster Cl_i are the same,

$$(\sigma_{T_j}^i)^2 = (\sigma_{C_j}^i)^2 =: (\sigma_{P_j}^i)^2,$$

like it is assumed in most of the parametric data analysis techniques, then for the variance of the estimated treatment effect $\tilde{\Theta}_j^i$ holds

$$\mathrm{var}(\tilde{\Theta}_j^i) = (\sigma_{P_j}^i)^2 \left(\frac{n_{T_j}^i + n_{C_j}^i}{n_{T_j}^i n_{C_j}^i} \right)$$

[5]. But before we get to this definition, we need an estimate for variance $(\sigma_{P_j}^i)^2$ which is specified in the following remark.

Remark 5.7.16. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$, *let* $(\tilde{\Sigma}_{T_j}^i)^2$ *be the estimator for the variance of the outcome in the treatment group with realization* $(\tilde{\sigma}_{T_j}^i)^2$. *Furthermore, let* $(\tilde{\Sigma}_{C_j}^i)^2$ *be the estimator for the variance in the control group, with realization* $(\tilde{\sigma}_{C_j}^i)^2$, *of study* j *in cluster* Cl_i. *Then*

$$(\tilde{\Sigma}_{P_j}^i)^2 := \frac{(n_{T_j}^i - 1)(\tilde{\Sigma}_{T_j}^i)^2 + (n_{C_j}^i - 1)(\tilde{\Sigma}_{C_j}^i)^2}{n_{T_j}^i + n_{T_j}^i - 2},$$

is the estimator for the variance of the pooled outcome of the treatment and control group of study j *in cluster* Cl_i *and*

$$(\tilde{\sigma}_{P_j}^i)^2 := \frac{(n_{T_j}^i - 1)(\tilde{\sigma}_{T_j}^i)^2 + (n_{C_j}^i - 1)(\tilde{\sigma}_{C_j}^i)^2}{n_{T_j}^i + n_{T_j}^i - 2},$$

the corresponding estimation, for $i = 1, \ldots, k$ *and* $j \in \mathcal{I}_{Cl_i}$.

Now the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j^i$ can

be formulated.

Remark 5.7.17. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$, let $(\tilde{\Sigma}^i_{P_j})^2$ be the estimator for the variance of the pooled outcome in the treatment and control group of study j in cluster Cl_i, with realization $(\tilde{\sigma}^i_{P_j})^2$. Then*

$$(\tilde{\Sigma}^i_j)^2 := (\tilde{\Sigma}^i_{P_j})^2 \left(\frac{n^i_{T_j} + n^i_{C_j}}{n^i_{T_j} n^i_{C_j}} \right)$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j$ of study j in cluster Cl_i,

$$(\tilde{\sigma}^i_j)^2 := (\tilde{\sigma}^i_{P_j})^2 \left(\frac{n^i_{T_j} + n^i_{C_j}}{n^i_{T_j} n^i_{C_j}} \right)$$

is the corresponding estimation and $\mathrm{var}(\tilde{\Theta}^i_j) \approx (\tilde{\sigma}^i_j)^2$, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

The second effect measure for cardinal data is the standardized difference between the mean outcome in the treatment and control group of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. This standardized difference is defined as follows.

Definition 5.7.18. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and let $\tilde{Y}^i_{T_j}$ be the estimator for the expected outcome in the treatment group, with estimation $\tilde{y}^i_{T_j}$, and let $\tilde{Y}^i_{C_j}$ be the estimator for the expected outcome in the control group, with estimation $\tilde{y}^i_{C_j}$, and $(\tilde{\Sigma}^i_{P_j})^2$ be the variance of the pooled outcome of the treatment and control group, with realization $(\tilde{\sigma}^i_{P_j})^2$, of study j in cluster Cl_i. Then*

$$\tilde{\Theta}^i_j := \frac{\tilde{Y}^i_{T_j} - \tilde{Y}^i_{C_j}}{\tilde{\Sigma}^i_{P_j}}$$

is the estimator for the treatment effect of study j. The corresponding estimation

$$\tilde{\theta}^i_j := \frac{\tilde{y}^i_{T_j} - \tilde{y}^i_{C_j}}{\tilde{\sigma}^i_{P_j}}$$

is called the standardized difference between the mean outcomes in the treatment and control group of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

As for the absolute difference, the estimated treatment effect $\tilde{\theta}_j^i$ is the realization of the approximately normally distributed random variable

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta^i, \mathrm{var}(\tilde{\Theta}_j^i)),$$

with unknown true treatment effect θ^i and variance $\mathrm{var}(\tilde{\Theta}_j^i)$. The variance of $\tilde{\Theta}_j^i$ can be approximated as follows [77].

Remark 5.7.19. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$. Then let n_j^i be the number of patients of study j in cluster Cl_i. Furthermore, let $n_{T_j}^i$ be the number of patients in the treatment group and $n_{C_j}^i$ the number of patients in the control group of study j in cluster Cl_i. Then*

$$(\tilde{\Sigma}_j^i)^2 := \frac{n_j^i}{n_{T_j}^i n_{C_j}^i}$$

is the estimator for the variance of the estimated treatment effect $\tilde{\Theta}_j^i$, $(\tilde{\sigma}_j^i)^2$ the corresponding estimation and $\mathrm{var}(\tilde{\Theta}_j^i) \approx (\tilde{\sigma}_j^i)^2$, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$.

5.7.3. Cluster-based fixed-effects model

In the following, we will introduce the new invented cluster-based fixed-effects model for the determination of the summary treatment effects of the patient collectives identified by the endpoint-oriented geometric clustering approach. We assume that all required assumptions, which are already described in Section 3.3.2, are fulfilled. Like it is discussed in the introductory paragraph of this section, in the cluster-based fixed-effects model we assume that there is one true fixed treatment effect θ^i for each patient collective i, for $i = 1, \ldots, k$. The estimation $\tilde{\theta}^i$ of the true treatment effect is calculated by the weighted aggregation of the treatment effect estimates of the single studies included in cluster Cl_i. For this aggregation let $\tilde{\Theta}_j^i$ be the independent approximately normally distributed

estimated treatment effect of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$,

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta^i, (\sigma_j^i)^2)$$

with unknown expected value θ^i, the true treatment effect in cluster Cl_i, and variance $(\sigma_j^i)^2$. For cluster Cl_i the cluster-based fixed-effects model is given by

$$\tilde{\theta}_j^i := \theta^i + \epsilon_j^i,$$

where $\tilde{\theta}_j^i$ is the observed treatment effect and ϵ_j^i the estimation error of study j in cluster Cl_i. The estimation error is a realization of the normally distributed random variable

$$\mathcal{E}_j^i \sim \mathcal{N}(0, (\sigma_j^i)^2),$$

with expected estimation error 0. The variance of $\mathcal{E}_j^i$ is denoted by $(\sigma_j^i)^2$ and is also called intra-study variance within study j in cluster Cl_i. In the cluster-based fixed-effects model the observed variance of $\tilde{\Theta}_j^i$ is also treated as if it were the true variance. From this it follows that the estimated treatment effect $\tilde{\Theta}_j^i$ of study j in cluster Cl_i is normally distributed,

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta^i, (\tilde{\sigma}_j^i)^2)),$$

with expected value θ^i, the true treatment effect of cluster Cl_i, and observed variance $(\tilde{\sigma}_j^i)^2$.

To get an estimation of the true treatment effect θ^i in cluster Cl_i we need to use the estimator $\tilde{\Theta}^i$ determined by the weighted mean of the existent estimated treatment effects of all studies included in cluster Cl_i. Like it is done in the already presented fixed-effects model in meta-analysis, for the aggregation we apply the inverse variance method. With this approach the estimated summary treatment effect in cluster Cl_i, for $i = 1, \ldots, k$, is described by

$$\tilde{\theta}^i = \frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\theta}_j^i w_j^i}{\sum_{j \in \mathcal{I}_{Cl_i}} w_j^i},$$

as a realization of the unbiased estimator

$$\tilde{\Theta}^i = \frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\Theta}^i_j w^i_j}{\sum_{j \in \mathcal{I}_{Cl_i}} w^i_j} \text{ with } w^i_j = \frac{1}{(\tilde{\sigma}^i_j)^2},$$

for $j \in \mathcal{I}_{Cl_i}$.

By using the inverse variance method, studies in cluster Cl_i with a smaller variance, by means of a more precise treatment effect, are given more weight than studies with larger variance. The estimator of the summary treatment effect of cluster Cl_i is also normally distributed,

$$\tilde{\Theta}^i \sim \mathcal{N}(\theta^i, \text{var}(\tilde{\Theta}^i),$$

with expected value θ^i and variance $\text{var}(\tilde{\Theta}^i)$. The variance can be determined by

$$\text{var}(\tilde{\Theta}^i) = \text{var}\left(\frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\Theta}^i_j w^i_j}{\sum_{j \in \mathcal{I}_{Cl_i}} w^i_j}\right) = \frac{1}{\sum_{j \in \mathcal{I}_{Cl_i}} w^i_j}. \tag{5.1}$$

With this information, we can specify the confidence interval for the true treatment effect in cluster Cl_i. Since

$$\frac{\tilde{\Theta}^i - \theta^i}{\sqrt{\text{var}(\tilde{\Theta}^i)}}$$

is standardized normally distributed, for the $(1-\alpha)$ confidence interval $\mathcal{I}^i$ of the true treatment effect θ^i of cluster Cl_i follows

$$\mathcal{I}^i = [\tilde{\theta}^i - z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}, \tilde{\theta}^i + z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}],$$

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution. With this result, we can conclude, that in $(1-\alpha)\cdot 100\%$ of all confidence intervals, built on the basis of random patients of cluster Cl_i, the true treatment effect θ^i is included,

$$P(\theta^i \in \mathcal{I}^i) = 1 - \alpha.$$

In a last step, we define the resulting cluster value independent on the chosen

treatment effect estimate.

Definition 5.7.20. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ and $\tilde{\theta}^i$ the estimated summary treatment effect of cluster Cl_i. Then the cluster value is defined by*

$$f_i(Cl_i) := \tilde{\theta}^i,$$

for $i = 1, \ldots, k$.

5.7.4. Cluster-based random-effects model

In this section, we will discuss the cluster-based random-effects model. This model is used if we assume that heterogeneity might still exist across the single studies in a patient collective due to some further random-effect. For the identification of the still existent heterogeneity we use the Q^i-test introduced in Section 5.7.6. We assume that all required assumptions, which are already described in Section 3.3.2, are fulfilled. In this model approach, we also assume that there is a treatment effect θ^i for each patient collective Cl_i, for $i = 1, \ldots, k$, which is estimated by aggregating all treatment effects of the single trials included in the corresponding cluster. Similar to the random-effects model described in Section 3.3.2, this new invented model allows the variation of the true treatment effect in the patient collective Cl_i across all included studies. Thus, we assume that the treatment effect θ_j^i of study j in cluster Cl_i is a realization of the random variable

$$\Theta_j^i \sim \mathcal{N}(\theta^i, (\tau^i)^2),$$

with expected value θ^i, the true treatment effect in cluster Cl_i, and variance $(\tau^i)^2$, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. In the cluster-based random-effects model for the realization of Θ_j^i holds

$$\theta_j^i := \theta^i + \nu_j^i,$$

with the study-specific random-effect ν_j^i of study j in cluster Cl_i which is a realization of the normally distributed random variable,

$$N \sim \mathcal{N}(0, (\tau^i)^2),$$

with expected random-effect 0 and variance $(\tau^i)^2$. We also call $(\tau^i)^2$ the between-study or inter-study variance within cluster Cl_i. It measures the difference between the treatment effect of study j and the treatment effect under all patients in cluster Cl_i. Furthermore, let $\tilde{\Theta}_j^i$ be an independent normally distributed treatment effect observation,

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta_j^i, (\sigma_j^i)^2),$$

with correspondent unknown expected value θ_j^i and variance $(\sigma_j^i)^2$ in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. The observed treatment effect of study j in cluster Cl_i can be described by

$$\tilde{\theta}_j^i = \theta_j^i + \epsilon_j^i,$$

where ϵ_j^i is the estimation error of study j in cluster Cl_i and a realization of the normally distributed random variable

$$\mathcal{E}_j^i \sim \mathcal{N}(0, (\sigma_j^i)^2),$$

with expected estimation error 0 and variance $(\sigma_j^i)^2$. The cluster-based random-effects model is then given by

$$\tilde{\theta}_j^i := \theta_j^i + \epsilon_j^i = \theta^i + \nu_j^i + \epsilon_j^i,$$

for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$, where ϵ_j^i is the estimation error and ν_j^i the random-effect of study j in cluster Cl_i. Due to the independence of $\mathcal{E}_j^i$ and N it follows that

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta^i, (\sigma_j^i)^2 + (\tau^i)^2).$$

$(\tau^i)^2$ is unknown and has to be estimated analogously to Section 3.3.2. With the help of the underlying data of the cluster-based fixed-effects model we get

$$(\tilde{\tau}^i)^2 = \max\left\{0; \frac{q^i - (n_i - 1)}{\sum_{j \in \mathcal{I}_{Cl_i}} w_j^i - \frac{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^2}{\sum_{j \in \mathcal{I}_{Cl_i}} w_j^i}}\right\},$$

where

$$q^i = \sum_{j \in \mathcal{I}_{Cl_i}} w_j^i (\tilde{\theta}_j^i - \hat{\theta}^i)^2$$

is the realization of the Q^i-statistic for cluster Cl_i, with summary treatment effect $\tilde{\theta}^i$ and weight w_j^i given by the cluster-based fixed-effects model and observed treatment effect $\tilde{\theta}_j^i$ of study j in cluster Cl_i. $n_i = |\mathcal{I}_{Cl_i}|$ denotes the number of studies included in cluster Cl_i. Due to the assumption that the true variance of the estimated treatment effect $\tilde{\Theta}_j^i$ of study j in cluster Cl_i corresponds to the variance $(\tilde{\sigma}_j^i)^2 + (\tilde{\tau}^i)^2$, we get

$$\tilde{\Theta}_j^i \sim \mathcal{N}(\theta^i, (w_j^i)^{-1} + (\tilde{\tau}^i)^2),$$

with $(w_j^i)^{-1} = (\tilde{\sigma}_j^i)^2$ defined like it is done for the cluster-based fixed-effects model. For the inverse variance method of the cluster-based random-effects model and the determination of the estimated summary treatment effect in cluster Cl_i the weights are given by

$$(w_j^i)^* = \frac{1}{(w_j^i)^{-1} + (\tilde{\tau}^i)^2}.$$

The unbiased maximum likelihood estimator of the summary treatment effect of cluster Cl_i is then given by

$$\tilde{\Theta}^i = \frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\Theta}_j^i (w_j^i)^*}{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^*},$$

with the realization

$$\tilde{\theta}^i = \frac{\sum_{j=1}^{n_i} \tilde{\theta}_j^i (w_j^i)^*}{\sum_{j=1}^{n_i} (w_j^i)^*}.$$

The estimator $\tilde{\Theta}^i$ of the summary treatment effect in cluster Cl_i is normally distributed

$$\tilde{\Theta}^i \sim \mathcal{N}(\theta^i, \mathrm{var}(\tilde{\Theta}^i)), \tag{5.2}$$

with expected value θ^i and variance $\mathrm{var}(\tilde{\Theta}^i)$ which can be determined by

$$\mathrm{var}(\tilde{\Theta}^i) = \mathrm{var}\left(\frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\Theta}_j^i (w_j^i)^*}{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^*} \right) = \frac{1}{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^*}.$$

Now, we can define the confidence interval of the true treatment effect θ^i of cluster Cl_i due to the observation that the quotient

$$\frac{\tilde{\Theta}^i - \theta^i}{\sqrt{\text{var}(\tilde{\Theta}^i)}}$$

is standardized normally distributed. For the $(1-\alpha)$ confidence interval follows

$$\mathcal{I}^i = [\tilde{\theta}^i - z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}, \tilde{\theta}^i + z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}],$$

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution. From this it follows that

$$P(\theta^i \in \mathcal{I}^i) = 1-\alpha.$$

The resulting cluster value for the cluster-based random-effects model is defined according to Definition 5.7.20.

5.7.5. Justification of different treatment effects across clusters

To justify the assumption that there is a true treatment effect θ^i in each cluster or patient collective Cl_i, we assess the heterogeneity across the clusters identified by the endpoint-oriented clustering algorithm. Therefore, we use a statistical test procedure in analogy to the Cochrans's χ^2 test, shown in Figure 5.1.

For the justification, our goal is to reject the null hypothesis formulated by

$$H_0 : \theta^1 = \cdots = \theta^k = \theta,$$

where θ^i, for $i = 1, \ldots, k$, is the underlying true treatment effect of the corresponding cluster Cl_i. We consequently assume that there is only one true effect for all clusters and therefore, for all patients independently of their characteristic values combinations. With this test we want to show, that the formulated null hypothesis can be rejected due to the identification of homogeneous pa-

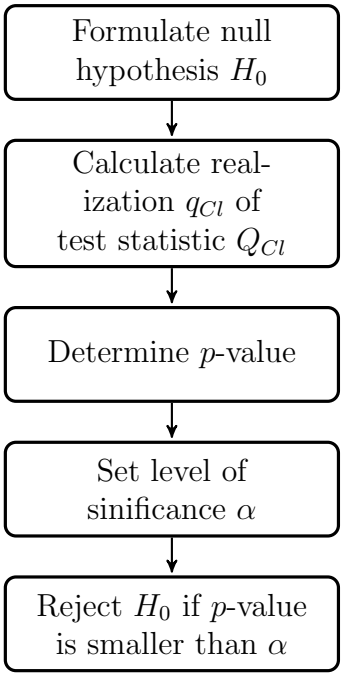

Figure 5.1.: Hypothesis test procedure for the justification of different treatment effects across clusters

tient collectives where patients with similar characteristic values combinations and therefore with similar true treatment effects are grouped. The alternative hypothesis is then given by

$$H_1 : \exists l \in \{1, \ldots, k\} : \theta^l \neq \theta^i, \forall i \in \{1, \ldots, k\}, i \neq l.$$

In the next step of the hypothesis test procedure, we need to specify a test statistic. As this hypothesis test is in analogy to the Cochrans's χ^2 test we will call the test statistic Q_{Cl}-statistic. For the assessment of the heterogeneity across the identified clusters, we have to calculate the realization of the test statistic

$$Q_{Cl} = \sum_{i=1}^{k} w^i (\tilde{\Theta}^i - \bar{\Theta})^2,$$

where $\bar{\Theta}$ is the estimator for the mean treatment effect of the defined clusters,

$$\bar{\Theta} = \frac{1}{k} \sum_{i=1}^{k} \tilde{\Theta}^i.$$

$(w^i)^{-1} = \text{var}(\tilde{\Theta}^i)$ denotes the variance of the estimated summary treatment effect of cluster Cl_i, calculated by equation 5.1 or 5.2, depending on the chosen cluster-based model. The realization of the Q_{Cl}-statistic is then given by

$$q_{Cl} = \sum_{i=1}^{k} w^i(\tilde{\theta}^i - \bar{\theta})^2,$$

with estimation

$$\bar{\theta} = \frac{1}{k}\sum_{i=1}^{k} \tilde{\theta}^i$$

of the mean treatment effect of the identified clusters. q_{Cl} is the weighted squared deviation of the summary treatment effects of the single clusters from the estimated mean treatment effect. As we know from Section 4.2.1, Q_{Cl} approximately follows a χ^2 distribution with $(k-1)$ degrees of freedom. After the calculation of q_{Cl}, we need to decide whether we reject the null hypothesis or not. Therefore we use the p-value,

$$pv = P(Q_{Cl} \geq q_{Cl}|H_0) = 1 - F_{\chi^2_{(k-1)}}(q_{Cl}),$$

where $F_{\chi^2_{(k-1)}}$ is the cumulative function of the χ^2 distribution with $(k-1)$ degrees of freedom. pv represents the probability to get the result q_{Cl} or higher under the assumption that the null hypothesis is true. For the actual decision, we use the level of significance α which should be chosen in advance. Is the p-value smaller than the pre-defined α, we can reject the null hypothesis. It indicates statistically significant heterogeneity of the treatment effect across the identified patient collectives. In other case, if the p-value is greater than α, one would suggest that the null hypothesis can not be rejected safely and the assumption that there is a true treatment effect in each patient collective can not be justified.

5.7.6. Assessment of heterogeneity within clusters

According to Section 4.2, we use the Q^i-test for the evaluation if there is still heterogeneity across the studies included in the identified patient collectives. For the conduction of the Q^i-test, we follow the test procedure shown in Figure 4.1.

The null hypothesis is formulated for all patient collectives by

$$H_0 : \theta_l^i = \theta_k^i, \forall l, k \in \mathcal{I}_{Cl_i} \text{ and } l \neq k,$$

where θ_j^i is the true treatment effect of study j in cluster Cl_i, for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. This implies that we assume the same treatment effect in all studies included in cluster Cl_i and $(\tau^i)^2 = 0$. For the conduction of the Q^i-test we use the Q^i-statistic,

$$Q^i = \sum_{j \in \mathcal{I}_{Cl_i}} w_j (\tilde{\Theta}_j^i - \tilde{\Theta}^i)^2$$

with realization

$$q^i = \sum_{j \in \mathcal{I}_{Cl_i}} w_j (\tilde{\theta}_j^i - \tilde{\theta}^i)^2.$$

$\tilde{\Theta}^i$ is the estimated summary treatment effect with realization $\tilde{\theta}^i$ in cluster Cl_i, calculated on the basis of one of the cluster-based fixed-effects model, and $\tilde{\Theta}_j^i$ with realization $\tilde{\theta}_j^i$ is the estimated treatment effect of study j in cluster Cl_i. The weight is denoted by w_j and is defined according to the chosen model. Q^i is approximately χ^2 distributed with $(n_i - 1)$ degrees of freedom, where n_i is the number of studies included in cluster Cl_i. The null hypothesis is rejected if the p-value

$$pv^i = P(Q^i \geq q^i | H_0) = 1 - F_{\chi^2_{(n_i-1)}}(q^i)$$

is smaller than the pre-defined α. $F_{\chi^2_{(n_i-1)}}$ is the cumulative function of the χ^2 distribution with $(n_i - 1)$ degrees of freedom, for $i = 1, \ldots, k$.

Since the Q^i-test only gives us the information about the presence or the absence of heterogeneity, we also use the Q^i-statistic-based $(l^i)^2$ index for the quantification of the heterogeneity in cluster Cl_i. It is defined by

$$(l^i)^2 := \max \left\{ 0, \frac{q^i - (n_i - 1)}{q^i} \right\} \cdot 100\%,$$

where q^i is the realization of the Q^i-statistic and $(n_i - 1)$ is the degree of freedom in cluster Cl_i. This index can be interpreted as it is listed in Table 4.2.

5.8. Cluster-based identification of heterogeneity

In this section, the endpoint-oriented geometric clustering approach is applied as an unsupervised learning approach for the identification of clinical and especially regional heterogeneity in the treatment effects within and across the identified patient collectives. Therefore, in the first step, we will have a look at the general terminology, like the definitions of different cluster values, adjusted to the examined characteristics for which heterogeneity is assumed. Based on the specification of the general terms, the hypothesis tests for the identification of heterogeneity can be formulated. For this analysis we use two different statistical hypothesis tests, the nonparametric χ^2 test and the parametric one or two tailed one sample test. The theory is based on the joint working paper of Pogarell, Brieden an Hinnenthal [21].

5.8.1. Clustering

For the cluster-based identification of heterogeneity, the entire patient data set

$$S = \{(x_j, y_j)\}_{j=1}^{n} = S^{all} \subset \mathbb{R}^d \times \Omega$$

is taken as training data set and is transformed to $\hat{S}$ like it is discussed in Section 5.5. It is important to consider also the administered drug as additional patient characteristic to identify how the constitution and the administered medication influence the response or outcome of a patient. On the transformed patient data set, the clustering Algorithm 4 is applied and we achieve a (k, l, u)-clustering

$$Cl = (Cl_1, \ldots, C_k)$$

as a partition of the heterogeneous individual patient data, the total population, into collectives with patients of similar combinations of their characteristic values.

5.8.2. Cluster-based analysis

Since it is assumed that there are different treatment effects for different patient collectives, the heterogeneity, and especially the regional heterogeneity, discussed in Section 4.1, in the treatment effects within and across the identified clusters has to be analyzed. For this analysis, we consider the unbiased estimator for the expected value of the outcome of patients or, in the binary case, the probability of success in cluster Cl_i as the corresponding cluster value in cluster Cl_i. Therefore, we need to define the patient data set of cluster Cl_i.

Definition 5.8.1. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^{n}$ and $\xi = (\xi_{ij}) \in \{0,1\}^{k \times n}$ the corresponding assignment of Cl. Then*

$$\hat{S}^i := \{(\hat{x}_j, y_j) | \xi_{ij} = 1\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega_{Cl_i} \subseteq \hat{S}$$

is the patients data set of cluster Cl_i. S^i denotes the non-translated patient data set of cluster Cl_i, for $i = 1, \dots, k$.

Another important term is the number of patients with specific characteristic values in cluster Cl_i.

Definition 5.8.2. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \dots, x_{jd}^i) = (a_1, \dots, a_d)$$

represented by $A_1, \dots, A_d$. Then

$$\kappa_i(a_l, \dots, a_m) := \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{x_{jl}^i, \dots, x_{jm}^i\}}(a_l, \dots, a_m)$$

is the number of patients with the characteristic values combination $A_l = a_l, \dots, A_m = a_m$, $l, \dots, m \in \{1, \dots, d\}$, in cluster Cl_i, for $i = 1, \dots, k$.

Cardinal data

For the assessment of the heterogeneity, the cluster value is defined by the realization of the unbiased estimator for the expected value of the patients' outcome in cluster Cl_i. For the definition of this estimator, we assume that the true outcome in cluster Cl_i,

$$Y^i \sim \mathcal{N}(y^i, (\sigma^i)^2),$$

is a normally distributed random variable with expected value y^i and standard deviation σ^i, for $i = 1, \ldots, k$. We also assume that the outcome of patient j,

$$Y^i_j \sim \mathcal{N}(y^i, (\sigma^i)^2),$$

for $j = 1, \ldots, \kappa_i$, in cluster Cl_i is also an independent normally distributed random variable with realization y^i_j, expected value y^i and standard deviation σ^i. The sample space of Y^i and Y^i_j is denoted by $\Omega_{Cl_i} \subseteq \Omega$. With these assumptions the unknown expected outcome

$$E(Y^i) = y^i$$

of patients in cluster Cl_i can be estimated by the unbiased estimator specified in the following theorem. For this approach, this estimator then also forms the cluster value in cluster Cl_i.

Theorem 5.8.3. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with normally distributed outcome and let $\hat{S}^i = \{(\hat{x}^i_j, y^i_j)\}_{j=1}^{\kappa_i}$ be the patient data set of cluster Cl_i. Then*

$$\tilde{Y}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} Y^i_j,$$

is the unbiased estimator for the expected value of the patients' outcome in cluster Cl_i and

$$f_i(Cl_i) = \tilde{y}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} y^i_j,$$

the corresponding estimation and cluster value, for $i = 1, \ldots, k$.

For the identification of heterogeneity between patients with different characteristic values combinations, e.g. the expected outcome of male and of female patients, we define the estimator for the conditional expected value

$$E(Y^i|A_l = a_l) =: y^i(a_l)$$

of the patients' outcome in cluster Cl_i, given characteristic value $A_l = a_l$.

Remark 5.8.4. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with normally distributed outcome and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \dots, x_{jd}^i) = (a_1, \dots, a_d)$$

represented by $A_1, \dots, A_d$. Then

$$\tilde{Y}^i(a_l) := \frac{1}{\kappa_i(a_l)} \sum_{j=1}^{\kappa_i} Y_j^i \mathbf{1}_{\{x_{jl}^i\}}(a_l),$$

is the estimator for the expected value of the patients' outcome given characteristic value $A_l = a_l$, $l \in \{1, \dots, d\}$, in cluster Cl_i and

$$\tilde{y}^i(a_l) := \frac{1}{\kappa_i(a_l)} \sum_{j=1}^{\kappa_i} y_j^i \mathbf{1}_{\{x_{jl}^i\}}(a_l),$$

the corresponding estimation, for $i = 1, \dots, k$.

For the conditional variance

$$\mathrm{var}(Y^i|A_l = a_l) =: (\sigma^i(a_l))^2$$

of the outcome of patients with given characteristic value $A_l = a_l$, we use the following estimator.

Remark 5.8.5. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with normally distributed outcome and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-*

translated patient data set of cluster Cl_i with d characteristic values

$$x_j^i = (x_{j1}^i, x_{j2}^i, \ldots, x_{jd}^i) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then

$$(\tilde{\Sigma}^i(a_l))^2 := \frac{1}{\kappa_i(a_l) - 1} \sum_{j=1}^{\kappa_i} (Y_j^i - \tilde{Y}^i(a_l))^2 \mathbf{1}_{\{x_{jl}^i\}}(a_l),$$

is the estimator for the variance of the patients' outcome given characteristic value $A_l = a_l$, $l \in \{1, \ldots, d\}$, in cluster Cl_i and

$$(\tilde{\sigma}^i(a_l))^2 := \frac{1}{\kappa_i(a_l) - 1} \sum_{j=1}^{\kappa_i} (y_j^i - \tilde{y}^i(a_l))^2 \mathbf{1}_{\{x_{jl}^i\}}(a_l),$$

the corresponding estimation, for $i = 1, \ldots, k$.

If we want to consider more than one characteristic value as condition for the estimation of the conditional expected value

$$E(Y_{Cl_i} | A_l = a_l, \ldots, A_m = a_m) =: y^i(a_l, \ldots, a_m)$$

and the conditional variance

$$\mathrm{var}(Y^i | A_l = a_l, \ldots, A_m = a_m) =: (\sigma^i(a_l, \ldots, a_m))^2$$

of the patents' outcome in cluster Cl_i, we can extend the last two specifications to the following.

Remark 5.8.6. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with normally distributed outcome and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \ldots, x_{jd}^i) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. *Then*

$$\tilde{Y}^i(a_l, \ldots, a_m) := \frac{\sum_{j=1}^{\kappa_i} Y_j^i \mathbb{1}_{\{x_{jl}^i, \ldots, x_{jm}^i\}}(a_l \ldots a_m)}{\kappa_i(a_l, \ldots, a_m)}$$

is the estimator for the expected value of the patients' outcome given characteristic value combination $A_l = a_l, \ldots, A_m = a_m$, $l, \ldots, m \in \{1, \ldots, d\}$, *in cluster* Cl_i *and*

$$\tilde{y}^i(a_l, \ldots, a_m) := \frac{\sum_{j=1}^{\kappa_i} y_j^i \mathbb{1}_{\{x_{jl}^i, \ldots, x_{jm}^i\}}(a_l, \ldots, a_m)}{\kappa_i(a_l, \ldots, a_m)}$$

the corresponding estimation, for $i = 1, \ldots, k$.

Remark 5.8.7. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of patient data set* $\hat{S}$ *with normally distributed outcome and let* $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ *be the non-translated patient data set of cluster* Cl_i *with* d *characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \ldots, x_{jd}^i) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. *Then*

$$(\tilde{\Sigma}^i(a_l, \ldots, a_m))^2 := \frac{\sum_{j=1}^{\kappa_i} (Y_j^i - \tilde{Y}^i(a_l))^2 \mathbb{1}_{\{x_{jl}^i, \ldots, x_{jm}^i\}}(a_l \ldots a_m)}{\kappa_i(a_l, \ldots, a_m) - 1},$$

is the estimator for the variance of the patients' outcome given characteristic value combination $A_l = a_l$, $l \in \{1, \ldots, d\}$, *in cluster* Cl_i *and*

$$(\tilde{\sigma}^i(a_l, \ldots, a_m))^2 := \frac{\sum_{j=1}^{\kappa_i} (y_j^i - \tilde{y}^i(a_l))^2 \mathbb{1}_{\{x_{jl}^i, \ldots, x_{jm}^i\}}(a_l, \ldots, a_m)}{\kappa_i(a_l, \ldots, a_m) - 1},$$

the corresponding estimation, for $i = 1, \ldots, k$.

Binary data

In the case of binary outcome Y^i, with sample space $\Omega_{Cl_i} = \{0, 1\}$, where 0 stands for failure and 1 for success, the random variable is Bernoulli distributed,

$$Y^i \sim \mathcal{B}(p^i),$$

with probability of success $p^i = P(Y^i = 1)$. The cluster value is then defined by the realization of the unbiased estimator of the probability of success in cluster Cl_i. Therefore, we assume that the outcome of patient j,

$$Y_j^i \sim \mathcal{B}(p^i),$$

for $j = 1, \ldots, \kappa_i$, in cluster Cl_i is also an independent Bernoulli distributed random variable with probability of success $p^i = P(Y_j^i = 1)$ and realization y_j^i. The sample space of Y^i and Y_j^i is denoted by $\Omega_{Cl_i} = \{0, 1\}$. With these assumptions, the unknown probability of success of patients in cluster Cl_i can be estimated by the unbiased estimator specified in the following theorem.

Theorem 5.8.8. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome and $\hat{S}^i = \{(\hat{x}_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ the patient data set of cluster Cl_i. Then*

$$\tilde{P}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(Y_j^i),$$

is the unbiased estimator for the probability of success in cluster Cl_i and

$$f_i(Cl_i) = \tilde{p}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(y_j^i),$$

the corresponding estimation and cluster value, for $i = 1, \ldots, k$.

For comparison of the probability of success of patients with different characteristic values combinations, e.g. the probability of success of young and elderly patients, we define the estimator for the conditional probability of success

$$P(Y^i = 1 | A_l = a_l) =: p^i(a_l)$$

in cluster Cl_i, given characteristic value $A_l = a_l$.

Remark 5.8.9. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-*

translated patient data set of cluster Cl_i with d characteristic values

$$x_j^i = (x_{j1}^i, x_{j2}^i, \ldots, x_{jd}^i) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then

$$\tilde{P}^i(a_l) := \frac{1}{\kappa_i(a_l)} \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(Y_j^i)\mathbb{1}_{\{x_{jl}^i\}}(a_l),$$

is the estimator of the expected value of the patients' outcome given characteristic value $A_l = a_l$, $l \in \{1, \ldots, d\}$, in cluster Cl_i and

$$\tilde{p}^i(a_l) := \frac{1}{\kappa_i(a_l)} \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(y_j^i)\mathbb{1}_{\{x_{jl}^i\}}(a_l),$$

the corresponding estimation, for $i = 1, \ldots, k$.

If we want to consider more than one characteristic value as condition for the estimation of the conditional probability of success

$$P(Y^i = 1 | A_l = a_l, \ldots, A_m = a_m) = p^i(a_l, \ldots, a_m)$$

in cluster Cl_i, we can extend the last remark to the following.

Remark 5.8.10. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \ldots, x_{jd}^i) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then

$$\tilde{P}^i(a_l, \ldots, a_m) := \frac{\sum_{j=1}^{\kappa_i} \mathbf{1}_{\{1\}}(Y_j^i)\mathbf{1}_{\{x_{jl}^i, \ldots, x_{jm}^i\}}(A_l, \ldots, A_m)}{\kappa_i(a_l, \ldots, a_m)},$$

is the estimator for the probability of success given characteristic value combina-

tion $A_l = a_l, \ldots, A_m = a_m$, $l, \ldots, m \in \{1, \ldots, d\}$, *in cluster* Cl_i *and*

$$\tilde{p}^i(a_l, \ldots, a_m) := \frac{\sum_{j=1}^{\kappa_i} \mathbf{1}_{\{1\}}(y_j^i)\mathbf{1}_{\{x_{jl}^i, \ldots, x_{jm}^i\}}(a_l, \ldots, a_m)}{\kappa_i(a_l, \ldots, a_m)},$$

the corresponding estimation, for $i = 1, \ldots, k$.

Proportion

For the comparison of the different shares of patients with specific characteristic values combinations in cluster Cl_i, e.g. the share of male patients compared to the share of female patients, we need an estimator for the proportion

$$\mathcal{P}(A_l = a_l) =: \rho^i(a_l),$$

$l \in \{1, \ldots, d\}$, of characteristic value $A_l = a_l$ in cluster Cl_i.

Remark 5.8.11. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of patient data set* $\hat{S}$ *and let* $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ *be the non-translated patient data set of cluster* Cl_i *with* d *characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \ldots, x_{jd}^i) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. *Then*

$$\tilde{\rho}^i(a_l) := \frac{\sum_{j=1}^{\kappa_i} \mathbf{1}_{\{x_{jl}^i\}}(a_l)}{\kappa_i},$$

is the estimation for the proportion of characteristic value $A_l = a_l$, $l \in \{1, \ldots, d\}$, *in cluster* Cl_i, *for* $i = 1, \ldots, k$.

It might also be of importance, to analyze the difference between the shares of patients considering more than one characteristic value, e.g. the share of youngsters under the male patients and the share of elderly under male patients in cluster Cl_i. Therefore, we use the estimator for the probability

$$\mathcal{P}(A_l = a_l | A_{f_1} = a_{f_1}, \ldots, A_{f_m} = a_{f_m}) =: \rho^i(a_l | a_f, \ldots, a_m),$$

for $l, f, \dots m \in \{1, \dots, d\}$ and $l \neq f, \dots, m$.

Remark 5.8.12. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ and let $S^i = \{(x_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j^i = (x_{j1}^i, x_{j2}^i, \dots, x_{jd}^i) = (a_1, \dots, a_d)$$

represented by $A_1, \dots, A_d$. Then

$$\tilde{\rho}^i(a_l | a_f, \dots, a_m) := \frac{\sum_{j=1}^{\kappa_i} \mathbf{1}_{\{x_{jl}^i, x_{jf}^i, \dots, x_{jm}^i\}}(a_l, a_f, \dots, a_m)}{\kappa_i(a_f, \dots, a_m)}$$

is the estimation for the proportion of characteristic value $A_l = a_l$, given characteristic value combination $A_f = a_f, \dots, A_m = a_m$, $l, f, \dots m \in \{1, \dots, d\}$ and $l \neq f, \dots, m$, in cluster Cl_i, for $i = 1, \dots, k$.

With these definitions, we can have a closer look at the hypotheses tests for the identification of heterogeneity.

5.8.3. Two sample hypothesis test

We begin with the description of the two sample t test which is used to determine if the expected outcome, due to an administered intervention, of patients with different characteristic values combinations in cluster Cl_i is equal, or if there is a significant heterogeneity between the analyzed outcome. The following theory is based on [61]. We formulate the null hypotheses that the expected outcome of patients with different characteristic values combinations is equal. Then the goal is to reject this formulated null hypotheses, to have an indication for heterogeneity in the patients' outcome. The two sample t test procedure is shown in Figure 5.2.

In a first approach, we analyze if the patients' outcome Y^i is independent from a pre-defined characteristic A_l, $l \in \{1, \dots, d\}$. E.g. we want to know, if the outcome of male patients differs to the outcome of female patients. In the first step of the procedure, we formulate the null hypothesis.

$$H_0 : y^i(a_{l_1}) = y^i(a_{l_2}),$$

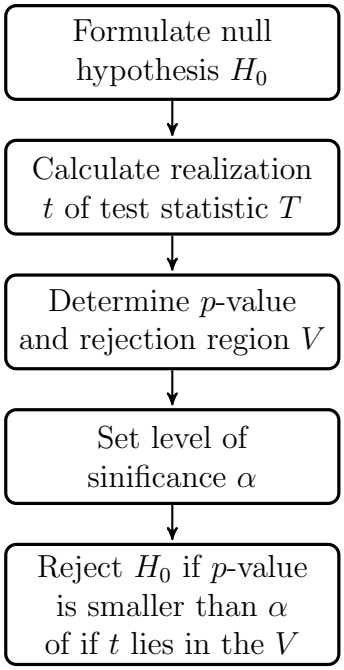

Figure 5.2.: Hypothesis test procedure for the two sample test

where $y^i(a_{l_f})$, $f = 1, 2$, is the expected outcome of patients with the characteristic values a_{l_1} and a_{l_2} in cluster Cl_i, for $i = 1, \ldots, k$. The alternative hypothesis is, that patients with the characteristic values a_{l_1} and a_{l_2} have different outcomes due to the administered medication, thus, that there is heterogeneity between the treatment effects. In the next step of the test procedure, we calculate the realization t of the test statistic T. We assume that the outcome of patients with characteristic value a_{l_1} is independent of the outcome of those with characteristic value a_{l_2}. The variances of the considered outcome of the two patient groups are unknown and we assume that they are unequal. The test statistic is then given by

$$T = \frac{\tilde{Y}^i(a_{l_1}) - \tilde{Y}^i(a_{l_2})}{\sqrt{\frac{(\tilde{\Sigma}^i(a_{l_1}))^2}{\kappa_i(a_{l_1})} + \frac{(\tilde{\Sigma}^i(a_{l_2}))^2}{\kappa_i(a_{l_1})}}}$$

and is an approximately standardized normally distributed random variable with realization

$$t = \frac{\tilde{y}^i(a_{l_1}) - \tilde{y}^i(a_{l_2})}{\sqrt{\frac{(\tilde{\sigma}^i(a_{l_1}))^2}{\kappa_i(a_{l_1})} + \frac{(\tilde{\sigma}^i(a_{l_2}))^2}{\kappa_i(a_{l_1})}}}.$$

With the pre-defined level of significance α, the null hypothesis can be rejected,

if realization t is located in the rejection region

$$V := \{t \in \mathbb{R} || t| \geq z[1 - \frac{\alpha}{2}]\},$$

where $z[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution. To evaluate if the null hypothesis can be rejected, we can also use the p-value, which is the probability to get the result t or an extremer value under the assumption that the null hypothesis is true. It is determined by

$$pv = P(T \geq t|H_0) = 1 - F_{\mathcal{N}}(t),$$

where $F_{\mathcal{N}}$ is the cumulative standard normal distribution function. If the p-value is smaller than α, the null hypothesis that the outcome of patients with different characteristic values is equal has to be rejected and statistically significant heterogeneity of the outcome can be indicated.

In a second approach, we analyze if the patients' outcome Y^i is independent from more than one pre-defined characteristics $A_l, \ldots, A_m$, $l, \ldots, m \in \{1, \ldots, d\}$. E.g. we want to know if the outcome of young male patients differs to the outcome of young female patients. In this case the null hypothesis is given by

$$H_0 : y^i(a_{l_1}, \ldots, a_{m_1}) = y^i(a_{l_2}, \ldots, a_{m_2}),$$

where $a_{l_f} \in \Omega_l, \ldots, a_{m_f} \in \Omega_m$, $f = 1, 2$, are the characteristic values of the characteristics $A_l, \ldots, A_m$, with sample spaces $\Omega_l, \ldots, \Omega_m$, $l, \ldots, m \in \{1, \ldots, d\}$. Then $y^i(a_{l_f}, \ldots, a_{m_f})$, $f = 1, 2$, is the expected outcome of patients with the characteristic values combinations $a_{l_1}, \ldots, a_{m_1}$ and $a_{l_2}, \ldots, a_{m_2}$ in cluster Cl_i, for $i = 1, \ldots, k$. The alternative hypothesis is that patients with the characteristic values combinations $a_{l_1}, \ldots, a_{m_1}$ and $a_{l_2}, \ldots, a_{m_2}$ have different outcomes due to the administered medication. Thus, that there is heterogeneity between the outcome.

The test statistic is then defined by

$$T = \frac{\tilde{Y}^i(a_{l_1}, \ldots, a_{m_1}) - \tilde{Y}^i(a_{l_2}, \ldots, a_{m_2})}{\sqrt{\frac{(\tilde{\Sigma}^i(a_{l_1},\ldots,a_{m_1}))^2}{\kappa_i(a_{l_1},\ldots,a_{m_1})} + \frac{(\tilde{\Sigma}^i(a_{l_2},\ldots,a_{m_2}))^2}{\kappa_i(a_{l_2},\ldots,a_{m_2})}}}.$$

Since we assume that the outcome of patients with characteristic values combination $a_{l_1}, \ldots, a_{m_1}$ is independent of the outcome of those with characteristic values combination $a_{l_2}, \ldots, a_{m_2}$, the test statistic is also approximately standardized normally distributed with realization

$$t = \frac{\tilde{y}^i(a_{l_1}, \ldots, a_{m_1}) - \tilde{y}^i(a_{l_2}, \ldots, a_{m_2})}{\sqrt{\frac{(\tilde{\sigma}^i(a_{l_1},\ldots,a_{m_1}))^2}{\kappa_i(a_{l_1},\ldots,a_{m_1})} + \frac{(\tilde{\sigma}^i(a_{l_2},\ldots,a_{m_2}))^2}{\kappa_i(a_{l_2},\ldots,a_{m_2})}}}.$$

The variances of the considered outcome of the two patient groups are unknown and we assume that they are unequal. In the next step, we set significance level α. The null hypotheses is then rejected and statistically significant heterogeneity of the outcome can be indicated if t is located in the rejection region

$$V := \{t \in \mathbb{R} || t | \geq z[1 - \frac{\alpha}{2}]\}.$$

or of the p-value, defined by

$$pv = P(T \geq t | H_0) = 1 - F_{\mathcal{N}}(t),$$

is smaller than α.

5.8.4. χ^2 test for independence

In case of Bernoulli distributed patients' outcome $Y^i \sim \mathcal{B}(p^i)$, with sample space $\Omega_{Cl_i} = \{0, 1\}$, we use the χ^2 hypothesis test procedure shown in Figure 5.3 for the identification of heterogeneity. This procedure is used for testing, if patients with different characteristic values respond differently to an administered intervention in cluster Cl_i. Therefore, we formulate the null hypothesis, that probability success p^i of patients with different characteristic values is equal. Then the goal is to reject this formulated null hypotheses to have an indication for heterogeneity in the patients' outcome. The following theory is based on [61].

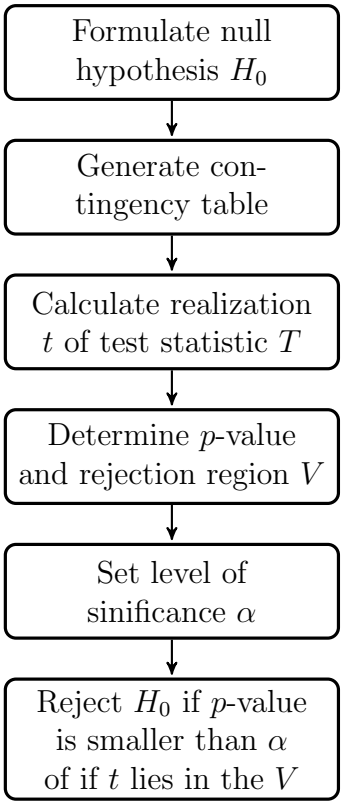

Figure 5.3.: Hypothesis test procedure for the χ^2 test

In a first approach, we analyze if the patients' outcome Y^i is independent from one pre-defined characteristic A_l, $l \in \{1, \ldots, d\}$. E.g. we want to know, if the response rate of male patients differs to the response rate of female patients. In the first step of the procedure, we formulate the null hypothesis,

$$H_0 : p^i(a_{l_1}) = p^i(a_{l_2}),$$

where $p^i(a_{l_f})$, $f = 1, 2$, is the probability of success of patients with the characteristic values a_{l_1} and a_{l_2} in cluster Cl_i, for $i = 1, \ldots, k$. The alternative hypothesis is that patients with the characteristic values a_{l_1} and a_{l_2} respond differently to the administered medication.

In the next step of the procedure, we have to formulate the test statistic. Therefore, we need contingency Table 5.2. The entries of this table are specified in the next remark.

	success	failure	
$A_l = a_{l_1}$	$a = s^i(a_{l_1})$	$b = f^i(a_{l_1})$	$a + b$
$A_l = a_{l_2}$	$c = s^i(a_{l_2})$	$d = f^i(a_{l_2})$	$c + d$
	$a + c$	$b + d$	$n = a + b + c + d$

Table 5.2.: Distribution of binary outcome on characteristic A_l in cluster Cl_i

Remark 5.8.13. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S^i = \{(x_j, y_j)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then for the entries of contingency Table 5.2 holds

$$s^i(a_{l_f}) := \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(y_j^i)\mathbb{1}_{\{x_{jl_f}\}}(a_{l_f})$$
$$f^i(a_{l_f}) := \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{x_{jl_f}\}}(a_{l_f}) - s^i(a_{l_f}),$$

for $f = 1, 2$ and $i = 1, \ldots, k$.

With Table 5.2, the test statistic is then given by

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(a + c)(c + d)(b + d)}. \tag{5.3}$$

χ^2 is the realization of an approximately χ_1^2 distributed random variable with one degree of freedom. With the pre-defined level of significance α, the null hypothesis will be rejected, if the test statistic lies in the rejection region

$$V := \{\chi^2 \in \mathbb{R} | \chi^2 > \chi_1^2[1 - \alpha]\}, \tag{5.4}$$

where $\chi_1^2[1 - \alpha]$ is the $(1 - \alpha)$-quantile of the χ_1^2 distribution with one degree of freedom. Another way to evaluate the result of the test statistic is the calculation of the p-value. As it is stated in 5.7.5, the p-value is the probability to get the

	success	failure	
$A_l = a_{l_1}, \ldots, A_m = a_{m_1}$	$a = s^i(a_{l_1}, \ldots, a_{m_1})$	$b = f^i(a_{l_1}, \ldots, a_{m_1})$	$a+b$
$A_l = a_{l_2}, \ldots, A_m = a_{m_2}$	$c = s^i(a_{l_2}, \ldots, a_{m_2})$	$d = f^i(a_{l_2}, \ldots, a_{m_2})$	$c+d$
	$a+c$	$b+d$	n

Table 5.3.: Distribution of binary outcome on characteristic $A_l, \ldots, A_m$ in cluster Cl_i

result χ^2 or higher under the assumption that the null hypothesis is true. It is determined by

$$pv = 1 - F_{\chi_1^2}(\chi^2), \tag{5.5}$$

where $F_{\chi_1^2}$ is the cumulative χ_1^2 distribution function. If the p-value is smaller than α, the null hypothesis that patients with different characteristic values respond equally has to be rejected and statistically significant heterogeneity of the treatment effects can be indicated.

If we want to consider more than one different characteristic value when analyzing different responses to a specific medication in cluster Cl_i, we need to adapt the null hypothesis to

$$H_0 : p^i(a_{l_1}, \ldots, a_{m_1}) = p^i(a_{l_2}, \ldots, a_{m_2}),$$

where $a_{l_f} \in \Omega_l, \ldots, a_{m_f} \in \Omega_m$, $f = 1, 2$, are the characteristic values of the characteristics $A_l, \ldots, A_m$, with the sample spaces $\Omega_l, \ldots, \Omega_m$, $l, \ldots, m \in \{1 \ldots, d\}$. Thus $p^i(a_{l_f}, \ldots, a_{m_f})$, $f = 1, 2$, is the conditional probability of success of the patients with characteristic values combinations $a_{l_1}, \ldots, a_{m_1}$ and $a_{l_2}, \ldots, a_{m_2}$. E.g. we test, if the response rate of young male patients is significantly different to the response rate of elderly female patients in cluster Cl_i. For the formulation of the test statistic (5.3) we use Table 5.3.

Remark 5.8.14. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S^i = \{(x_j, y_j)\}_{j=1}^{\kappa_i}$ be the non-*

translated patient data set of cluster Cl_i with d characteristic values

$$x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then for the entries of contingency Table 5.3 holds

$$s^i(a_{l_f}, \ldots, a_{m_f}) := \sum_{j=1}^{\kappa_i} \mathbf{1}_{\{1\}}(y_j^i)\mathbf{1}_{\{x_{jl_f},\ldots,x_{jm_f}\}}(a_{l_f}, \ldots, a_{m_f})$$

$$f^i(a_{l_f}, \ldots, a_{m_f}) := \sum_{j=1}^{\kappa_i} \mathbf{1}_{\{x_{jl_f},\ldots,x_{jm_f}\}}(a_{l_f}, \ldots, a_{m_f}) - s^i(a_{l_f}, \ldots, a_{m_f}),$$

for $f = 1, 2$ and $i = 1, \ldots, k$.

With these entries of the contingency table we can formulate the test statistic for the given null hypothesis according to (5.3). We reject the null hypothesis, if the result of the test statistic lies in rejection region V (5.4) or if the p-value (5.5) is smaller than the pre-defined α.

The χ^2 test can also be applied to identify heterogeneity between the share of patients with a specific characteristic value $a_l \in \Omega_l$ in cluster Cl_i and the total population by means of all patients participated in the considered studies. E.g. we can test, if the share of male patients in the total population differ significantly from the share of male patients in cluster Cl_i. Therefore we formulate the null hypothesis

$$H_0 : \rho^i(a_l) = \rho(a_l),$$

where $\rho(a_l)$ is the share of patients with characteristic value $A_l = a_l$ in the total population and $\rho^i(a_l)$ the share in cluster Cl_i. For the calculation of the test statistic (5.3), we take the distribution information found in Table 5.4. The distribution in cluster Cl_i is defined in the following.

Remark 5.8.15. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S^i = \{(x_j, y_j)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$$

	cluster Cl_i	Total population	
$A_l = a_{l_1}$	a	b	$a+b$
else	c	d	$c+d$
	$a+c$	$b+d$	$n = a+b+c+d$

Table 5.4.: Share of characteristic value a_l in cluster Cl_i and total population

represented by $A_1, \ldots, A_d$. Then for the entries of contingency Table 5.4 holds

$$a := \sum_{j=1}^{\kappa_i} \mathbf{1}_{\{x_{jl}\}}(a_l)$$
$$c := \kappa_i - a.$$

For the total population the entries of the contingency table are defined as follows.

Remark 5.8.16. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S = \{(x_j, y_j)\}_{j=1}^n$ be the non-translated patient data set of with d characteristic values*

$$x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then for the entries of contingency Table 5.4 holds

$$b := \sum_{j=1}^{n} \mathbf{1}_{\{x_{jl}\}}(a_l)$$
$$d := n - b.$$

We reject the null hypothesis if the result of the test statistic (5.3) lies in V (5.4) or if the p-value (5.5) is smaller than α.

In a last application, the test can be used to analyze if there is heterogeneity between the shares of patients with characteristic value $a_{l_1} \in \Omega_l$ given $a_{f_1}, \ldots, a_{m_1}$ and patients with characteristic value $a_{l_2} \in \Omega_l$ given $a_{f_2}, \ldots, a_{m_2}$ in cluster Cl_i. E.g. we can analyze, if there is a statistically significant difference between the

	$A_l = a_{l_1}$	$A_l = a_{l_1}$	
$A_f = a_{f_1}, \ldots, A_m = a_{m_1}$	a	b	$a+b$
$A_f = a_{f_2}, \ldots, A_m = a_{m_2}$	c	d	$c+d$
	$a+c$	$b+d$	n

Table 5.5.: Share of characteristic A_l under patients with characteristic $A_f, \ldots, A_m$ in cluster Cl_i

share of young male patients and elderly male patients in cluster Cl_i. Therefore, we formulate the null hypothesis

$$H_0 : \rho^i(a_{l_1}|a_{f_1}, \ldots, a_{m_1}) = \rho^i(a_{l_2}|a_{f_2}, \ldots, a_{m_2}).$$

The test statistic (5.3) is then calculated based on contingency Table 5.5.

Remark 5.8.17. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of patient data set $\hat{S}$ with Bernoulli distributed outcome and let $S^i = \{(x_j, y_j)\}_{j=1}^{\kappa_i}$ be the non-translated patient data set of cluster Cl_i with d characteristic values*

$$x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) = (a_1, \ldots, a_d)$$

represented by $A_1, \ldots, A_d$. Then for the entries of contingency Table 5.5 holds

$$\begin{aligned}
a &:= \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{x_{jl_1}, x_{jf_1}, \ldots, x_{jm_1}\}}(a_{l_1}, a_{f_1}, \ldots, a_{m_1}) \\
b &:= \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{x_{jf_1}, \ldots, x_{jm_1}\}}(a_{f_1}, \ldots, a_{m_1}) - a \\
c &:= \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{x_{jl_2}, x_{jf_2}, \ldots, x_{jm_2}\}}(a_{l_2}, a_{f_2}, \ldots, a_{m_2}) \\
d &:= \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{x_{jf_2}, \ldots, x_{jm_2}\}}(a_{f_2}, \ldots, a_{m_2}) - c
\end{aligned}$$

As for the previous test, the null hypothesis is rejected if the result of the test statistic (5.3) lies in V (5.4) or if the p-value (5.5) is smaller than α.

5.9. Cluster-based prediction of treatment effects

In the medical sector, it is important to forecast the efficacy of medical intervention. And since the assumption is reliable that there are different treatment effects for patients with different characteristic values combinations, in this section the geometric clustering approach is examined as a supervised learning approach for the prediction of the efficacy of medical interventions on patient collectives. Furthermore, an assessment method for the reliability of these predictive effects is presented. This might be important, e.g. for a more patient-oriented medication in terms of evidence-based medicine. Therefore, in Section 5.9.2, we will have a look how the predictive value for each cluster can be determined based on the available patient data. In this context, we will interpret the accuracy of the calculated values by analyzing the corresponding confidence intervals. Then, in Section 5.9.3, we will evaluate the reliability of the predicted values by using a new invented hypothesis test procedure which will be applied to each patient collective. The following theory is based on the joint working paper of Brieden, Öllinger and Hinnenthal [20].

5.9.1. Clustering

For the prediction of the efficacy of medical interventions, the patient data set $S^{all} = \{(x_j, y_j)\}_{j=1}^{N} \subset \mathbb{R}^d \times \Omega$ is divided into a training

$$S = \{(x_j, y_j)\}_{j=1}^{n} \subset \mathbb{R}^d \times \Omega$$

and testing patient data set

$$S^{te} = \{(x_j^{te}, y_j^{te})\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega.$$

It is common to use 80% of the available data for the training data set and respectively 20% of the data for the testing data set. The choice of this ratio is important with regard to the comparability of different, already existent predictive algorithms. After this division, the training and the testing data set are

transformed to $\hat{S}$ and $\hat{S}^{te}$ with respect to the outcome variable like it is discussed in Section 5.5. It is important to consider the medication as additional patient characteristic, to identify how the constitution and the administered intervention influence the response or outcome of a patient. The goal is the prediction of the outcome of patients with specific characteristic values combinations and specific administered medication, e.g. the response of young male patients, if they are medicated with drug A. Then the clustering Algorithm 4 is applied and we get a (k, l, u)-clustering

$$Cl = (Cl_1, \ldots, Cl_k)$$

as a partition of the transformed training data set $\hat{S}$. Each resulting cluster represents a collective with patients of similar combinations of their characteristic values which are identified by the clustering algorithm.

5.9.2. Cluster-based analysis

Since we assume that there is a treatment effect for each identified patient collective, we want to do predictive analysis for each cluster. Therefore, we firstly define the patient data set of cluster Cl_i, for $i = 1, \ldots, k$.

Definition 5.9.1. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S} = \{(\hat{x}_j, y_j)\}_{j=1}^n$ and $\xi = (\xi_{ij}) \in \{0, 1\}^{k \times n}$ the corresponding assignment of Cl. Then*

$$\hat{S}^i := \{(\hat{x}_j, y_j) | \xi_{ij} = 1\}_{j=1}^n \subset \mathbb{R}^d \times \Omega_{Cl_i} \subseteq \hat{S}$$

is the patients data set of cluster Cl_i. S^i denotes the non-translated patient data set of cluster Cl_i.

Cardinal data

In the next step, we define the cluster value according to Definition 5.6.6. For the prediction of the efficacy of a medical intervention for a patient collective, we use the unbiased estimator of the expected patients' outcome in cluster Cl_i.

Therefore, we assume that there is a true normally distributed outcome

$$Y^i \sim \mathcal{N}(y^i, (\sigma^i)^2),$$

with expected value y^i and variance $(\sigma^i)^2$ in cluster Cl_i. Furthermore, let

$$Y_j^i \sim \mathcal{N}(y^i, (\sigma^i)^2)$$

be an independent normally distributed random variable for the outcome of patient j in the training patient data set in cluster Cl_i with outcome y_j^i, expected outcome y^i and unknown variance $(\sigma^i)^2$. The sample space of Y^i and Y_j^i is denoted by Ω_{Cl_i}, for $i = 1, \ldots, k$ and $j = 1, \ldots, \kappa_i$. Then the unbiased estimators for the expected value and the unknown variance of Y^i are defined as follows.

Theorem 5.9.2. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome and $\hat{S}^i = \{(\hat{x}_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ the patient data set of cluster Cl_i. Then*

$$\tilde{Y}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} Y_j^i,$$

is the unbiased estimator for the expected value of the patients' outcome in cluster Cl_i and

$$f_i(Cl_i) = \tilde{y}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} y_j^i,$$

the corresponding estimation, for $i = 1, \ldots, k$.

Theorem 5.9.3. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome and $\hat{S}^i := \{(\hat{x}_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ the patient data set of cluster Cl_i. Then*

$$(\tilde{\Sigma}^i)^2 := \frac{1}{\kappa_i - 1} \sum_{j=1}^{\kappa_i} (Y_j^i - \tilde{Y}^i)^2,$$

is the unbiased estimator for the unknown variance of the patients' outcome in

cluster Cl_i and

$$(\tilde{\sigma}^i)^2 = \frac{1}{\kappa_i - 1} \sum_{j=1}^{\kappa_i} (y_j^i - \tilde{y}^i)^2,$$

the corresponding estimation, for $i = 1, \ldots, k$.

Based on these estimations, in the next step, we define the $(1 - \alpha)$ confidence interval $\mathcal{I}^i$ of the expected outcome y^i of patients in cluster Cl_i with the predefined significance level α. The unbiased estimator $\tilde{Y}^i$ is normally distributed with expected value

$$E(\tilde{Y}^i) = y^i$$

and standard deviation

$$\sigma_{\tilde{Y}^i} := \frac{\tilde{\sigma}^i}{\sqrt{\kappa_i}}.$$

Then, if the number of patients κ_i in the training patient data set in cluster Cl_i is sufficiently large ($\kappa_i > 30$), due to the central limit theorem the quotient

$$\frac{\tilde{Y}^i - y^i}{\sigma_{\tilde{Y}^i}} \tag{5.6}$$

is approximately standardized normally distributed. For the $(1 - \alpha)$ confidence interval $\mathcal{I}^i$ of the expected outcome y^i in cluster Cl_i then follows

$$\mathcal{I}^i = [\tilde{y}^i - z[1 - \frac{\alpha}{2}]\sigma_{\tilde{Y}^i}, \tilde{y}^i + [1 - \frac{\alpha}{2}]\sigma_{\tilde{Y}^i}],$$

where $z[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution.

In case of a small random sample, Quotient (5.6) is student's t-distributed with $(\kappa_i - 1)$ degrees of freedom. For the $(1 - \alpha)$ confidence interval $\mathcal{I}^i$ of the expected outcome y^i in cluster Cl_i follows

$$\mathcal{I}^i = [\tilde{y}^i - t_{(\kappa_i - 1)}[1 - \frac{\alpha}{2}]\sigma_{\tilde{Y}^i}, \tilde{y}^i + t_{(\kappa_i - 1)}[1 - \frac{\alpha}{2}]\sigma_{\tilde{Y}^i}],$$

where $t_{(\kappa_i - 1)}[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the student's t-distribution. From this it follows that in $(1 - \alpha) \cdot 100\%$ of all cases the expected value y^i lies in $\mathcal{I}^i$, $P(y^i \in \mathcal{I}^i) = 1 - \alpha$.

Binary data

For the prediction of the efficacy of a medical intervention for a patient with the two possible outcomes success and failure, we use the unbiased estimator of the probability of success in cluster Cl_i. Therefore, we assume that there is a true Bernoulli distributed outcome

$$Y^i \sim \mathcal{B}(p^i),$$

with probability of succes $p^i = P(Y^i = 1)$ for all patients in cluster Cl_i. Furthermore, let

$$Y_j^i \sim \mathcal{B}(p^i)$$

be an independent Bernoulli distributed random variable for the outcome of patient j, $j = 1 \ldots, \kappa_i$, in the training patient data set in cluster Cl_i with realization y_j^i and unknown probability of success

$$p^i = P(Y_j^i = 1).$$

The sample space of Y^i and Y_j^i is denoted by $\Omega_{Cl_i} = \{0, 1\}$, for $i = 1, \ldots, k$ and $j = 1, \ldots, \kappa_i$. For Bernoulli distributed random variables holds

$$E(Y_j^i) = p^i \text{ and } \operatorname{var}(Y_j^i) = p^i(1 - p^i).$$

Then the estimators for the probability of success and the variance of Y_j^i in cluster Cl_i are defined as follows.

Theorem 5.9.4. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome and $\hat{S}^i := \{(\hat{x}_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ the patient data set of cluster Cl_i. Then*

$$\tilde{P}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} \mathbf{1}_{\{1\}}(Y_j^i),$$

is the unbiased estimator for the patients' probability of success in cluster Cl_i and

$$f_i(Cl_i) = \tilde{p}^i := \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(y_j^i),$$

the corresponding estimation, for $i = 1, \ldots, k$.

Remark 5.9.5. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$ *with Bernoulli distributed outcome and* $\hat{S}^i := \{(\hat{x}_j^i, y_j^i)\}_{j=1}^{\kappa_i}$ *the patient data set of cluster* Cl_i*. Then*

$$(\tilde{\Sigma}^i)^2 := \tilde{P}^i(1 - \tilde{P}^i),$$

is the estimator for the unknown variance of the patients' probability of success in cluster Cl_i *and*

$$(\tilde{\sigma}^i)^2 = \tilde{p}^i(1 - \tilde{p}^i),$$

the corresponding estimation, for $i = 1, \ldots, k$.

Also in the binary case, we determine the $(1 - \alpha)$ confidence interval $\mathcal{I}^i$ of the true probability of success p^i of patients in cluster Cl_i with the pre-defined significance level α. The unbiased estimator $\tilde{P}^i$ is normally distributed with expected value

$$E(\tilde{P}^i) = p^i$$

and standard deviation

$$\sigma_{\tilde{P}^i} := \frac{\tilde{\sigma}^i}{\sqrt{\kappa_i}}.$$

Then due to the central limit theorem, the quotient

$$\frac{\tilde{P}^i - p^i}{\sigma_{\tilde{P}^i}}$$

is approximately standardized normally distributed and for the $(1-\alpha)$ confidence interval $\mathcal{I}^i$ of the true probability of success p^i in cluster Cl_i follows

$$\mathcal{I}^i = [\tilde{p}^i - z[1 - \frac{\alpha}{2}]\sigma_{\tilde{P}^i}, \tilde{p}^i + z[1 - \frac{\alpha}{2}]\sigma_{\tilde{P}^i}],$$

where $z[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution. From this it follows that $P(p^i \in \mathcal{I}^i) = 1 - \alpha$.

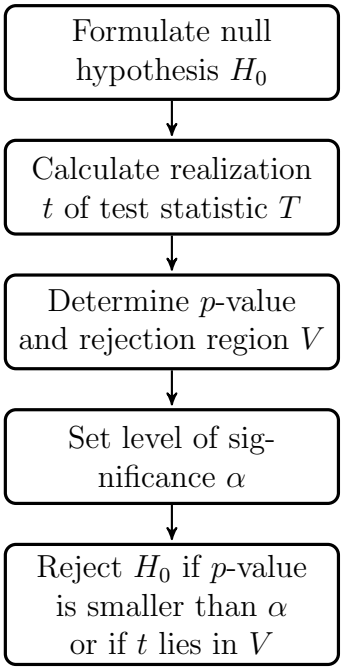

Figure 5.4.: Hypothesis test procedure for the right and left tailed tests

Finally it can be concluded, that the cluster value $f_i(Cl_i)$ of cluster Cl_i, for $i = 1, \ldots, k$, is then the predictive outcome or the predictive probability of responding to a medication for each patient who would be assigned to this corresponding cluster due to his characteristic values combination. In the next step, we need to analyze the reliability of this predictive value in the training patient data set.

5.9.3. Statistical evaluation

For the evaluation of the reliability of the predictive value, by means of the predicted outcome or the predicted response to an administered medication in cluster Cl_i we use the hypothesis test procedure shown in Figure 5.4. This procedure is relying on the patient collectives, $Cl_1, \ldots, Cl_k$, determined by the endpoint-oriented geometric clustering approach and consists of left and right tailed hypothesis tests for each cluster based on the cluster values.

The geometric cluster Algorithm 4 separates the underlying training patient data set into k clusters $Cl_1, \ldots, Cl_k$. Then for each cluster the cluster value, by means of the expected outcome or the probability of responding to an adminis-

tered medication in the binary case, is calculated according to Definition 5.9.2 or 5.9.4. For the evaluation of the reliability of the predictive values, we use the non-restricting condition that the clusters are sorted and renumbered with regard to their cluster values. Thus, for the cluster values of $Cl_1, \ldots, Cl_k$ holds

$$f_1(Cl_1) \leq f_2(Cl_2) \leq \cdots \leq f_k(Cl_k).$$

Cardinal data

The main assumption for the evaluation states that the true prediction y^i for the normally distributed outcome in cluster Cl_i satisfies inequation

$$f_i(Cl_i) \cdot \delta_l^i \leq y^i \leq f_i(Cl_i) \cdot \delta_u^i,$$

for $i = 2, \ldots, k-1$, and for the left and right border it holds

$$y^1 \leq f_1(Cl_1) \cdot \delta_u^1 \text{ and } y^k \geq f_k(Cl_k) \cdot \delta_l^k,$$

where δ_l^i is a lower and δ_u^i an upper parameter for the adjustment of $f_i(Cl_i)$. In the first step of the hypothesis test procedure, we formulate the null hypotheses. Like it is common for hypothesis testing, the goal is to reject the null hypothesis. Therefore, we set for the right tailed hypothesis test

$$H_{0,l}^i : y^i \leq f_i(Cl_i) \cdot \delta_l^i =: \tilde{y}_l^i, \text{ for } i = 2, \ldots, k,$$

which indicates the assumption that the true patients' outcome y^i is at most $f_i(Cl_i) \cdot \delta_l^i$. Furthermore, we formulate the null hypothesis of the left tailed hypothesis test,

$$H_{0,u}^i : y^i \geq f_i(Cl_i) \cdot \delta_u^i =: \tilde{y}_u^i, \text{ for } i = 1, \ldots, k-1,$$

which stand for the assumption that y^i is at least $f_i(Cl_i) \cdot \delta_u^i$. Thus, for each cluster we obtain a set of null hypotheses which can be summarized to one big set of hypotheses.

Definition 5.9.6. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the trans-*

formed patient data set $\hat{S}$ and $f_1(Cl_1), \ldots, f_k(Cl_k)$ the corresponding sorted cluster values. Then the set

$$\mathcal{H}_0 := \{H^1_{0,u}, (H^2_{0,l}, H^2_{0,u}), \ldots, (H^{(k-1)}_{0,l}, H^{(k-1)}_{0,u}), H^k_{0,l}\}$$

of the left and right tailed hypotheses tests is called set of null hypotheses for clustering Cl. The corresponding set

$$\Delta := \{\delta^1_u, (\delta^2_l, \delta^2_u), \ldots, (\delta^{(k-1)}_l, \delta^{(k-1)}_u), \delta^k_l\},$$

is called set of parameters for clustering Cl.

For the final determination of the set of null hypotheses for clustering Cl, we need to define the adjusting parameters δ^i_l and δ^i_u for the corresponding hypotheses. There are different approaches for the setting of these parameters. The first one is based on the convex combination of the neighbored cluster values.

Definition 5.9.7. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$, $f_1(Cl_1), \ldots, f_k(Cl_k)$ the corresponding sorted cluster values and $\mathcal{H}_0$ the set of null hypotheses for clustering Cl. Then the parameters of set Δ are given by*

$$\delta^i_u := 1 - \beta^i + \frac{\beta^i f_{i+1}(Cl_{i+1})}{f_i(Cl_i)}, \text{ for } i = 1, \ldots, k-1$$
$$\delta^i_l := 1 - \beta^i + \frac{\beta^i f_{i-1}(Cl_{i-1})}{f_i(Cl_i)}, \text{ for } i = 2, \ldots, k,$$

with $\beta^i \in [0, 1]$.

A second approach for the determination of the parameters of set Δ is the use of the borders of the confidence intervals defined in Section 5.9.1. If we have a sufficiently large random sample ($\kappa_i > 30$) and the patients outcome is normally distributed, we use the following parameters.

Definition 5.9.8. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome, $f_1(Cl_1), \ldots, f_k(Cl_k)$*

the corresponding sorted cluster values and $\mathcal{H}_0$ the set of null hypotheses for clustering Cl. Then the parameters of set Δ are given by

$$\delta_u^i := 1 + \frac{\beta^i z[1-\frac{\alpha}{2}]\tilde{\sigma}^i}{f_i(Cl_i)\sqrt{\kappa_i}}, \text{ for } i = 1, \ldots, k-1$$
$$\delta_l^i := 1 - \frac{\beta^i z[1-\frac{\alpha}{2}]\tilde{\sigma}^i}{f_i(Cl_i)\sqrt{\kappa_i}}, \text{ for } i = 2, \ldots, k.$$

if $\kappa_i > 30$, with $\beta^i \in \mathbb{R}$, the $(1-\frac{\alpha}{2})$-quantile $z[1-\frac{\alpha}{2}]$ of the standard normal distribution and the estimation for the variance $\tilde{\sigma}^i$ of the patients' outcome in the training patient data set in cluster Cl_i.

If the number of patients in the training data set in cluster Cl_i is small, we use the quantile of the student's t-distribution with $(\kappa_i - 1)$ degrees of freedom.

Definition 5.9.9. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome, $f_1(Cl_1), \ldots, f_k(Cl_k)$ the corresponding sorted cluster values and $\mathcal{H}_0$ the set of null hypotheses for clustering Cl. Then the parameters of set Δ are given by*

$$\delta_u^i := 1 + \frac{\beta^i t_{(\kappa_i-1)}[1-\frac{\alpha}{2}]\tilde{\sigma}^i}{f_i(Cl_i)\sqrt{\kappa_i}}, \text{ for } i = 1, \ldots, k-1$$
$$\delta_l^i := 1 - \frac{\beta^i t_{(\kappa_i-1)}[1-\frac{\alpha}{2}]\tilde{\sigma}^i}{f_i(Cl_i)\sqrt{\kappa_i}}, \text{ for } i = 2, \ldots, k.$$

if $\kappa_i \leq 30$, with $\beta^i \in \mathbb{R}$, the $(1-\frac{\alpha}{2})$-quantile $t_{(\kappa_i-1)}[1-\frac{\alpha}{2}]$ of the student's t-distribution and the estimation or the variance $\tilde{\sigma}^i$ of the patients' outcome in the training patient data set in cluster Cl_i.

Like it is stated above, the cluster value is estimated on the basis of the underlying training patient data set. For the evaluation of the reliability of this estimation, we need to formulate the test statistic. Therefore, in the next step of the hypothesis test procedure, we determine the cluster value on the basis of the testing patients data set. But first of all, we define the testing data set of cluster Cl_i.

Definition 5.9.10. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$, $\hat{S}^{te} = \{(\hat{x}_j^{te}, y_j^{te})\}_{j=1}^{n^{te}}$ be the testing patient data set and $\{Cl(x_j^{te})\}_{j=1}^{n^{te}}$ be the set of cluster assignment vectors for the testing patient data set. Then*

$$\hat{S}_{Cl_i}^{te} := \{(\hat{x}_j^{te}, y_j^{te}) | Cl_i(x_j^{te}) = 1\}_{j=1}^{n^{te}} \subset \mathbb{R}^d \times \Omega_{Cl_i} \subseteq \hat{S}^{te}$$

is the testing patient data set of cluster Cl_i and $\kappa_i^{te} := |\hat{S}_{Cl_i}^{te}|$ the number of patients in the testing data set of cluster Cl_i, for $i = 1, \ldots, k$.

For the calculation of the cluster value of the testing data set of cluster Cl_i, let

$$Y_j^{i,te} \sim \mathcal{N}(y^i, (\sigma^i)^2)$$

be a independent normally distributed random variable for the outcome of patient j, $j = 1, \ldots, \kappa_i^{te}$, in the testing patient data set in cluster Cl_i with realization $y_j^{i,te}$, expected value y^i and unknown variance $(\sigma^i)^2$. The sample space of $Y_j^{i,te}$ is denoted by Ω_{Cl_i}, for $i = 1, \ldots, k$ and $j = 1, \ldots, \kappa_i^{te}$. Then the unbiased estimator for the expected value is defined as follows.

Theorem 5.9.11. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome and $\hat{S}_{Cl_i}^{te} = \{(\hat{x}_j^{i,te}, y_j^{i,te})\}_{j=1}^{\kappa_i^{te}}$ be the testing patient data set of cluster Cl_i. Then*

$$\tilde{Y}^{i,te} := \frac{1}{\kappa_i^{te}} \sum_{j=1}^{\kappa_i^{te}} Y_j^{i,te},$$

is the unbiased estimator for the expected value of the patients' outcome in cluster Cl_i and

$$f_i^{te}(Cl_i) = \tilde{y}^{i,te} := \frac{1}{\kappa_i^{te}} \sum_{j=1}^{\kappa_i^{te}} y_j^{i,te}$$

the corresponding estimation and cluster value of the testing patient data set of cluster Cl_i, for $i = 1, \ldots, k$.

This unbiased estimator in cluster Cl_i, $i = 1, \ldots, k$, is normally distributed

with expected value

$$E(\tilde{Y}^{i,te}) = y^i$$

and standard deviation

$$\sigma_{\tilde{Y}^{i,te}} = \frac{\sigma^i}{\sqrt{\kappa_i^{te}}}.$$

Since the standard deviation σ^i is unknown, we need to use an estimator which is given in the next theorem.

Theorem 5.9.12. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome and $\hat{S}^{te}_{Cl_i} = \{(\hat{x}_j^{i,te}, y_j^{i,te})\}_{j=1}^{\kappa_i^{te}}$ be the testing patient data set of cluster Cl_i. Then*

$$(\tilde{\Sigma}^{i,te})^2 := \frac{1}{\kappa_i^{te} - 1} \sum_{j=1}^{\kappa_i^{te}} (Y_j^{i,te} - \tilde{Y}^{i,te})^2,$$

is the unbiased estimator for the unknown variance of the patients' outcome in cluster Cl_i and

$$(\tilde{\sigma}^{i,te})^2 = \frac{1}{\kappa_i^{te} - 1} \sum_{j=1}^{\kappa_i^{te}} (y_j^{i,te} - \tilde{y}^{i,te})^2,$$

the corresponding estimation, for $i = 1, \ldots, k$.

In case of normally distributed patients' outcome the test statistics T_l^i and T_u^i for the null hypotheses of set $\mathcal{H}_0$ and the corresponding realizations t_l^i and t_u^i are defined in the following.

Definition 5.9.13. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome. Then the test statistics of set $\mathcal{H}_0$ are given by*

$$T_u^i := \frac{\tilde{Y}^{i,te} - \tilde{y}_u^i}{\frac{\tilde{\sigma}^{i,te}}{\sqrt{\kappa_i^{te}}}}, \text{ for } i = 1, \ldots, k-1$$

$$T_l^i := \frac{\tilde{Y}^{i,te} - \tilde{y}_l^i}{\frac{\tilde{\sigma}^{i,te}}{\sqrt{\kappa_i^{te}}}}, \text{ for } i = 2, \ldots, k,$$

with the estimator $\tilde{Y}^{i,te}$ *for the patients' outcome and the estimation* $\tilde{\sigma}^{i,te}$ *for the variance in cluster* Cl_i*. The set of test statistics of* $\mathcal{H}_0$ *is denoted by*

$$\mathcal{T} := \{T_u^1, (T_l^2, T_u^2), \ldots, (T_l^{(k-1)}, T_u^{(k-1)}), T_l^k\}.$$

Remark 5.9.14. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$ *with normally distributed outcome. Then the realizations of the test statistics in* $\mathcal{T}$ *of set* $\mathcal{H}_0$ *are given by*

$$t_u^i := \frac{f_i^{te}(Cl_i) - \tilde{y}_u^i}{\frac{\tilde{\sigma}^{i,te}}{\sqrt{\kappa_i^{te}}}}, \text{ for } i = 1, \ldots, k-1$$
$$t_l^i := \frac{f_i^{te}(Cl_i) - \tilde{y}_l^i}{\frac{\tilde{\sigma}^{i,te}}{\sqrt{\kappa_i^{te}}}}, \text{ for } i = 2, \ldots, k$$

with the estimation $f_i^{te}(Cl_i)$ *of the patients' outcome in cluster* Cl_i*.*

Since the standard deviation of $\tilde{Y}^{i,te}$ has to be estimated, the test statistics are quotients built from two random variables. For a sufficient large number of random samples κ_i^{te} in the testing patient data set in cluster Cl_i, the test statistic converges to the standard normal distribution due to the central limit theorem. Now, the null hypotheses of $\mathcal{H}_0$ can be rejected, if the realization of the test statistics lies in the rejection region. In the case of standardized normally distributed test statistics in $\mathcal{T}$ the rejection region is defined as follows.

Definition 5.9.15. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$ *with normally distributed outcome and let* $\mathcal{T}$ *be the set of test statistics of* $\mathcal{H}_0$ *for clustering* Cl*. The refection regions for the test statistics in* $\mathcal{T}$ *are given by*

$$V_u^i := \{T_u^i \in \mathbb{R} | T_u^i < z[\alpha]\}, \text{ for } i = 1, \ldots, k-1$$
$$V_l^i := \{T_l^i \in \mathbb{R} | T_u^i > z[1-\alpha]\}, \text{ for } i = 2, \ldots, k,$$

if $\kappa_i^{te} > 30$*.* $z[\alpha]$ *is the* α*- and* $z[1-\alpha]$ *is the* $(1-\alpha)$*-quantile of the standard normal distribution. The set of the rejection regions is denoted by*

$$\mathcal{V} := \{V_u^1, (V_l^2, V_u^2), \ldots, (V_l^{(k-1)}, V_u^{(k-1)}), V_l^k\}.$$

If the number of patients is not sufficiently large we need to use the quantiles of the student's t-distribution. This approach is specified in the following definition.

Definition 5.9.16. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome and let $\mathcal{T}$ be the set of test statistics of $\mathcal{H}_0$ for clustering Cl. The refection regions for the test statistics in $\mathcal{T}$ are given by*

$$V_u^i := \{T_u^i \in \mathbb{R} | T_u^i < t_{(\kappa_i^{te}-1)}[\alpha]\}, \text{ for } i = 1, \dots, k-1$$
$$V_l^i := \{T_l^i \in \mathbb{R} | T_u^i > t_{(\kappa_i^{te}-1)}[1-\alpha]\}, \text{ for } i = 2, \dots, k,$$

if $\kappa_i^{te} \leq 30$. $t_{(\kappa_i^{te}-1)}[\alpha]$ is the α- and $t_{(\kappa_i^{te}-1)}[1-\alpha]$ is the $(1-\alpha)$-quantile of the student's t-distribution with $(\kappa_i^{te}-1)$ degrees of freedom. The set of the rejection regions is denoted by

$$\mathcal{V} := \{V_u^1, (V_l^2, V_u^2), \dots, (V_l^{(k-1)}, V_u^{(k-1)}), V_l^k\}.$$

The more intuitive way to evaluate the realization of the test statistics is the use of the p-value already discussed in the previous sections. We reject a null hypothesis of set $\mathcal{H}_0$ if the corresponding p-value is smaller than the predefined α. Hence, if the probability to get the realization of the test statistic of set $\mathcal{T}$ or an extremer value, under the assumption that the null hypothesis is true, is smaller than α we reject the null hypothesis. Thus the smaller the p-value the stronger the presumption against the null hypothesis in set $\mathcal{H}_0$. For this approach, we also differentiate between small and large random samples.

Definition 5.9.17. *Let $Cl = (Cl_1, \dots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with normally distributed outcome and let $\mathcal{T}$ be the set of test statistics of $\mathcal{H}_0$ for clustering Cl. The p-values for the test statistics in $\mathcal{T}$ are then given by*

$$pv_u^i := P(T_u^i < t_u^i | H_{0,u}^i) = F_{\mathcal{N}}(t_u^i), \text{ for } i = 1, \ldots, k-1$$
$$pv_l^i := P(T_l^i > t_l^i | H_{0,l}^i) = 1 - F_{\mathcal{N}}(t_l^i), \text{ for } i = 2, \ldots, k,$$

if $\kappa_i^{te} > 30$. $F_{\mathcal{N}}$ *denotes the cumulative distribution function of the standard normal distribution. The set of the p-values is denoted by*

$$\mathcal{P} := \{pv_u^1, (pv_l^2, pv_u^2), \ldots, (pv_l^{(k-1)}, pv_u^{(k-1)}), pv_l^k\}$$

Definition 5.9.18. *Let* $Cl = (Cl_1, \ldots, Cl_k)$ *be a* (k, l, u)*-clustering of the transformed patient data set* $\hat{S}$ *with normally distributed outcome and let* $\mathcal{T}$ *be the set of test statistics of* $\mathcal{H}_0$ *for clustering* Cl*. The p-values for the test statistics in* $\mathcal{T}$ *are given by*

$$pv_u^i := P(T_u^i < t_u^i | H_{0,u}^i) = F_{t_{(\kappa_i^{te}-1)}}(t_u^i), \text{ for } i = 1, \ldots, k-1$$
$$pv_l^i := P(T_l^i > t_l^i | H_{0,l}^i) = 1 - F_{t_{(\kappa_i^{te}-1)}}(t_l^i), \text{ for } i = 2, \ldots, k,$$

if $\kappa_i^{te} \leq 30$. $F_{t_{(\kappa_i^{te}-1)}}$ *denotes the cumulative distribution function of the student's t-distribution with* $(\kappa_i^{te} - 1)$ *degrees of freedom. The set of the p-values is denoted by*

$$\mathcal{P} := \{pv_u^1, (pv_l^2, pv_u^2), \ldots, (pv_l^{(k-1)}, pv_u^{(k-1)}), pv_l^k\}.$$

With set $\mathcal{P}$ we now have an indication how plausible the predictive value $f_i(Cl_i)$ in cluster Cl_i, for $i = 1, \ldots, k$, is. The elements of $\mathcal{P}$ can be seen as probability for trusting the predictive values.

Binary data

Also for patients' outcome with only two occurrences the assumption is that the true prediction p^i for the probability of success in cluster Cl_i satisfies inequation

$$f_i(Cl_i)\delta_l^i \leq p^i \leq f_i(Cl_i)\delta_u^i,$$

for $i = 2, \ldots, k-1$, and for the left and right border holds

$$p^1 \leq f_1(Cl_1)\delta_u^1 \text{ and } p^k \geq f_k(Cl_k)\delta_l^k,$$

where δ_l^i is a lower and δ_u^i an upper parameter for the adjustment of $f_i(Cl_i)$. In the binary case the right tailed hypothesis test is given by

$$H_{0,l}^i : p^i \leq f_i(Cl_i)\delta_l^i =: \tilde{p}_l^i, \text{ for } i = 2, \ldots, k,$$

which indicates the assumption that the true probability of success p^i is at most $\tilde{p}_l^i$. The null hypothesis of the left tailed hypothesis test is then formulated by

$$H_{0,u}^i : p^i \geq f_i(Cl_i)\delta_u^i =: \tilde{p}_u^i, \text{ for } i = 1, \ldots, k-1,$$

which stand for the assumption that p^i is at least $\tilde{p}_u^i$. These null hypotheses and the adjusting parameters can be summarized to set $\mathcal{H}_0$ and set Δ like it is already specified in Definition 5.9.6.

The parameter of set Δ can be chosen according to Definition 5.9.7 or as adjusted borders of the confidence interval defined in Section 5.9.1.

Definition 5.9.19. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome, $f_1(Cl_1), \ldots, f_k(Cl_k)$ the corresponding sorted cluster values and $\mathcal{H}_0$ the set of hypotheses of clustering Cl. Then the parameters of set Δ are given by*

$$\delta_u^i := 1 + \frac{\beta^i z[1-\frac{\alpha}{2}]\tilde{\sigma}^i}{f_i(Cl_i)\sqrt{\kappa_i}}, \text{ for } i = 1, \ldots, k-1$$
$$\delta_l^i := 1 - \frac{\beta^i z[1-\frac{\alpha}{2}]\tilde{\sigma}^i}{f_i(Cl_i)\sqrt{\kappa_i}}, \text{ for } i = 2, \ldots, k.$$

with $\beta^i \in \mathbb{R}$, the $(1-\frac{\alpha}{2})$-quantile $z[1-\frac{\alpha}{2}]$ of the standard normal distribution and the standard deviation of the patients' probability of success $\tilde{\sigma}^i$ in cluster Cl_i.

For the evaluation of the reliability in the binary case, we also need to formulate the corresponding test statistics of set $\mathcal{H}_0$ with the help of the testing patient data set $\hat{S}_{Cl_i}^{te}$ of cluster Cl_i, for $i = 1, \ldots, k$, specified in Definition 5.9.10. Therefore,

let

$$Y_j^{i,te} \sim \mathcal{B}(p^i)$$

be a independent Bernoulli distributed random variable for the probability of success of patient j, $j = 1, \ldots, \kappa_i^{te}$ in the testing patient data set in cluster Cl_i with realization $y_j^{i,te}$, expected value p^i and unknown variance $p^i(1-p^i)$. The sample spaces of $Y_j^{i,te}$ is denoted by $\Omega_{Cl_i} = \{0, 1\}$, for $i = 1, \ldots, k$ and $j = 1, \ldots, \kappa_i^{te}$, . The unbiased estimator for the unknown probability of success is given in the next theorem.

Theorem 5.9.20. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome and $\hat{S}_{Cl_i}^{te} = \{(\hat{x}_j^{i,te}, y_j^{i,te})\}_{j=1}^{\kappa_i^{te}}$ the testing patient data set of cluster Cl_i. Then*

$$\tilde{P}^{i,te} := \frac{\sum_{j=1}^{\kappa_i^{te}} 1_{\{1\}}(Y_j^{i,te})}{\kappa_i^{te}},$$

is the unbiased estimator of the expected value of the patients' outcome in cluster Cl_i and

$$f_i^{te}(Cl_i) = \tilde{p}^{i,te} := \frac{\sum_{j=1}^{\kappa_i^{te}} 1_{\{1\}}(y_j^{i,te})}{\kappa_i^{te}}$$

the corresponding estimation and cluster value of the testing patient data set of cluster Cl_i, for $i = 1, \ldots, k$.

Then the test statistics T_l^i and T_u^i for the null hypotheses of set $\mathcal{H}_0$ and the corresponding realizations t_l^i and t_u^i are defined in the following.

Definition 5.9.21. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome. Then the test statistics of set $\mathcal{H}_0$ are given by*

$$T_u^i := \frac{\tilde{P}^{i,te} - \tilde{p}_u^i}{\sqrt{\frac{\tilde{p}_u^i(1-\tilde{p}_u^i)}{\kappa_i^{te}}}}, \text{ for } i = 1, \ldots, k-1$$

$$T_l^i := \frac{\tilde{P}^{i,te} - \tilde{p}_l^i}{\sqrt{\frac{\tilde{p}_l^i(1-\tilde{p}_l^i)}{\kappa_i^{te}}}}, \text{ for } i = 2, \ldots, k,$$

with the estimator $\tilde{P}^{i,te}$ of the probability of success in cluster Cl_i, for $i = 1, \ldots, k$. The set of test statistics of $\mathcal{H}_0$ is denoted by

$$\mathcal{T} := \{T_u^1, (T_l^2, T_u^2), \ldots, (T_l^{(k-1)}, T_u^{(k-1)}), T_l^k\}.$$

Remark 5.9.22. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$. Then the realizations of the test statistics in $\mathcal{T}$ of set $\mathcal{H}_0$ are given by*

$$t_u^i := \frac{f^{te}(Cl_i) - \tilde{p}_u^i}{\sqrt{\frac{\tilde{p}_u^i(1-\tilde{p}_u^i)}{\kappa_i^{te}}}}, \text{ for } i = 1, \ldots, k-1$$
$$t_l^i := \frac{f^{te}(Cl_i) - \tilde{p}_l^i}{\sqrt{\frac{\tilde{p}_l^i(1-\tilde{p}_l^i)}{\kappa_i^{te}}}}, \text{ for } i = 2, \ldots, k$$

with the estimation $f^{te}(Cl_i)$ of the probability of success in cluster Cl_i, for $i = 1, \ldots, k$.

Due to the central limit theorem the test statistics of set $\mathcal{T}$ are approximately standardized normally distributed. According to this knowledge the null hypotheses of $\mathcal{H}_0$ can be rejected if the realization of the test statistics are located in the rejection region defined in the following.

Definition 5.9.23. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome and let $\mathcal{T}$ be the set of test statistics of $\mathcal{H}_0$ for clustering Cl. The refection regions for the test statistics in $\mathcal{T}$ are then given by*

$$V_u^i := \{T_u^i \in \mathbb{R} | T_u^i < z[\alpha]\}, \text{ for } i = 1, \ldots, k-1$$
$$V_l^i := \{T_l^i \in \mathbb{R} | T_u^i > z[1-\alpha]\}, \text{ for } i = 2, \ldots, k,$$

where $z[\alpha]$ is the α and $z[1-\alpha]$ is the $(1-\alpha)$-quantile of the standard normal distribution. The set of the rejection regions is denoted by

$$\mathcal{V} := \{V_u^1, (V_l^2, V_u^2), \ldots, (V_l^{(k-1)}, V_u^{(k-1)}), V_l^k\}.$$

Here also the p-value is the approach for the evaluation of the realization of

the test statistic. We reject a null hypothesis of set $\mathcal{H}_0$ if the probability to get the realization of the test statistic of set $\mathcal{T}$ or an extremer value, under the assumption that the null hypothesis is true, is smaller than the pre-defined α.

Definition 5.9.24. *Let $Cl = (Cl_1, \ldots, Cl_k)$ be a (k, l, u)-clustering of the transformed patient data set $\hat{S}$ with Bernoulli distributed outcome and let $\mathcal{T}$ be the set of test statistics of $\mathcal{H}_0$ for clustering Cl. The p-values for the test statistics in $\mathcal{T}$ are given by*

$$pv_u^i := P(T_u^i < t_u^i | H_{0,u}^i) = F_{\mathcal{N}}(t_u^i), \text{ for } i = 1, \ldots, k-1$$
$$pv_l^i := P(T_l^i > t_l^i | H_{0,l}^i) = 1 - F_{\mathcal{N}}(t_l^i), \text{ for } i = 2, \ldots, k,$$

where $F_{\mathcal{N}}$ denotes the cumulative distribution function of the standard normal distribution. The set of the p-values is denoted by

$$\mathcal{P} := \{pv_u^1, (pv_l^2, pv_u^2), \ldots, (pv_l^{(k-1)}, pv_u^{(k-1)}), pv_l^k\}$$

The new approach for the cluster-based prediction of treatment effects enables the computation of a probability for trusting the predicted outcome of a patient in the identified collectives. It enhances the fundamental concept of predicting the efficacy of a medical intervention by an additional measure of trust. Especially when the treatment effect of different patient collectives are compared, a trade-off between accuracy and reliability is an upcoming effect. This can be seen in the practical part of this thesis presented in Chapter 6.

6. Practical application - Empirical results

For the practical application and comparison of the meta-analysis approach described in Chapter 3 and the new invented cluster-based approaches described in Chapter 5, we analyzed individual patient data of 12 equally conducted RCTs including 6010 patients treated with three different antidepressants and placebo. Thereby, two of the administered drugs are already approved standard therapies of depression and the third drug, which is available in four different dosages, is a new invented therapy with not sufficiently proven benefit but with the potential of a necessary treatment alternative. Therefore, this new drug was analyzed due to the additional benefit for patients in terms of health economic evaluation.

In Section 6.1, a detailed description of the individual patient data of the available clinical trials is given. In Section 6.2, the results of the conducted meta-analysis of the available data are presented. Then, in Section 6.3, the results of the new invented cluster-based meta-analysis, applied on the same individual data, are demonstrated. In terms of health economic evaluation, the results of the comparison of the both approaches, with regard to the additional benefit of the new introduced drug, are shown. In Section 6.4, we will have a closer look on the clinical heterogeneity identified by the endpoint-oriented geometric clustering approach. In the last Section 6.5, we used this new approach for the prediction of the efficacy of the three different antidepressant for the identified patient collectives. As level of significance, we used $\alpha = 0.1$ for all analyses.

6.1. Data description

Altogether, individual patient data from $N = 6010$ participants in 12 RCTs have been available to assess the efficacy of three different antidepressants in the acute treatment of depression. Hereafter, the $N = 6010$ participating patients represent the patient data set

$$S^{all} = \{(x_j, y_j)\}_{j=1}^{6010}.$$

This data set can also be seen as total population and can be described in statistical terms as marginal distribution. The goal of the conducted trials was the comparison of the new introduced antidepressant, here called A, with its dosages 5mg, 10mg, 15mg and 20mg, to already established treatments, here called C1 and C2.

The following characteristics have been available on an individual patient level:

- Name of the study (study)
- Region where the study has been conducted (region)
- Administered medication (treatment)
- Body Mass Index at the beginning of the study (BMI): The BMI is a value derived from the weight and height of an individual; it is defined as the body mass divided by the square of the body height and is universally expressed in units of $[kg/m^2]$
- MADRS total score at the beginning of the study (MADRS): MADRS is the abbreviation of Montgomery-Asberg Depression Rating Scale and is a ten-item diagnostic questionnaire which psychiatrists use to measure the severity of depressive episodes of patients with mood disorders; the overall score ranges from 0 to 60, where a higher score indicates more severe depression
- CGI-S total score at the beginning of the study (CGI-S): CGI-S is the abbreviation of Clinical Global Impression - Severity Scale and is a 7-point scale that requires the clinician to rate the severity of the patient's illness at the time of assessment, relative to the clinician's past experience with

patients who have the same diagnosis. Considering total clinical experience, a patient is assessed on severity of mental illness at the time of rating: 1, normal, not at all ill; 2, borderline mentally ill; 3, mildly ill; 4, moderately ill; 5, markedly ill; 6, severely ill; or 7, extremely ill

- Duration of the depressive episode in days at the beginning of the study (duration)
- Age (age)
- Gender (sex)
- For drop-outs: Primary Reason for withdrawal (withdrawal)

The individual binary outcome (endpoint) was measured as follows:

- Response (yes/no): a patient has responded, if the MADRS total score has been reduced by at least 50% at the end of the study
- Remission (yes/no): a patient has remitted, if the MADRS total score has been reduced to ≤ 9 points at the end of the study

Based on the available data, we were able to conduct a meta-analysis. The results are discussed in the next section.

6.2. Meta-analysis

In this section, we will present the results of the conducted meta-analysis according to the described theory in Chapter 3. The goal was the assessment of the efficacy of the three different antidepressants A, C1 and C2, on the basis of the response shown by the participating patients. The correspondent control group was treated with placebo. For the comparison of the efficacy of the different interventions, we conducted a meta-analysis for each medication. In Table 6.1 the distribution of the 6010 participating patients in studies and treatments is shown.

Study	A5mg	A10mg	A15mg	A20mg	C1	C2	PBO	Total
T11	107	97				110	102	416
T12	150	147			143		145	585
T13	150				144		142	436
T14			141	146	142		153	582
T15		187		200			189	576
T21	282						275	557
T22	150				141		139	430
T23	137	137					137	411
T24			139	140	142		150	571
T25		149		141			152	442
T26		142	137				144	423
T27	141	147		147			146	581
Total	1117	1006	417	774	570	252	1874	6010

Table 6.1.: Meta-analysis: Distribution of patients in study and treatment

6.2.1. Treatment effect estimates

Due to the recorded binary outcome for each patient in the treatment and the control group of study j, $j = 1, \ldots, n_{st}$, with the two occurrences "response" (= 1) and "no response" (= 0), for each administered drug we used the transformed Risk Ratio for the estimation of the treatment effect. Thereby, n_{st} is the number of studies in which the correspondent drug was administered. The results of the meta-analyses on the basis of the transformed Odds Ratio and the Risk Difference can be found in Appendix A. For the approximation of the transformed Risk Ratio, we had to estimate the probability of success or in our case the response rate

$$\tilde{p}_{T_j} = \frac{s_{T_j}}{n_{T_j}}$$

in the treatment group and the response rate

$$\tilde{p}_{C_j} = \frac{s_{C_j}}{n_{C_j}}$$

in the control group. The estimations are based on Table 6.2. The table of each medication can be found in Appendix A. With the help of the probabilities, the transformed Risk Ratio,

$$\tilde{\theta}_j = \ln\left(\frac{\tilde{p}_{T_j}}{\tilde{p}_{C_j}}\right)$$

	Response	No response	
Treatment group	s_{T_j}	f_{T_j}	n_{T_j}
Control group	s_{C_j}	f_{C_j}	n_{C_j}
	s_j	f_j	n_j

Table 6.2.: Meta-analysis: 2×2-table of binary outcome of study j

Treatment group: A5mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.6729	0.3787	0.0160	0.1705	0.5868
T12	2	0.4621	0.5733	0.2158	0.0130	0.0283	0.4032
T13	3	0.3732	0.5867	0.4522	0.0165	0.2408	0.6637
T21	4	0.4764	0.4894	0.0269	0.0077	-0.1174	0.1712
T22	5	0.3525	0.4400	0.2217	0.0217	-0.0206	0.4640
T23	6	0.2409	0.4526	0.6306	0.0318	0.3372	0.9241
T27	7	0.3973	0.4965	0.2229	0.0176	0.0048	0.4410

Table 6.3.: Meta-analysis: Transformed Risk Ratio, antidepressant A5mg

and the correspondent variance

$$\tilde{\sigma}_j^2 = \frac{1}{s_{T_j}} - \frac{1}{n_{T_j}} + \frac{1}{s_{C_j}} - \frac{1}{n_{C_j}}$$

for study j were estimated. Additionally, the $(1-\alpha)$ confidence interval

$$\mathcal{I}_j = [\tilde{\theta}_j - z[1-\frac{\alpha}{2}]\tilde{\sigma}_j, \tilde{\theta}_j + z[1-\frac{\alpha}{2}]\tilde{\sigma}_j],$$

for the treatment effect of study j, for $j = 1, \ldots, n_{st}$, was calculated, where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

The results of medication A5mg can be found in Table 6.3. E.g. for study $j = 6$ (T23) we got

$$\tilde{p}_{T_6} = 0.4526$$

and

$$\tilde{p}_{C_6} = 0.2409,$$

with a Risk Ratio of

$$\frac{\tilde{p}_{T_6}}{\tilde{p}_{C_6}} = \frac{0.4526}{0.2409} = 1.8788.$$

This implies that the response under the patients treated with A5mg is 1.8788 times higher than the response under patients with no verum treatment. The resulting transformed Risk Ratio is

$$\tilde{\theta}_6 = 0.6306.$$

The $(1 - \alpha)$ confidence interval is given by

$$\mathcal{I}_6 = [0.3372, 0.9241],$$

for $\alpha = 0.1$. The transformed Risk Ratios in the single studies vary from $\tilde{\theta}_4 = 0.0269$ to $\tilde{\theta}_6 = 0.6306$. This is due to the high variation in the control groups of the seven studies. In the treatment groups, the probabilities of success or the response rates seem to be equal. Nevertheless, this results in a high variation of the transformed Risk Ratio. Also the confidence intervals show less overlapping which leads to the assumption, that there is heterogeneity which had to explained.

The results for the further dosages of medication A, dosage A10mg, A15mg and A20mg can be found in Table 6.4, Table 6.5 and Table 6.6. The results of the already established depression therapies C1 and C2 are listed in Table 6.7 and Table 6.8. All medications show correspondent results in the analysis of the treatment estimates of the single trials.

6.2.2. Fixed-effects model

For the weighted aggregation of the treatment effects in the single studies, we firstly used the fixed-effects model. It is assumed that there is one true treatment effect θ for each administered medical intervention and that the variation in the treatment effects across trials is due to estimation errors. The estimated treatment effect of study j is given by

$$\tilde{\theta}_j := \theta + \epsilon_j,$$

Treatment group: A10mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.6907	0.4048	0.0161	0.1962	0.6134
T12	2	0.4621	0.5714	0.2124	0.0131	0.0239	0.4009
T15	3	0.3069	0.4920	0.4720	0.0175	0.2546	0.6894
T23	4	0.2409	0.4964	0.7230	0.0304	0.4362	1.0098
T25	5	0.2961	0.3356	0.1253	0.0289	-0.1545	0.4051
T26	6	0.3403	0.3803	0.1111	0.0249	-0.1486	0.3709
T27	7	0.3973	0.5442	0.3148	0.0161	0.1061	0.5234

Table 6.4.: Meta-analysis: Transformed Risk Ratio, antidepressant A10mg

Treatment group: A15mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T14	1	0.3333	0.5816	0.5566	0.0182	0.3348	0.7783
T24	2	0.3933	0.4604	0.1575	0.0187	-0.0675	0.3825
T26	3	0.3403	0.3723	0.0898	0.0258	-0.1742	0.3539

Table 6.5.: Meta-analysis: Transformed Risk Ratio, antidepressant A15mg

where ϵ_j is the estimation error, for $j = 1, \ldots, n_{st}$. With the inverse variance method, the summary treatment effect is given by

$$\hat{\theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\theta}_j w_j}{\sum_{i=j}^{n_{st}} w_j},$$

with weights

$$w_j = \frac{1}{\tilde{\sigma}_j^2},$$

for $j = 1, \ldots, n_{st}$. The variance was calculated by

$$\text{var}(\hat{\Theta}) = \frac{1}{\sum_{j=1}^{n_{st}} w_j}.$$

Then for the $(1-\alpha)$ confidence interval $\mathcal{I}$ of the true treatment effect θ follows

$$\mathcal{I} = [\hat{\theta} - z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\hat{\Theta})}, \hat{\theta} + z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\hat{\Theta})}],$$

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

Treatment group: A20mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T14	1	0.3333	0.6233	0.6259	0.0172	0.4101	0.8417
T15	2	0.3069	0.6000	0.6705	0.0153	0.4671	0.8738
T24	3	0.3933	0.4500	0.1346	0.0190	-0.0922	0.3614
T25	4	0.2961	0.4113	0.3289	0.0258	0.0647	0.5931
T27	5	0.3973	0.5102	0.2502	0.0169	0.0362	0.4642

Table 6.6.: Meta-analysis: Transformed Risk Ratio, antidepressants A20mg

Treatment group: C1

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T12	1	0.4621	0.5944	0.2518	0.0128	0.0658	0.4379
T13	2	0.3732	0.6875	0.6108	0.0150	0.4095	0.8122
T14	3	0.3333	0.7535	0.8156	0.0154	0.6117	1.0196
T22	4	0.3525	0.5319	0.4114	0.0195	0.1820	0.6408
T24	5	0.3933	0.5634	0.3593	0.0157	0.1529	0.5657

Table 6.7.: Meta-analysis: Transformed Risk Ratio, antidepressant C1

For each medication, the estimated summary treatment effect, the variance and the $(1-\alpha)$ confidence interval, for $\alpha = 0.1$, is shown in Table 6.9. E.g. the summary treatment effect for medication A5mg is

$$\hat{\theta} = 0.2452,$$

with variance

$$\text{var}(\hat{\Theta}) = 0.0022.$$

The $(1-\alpha)$ confidence interval is presented by

$$\mathcal{I} = [0.1688, 0.3215],$$

for $\alpha = 0.1$. This summary treatment effect is interpreted as the estimated true treatment effect for all patients. When looking back at the result of study 6, this estimation is even not lying in the confidence interval $\mathcal{I}_6$. This leads to the assumption that there is unexplained heterogeneity across the considered trials.

As can also be seen in Table 6.3 to Table 6.8, the treatment effects in the single

Treatment group: C2

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.7273	0.4564	0.0149	0.2557	0.6570

Table 6.8.: Meta-analysis: Transformed Risk Ratio, antidepressant C2

Fixed-effects model

Drug	$\hat{\theta}$	var($\hat{\Theta}$)	$\mathcal{I}$		q	pv	l^2
A5mg	0.2452	0.0022	0.1688	0.3215	14.6834	0.0229	59.14%
A10mg	0.3299	0.0027	0.2438	0.4159	11.0165	0.0879	45.54%
A15mg	0.2888	0.0068	0.1532	0.4243	6.4023	0.0407	68.76%
A20mg	0.4192	0.0036	0.3198	0.5185	4.1820	0.3819	4.35%
C1	0.4847	0.0031	0.3935	0.5760	13.6952	0.0083	70.79%
C2	0.4564	0.0149	0.2557	0.6570	0.0000		

Table 6.9.: Meta-analysis: Risk Ratio estimated by the fixed-effects model

trials vary for each medication. This might be due to estimation errors in the single trials but might have further reasons. Therefore, a common way to identify heterogeneity is the use and evaluation of the realization of the Q-statistic,

$$q = \sum_{j=1}^{n_{st}} w_j(\tilde{\theta}_j - \hat{\theta})^2.$$

For the evaluation of the result of the Q-statistic, we used the p-value

$$pv = 1 - F_{\chi^2_{(n_{st}-1)}}(q),$$

where $F_{\chi^2_{(n_{st}-1)}}$ is the cumulative function of the χ^2 distribution with $(n_{st}-1)$ degrees of freedom. Another way to assess heterogeneity is e.g. the Q related l^2 index

$$l^2 = \max\left\{0, \frac{q-(n_{st}-1)}{q}\right\} \cdot 100\%.$$

When considering medication A5mg, for the realization of the Q-statistic followed

$$q = 14.6834.$$

For the evaluation of this result, we calculated the p-value

$$pv = 1 - F_{\chi^2_2}(14.6834) = 0.0229.$$

This indicates that, under the assumption that the treatment effects are equal in the seven studies, the probability to get this result or a higher value for q, is only 2.29%. For $\alpha = 0.1$ this implies, that we can reject the null hypothesis that the treatment effects in the single trials are equal. Another way for the assessment of heterogeneity is the l^2 index. The result for medication A5mg is

$$l^2 = 59.14.\%$$

According to Table 4.2, this result can be interpreted as there may be substantial heterogeneity across the single trials. Therefore, it is reliable to reject the hypothesis that the treatment effects of the single trials are equal for each administered medication. Similar results can be found for all other medication, except for A20mg. Here, we have no statistical significant heterogeneity. The p-value is $pv = 0.3819$.

6.2.3. Random-effects model

Since there is heterogeneity that cannot readily be explained for all medications, except A20mg, we used the random-effects model. This model allows the variation of the true treatment effect across studies, i.e. that there is a treatment effect θ_j in each study. The estimated treatment effect is given by

$$\tilde{\theta}_j := \theta_j + \epsilon_j = \theta + \nu_j + \epsilon_j,$$

for $j = 1, \ldots, n_{st}$, where ϵ_j is the estimation error and ν_j is the random-effect of study j. Therefore, we assume that the variation of the estimated treatment effects in the single trials are due estimation errors and the already mentioned random-effects. With the inverse variance method, the summary treatment effect is given by

$$\hat{\theta} = \frac{\sum_{j=1}^{n_{st}} \tilde{\theta}_j w_j^*}{\sum_{j=1}^{n_{st}} w_j^*}.$$

with weights

$$w_j^* = \frac{1}{w_j^{-1} + \hat{\tau}^2}$$

for $j = 1, \ldots, n_{st}$. The variance is given by

$$\text{var}(\hat{\Theta}) = \frac{1}{\sum_{j=1}^{n_{st}} w_j^*}$$

and the inter-variance is estimated by

$$\hat{\tau}^2 = \max \left\{ 0; \frac{q - (n_{st} - 1)}{\sum_{j=1}^{n_{st}} w_j - \frac{\sum_{j=1}^{n_{st}} w_j^2}{\sum_{j=1}^{n_{st}} w_j}} \right\}.$$

Then for the $(1 - \alpha)$ confidence interval of the true treatment effect θ follows

$$\mathcal{I} = [\hat{\theta} - z[1 - \frac{\alpha}{2}]\sqrt{\text{var}(\hat{\Theta})}, \hat{\theta} + z[1 - \frac{\alpha}{2}]\sqrt{\text{var}(\hat{\Theta})}],$$

where $z[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the standardized normal distribution.

The estimated summary treatment effect, the variance and the $(1 - \alpha)$ confidence interval for each medication is shown in Table 6.10. E.g. the summary treatment effect for medication A5mg is given by

$$\hat{\theta} = 0.2476,$$

with variance

$$\text{var}(\hat{\Theta}) = 0.0022.$$

For the $(1 - \alpha)$ confidence interval resulted

$$\mathcal{I} = [0.1698, 0.3255],$$

for $\alpha = 0.1$. When analyzing all results of the random-effects model, by using the indirect comparison, the new invented medication A, with its dosages 5mg, 10mg, 15mg and 20mg, don't show an additional benefit for all patients in comparison to medication C1 or C2. The new invented medication with the highest

Random-effects model				
Drug	$\tilde{\theta}$	var($\tilde{\Theta}$)	$\mathcal{I}$	
A5mg	0.2476	0.0022	0.1698	0.3255
A10mg	0.3300	0.0028	0.2433	0.4167
A15mg	0.2870	0.0075	0.1447	0.4294
A20mg	0.4177	0.0040	0.3138	0.5216
C1	0.4854	0.0034	0.3900	0.5808
C2	0.4564	0.0149	0.2557	0.6570

Table 6.10.: Meta-analysis: Risk Ratio estimated by the random-effects model

treatment effect is A20mg with $\tilde{\theta} = 0.4177$ in comparison to medication C2 with a treatment effect of $\tilde{\theta} = 0.4854$.

Note that a random-effects model does not 'take account' of heterogeneity in the sense, that it is no longer an issue. There might be heterogeneity due to e.g. socio-demographic, biographical or clinical parameters, which hasn't been considered. The geometric clustering approach deals with the classification of patient data into patient collectives with similar characteristic values combinations. With this approach, this heterogeneity can be taken into account.

6.3. Cluster-based meta-analysis

According to the new theory described in Chapter 5, in this section the results of the conducted cluster-based meta-analysis are presented. The clustering approach has been applied on the same clinical trials available for the meta-analysis presented in Section 6.2, respectively on the data set S^{all} described in Section 6.1. Since the cluster-based meta-analysis is based on an unsupervised clustering approach, we used all available data for the identification of the patient collectives. In the following, the efficacy of the three different antidepressants A, C1 and C2 has been assessed on patient collectives classified by the endpoint-oriented geometric clustering algorithm. We assume that there is one true treatment effect for each patient collective. By using this approach, we have been able to address heterogeneity due to socio-demographic, biographical or clinical parameters, which is not considered in the common meta-analysis approach.

A_i	Characteristic	Values
A_1	Region	EU, non-EU
A_2	BMI	13 to 68.41 kg/m^2
A_3	MADRS	13 to 52
A_4	CGI-S	3 to 7
A_5	Duration	28 to 7976 days
A_6	Sex	Female, Male
A_7	Age	18 to 88 years
A_8	Withdrawal	Adverse events, Lack of efficacy, Lost to follow-up, Withdrawal of consent
	Outcome	Values
	Response	yes, no
	Remission	yes, no

Table 6.11.: Cluster-based meta-analysis: Independent and dependent variables used for the clustering approach

6.3.1. Clustering

For the application of the clustering approach explained in Chapter 5, to identify patient collectives as a partition of the data set

$$S = S^{all} = \{x_j, y_j\}_{j=1}^{6010},$$

in the first step, the available characteristics with their characteristic values had to be picked, which might have an influence on the response of a patient. For the cluster-based meta-analysis we have chosen those random variables listed in Table 6.11. Since the random variables 'BMI', 'duration' and 'age' have plenty of values and therefore, the number of value combinations would increase proportionally, we classified those characteristics like it is described in Section 5.5. In Table 6.12, the characteristic values of those three variables are divided into five classes of equal class density. Since the attributes 'treatment', 'region', 'sex', and 'withdrawal' have a nominal level of scale, we needed to transform the existing data according to Section 5.5. With this transformed data, we had been able to apply the clustering algorithm to get a predefined number of patient collectives. The parameter setup for the clustering consisted of $k = 6$ clusters, lower bounds $l_i = 100$, for $i = 1, \ldots, 6$, and no upper bounds. In this thesis, the setting $k = 6$

A_i	Characteristic	Class number	Class
A_2	BMI	1	$]0, 22.53]$
		2	$]22.53, 25.28]$
		3	$]25.53, 28.39]$
		4	$]28.39, 32.91]$
		5	$]32.91, \infty]$
A_5	Duration (in days)	1	$]0, 113]$
		2	$]113, 142]$
		3	$]142, 197]$
		4	$]197, 323]$
		5	$]323, \infty]$
A_7	Age (in years)	1	$]0, 34]$
		2	$]34, 44]$
		3	$]44, 51]$
		4	$]51, 60]$
		5	$]60, \infty]$

Table 6.12.: Cluster-based meta-analysis: Classification of random variables

Cluster Cl_i	1	2	3	4	5	6	all
Patients κ_i	1096	476	1400	1948	335	755	6010
Response $\tilde{p}^i$	0.6086	0.1639	0.4800	0.5524	0.1791	0.4689	0.4837

Table 6.13.: Cluster-based meta-analysis: Clustering results with respect to 'response' for defining 6 clusters

has been arisen due to corresponding preliminary analysis. In general, the cluster number can be set freely in principle, but should be selected goal-oriented. Each cluster is uniquely defined by combinations of those characteristic values of the various attributes that have been identified as similar by the algorithm. The clustering results can be found in Table 6.13. E.g. in cluster Cl_1, $\kappa_1 = 1096$ patients are grouped with a response rate of $\tilde{p}^1 = 0.6086$. A detailed presentation of the results is given in Appendix B.

i / Treatment	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
eu														
A10mg	0.6216	0.1688	0.1905	0.1963					0.1111	0.1525	0.7895	0.1293	0.5769	0.1661
A15mg	0.6667	0.0903	0.0714	0.1308					0.0000	0.1695	0.8333	0.1225	0.5816	0.1001
A20mg	0.6991	0.1031	0.0909	0.1028					0.0000	0.1695	0.9167	0.0816	0.6233	0.1036
A5mg	0.6122	0.1788	0.0000	0.1122					0.4000	0.0848	0.3871	0.2109	0.5492	0.1732
C1	0.8087	0.1049	0.2857	0.0654					0.5000	0.0678	0.6250	0.1088	0.7535	0.1008
PBO	0.3967	0.2737	0.0667	0.2804					0.1333	0.2542	0.1702	0.3197	0.3342	0.2782
C2	0.8523	0.0803	0.1667	0.1122					0.0000	0.1017	0.7500	0.0272	0.7273	0.07807
non eu														
A10mg			0.1129	0.1680	0.4498	0.1493	0.5556	0.1756	0.2381	0.1522	0.5043	0.1924	0.4663	0.1678
A15mg			0.2105	0.0515	0.3761	0.0836	0.5811	0.0380	0.2333	0.1087	0.4722	0.0592	0.4167	0.0600
A20mg			0.2046	0.1192	0.5421	0.1357	0.6000	0.1335	0.1795	0.1413	0.4316	0.1563	0.5032	0.1365
A5mg			0.1667	0.2114	0.4826	0.1850	0.6228	0.2028	0.1556	0.1630	0.5938	0.1579	0.5132	0.1897
C1			0.2833	0.1626	0.6962	0.1129	0.6894	0.1206	0.2046	0.1594	0.5616	0.1201	0.5947	0.1239
PBO			0.1509	0.2873	0.4197	0.3336	0.4346	0.3296	0.1711	0.2754	0.3403	0.3141	0.3839	0.3221
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table 6.14.: Cluster-based meta analysis: Response rate in clusters classified by region and treatment

6.3.2. Cluster-based analysis

Since the number of patients in some trials of the clusters is low, see Table B.7 in the Appendix, and therefore the statistical power is decreased, in a first step, we have analyzed the patient collectives, identified by the clustering algorithm, without differentiating between the single trials. The response rates and the conditional response rates have been calculated according to the introduced theory in Section 5.8 and the treatment effect estimates according to the theory described in Section 5.7.

Cluster Cl_2 and Cl_5 include all participants who have dropped out of a study. In both clusters, EU and non-EU patients are grouped. In cluster Cl_1, only EU patients and in cluster Cl_3 and Cl_4, only non-EU patients are included. In cluster Cl_6, EU and non-EU patients are grouped with a higher share of non-EU participants in comparison to the marginal distribution,

$$\tilde{\rho}^6(\text{non-EU}) = 0.8053$$
$$\tilde{\rho}(\text{non-EU}) = 0.7656.$$

The mean response rates in cluster Cl_6 stratified by region also differ,

$$\tilde{p}_1^6(\text{EU}) = 0.5034$$
$$\tilde{p}_1^6(\text{non-EU}) = 0.4605.$$

In Table 6.14, all response rates stratified by region and medication are presented. This table shows regional differences in the response rates for all medications. In cluster Cl_1 the response rates are very high for all medications and the PBO response is low which leads to high transformed Risk Ratios,

$$\begin{aligned}
\tilde{\theta}^1(\text{A5mg, EU}) &= 0.4340 \\
\tilde{\theta}^1(\text{A10mg, EU}) &= 0.4492 \\
\tilde{\theta}^1(\text{A15mg, EU}) &= 0.5192 \\
\tilde{\theta}^1(\text{A20mg, EU}) &= 0.5667 \\
\tilde{\theta}^1(\text{C1, EU}) &= 0.7123 \\
\tilde{\theta}^1(\text{C2, EU}) &= 0.7648,
\end{aligned}$$

where the transformed Risk Ratios are defined by

$$\tilde{\theta}^i(\text{drug, region}) = \ln\left(\frac{\tilde{p}^i(\text{drug, region})}{\tilde{p}^i(\text{PBO, region})}\right),$$

for $i = 1, \ldots, k$. Medication C2 shows the highest and A5mg the lowest transformed Risk Ratio. In comparison to the EU cluster Cl_1, in the non-EU clusters Cl_3 and Cl_4 the response rates in the control groups are higher and those of the treatment groups are lower. This results in lower transformed Risk Ratios, which can be seen in the following. For cluster Cl_3, we have

$$\begin{aligned}
\tilde{\theta}^3(\text{A5mg, non-EU}) &= 0.1397 \\
\tilde{\theta}^3(\text{A10mg, non-EU}) &= 0.0692 \\
\tilde{\theta}^3(\text{A15mg, non-EU}) &= -0.1098 \\
\tilde{\theta}^3(\text{A20mg, non-EU}) &= 0.2559 \\
\tilde{\theta}^3(\text{C1, non-EU}) &= 0.5061.
\end{aligned}$$

The transformed Risk Ratios are significantly lower than in cluster Cl_1. Medication C1 shows the highest and A15mg the lowest treatment effect. In cluster Cl_4 the results are

$$\begin{aligned}
\tilde{\theta}^4(\text{A5mg, non-EU}) &= 0.3598 \\
\tilde{\theta}^4(\text{A10mg, non-EU}) &= 0.2456 \\
\tilde{\theta}^4(\text{A15mg, non-EU}) &= -0.2905 \\
\tilde{\theta}^4(\text{A20mg, non-EU}) &= 0.3226 \\
\tilde{\theta}^4(\text{C1, non-EU}) &= 0.4614.
\end{aligned}$$

Here, also for medication C1 we have the highest and for A10mg the lowest transformed Risk Ratio. In the mixed cluster Cl_6, there are also significant regional differences. The placebo treated patients in the EU regions show a lower response rate than in the control group in the non-EU group. The treatment group in the EU regions shows a higher response than those patients in the non-EU regions. This results in significant higher treatment effects in the EU group in cluster Cl_6. The results are

$$\begin{aligned}
\tilde{\theta}^6(\text{A5mg, EU}) &= 0.8216 & \tilde{\theta}^6(\text{A5mg, non-EU}) &= 0.5566 \\
\tilde{\theta}^6(\text{A10mg, EU}) &= 1.5343 & \tilde{\theta}^6(\text{A10mg, non-EU}) &= 0.3932 \\
\tilde{\theta}^6(\text{A15mg, EU}) &= 1.5884 & \tilde{\theta}^6(\text{A15mg, non-EU}) &= 0.3276 \\
\tilde{\theta}^6(\text{A20mg, EU}) &= 1.6837 & \tilde{\theta}^6(\text{A20mg, non-EU}) &= 0.2376 \\
\tilde{\theta}^6(\text{C1, EU}) &= 1.3007 & \tilde{\theta}^6(\text{C1, non-EU}) &= 0.5010 \\
\tilde{\theta}^6(\text{C1, EU}) &= 1.4830. & \tilde{\theta}^6(\text{C2, non-EU}) &= n/a.
\end{aligned}$$

For the EU patients in cluster Cl_6, the most efficient medication is indeed A20mg and for the non-EU group the highest treatment effect has medication A5mg, which implies an additional benefit for the new invented medication A. These results are very interesting due to the findings of the meta-analysis approach in Section 6.2, where medication A seems to have no additional benefit. We can conclude that there is very high regional heterogeneity in the PBO response rates. Non-EU patients showed a higher and EU patients a lower PBO response. A more detailed examination of the heterogeneity can be found in the Section 6.4.

	Response	No response
Treatment group	$s^i_{T_j}$	$f^i_{T_j}$
Control group	$s^i_{C_j}$	$f^i_{C_j}$

Table 6.15.: Cluster-based meta-analysis: 2×2-table of binary outcome of study j in cluster Cl_i

When analyzing the patients characteristics in the corresponding clusters, it is noticeable that in the EU cluster Cl_1 more patients with a lower BMI, a higher depression at baseline and a shorter duration of the last episode can be found. In the non-EU cluster Cl_3 patients with a higher BMI, a lower depression and longer duration of the last espisode are represented. And in the non-EU cluster Cl_4 patients with a higher BMI, a lower depression and a shorter duration of the last episode are grouped. In cluster Cl_6 there is no conspicuousness concerning the distribution of the BMI and the duration of the last depression episode, but we find more patients with a lower depression. In the drop-out cluster Cl_2, EU and non-EU patients are grouped with a rather low BMI, a rather low depression and a short duration of the last episode. In cluster Cl_5, EU and non-EU patients are included with a rather low depression, a long duration of the last episode and the share of male patients was increased. It is noteworthy that the participants in Cl_2 and Cl_5 are rather young. As we can see, socio-demographic, biographical and clinical parameters vary across the identified patient collectives.

Although the number of patients is low in some trials of the identified clusters, we will have a look at the results of the cluster-based meta-analysis in the next section.

6.3.3. Treatment effect estimates

In this section, we will present the results of the cluster-based meta-analysis. Therefore, in the first step, we analyzed the treatment effects in the single trials of the clusters. Since the recorded outcome for each patient in the treatment group and in the control group of study j in cluster Cl_i, $j \in \mathcal{I}_{Cl_i}$ and $i = 1, \ldots, k$, is binary, with the two occurrences "response" $(= 1)$ and "no response" $(= 0)$, for each administered drug, we used the transformed Risk Ratio for the estimation

of the treatment effect. The results of the cluster-based meta-analyses on the basis of the transformed Odds Ratio and the Risk Difference can be found in Appendix B. The set of indices of the studies included in cluster Cl_i, in which the correspondent drug is administered, is denoted by $\mathcal{I}_{Cl_i}$. . For the approximation of the transformed Risk Ratio of study j in cluster Cl_i, we had to estimate the response rate

$$\tilde{p}^i_{T_j} = \frac{s^i_{T_j}}{s^i_{T_j} + f^i_{T_j}}$$

in the treatment group and the response rate

$$\tilde{p}^i_{C_j} = \frac{s^i_{C_j}}{s^i_{C_j} + f^i_{C_j}}$$

in the control group. The estimations are based on Table 6.15. The table of each medication can be found in Appendix B. Based on the response rates, the transformed Risk Ratio

$$\tilde{\theta}^i_j = \ln\left(\frac{\tilde{p}^i_{T_j}}{\tilde{p}^i_{C_j}}\right),$$

and the correspondent variance

$$(\tilde{\sigma}^i_j)^2 = \frac{1}{s^i_{T_j}} - \frac{1}{s^i_{T_j} + f^i_{T_j}} + \frac{1}{s^i_{C_j}} - \frac{1}{s^i_{C_j} + f^i_{C_j}}$$

for study j in cluster Cl_i were estimated. Additionally the $(1-\alpha)$ confidence interval

$$\mathcal{I}^i_j = [\tilde{\theta}^i_j - z[1-\frac{\alpha}{2}]\tilde{\sigma}^i_j, \tilde{\theta}^i_j + z[1-\frac{\alpha}{2}]\tilde{\sigma}^i_j],$$

for the treatment effect of study j in cluster Cl_i was calculated, where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

The results of medication A5mg can be found in Table 6.16. E.g. study T12 is represented in all clusters except cluster Cl_1, which implies, that this study was conducted in non-EU regions. In the non-EU cluster Cl_3, Cl_4 and in the mixed cluster Cl_6 the response rates in the treatment group are given by

$$\tilde{p}^3_{T_2} = 0.5667,\ \tilde{p}^4_{T_2} = 0.7162 \text{ and } \tilde{p}^6_{T_2} = 0.5294$$

Treatment group: A5mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.7204	0.2502	0.0137	0.0575	0.4428
	T23	8	0.2925	0.5146	0.5650	0.0320	0.2709	0.8592
2	T12	2	0.1333	0.3529	0.9734	0.5412	-0.2366	2.1835
	T21	6	0.2308	0.0769	-1.0986	0.5897	-2.3618	0.1645
	T22	7	0.1111	0.1765	0.4626	1.1634	-1.3115	2.2368
3	T12	2	0.4857	0.5667	0.1542	0.0557	-0.2342	0.5425
	T13	3	0.3333	0.4878	0.3808	0.0769	-0.0753	0.8369
	T21	6	0.4944	0.5165	0.0437	0.0218	-0.1990	0.2865
	T22	7	0.4634	0.3778	-0.2043	0.0648	-0.6232	0.2145
	T27	12	0.4490	0.4615	0.0276	0.0475	-0.3308	0.3860
4	T12	2	0.5313	0.7162	0.2987	0.0191	0.0712	0.5263
	T13	3	0.4583	0.6914	0.4111	0.0219	0.1675	0.6546
	T21	6	0.5538	0.5680	0.0252	0.0123	-0.1570	0.2075
	T22	7	0.3279	0.5789	0.5686	0.0464	0.2144	0.9228
	T27	12	0.4286	0.5690	0.2834	0.0369	-0.0325	0.5992
5	T12	2	0.1429	0.0833	-0.5390	1.7738	-2.7297	1.6517
	T21	6	0.1538	0.0833	-0.6131	1.3397	-2.5170	1.2908
	T22	7	0.2000	0.2727	0.3102	0.6424	-1.0082	1.6285
6	T12	2	0.5417	0.5294	-0.0229	0.0875	-0.5096	0.4638
	T13	3	0.4118	0.7143	0.5508	0.1126	-0.0011	1.1028
	T21	6	0.4118	0.6071	0.3883	0.1071	-0.1501	0.9267
	T22	7	0.3889	0.5000	0.2513	0.1373	-0.3582	0.8608
	T23	8	0.0952	0.3077	1.1727	0.5389	-0.0348	2.3802
	T27	12	0.4000	0.6471	0.4810	0.0821	0.0097	0.9522

Table 6.16.: Cluster-based meta-analysis: Risk Ratio, antidepressant A5mg

and the response rates in the control group are

$$\tilde{p}^3_{C_2} = 0.4857,\ \tilde{p}^4_{C_2} = 0.5313 \text{ and } \tilde{p}^6_{C_2} = 0.5417.$$

For the transformed Risk Ratios we got

$$\tilde{\theta}^3_2 = 0.1542,\ \tilde{\theta}^4_2 = 0.2987 \text{ and } \tilde{\theta}^6_2 = -0.0229.$$

The correspondent $(1-\alpha)$ confidence intervals were then given by

$$\begin{aligned}\mathcal{I}^3_2 &= [-0.2342, 0.5425],\\ \mathcal{I}^4_2 &= [0.0712, 0.5263] \text{ and}\\ \mathcal{I}^6_2 &= [-0.5096, 0.4638]\end{aligned}$$

for $\alpha = 0.1$. In study T12 we see, that there is heterogeneity in the treatment effects across the patient collectives Cl_3, Cl_4 and Cl_6 for medication A5mg.

As we have seen in the cluster-based analysis, in the EU cluster Cl_1, the mean PBO response is not so high in comparison to the non-EU cluster Cl_3 and Cl_4. This results in higher Risk Ratios for the single trials in cluster Cl_1, especially for study T23. There we have a Risk Ratio of

$$\tilde{\theta}^1_8 = 0.5650.$$

When analyzing the non-EU cluster Cl_3, the transformed Risk Ratios in the single studies vary from

$$\tilde{\theta}^3_7 = -0.2043 \text{ to } \tilde{\theta}^3_3 = 0.3808.$$

The response rates of the control and the treatment groups don't vary as much as in the meta-analysis approach like it is assumed due to the cluster-based approach. For the control group the mean response rate is higher then in the EU

cluster Cl_1 and the values vary from

$$\tilde{p}^3_{C_3} = 0.3333 \text{ to } \tilde{p}^3_{C_6} = 0.4944.$$

For the treatment group we have a lower response than the EU cluster and the values vary from

$$\tilde{p}^3_{T_7} = 0.3778 \text{ to } \tilde{p}^3_{T_2} = 0.5667.$$

The same effects can be observed in the non-EU cluster Cl_4. Although the response rates in the treatment groups are high, the Risk Ratio vary across the included trials from

$$\tilde{\theta}^4_6 = 0.0252 \text{ to } \tilde{\theta}^4_7 = 0.5686.$$

This can be explained by to the response rates in the control groups. In the mean the rate is very high which could also be observed in the cluster-based analysis, except for study T22. There, we have

$$\tilde{p}^4_{C_7} = 0.3279 \text{ and } \tilde{p}^4_{T_7} = 0.5789,$$

which results in a high transformed Risk Ratio of

$$\tilde{\theta}^4_7 = 0.5686$$

.

In the mixed cluster Cl_6, the Risk Ratios are higher. They vary from

$$\tilde{\theta}^6_2 = -0.0229 \text{ to } \tilde{\theta}^6_8 = 0.1.1727.$$

This is a result of higher response rates in the treatment and response rates in the control groups which are similar to those in cluster Cl_3.

The results for the further dosages of medication A, dosage A10mg, A15mg and A20mg can be found in Table 6.17, Table 6.18 and Table 6.19. The results of the already established depression therapy C1 and C2 are listed in Table 6.20 and Table 6.21.

We can conclude that in all clusters we have overlapping confidence intervals for all medications, except in cluster Cl_4, which leads to the assumption, that the existing heterogeneity across trials has been considered in the cluster-based approach. Furthermore, for all medications we have a high variation in the single trials across the patient collectives and within a cluster the variation across trials seems to be less pronounced than in the meta-analysis approach.

Treatment group: A10mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.7750	0.3232	0.0132	0.1344	0.5120
	T23	8	0.2925	0.5048	0.5458	0.0322	0.2508	0.8408
2	T11	1	0.1000	0.2000	0.6931	1.3000	-1.1823	2.5686
	T12	2	0.1333	0.1154	-0.1446	0.7282	-1.5482	1.2591
	T15	5	0.2105	0.2727	0.2589	0.4398	-0.8320	1.3497
	T25	10	0.4000	0.1111	-1.2809	1.1889	-3.0744	0.5126
3	T12	2	0.4857	0.4783	-0.0155	0.0777	-0.4739	0.4430
	T15	5	0.5455	0.4688	-0.1515	0.0607	-0.5567	0.2536
	T25	10	0.2593	0.3654	0.3431	0.0863	-0.1401	0.8264
	T26	11	0.2787	0.4038	0.3709	0.0708	-0.0668	0.8087
	T27	12	0.4490	0.5600	0.2210	0.0408	-0.1111	0.5530
4	T12	2	0.5313	0.7532	0.3492	0.0180	0.1282	0.5701
	T15	5	0.2844	0.5263	0.6155	0.0310	0.3260	0.9050
	T25	10	0.4074	0.3393	-0.1830	0.0617	-0.5916	0.2256
	T26	11	0.4400	0.5000	0.1278	0.0518	-0.2464	0.5021
	T27	12	0.4286	0.5965	0.3306	0.0357	0.0199	0.6413
5	T25	10	0.3750	0.3125	-0.1823	0.3458	-1.1496	0.7850
	T26	11	0.3636	0.2727	-0.2877	0.4015	-1.3299	0.7546
6	T12	2	0.5417	0.7059	0.2648	0.0598	-0.1373	0.6669
	T15	5	0.2381	0.4815	0.7042	0.1923	-0.0170	1.4254
	T23	8	0.0952	0.7647	2.0831	0.4705	0.9549	3.2113
	T25	10	0.1290	0.3750	1.0669	0.3219	0.1336	2.0001
	T26	11	0.3125	0.3438	0.0953	0.1972	-0.6350	0.8257
	T27	12	0.4000	0.6800	0.5306	0.0688	0.0991	0.9621

Table 6.17.: Cluster-based meta-analysis: Risk Ratio, antidepressant A10mg

6.3.4. Cluster-based fixed-effects model

For each cluster we have applied the cluster-based fixed effect model for the weighted aggregation of the treatment effects in the single studies which are included in the corresponding cluster. It is assumed that there is one true treatment effect θ^i for each administered medical intervention in each cluster. We assume that the variation of the treatment effects across clusters is due to socio-demographic, biographical and/or clinical parameters. Whereas the variation of the treatment effects across trials within a cluster is due to estimation errors. Like it is defined in Section 5.7.3, for cluster Cl_i the fixed-effects model is given

Treatment group: A15mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.6667	0.5754	0.0199	0.3431	0.8076
2	T14	4	0.0833	0.0714	-0.1542	1.8452	-2.3885	2.0802
	T26	11	0.1667	0.2000	0.1823	1.2333	-1.6444	2.0090
3	T24	9	0.4848	0.4107	-0.1659	0.0417	-0.5019	0.1700
	T26	11	0.2787	0.3443	0.2113	0.0737	-0.2351	0.6577
4	T24	9	0.4565	0.6500	0.3533	0.0393	0.0271	0.6796
	T26	11	0.4400	0.5000	0.1278	0.0549	-0.2574	0.5131
5	T24	9	0.1000	0.2353	0.8557	1.0912	-0.8625	2.5739
	T26	11	0.3636	0.2308	-0.4547	0.4155	-1.5150	0.6055
6	T14	4	0.2857	0.8333	1.0704	0.1302	0.4770	1.6639
	T24	9	0.2941	0.5294	0.5878	0.1935	-0.1357	1.3113
	T26	11	0.3125	0.4211	0.2982	0.2099	-0.4554	1.0517

Table 6.18.: Cluster-based meta-analysis: Risk Ratio, antidepressant A15mg

by

$$\tilde{\theta}^i_j := \theta^i + \epsilon^i_j,$$

where $\tilde{\theta}^i_j$ is the observed treatment effect and ϵ^i_j the estimation error of study j in cluster Cl_i. With the inverse variance method, the estimated summary treatment effect in cluster Cl_i was calculated by

$$\tilde{\theta}^i = \frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\theta}^i_j w^i_j}{\sum_{j \in \mathcal{I}_{Cl_i}} w^i_j},$$

with weights

$$w^i_j = \frac{1}{\text{var}(\tilde{\Theta}^i_j)},$$

for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. The variance was calculated by

$$\text{var}(\tilde{\Theta}^i) = \frac{1}{\sum_{j \in \mathcal{I}_{Cl_i}} w^i_j}.$$

Treatment group: A20mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.6991	0.6229	0.0187	0.3980	0.8478
2	T14	4	0.0833	0.0909	0.0870	1.8258	-2.1355	2.3095
	T15	5	0.2105	0.1429	-0.3878	0.6259	-1.6891	0.9136
	T25	10	0.4000	0.2500	-0.4700	0.6750	-1.8214	0.8814
3	T15	5	0.5455	0.5897	0.0781	0.0431	-0.2634	0.4195
	T24	9	0.4848	0.5600	0.1441	0.0318	-0.1493	0.4375
	T25	10	0.2593	0.4211	0.4849	0.0770	0.0284	0.9415
	T27	12	0.4490	0.6364	0.3488	0.0380	0.0280	0.6696
4	T15	5	0.2844	0.7217	0.9313	0.0264	0.6638	1.1987
	T24	9	0.4565	0.4390	-0.0391	0.0570	-0.4319	0.3538
	T25	10	0.4074	0.5319	0.2667	0.0457	-0.0848	0.6181
	T27	12	0.4286	0.5263	0.2054	0.0396	-0.1219	0.5328
5	T24	9	0.1000	0.1875	0.6286	1.1708	-1.1512	2.4084
	T25	10	0.3750	0.2727	-0.3185	0.4508	-1.4228	0.7859
6	T14	4	0.2857	0.9167	1.1658	0.1266	0.5804	1.7511
	T15	5	0.2381	0.4583	0.6549	0.2016	-0.0837	1.3935
	T24	9	0.2941	0.4545	0.4353	0.1957	-0.2924	1.1630
	T25	10	0.1290	0.2222	0.5436	0.4122	-0.5124	1.5996
	T27	12	0.4000	0.5161	0.2549	0.0802	-0.2110	0.7208

Table 6.19.: Cluster-based meta-analysis: Risk Ratio, antidepressant A20mg

Treatment group: C1

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.8087	0.7685	0.0169	0.5544	0.9826
2	T12	2	0.1333	0.1905	0.3567	0.6357	-0.9548	1.6681
	T14	4	0.0833	0.2857	1.2321	1.2738	-0.6243	3.0886
	T22	7	0.1111	0.2778	0.9163	1.0333	-0.7558	2.5883
3	T12	2	0.4857	0.6364	0.2701	0.0562	-0.1199	0.6602
	T13	3	0.3333	0.7692	0.8362	0.0590	0.4368	1.2357
	T22	7	0.4634	0.7027	0.4163	0.0397	0.0887	0.7439
	T24	9	0.4848	0.6667	0.3185	0.0244	0.0614	0.5756
4	T12	2	0.5313	0.6986	0.2739	0.0197	0.0430	0.5047
	T13	3	0.4583	0.8358	0.6008	0.0193	0.3720	0.8296
	T22	7	0.3279	0.5714	0.5555	0.0470	0.1989	0.9121
	T24	9	0.4565	0.5897	0.2561	0.0437	-0.0879	0.6000
5	T12	2	0.1429	0.1111	-0.2513	1.7460	-2.4248	1.9222
	T14	4	0.2500	0.5000	0.6931	0.6250	-0.6072	1.9935
	T22	7	0.2000	0.1538	-0.2624	0.8231	-1.7546	1.2299
	T24	9	0.1000	0.2857	1.0498	1.0786	-0.6584	2.7581
6	T12	2	0.5417	0.8333	0.4308	0.0464	0.0766	0.7850
	T13	3	0.4118	0.4091	-0.0065	0.1497	-0.6429	0.6299
	T14	4	0.2857	0.6250	0.7828	0.1565	0.1320	1.4336
	T22	7	0.3889	0.5882	0.4138	0.1285	-0.1757	1.0034
	T24	9	0.2941	0.4375	0.3971	0.2215	-0.3771	1.1713

Table 6.20.: Cluster-based meta-analysis: Risk Ratio, antidepressant C1

Treatment group: C2

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.8523	0.4182	0.0115	0.2417	0.5947
2	T11	1	0.1000	0.1667	0.5108	1.3167	-1.3766	2.3982

Table 6.21.: Cluster-based meta-analysis: Risk Ratio, antidepressant C2

Then, for the $(1-\alpha)$ confidence interval $\mathcal{I}^i$ of the true treatment effect θ^i in cluster Cl_i we got

$$\mathcal{I}^i = [\tilde{\theta}^i - z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}, \tilde{\theta}^i + z[1-\frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}],$$

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

The estimated summary treatment effect, the variance of this estimated treatment effect and the $(1-\alpha)$ confidence interval for each medication is shown in Table 6.22. E.g. the summary treatment effect for medication A5mg in the single clusters Cl_1, Cl_3, Cl_5 and Cl_6 are given by

$$\hat{\theta}^1 = 0.3447,\ \hat{\theta}^3 = 0.0631,\ \hat{\theta}^4 = 0.2472 \text{ and } \hat{\theta}^6 = 0.3556,$$

with correspondent variances

$$(\tilde{\sigma}^1)^2 = 0.0096,\ (\tilde{\sigma}^3)^2 = 0.0088,\ (\tilde{\sigma}^4)^2 = 0.0044 \text{ and } (\tilde{\sigma}^6)^2 = 0.0196.$$

The $(1-\alpha)$ confidence intervals are therefore given by

$$\begin{aligned}
\mathcal{I}^1 &= [0.1835, 0.5058],\\
\mathcal{I}^3 &= [-0.0914, 0.2176],\\
\mathcal{I}^4 &= [0.1383, 0.3562] \text{ and}\\
\mathcal{I}^6 &= [0.1252, 1.5860].
\end{aligned}$$

When looking at Table 6.16 to Table 6.21 it is noticeable, that the treatment effects in the single trials of a cluster do not vary as much as in the meta-analysis for each medication. For the verification, we calculated the realization of the Q^i-statistic for cluster Cl_i,

$$q^i = \sum_{j \in \mathcal{I}_{Cl_i}} w_j^i (\tilde{\theta}_j^i - \hat{\theta}^i)^2,$$

for $i = 1, \ldots, k$. The results can be found in Table 6.22. For the evaluation of

Cluster-based fixed-effects model

Drug	i	$\tilde{\theta}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$		q^i	pv^i	$(l^i)^2$
A5mg	1	0.3447	0.0096	0.1835	0.5058	2.1691	0.1408	52.47%
	2	0.0758	0.2271	-0.7081	0.8596	3.9563	0.1383	49.45%
	3	0.0631	0.0088	-0.0914	0.2176	2.6078	0.6254	0.00%
	4	0.2472	0.0044	0.1383	0.3562	7.6385	0.1058	46.30%
	5	-0.0972	0.3488	-1.0687	0.8743	0.5670	0.7531	0.00%
	6	0.3556	0.0196	0.1252	0.5860	3.4944	0.6242	0.00%
A10mg	1	0.3879	0.0093	0.2288	0.5469	1.0928	0.2958	8.48%
	2	-0.0293	0.1902	-0.7467	0.6880	1.9262	0.5879	0.00%
	3	0.1500	0.0126	-0.0344	0.3343	3.0962	0.5419	0.00%
	4	0.3173	0.0066	0.1835	0.4510	7.6808	0.1040	46.70%
	5	-0.2311	0.1858	-0.9401	0.4779	0.0149	0.9030	0.00%
	6	0.5138	0.0214	0.2733	0.7543	8.3031	0.1403	38.63%
A15mg	1	0.5754	0.0199	0.3431	0.8076	0.0000		
	2	0.0475	0.7392	-1.3667	1.4618	0.0368	0.8479	0.00%
	3	-0.0295	0.0266	-0.2980	0.2389	1.2335	0.2667	18.68%
	4	0.2592	0.0229	0.0102	0.5081	0.5398	0.4625	0.00%
	5	-0.0934	0.3009	-0.9957	0.8089	1.1397	0.2857	12.26%
	6	0.7199	0.0568	0.3281	1.1118	1.8818	0.3903	0.00%
A20mg	1	0.6229	0.0187	0.3980	0.8478	0.0000		
	2	-0.3497	0.2757	-1.2134	0.5140	0.1282	0.9379	0.00%
	3	0.2322	0.0106	0.0625	0.4019	1.9819	0.5762	0.00%
	4	0.4445	0.0098	0.2821	0.6070	15.1980	0.0017	64.92%
	5	-0.0552	0.3255	-0.9936	0.8832	0.5531	0.4570	0.00%
	6	0.5836	0.0304	0.2966	0.8706	4.1644	0.3842	3.91%
C1	1	0.7685	0.0169	0.5544	0.9826	0.0000		
	2	0.7262	0.3007	-0.1758	1.6281	0.4507	0.7982	0.00%
	3	0.4214	0.0099	0.2577	0.5852	3.7595	0.2886	20.03%
	4	0.4272	0.0068	0.2914	0.5631	3.7719	0.2872	20.27%
	5	0.3754	0.2318	-0.4165	1.1672	1.3024	0.7286	0.00%
	6	0.4104	0.0213	0.1703	0.6505	2.0567	0.7253	0.00%
C2	1	0.4182	0.0115	0.2417	0.5947	0.0000		
	2	0.5108	1.3167	-1.3766	2.3982	0.0000		

Table 6.22.: Cluster-based meta-analysis: Risk Ratio estimated by the cluster-based fixed-effects model

the results, we used the p-value

$$pv^i = 1 - F_{\chi^2_{(n^i-1)}}(q^i),$$

where $F_{\chi^2_{(n^i-1)}}$ is the cumulative function of the χ^2 distribution with $(n^i - 1)$ degrees of freedom and the number of studies n^i in cluster Cl_i. Another way to assess heterogeneity is the use of e.g. the Q^i related $(l^i)^2$ index

$$(l^i)^2 = \max\left\{0, \frac{q^i - (n^i - 1)}{q^i}\right\} \cdot 100\%.$$

When considering medication A5mg, for the realization of the Q^3-statistic of cluster Cl_3 followed

$$q^3 = 2.6078.$$

For the evaluation of this result we calculated the p-value

$$pv^3 = 1 - F_{\chi^2_4}(2.6078) = 0.6254.$$

This implies that the probability to get this result or a higher value for q^3 is 62.54%, under the assumption that the treatment effects are equal in the five studies in cluster Cl_3. This means that we can not safely reject the hypothesis that the treatment effects in the single trials are equal. The result for the $(l^3)^2$ index for medication A5mg is

$$(l^3)^2 = 0.0000\%.$$

According to Table 4.2, this result can be interpreted that there is no considerable heterogeneity across the single trials in cluster Cl_3. As can be seen, there are also results which can be interpreted that there is still heterogeneity within clusters, e.g for medication A5mg in cluster Cl_1 or Cl_4. This heterogeneity might be due to further random-effects across trials within a cluster. To address this specific heterogeneity, we use the cluster-based random-effects approach. The results are discussed in the next section.

Cluster-based fixed-effects model		
Drug	q_{Cl}	pv
A5mg	8.1732	0.1469
A10mg	13.4022	0.0199
A15mg	12.6795	0.0266
A20mg	16.9318	0.0046
C1	6.7260	0.2418
C2	0.1878	0.9992

Table 6.23.: Cluster-based meta-analysis: Risk Ratio estimated by the cluster-based fixed-effects model, Q_{Cl}-statistic

But first, for the further justification of the assumption that there is one true treatment effect for each patient collective, we used the realization of the Q_{Cl}-statistic

$$q_{Cl} = \sum_{i=1}^{k} w^i(\tilde{\theta}^i - \bar{\theta})^2,$$

where $\tilde{\theta}^i$ is the estimated summary treatment effect in cluster Cl_i, $(w^i)^{-1} = \text{var}(\tilde{\Theta}^i)$ is the corresponding variance and $\bar{\theta}$ the mean treatment effect of all clusters. The results are shown in Table 6.23. For the evaluation of the results of q_{Cl}, we used the p-value

$$pv = 1 - F_{\chi^2_{(k-1)}}(q_{Cl}).$$

E.g. for medication A10mg, the realization of the Q_{Cl}-statistic is given by

$$q_{Cl} = 13.4022.$$

The corresponding p-value is

$$pv = 0.0199,$$

This implies that the probability to get this result for q_{Cl} or a higher value is only 1.99%, under the assumption that the true treatment effects are equal in each cluster. Therefore, the null hypotheses, that there is only one true treatment effect for all patients, can be rejected. This implies that there are significant differences between the treatment effects across clusters and that it is reliable to assume that there is one true treatment effect for each patient collective. Similar

results can be found for medication A5mg, A15mg and A20mg. The p-value for the realization of the Q_{Cl}-statsitic for medication C1 is $pv = 0.2418$. With a look at Table 6.22, this result seems to be plausible, since the treatment effects are more equal in the patient collectives than for medication A. And since medication C2 was only administered in two clusters, where the treatment effects seem to be equal, the high p-value of $pv = 0.9992$ is also plausible.

6.3.5. Cluster-based random-effects model

For the remaining heterogeneity that cannot readily be explained, we used the random-effects model. As for the meta-analysis approach, this model allows the variation of the true treatment effect across studies within a cluster. The variation in the effects within a cluster can be explained with the random-effect. The estimated treatment effect of study j in cluster Cl_i is given by

$$\tilde{\theta}_j^i := \theta_j^i + \epsilon_j^i = \theta^i + \nu_j^i + \epsilon_j^i,$$

for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$, where ϵ_j^i is the estimation error and ν_j^i is the random-effect of study j in cluster Cl_i. With the inverse variance method, the summary treatment effect of cluster Cl_i is given by

$$\tilde{\theta}^i = \frac{\sum_{j \in \mathcal{I}_{Cl_i}} \tilde{\theta}_j^i (w_j^i)^*}{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^*},$$

with weights

$$(w_j^i)^* = \frac{1}{(w_j^i)^{-1} + (\tilde{\tau}^i)^2},$$

for $i = 1, \ldots, k$ and $j \in \mathcal{I}_{Cl_i}$. The variance was calculated by

$$\text{var}(\tilde{\Theta}^i) = \frac{1}{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^*}$$

and the inter-variance was estimated by

$$(\tilde{\tau}^i)^2 = \max \left\{ 0; \frac{q^i - (n^i - 1)}{\sum_{j \in \mathcal{I}_{Cl_i}} w_j^i - \frac{\sum_{j \in \mathcal{I}_{Cl_i}} (w_j^i)^2}{\sum_{j \in \mathcal{I}_{Cl_i}} w_j^i}} \right\}.$$

Then, for the $(1-\alpha)$ confidence interval of the true treatment effect θ^i of cluster Cl_i follows

$$\mathcal{I}^i = [\tilde{\theta}^i - z[1 - \frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}, \tilde{\theta}^i + z[1 - \frac{\alpha}{2}]\sqrt{\text{var}(\tilde{\Theta}^i)}],$$

where $z[1 - \frac{\alpha}{2}]$ is the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution.

The estimated summary treatment effect, the variance and the $(1-\alpha)$ confidence interval for each medication and each patient collective is shown in Table 6.24. E.g. the summary treatment effects for medication A5mg in cluster Cl_1, Cl_3, Cl_4 and Cl_6 are given by

$$\hat{\theta}^1 = 0.3466,\ \hat{\theta}^3 = 0.0631,\ \hat{\theta}^4 = 0.2488 \text{ and } \hat{\theta}^6 = 0.3556,$$

with correspondent variances

$$(\tilde{\sigma}^1)^2 = 0.0100,\ (\tilde{\sigma}^3)^2 = 0.0088,\ (\tilde{\sigma}^4)^2 = 0.0045 \text{ and } (\tilde{\sigma}^6)^2 = 0.0196.$$

The $(1-\alpha)$ confidence intervals are therefore given by

$$\begin{aligned}
\mathcal{I}^1 &= [0.1820, 0.5112], \\
\mathcal{I}^3 &= [-0.0914, 0.2176], \\
\mathcal{I}^4 &= [0.1385, 0.3590] \text{ and} \\
\mathcal{I}^6 &= [0.1252, 1.5860].
\end{aligned}$$

For the justification of the assumption that there is a treatment effect for each patient collective, we also used the realization of the Q_{Cl}-statistic for the cluster-based random-effects model,

$$q_{Cl} = \sum_{i=1}^{k} w^i(\tilde{\theta}^i - \bar{\theta})^2.$$

The results can be found in Table 6.25. For the evaluation of the results of q_{Cl}, we used the p-value

$$pv = 1 - F_{\chi^2_{(k-1)}}(q_{Cl}).$$

E.g. for medication A10mg, the realization of the Q_{Cl}-statistic is given by

$$q_{Cl} = 15.1587,$$

with corresponding p-value

$$pv = 0.0097.$$

The probability to get this result for q_{Cl} or a higher value is only 0.97%, under the assumption that the true treatment effects are equal in each cluster. Therefore the null hypothesis, that there is only one true treatment effect for all patients, can be rejected. This implies that there are significant differences between the treatment effects and that it is reliable to assume that there is one true treatment effect for each patient collective. Similar results can be found for the other dosages of mediaction A.

Cluster-based random-effects model

RR	i	$\tilde{\theta}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$	
A 5mg	1	0.3466	0.0100	0.1820	0.5112
	2	0.0758	0.2271	-0.7081	0.8596
	3	0.0631	0.0088	-0.0914	0.2176
	4	0.2488	0.0045	0.1385	0.3590
	5	-0.0972	0.3488	-1.0687	0.8743
	6	0.3556	0.0196	0.1252	0.5860
A 10mg	1	0.3879	0.0093	0.2288	0.5469
	2	-0.0293	0.1902	-0.7467	0.6880
	3	0.1500	0.0126	-0.0344	0.3343
	4	0.3158	0.0069	0.1796	0.4521
	5	-0.2311	0.1858	-0.9401	0.4779
	6	0.0845	0.0037	-0.0150	0.1840
A 15mg	1	0.5754	0.0199	0.3431	0.8076
	2	0.0475	0.7392	-1.3667	1.4618
	3	-0.0294	0.0267	-0.2983	0.2396
	4	0.2592	0.0229	0.0102	0.5081
	5	-0.0934	0.3009	-0.9957	0.8089
	6	0.7199	0.0568	0.3281	1.1118
A 20mg	1	0.6229	0.0187	0.3980	0.8478
	2	-0.3497	0.2757	-1.2134	0.5140
	3	0.2322	0.0106	0.0625	0.4019
	4	0.3996	0.0168	0.1867	0.6126
	5	-0.0552	0.3255	-0.9936	0.8832
	6	0.5836	0.0304	0.2966	0.8706
C1	1	0.7685	0.0169	0.5544	0.9826
	2	0.7262	0.3007	-0.1758	1.6281
	3	0.4215	0.0099	0.2575	0.5856
	4	0.4272	0.0068	0.2912	0.5632
	5	0.3754	0.2318	-0.4165	1.1672
	6	0.4104	0.0213	0.1703	0.6505
C2	1	0.4182	0.0115	0.2417	0.5947
	2	0.5108	1.3167	-1.3766	2.3982

Table 6.24.: Cluster-based meta-analysis: Risk Ratio estimated by the cluster-based random-effects model

Cluster-based random-effects model		
Drug	q_{Cl}	pv
A5mg	8.0847	0.1516
A10mg	15.1587	0.0097
A15mg	12.6649	0.0267
A20mg	14.8588	0.0110
C1	6.7183	0.2424
C2	0.1878	0.9992

Table 6.25.: Cluster-based meta-analysis: Risk Ratio estimated by the cluster-based random-effects model, Q_{Cl}-statistic

Drug	$\tilde{\theta}$(drug,C1)	$\tilde{\sigma}^2$(drug,C1)
A5mg	-0.2378	0.0056
A10mg	-0.1554	0.0061
A15mg	-0.1983	0.0109
A20mg	-0.0676	0.0074

Table 6.26.: Meta-analysis random-effects model: Indirect comparison of the Risk Ratios of medication A5mg, A10mg, A15mg and A20mg with C1

6.3.6. Indirect comparison

For the comparison of the results of the meta-analysis and cluster-based meta-analysis presented in Section 6.2 and Section 6.3, we used the indirect comparison method. Since medication C1 is one of the already established standard depression therapy, and is represented in each cluster, we calculated the difference between the treatment effects of medication A, with its different dosages, and C1 estimated by the random-effects approaches. The estimated summary treatment effects and its confidence intervals ($\alpha = 0.1$) for each medication derived by the random-effects model and the cluster-based random-effects model are illustrated in Figure 6.1. It is noticeable that for the drop-out cluster Cl_2 and Cl_5, the summary treatment effects are low with wide ranged confidence intervals. For all other clusters, the differences in the effects of the analyzed medication is clearly obvious.

For the indirect comparison of the results of the meta-analysis, we calculated the differences of the treatment effects of medication A and its different dosages

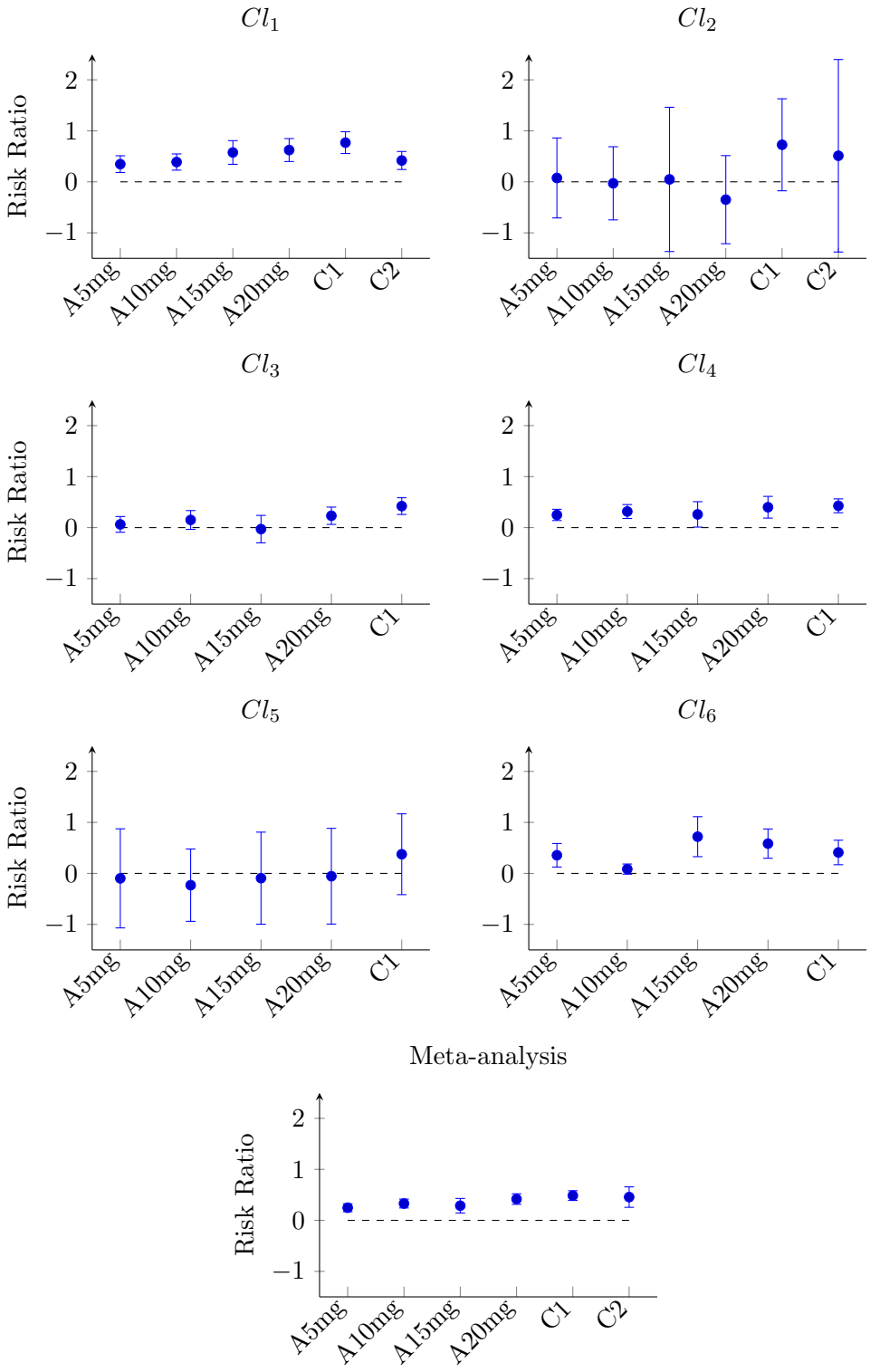

Figure 6.1.: Confidence intervals of Risk Ratio estimated by the random-effects model and the cluster-based random-effects model

Drug	i	$\tilde{\theta}^i$(drug,C1)	$(\tilde{\sigma}^i)^2$(drug,C1)
A5mg	1	-0.4238	0.0265
	2	-0.6504	0.5278
	3	-0.3583	0.0187
	4	-0.1800	0.0112
	5	-0.4726	0.5806
	6	-0.0548	0.0409
A10mg	1	-0.3806	0.0263
	2	-0.7555	0.4909
	3	-0.2715	0.0225
	4	-0.1099	0.0134
	5	-0.6065	0.4176
	6	**0.1034**	0.0427
A15mg	1	-0.1931	0.0369
	2	-0.6786	1.0399
	3	-0.4510	0.0365
	4	-0.1681	0.0297
	5	-0.4687	0.5327
	6	**0.3095**	0.0781
A20mg	1	-0.1456	0.0356
	2	-1.0758	0.5764
	3	-0.1892	0.0206
	4	**0.0173**	0.0166
	5	-0.4306	0.5572
	6	**0.1732**	0.0517

Table 6.27.: Cluster-based meta-analysis random-effects model: Indirect comparison of the Risk Ratios of medication A5mg, A10mg, A15mg and A20mg with C1

and the treatment effect of medication C1,

$$\tilde{\theta}(\text{drug,C1}) = \tilde{\theta}(\text{drug}) - \tilde{\theta}(\text{C1})$$

and the variances

$$\tilde{\sigma}^2(\text{drug,C1}) = \tilde{\sigma}^2(\text{drug}) + \tilde{\sigma}^2(\text{C1}).$$

The difference can also be treated as additional benefit for the patients treated with the corresponding intervention. The results are shown in Table 6.26. It can be seen that there is no additional benefit for patients who have been treated with medication A, when applying the meta-analysis approach. In terms of health economic evaluation, in the worst case, this would imply that medication A has to be taken off the market due to no considerable additional benefit for the patients in the depression therapy.

For the indirect comparison of the results of the cluster-based meta-analysis, we also calculated the cluster-based difference for each dosage of medication A and C1,

$$\tilde{\theta}^i(\text{drug,C1}) = \tilde{\theta}^i(\text{drug}) - \tilde{\theta}^i(\text{C1})$$

and the variance

$$(\tilde{\sigma}^i)^2(\text{drug,C1}) = (\tilde{\sigma}^i)^2(\text{drug}) + (\tilde{\sigma}^i)^2(\text{C1}),$$

for $i = 1, \ldots, k$. The results can be found in Table 6.27. Dosage A5mg has no additional benefit in all clusters, but the other dosages show positive results. Medication A10mg has an additional benefit in cluster Cl_6,

$$\tilde{\theta}^6(\text{A10mg,C1}) = 0.1034,$$

with variance

$$(\tilde{\sigma}^6)^2(\text{A10mg,C1}) = 0.0427,$$

and medication A15mg has a higher treatment effect than the standard therapy C1, also in cluster Cl_6,

$$\tilde{\theta}^6(\text{A15mg,C1}) = 0.3095,$$

with variance

$$(\tilde{\sigma}^6)^2(\text{A15mg,C1}) = 0.0781.$$

For medication A20mg the cluster-based random-effects model indicated an additional benefit for patient collective Cl_4 with treatment difference

$$\tilde{\theta}^4(\text{A20mg,C1}) = 0.0173$$

and for patient collective Cl_6 with

$$\tilde{\theta}^6(\text{A20mg,C1}) = 0.1732.$$

This implies that in the mixed cluster Cl_6, medication A has a higher effect than C1 as well as in the non-EU cluster Cl_4. Both collectives include patients with lower depression. With regard to health economic evaluation, this means that medication A is a considerable alternative in the treatment of depression for patients with characteristic values combinations similar to those of cluster Cl_4 and Cl_6.

6.4. Cluster-based identification of heterogeneity

In Chapter 4, we carved out that clinical and especially regional heterogeneity is one important factor which has to be explained if there are distinctive country effects in multinational studies or if there is evidence of increased heterogeneity in meta-analyses. Transferred to the data of these existing drug studies, this means that in particular the baseline data should be considered in terms of e.g. socio-demographic and biographical parameters. Due to the complexity of the existing data, the process of endpoint-oriented cluster optimization was used to explain regional heterogeneity. Within this framework, different clusters, which ware the outcome of a multivariate cluster optimization with respect to response data, have been analyzed in detail. The aim of this analysis was to identify factors explaining the heterogeneity in global conducted placebo-controlled clinical trials. For the identification, we used the χ^2 tests of independence introduced in Section 5.8.4.

A_i	Characteristic	Values
A_1	region	EU, non-EU
A_2	treatment	PBO (placebo), A (5mg, 10mg, 15mg, 20mg), C1, C2
A_3	BMI	13 to 68.41 kg/m^2
A_4	MADRS	13 to 52
A_5	CGI-S	3 to 7
A_6	duration	28 to 7976 days
A_7	sex	female, male
A_8	age	18 to 88 years
A_9	withdrawal	adverse events, lack of efficacy, lost to follow-up, withdrawal of consent
	Outcome	Values
	Response	yes, no
	Remission	yes, no

Table 6.28.: Cluster-based identification of heterogeneity: Independent and dependent variables used for the clustering approach

6.4.1. Clustering

For the application of the endpoint-oriented clustering approach explained in Chapter 5, to identify patient collectives as a partition of data set

$$S = S^{all} = \{x_j, y_j\}_{j=1}^{6010},$$

the characteristics in the available trials were picked, which value combinations might have an influence on the outcome of a patient. There are different multivariate quantitative methods for the selection of influencing factors, the interested reader is referred to [2] for more information. For the cluster-based explanation of clinical heterogeneity, we have chosen those random variables listed in Table 6.28. Since the random variable 'BMI', 'duration' and 'age' have plenty of values and therefore the number of value combinations would increase proportionally, we classified those characteristics, like it is described in Section 5.5. In Table 6.29, the characteristic values of those three variables are divided into 5 classes of equal class density. For the evaluation of the medical intervention, we consider the outcome variables 'response' and 'remission' which are binary with the two oc-

A_i	Characteristic	Class number	Class
A_3	BMI	1	$]0, 22.53]$
		2	$]22.53, 25.28]$
		3	$]25.53, 28.39]$
		4	$]28.39, 32.91]$
		5	$]32.91, \infty]$
A_6	duration (in days)	1	$]0, 113]$
		2	$]113, 142]$
		3	$]142, 197]$
		4	$]197, 323]$
		5	$]323, \infty]$
A_8	age (in years)	1	$]0, 34]$
		2	$]34, 44]$
		3	$]44, 51]$
		4	$]51, 60]$
		5	$]60, \infty]$

Table 6.29.: Cluster-based identification of heterogeneity: Classification of random variables

Cluster Cl_i	1	2	3	4	5	6	all
Patients κ_i	496	584	2388	1641	674	227	6010
Response $\tilde{p}_1^i$	0.5262	0.1798	0.5674	0.4052	0.7240	0.1454	0.4837
Remission $\tilde{p}_2^i$	0.3327	0.0873	0.3585	0.2681	0.4941	0.0617	0.3093

Table 6.30.: Cluster-based identification of heterogeneity: Clustering results with respect to 'response' for defining 6 clusters

currences 'response/remission' (= 1) and 'no response/no remission' (= 0). Then the transformation method explained in Section 5.5 was applied with respect to the target variable 'response'. With this transformed data, we had been able to apply the clustering algorithm to get a predefined number of patient collectives. The parameter setup for the clustering approach consisted of $k = 6$ clusters, lower bounds $l_i = 100$, for $i = 1, \ldots, 6$, and no upper bounds. In this thesis, the setting $k = 6$ has been arisen due to corresponding preliminary analysis. Thereby, it is selectable freely in principle but should be set goal-oriented. Each cluster is uniquely defined by combinations of the characteristic values of the various attributes that have been identified as similar by the algorithm. The distribution of the 6010 participating patients into the six patient collectives is shown in Table

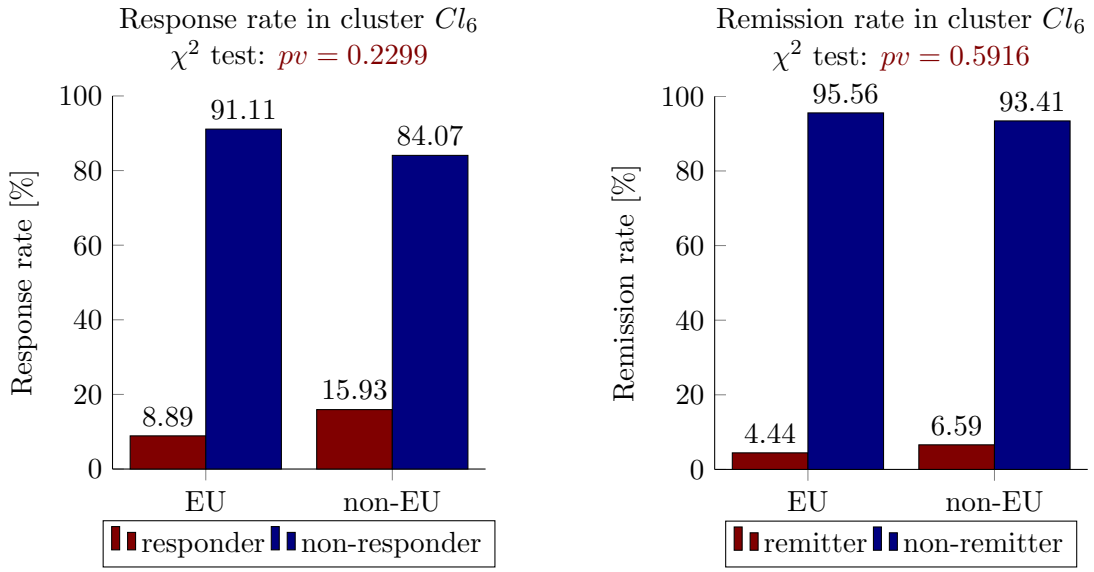

Figure 6.2.: Cluster-based identification of heterogeneity: Cluster Cl_6, response rate and remission rate stratified by region

6.30. 'Response' is referred to the response rate and means e.g. that 40.52% of those patients who are assigned to cluster Cl_4, due to their individual combination of characteristic values, have actually responded. In addition, for each cluster, which has been identified by the target response, the remission rates have been determined. The group of remitters per cluster has been analyzed analogously to the group of responders and has been evaluated in the relevant sections under specific consideration. The following results are based on the analysis of [21].

6.4.2. Cluster-based analysis

Cluster Cl_6: Placebo dropouts

This cluster includes $\kappa_6 = 227$ placebo treated patients with a response rate of

$$\tilde{p}_1^6 = 0.1454$$

and a remission rate of $\tilde{p}_2^6 = 0.0617$. Despite these rates, all individuals of this cluster have dropped out of the study. As shown in Figure 6.2, there are different

response rates between EU and non-EU patients. The response rates are

$$\tilde{p}_1^6(\text{EU}) = 0.0889 \text{ and } \tilde{p}_1^6(\text{non-EU}) = 0.1593,$$

with $pv = 0.2299$. The remission rate is $\tilde{p}_2^6(\text{EU}) = 0.0444$ in the EU, compared to $\tilde{p}_2^6(\text{non-EU}) = 0.0659$ in the non-EU regions, with $pv = 0.5916$.

The main reasons for withdrawal of the responding and remitting patients are of further interest. It is noticeable that in the EU regions nobody drops out because of 'adverse events' and all drop-outs are due to 'lost to follow-up',

$$\tilde{\rho}^6(\text{adverse events}|\text{EU}) = 0 \text{ and}$$
$$\tilde{\rho}^6(\text{lost to follow-up}|\text{EU}) = 1.$$

From a clinical point of view, this constellation appears to be understandable and obvious for a placebo group with response and remission. The non-EU group shows significant differences in the reasons for withdrawal. The share of patients with reasons 'adverse events' and 'lost to follow-up' are

$$\tilde{\rho}^6(\text{adverse events}|\text{non-EU}) = 0.4138 \text{ and}$$
$$\tilde{\rho}^6(\text{lost to follow-up}|\text{non-EU}) = 0.5862.$$

Thus, for the available sample can be concluded that the reasons for withdrawal are regionally different.

Also in terms of gender composition, regional descriptive differences can be found in cluster Cl_6. In the EU regions, male patients show no response in comparison to non-EU men,

$$\tilde{p}_1^6(\text{EU, male}) = 0 \text{ and } \tilde{p}_1^6(\text{non-EU, male}) = 0.1333.$$

For women, however, the response rates are balanced with

$$\tilde{p}_1^6(\text{EU, female}) = 0.1176 \text{ and } \tilde{p}_1^6(\text{non-EU, female}) = 0.1776.$$

The male patients show consequently no remission, $\tilde{p}_2^6(\text{EU, male}) = 0$, in the EU

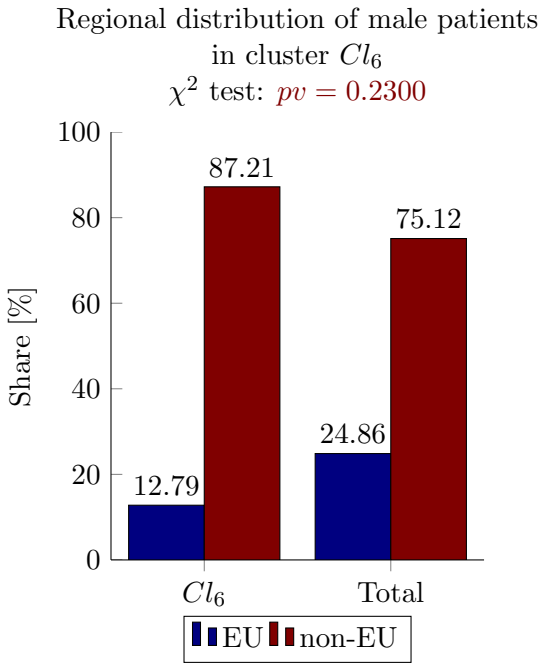

Figure 6.3.: Cluster-based identification of heterogeneity: Cluster Cl_6 vs. total population, share of male patients

regions compared to the non-EU men with remission rate $\tilde{p}_2^6$(non-EU, male) = 0.0667. For women the remission rates are also balanced with $\tilde{p}_2^6$(EU, female) = 0.0559 in comparison to $\tilde{p}_2^6$(non-EU, female) = 0.0654. It is interesting that in cluster Cl_6 the share of non-EU patients under the male participants is descriptively increased versus the corresponding marginal distribution (of the total sample), as can be seen in Figure 6.3. The shares are

$$\tilde{\rho}^6(\text{non-EU}|\text{male}) = 0.8721 \text{ and } \tilde{\rho}(\text{non-EU}|\text{male}) = 0.7512,$$

with $pv = 0.2300$.

In summary, in cluster Cl_6 a higher drop-out rate due to 'adverse events' in the non-EU countries and a descriptive higher placebo response and placebo remission can be found. These effects occur primarily due to the male patients, whereupon the percentage of male patients was significantly increased in the non-EU regions.

Non-responder	Responder
EU:	EU:
0.5094 adverse events	0.6667 adverse events
0.2547 lack of efficacy	[0 lack of efficacy (or N/A)]
0.0472 lost to follow-up	0.0667 lost to follow-up
0.1887 withdrawal of consent	0.2667 withdrawal of consent
non-EU:	non-EU:
0.4799 adverse events	0.3778 adverse events
0.1260 lack of efficacy	[0 lack of efficacy (or N/A)]
0.1930 lost to follow-up	0.3556 lost to follow-up
0.2011 withdrawal of consent	0.2667 withdrawal of consent

Table 6.31.: Cluster-based identification of heterogeneity: Cluster Cl_2, reasons for withdrawal of non-responder stratified by region

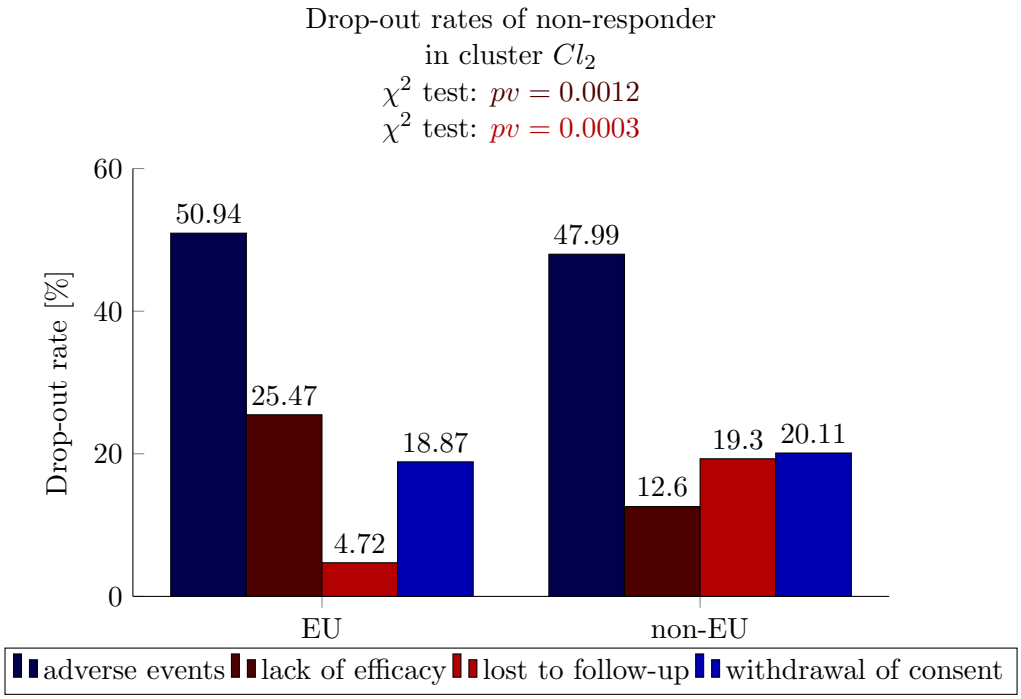

Figure 6.4.: Cluster-based identification of heterogeneity: Cluster Cl_2, reasons for withdrawal of non-responder stratified by region

Non-remitter	Remitter
EU:	EU:
0.5175 adverse events	0.7143 adverse events
0.2368 lack of efficacy	[0 lack of efficacy (or N/A)]
0.0439 lost to follow-up	0.1429 lost to follow-up
0.2018 withdrawal of consent	0.1429 withdrawal of consent
non-EU:	non-EU:
0.4749 adverse events	0.3182 adverse events
0.1122 lack of efficacy	[0 lack of efficacy (or N/A)]
0.2005 lost to follow-up	0.4545 lost to follow-up
0.2124 withdrawal of consent	0.2273 withdrawal of consent

Table 6.32.: Cluster-based identification of heterogeneity: Cluster Cl_2, reasons for withdrawal of non-remitter stratified by region

Cluster Cl_2: Verum drop-outs

The cluster of the verum (medication A, C1 and C2) drop-outs includes $\kappa_2 = 584$ patients with a response rate of

$$\tilde{p}_1^2 = 0.1798$$

and a remission rate of $\tilde{p}_2^2 = 0.0873$. This cluster is important because here regional differences can contribute particularly to influence study results as a whole. A high drop-out rate of verum treated patients due to causes 'adverse events', 'lost to follow-up' and 'lack of efficacy' reduces the verum-placebo difference. This is of importance because in cluster Cl_6 it was found that the placebo response and placebo remission in non-EU regions appears to be higher compared to the EU study centers. Also in cluster Cl_2, the response rate shows differences between EU and non-EU patients. The response rates are

$$\tilde{p}_1^2(\text{EU}) = 0.1944 \text{ and } \tilde{p}_1^2(\text{non-EU}) = 0.1240,$$

with $pv = 0.0725$. The remission rate is $\tilde{p}_2^2(\text{EU}) = 0.0579$ in the EU regions compared to $\tilde{p}_2^2(\text{non-EU}) = 0.0950$ in the non-EU regions, with $pv = 0.0598$.

The primary reasons for withdrawal of non-responders and non-remitters are of particular interest. One would expect that the majority drops out due to 'lack

of efficacy' and 'adverse events'. Here, however, the differences between EU and non-EU patients are shown in Table 6.31. With regard to non-responders, the data of the EU-group seams plausible by having especially 'adverse events' or 'lack of efficacy' as reasons for withdrawal. On the other hand the non-EU group has a significant lower rate of 'lack of efficacy',

$$\tilde{\rho}^2(\text{lack of efficacy}|\text{no response, EU}) = 0.2547$$
$$\tilde{\rho}^2(\text{lack of efficacy}|\text{no response, non-EU}) = 0.1260,$$

with $pv = 0.0012$, and a higher rate of persons not remaining in the study due to 'lost to follow-up',

$$\tilde{\rho}^2(\text{lost to follow-up}|\text{no response, EU}) = 0.0470$$
$$\tilde{\rho}^2(\text{lost to follow-up}|\text{no response, non-EU}) = 0.1930,$$

with $pv = 0.0003$, as can be seen in Figure 6.4. This unequal distributed group of 'lost to follow-up' tend to affect the study results in an adverse way. At a higher retention rate the patient would possibly still have responded. Upon further review of the cluster data, there is generally a higher rate of 'lost to follow-up' drop-outs in the group of non-EU patients, even in the group of responders. Here, in this cluster, the response rates are

$$\tilde{\rho}^2(\text{lost to follow-up}|\text{response, EU}) = 0.0667$$
$$\tilde{\rho}^2(\text{lost to follow-up}|\text{response, non-EU}) = 0.3567,$$

with $pv = 0.0257$. As can be seen from Table 6.32, correspondent distributions of the reasons for withdrawal can be found among the non-remitters. The non-EU group shows a significant lower rate of 'lack of efficacy',

$$\tilde{\rho}^2(\text{lack of efficacy}|\text{no remission, EU}) = 0.2368$$
$$\tilde{\rho}^2(\text{lack of efficacy}|\text{no remission, non-EU}) = 0.1122,$$

with $pv = 0.0006$, and a higher rate for patients with primary reason 'lost to

follow-up',

$$\tilde{\rho}^2(\text{lost to follow-up}|\text{no remission, EU}) = 0.0439$$
$$\tilde{\rho}_2^2(\text{lost to follow-up}|\text{no remission, non-EU}) = 0.2005,$$

with $pv = 0.0001$. There is also a higher rate for patients with primary reason for withdrawal 'lost to follow-up' in the group of non-EU patients, also among the remitters,

$$\tilde{\rho}^2(\text{lost to follow-up}|\text{remission, EU}) = 0.1429$$
$$\tilde{\rho}^2(\text{lost to follow-up}|\text{remission, non-EU}) = 0.4545,$$

with $pv = 0.0806$.

Upon the review of socio-demographic data, cluster Cl_2 includes a disproportionate number of young patients (age class 1: to 34 years). In this class there is a clear regional difference:

$$\tilde{\rho}^2(\text{age class 1}|\text{EU}) = 0.2397 \text{ and } \tilde{\rho}^2(\text{age class 1}|\text{non-EU}) = 0.2829.$$

By stratification of this data, considering the reasons for withdrawal, it appears that especially young patients show a higher drop-out rate due to 'lost to follow-up' which is again more pronounced in the non-EU group. The shares of young patients with primary reason for withdrawal 'lost to follow-up' are

$$\tilde{\rho}^2(\text{lost to follow-up}|\text{age class 1, EU}) = 0.0690$$
$$\tilde{\rho}^2(\text{lost to follow-up}|\text{age class 1, non-EU}) = 0.3893,$$

with $pv = 0.0009$, as can be seen in Figure 6.5.

In summary, in cluster Cl_2 there is a descriptively high number of young patients in age class 1 (to 34 years) who show differences between the regions (EU vs. non-EU). With stratification by reasons for withdrawal it appears that especially young patients show a higher drop-out rate due to 'lost to follow-up'. This effect is more pronounced in the non-EU group. Altogether the primary reasons

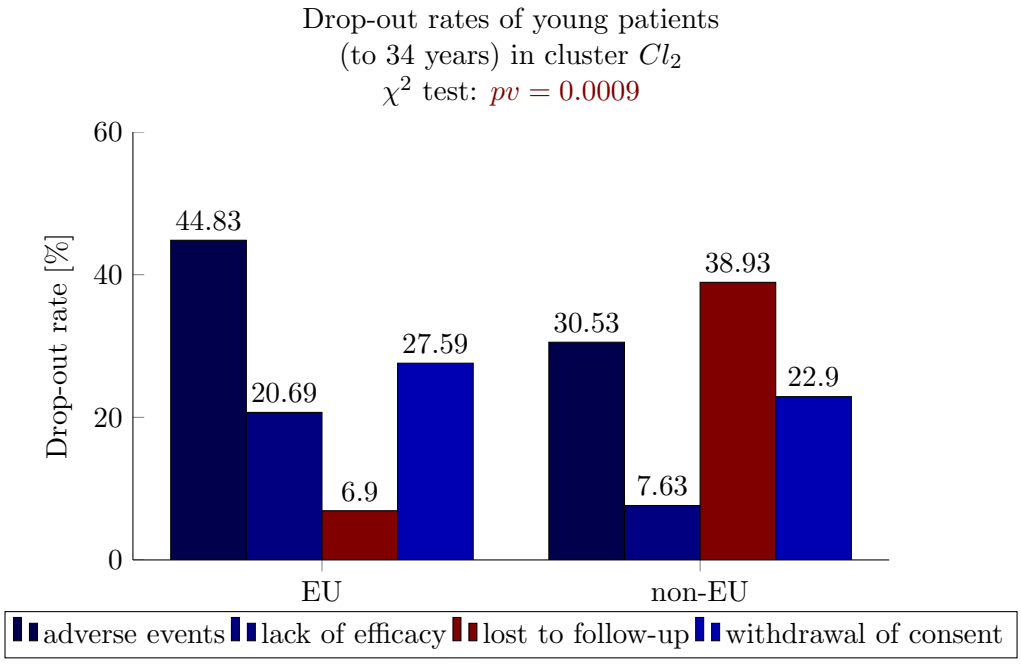

Figure 6.5.: Cluster-based identification of heterogeneity: Cluster Cl_2, reasons for withdrawal of young patients stratified by region

for withdrawal in the EU group is clinical plausible, because especially drop-outs due to 'adverse events' or 'lost to follow-up' have been reported. Compared to this, the non-EU group show higher rates of 'lost to follow-up'. Verum responders and remitters of cluster Cl_2 outside the EU drop-out more often due to 'lost to follow-up'.

Cluster Cl_4: Placebo completer

Also of interest are the placebo-treated patients who have gone through the study until the end. This group is found in cluster Cl_4 with $\kappa_4 = 1641$ patients with a response rate of

$$\tilde{p}_1^4 = 0.4052.$$

Descriptive, the response rate shows the following regional differences:

$$\tilde{p}_1^4(\text{EU}) = 0.3671 \text{ and } \tilde{p}_1^4(\text{non-EU}) = 0.4154,$$

with $pv = 0.1034$. Furthermore $\tilde{p}_2^4(\text{EU}) = 0.2283$ of the EU patients and $\tilde{p}_2^4(\text{non-EU}) = 0.2779$ of the non-EU patients remitted ($pv = 0.0599$). Over-

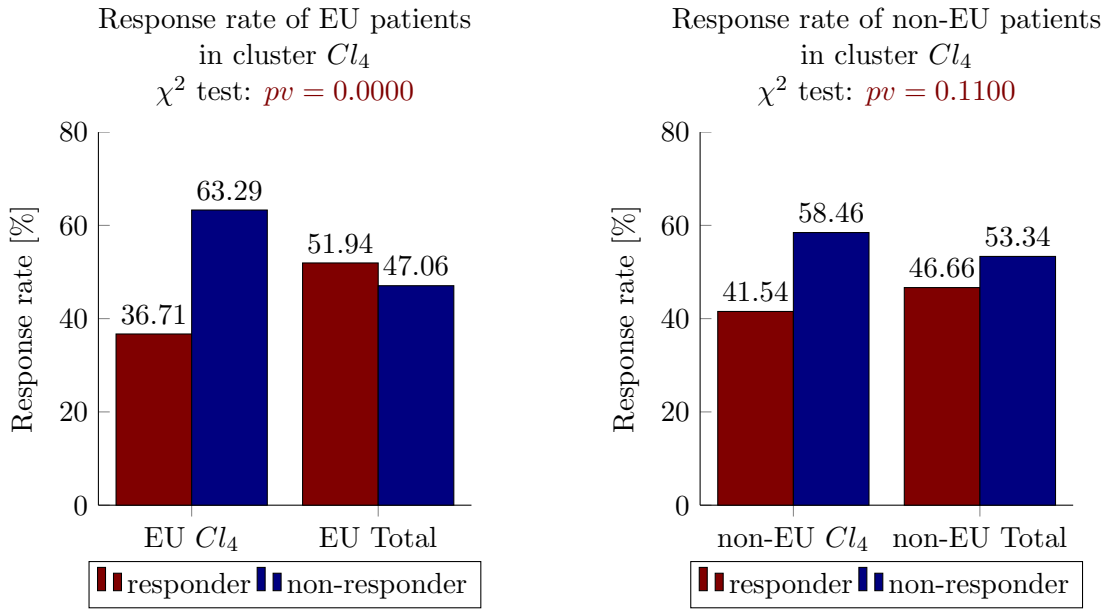

Figure 6.6.: Cluster-based identification of heterogeneity: Cluster Cl_4 vs. total population, response rate stratified by region

all there is a high placebo response rate in comparison to a response rate of the total population of

$$\tilde{p}_1 = 0.4837.$$

With regard to the differences of the response rates in cluster Cl_4 and the corresponding marginal distribution (of the total population) stratified by region there are in turn significant differences in cluster Cl_4 in comparison to the marginal distribution. For the EU regions holds

$$\tilde{p}_1^4(\text{EU}) = 0.3671 \text{ and } \tilde{p}_1(\text{EU}) = 0.5394,$$

with $pv = 0.0000$, and for the non-EU region we have

$$\tilde{p}_1^4(\text{non-EU}) = 0.4154 \text{ and } \tilde{p}_1(\text{non-EU}) = 0.4666,$$

with $pv = 0.0011$. These effects can be seen in Figure 6.6. Differences in placebo response with regard to verum-placebo differences pooled analyses are important as it can be derived by the comparison of the placebo response data in cluster Cl_4 and the response rates of the verum cluster Cl_1 and Cl_3 (see below). In

the verum cluster Cl_1 and Cl_3, significant response differences between EU and non-EU regions are shown,

$$\tilde{p}_1^1(\text{EU}) = 0.6495 \text{ and } \tilde{p}_1^1(\text{non-EU}) = 0.4962,\ pv = 0.0067$$
$$\tilde{p}_1^3(\text{EU}) = 0.6412 \text{ and } \tilde{p}_1^3(\text{non-EU}) = 0.5439,\ pv < 0.0010.$$

Lower non-EU verum response rates (cluster Cl_1 and cluster Cl_3) front higher non-EU placebo response rates (cluster Cl_4), as indication for lower verum-placebo differences in the non-EU regions. This disparity is also evident in the relation EU response to non-EU-response in clusters Cl_1 and Cl_3 compared to cluster Cl_4. The relationship EU responders to non-EU responders is higher than 1 in cluster Cl_1 and Cl_3 and lower than 1 in cluster Cl_4. Also in the placebo cluster Cl_6, the response relationship EU to non-EU is lower than 1.

Similar results can be identified for the remission data. There is a high placebo remission rate of $\tilde{p}_2^4 = 0.2681$, compared to a remission rate of the total population of $\tilde{p}_2 = 0.3093$. With regard to the differences of the remission rates in cluster Cl_4 and the marginal distribution stratified by region, there are also differences: Among the EU patients a remission rate of $\tilde{p}_2^4(\text{EU}) = 0.2283$ can be found in placebo cluster Cl_4, in comparison to the marginal distribution, here we have a remission rate of $\tilde{p}_2(\text{EU}) = 0.3620$ ($pv = 0.0000$). This difference is more pronounced than in the non-EU group with a remission rate of $\tilde{p}_2^4(\text{non-EU}) = 0.2788$ in placebo cluster Cl_4 in comparison to the marginal distribution with a remission rate of $\tilde{p}_2(\text{non-EU}) = 0.2932$ ($pv = 0.3120$). These effects can be seen in Figure 6.7. Also in the verum cluster Cl_1 and Cl_3, there are significant differences in the remission rates between EU and non-EU regions,

$$\tilde{p}_2^1(\text{EU}) = 0.4742 \text{ and } \tilde{p}_2^1(\text{non-EU}) = 0.2982,\ pv = 0.0010$$
$$\tilde{p}_2^3(\text{EU}) = 0.4159 \text{ and } \tilde{p}_2^3(\text{non-EU}) = 0.3401,\ pv = 0.0009.$$

With regard to the age classes the response rate within age class 1 (patients to 34 years) is

$$\tilde{p}_1^4(\text{age class 1}) = 0.4915.$$

In all other age classes, the response rates are between 0.3600 and 0.3900. Corre-

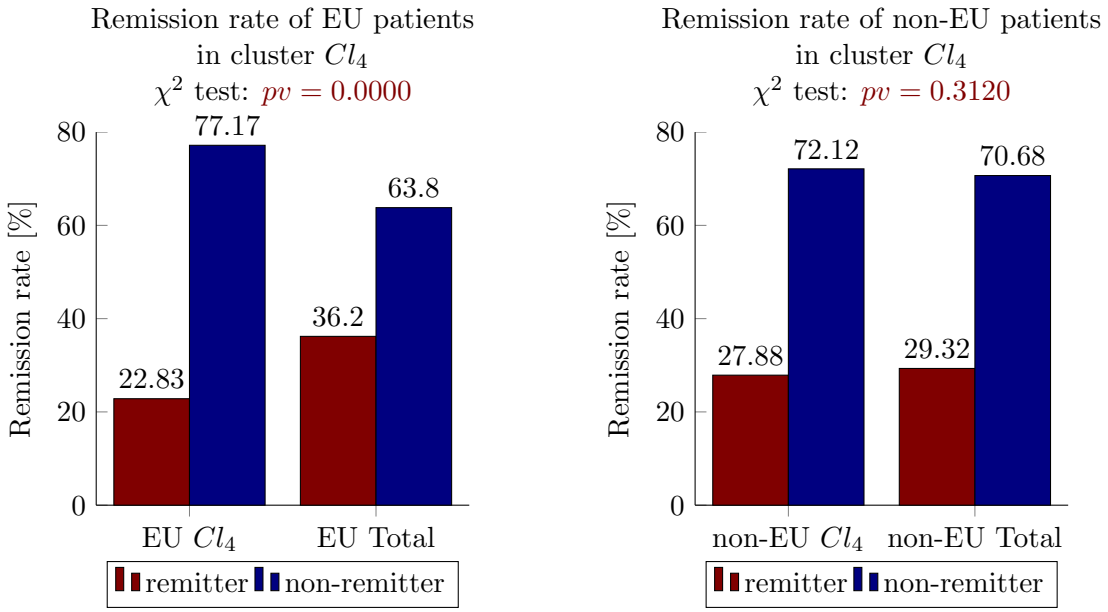

Figure 6.7.: Cluster-based identification of heterogeneity: Cluster Cl_4 vs. total population, remission rate stratified by region

sponding data are also found for the remission rates. In age class 1 the remission rate is $\tilde{p}_2^4(\text{age class 1}) = 0.3531$, in the other age classes the remission rates are between 0.2218 and 0.2807. A closer analysis of age class 1 shows

$$\tilde{\rho}^4(\text{EU}|\text{age class 1}) = 0.1780,$$

but

$$\tilde{\rho}^4(\text{non-EU}|\text{age class 1}) = 0.8220,$$

see Figure 6.8. In comparison to the marginal distribution the share of non-EU patients in age class 1 is descriptively higher in this cluster,

$$\tilde{\rho}^4(\text{non-EU}|\text{age class 1}) = 0.8220$$
$$\tilde{\rho}(\text{non-EU}|\text{age class 1}) = 0.7919.$$

This is of relevance, because the further stratification by response in this age class

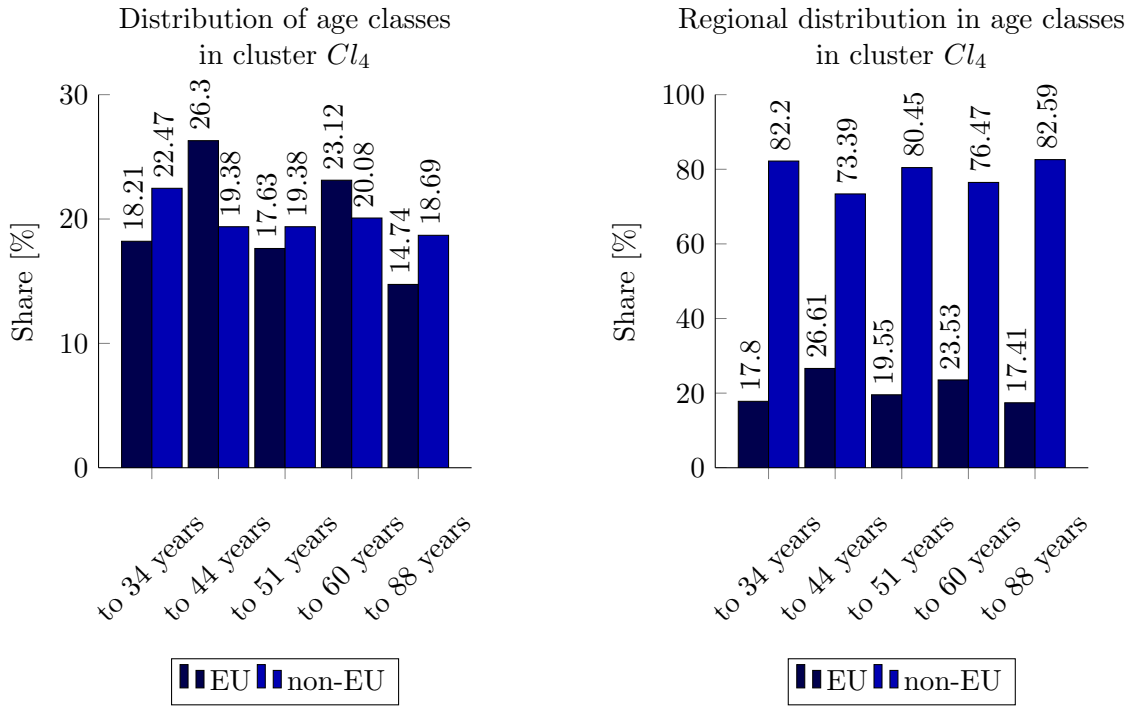

Figure 6.8.: Cluster-based identification of heterogeneity: Cluster Cl_4, distribution of age classes stratified by region

shows also regional differences, as can be seen in Figure 6.9. The results are

$$\tilde{p}_1^4(\text{age class 1, EU}) = 0.3651$$
$$\tilde{p}_1^4(\text{age class 1, non-EU}) = 0.6189$$

with $pv = 0.0268$. The stratification by remission in this age class shows the following differences: $\tilde{p}_2^4(\text{age class 1, EU}) = 0.2540$ of the young patients of EU regions have remitted in comparison to $\tilde{p}_2^4(\text{age class 1, non-EU}) = 0.3746$ of the young patients of non-EU regions ($pv = 0.0694$). Age class 1 is over-represented in the group of placebo responders in the non-EU centers and shows a higher placebo response rate and respectively a higher placebo remission rate. In the other age classes these differences can not be found, see Figure 6.10. In age class 1 the described regional effect is mainly due to the young man with a the response

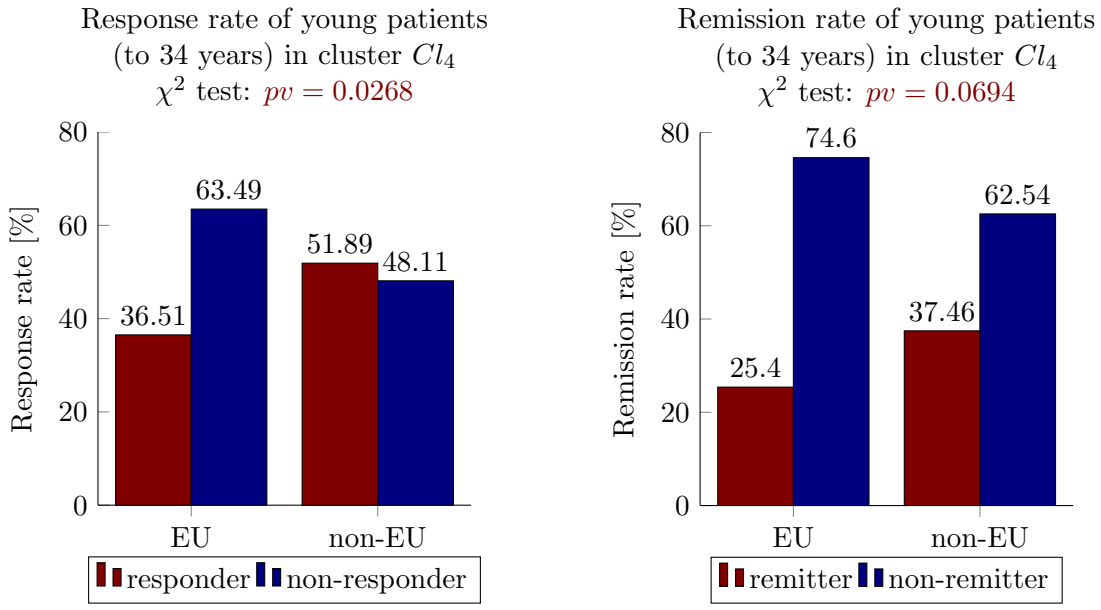

Figure 6.9.: Cluster-based identification of heterogeneity: Cluster Cl_4, response rate and remission rate in age class 1 stratified by region

rates

$$\tilde{p}_1^4(\text{male, age class 1, EU}) = 0.1905$$
$$\tilde{p}_1^4(\text{male, age class 1, non-EU}) = 0.5378,$$

with $pv = 0.0036$. For young female patients aged to 34 years, the ratio is well-balanced,

$$\tilde{p}_1^4(\text{female, age class 1, EU}) = 0.4524$$
$$\tilde{p}_1^4(\text{female, age class 1, non-EU}) = 0.5081,$$

with $pv = 0.5144$, as can be seen in Figure 6.11. For the remission rates the differences among the young men were more significant: $\tilde{p}_2^4(\text{male, age class 1, EU}) = 0.0952$ in comparison to $\tilde{p}_2^4(\text{male, age class 1, non-EU}) = 0.3773$ ($pv = 0.0121$). For the young female patients aged to 34 years, the ratio was also well-balanced with regard to remission, $\tilde{p}_2^4(\text{female, age class 1, EU}) = 0.3333$ in comparison to $\tilde{p}_2^4(\text{female, age class 1, non-EU}) = 0.3729$ ($pv = 0.6301$), as can be seen in Figure

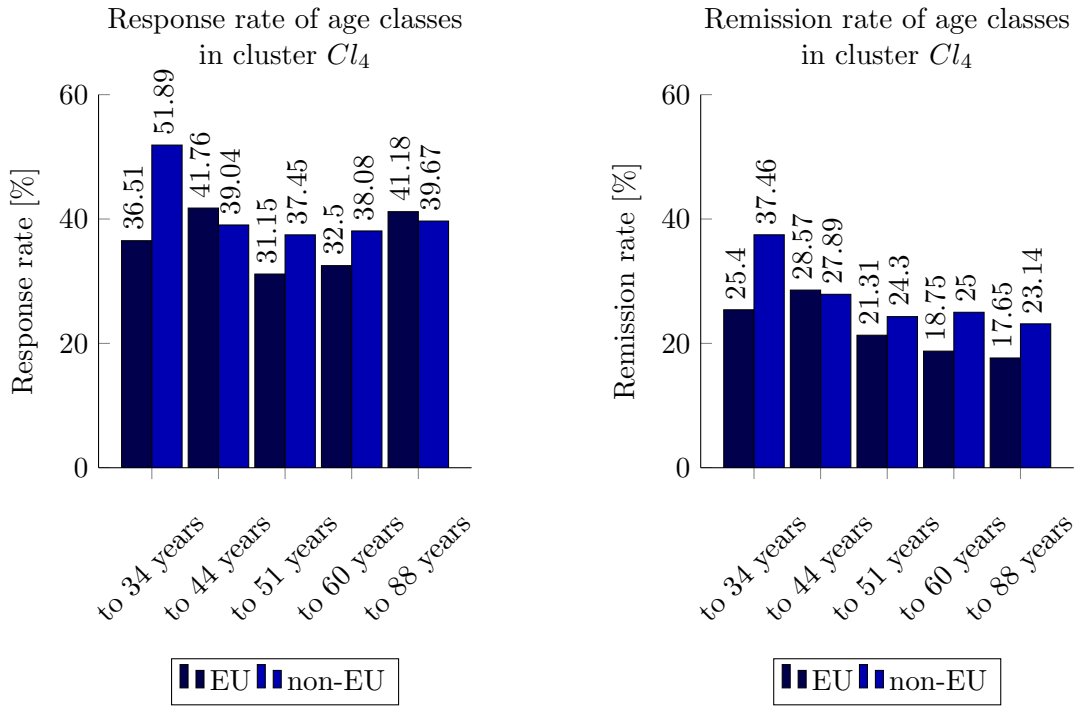

Figure 6.10.: Cluster-based identification of heterogeneity: Cluster Cl_4, response and remission rate in age classes stratified by region

6.12. The share of non-EU patients under young male participants is increased,

$$\tilde{\rho}^4(\text{non-EU}|\text{male, age class 1}) = 0.8346$$

with tendency towards significance in comparison to the marginal distribution

$$\tilde{\rho}(\text{non-EU}|\text{male, age class 1}) = 0.7571,$$

with $pv = 0.0663$.

The data in the placebo cluster Cl_4 indicates significant regional differences with regard to the recruitment behavior of younger patients - as shown in the analysis of cluster Cl_2 - they also have a higher drop-out rate respectively a lower retention rate. It is striking, especially the high placebo response / placebo remission of the younger age class which is expected to cause a reduction in the verum-placebo response. The stratification of the data by gender shows that the

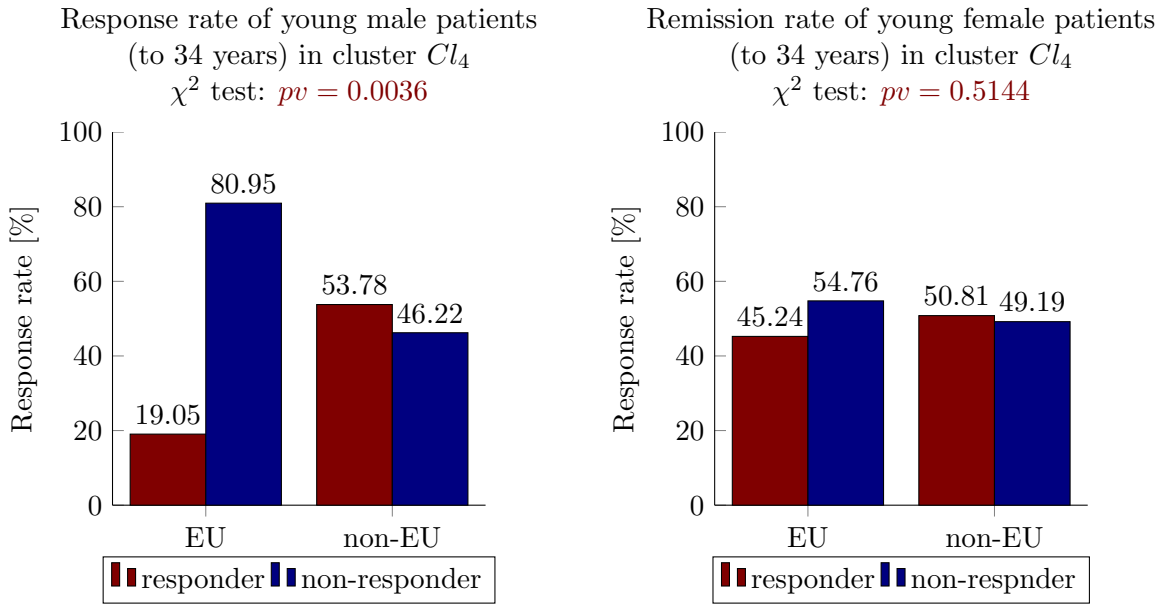

Figure 6.11.: Cluster-based identification of heterogeneity: Cluster Cl_4, response rate in age class 1 stratified by gender and region

ratio women to men in age class 1 is about two to one. Here, there are no significant differences between the regions; however, young male patients in the non-EU region are over-represented and have also a higher placebo-response/placebo remission rate. Young male patients of non-EU regions contribute significantly to the response/ remission rate in this cluster.

Cluster Cl_5

Cluster Cl_5 includes patients treated with verum medication C1 and C2, who have gone through the study until the end. In this cluster, $\kappa_5 = 674$ patients are included with a response rate of

$$\tilde{p}_1^5 = 0.7240$$

and a remission rate of $\tilde{p}_2^5 = 0.4949$. Stratification by region shows a lower share of non-EU patients in comparison to the marginal distribution,

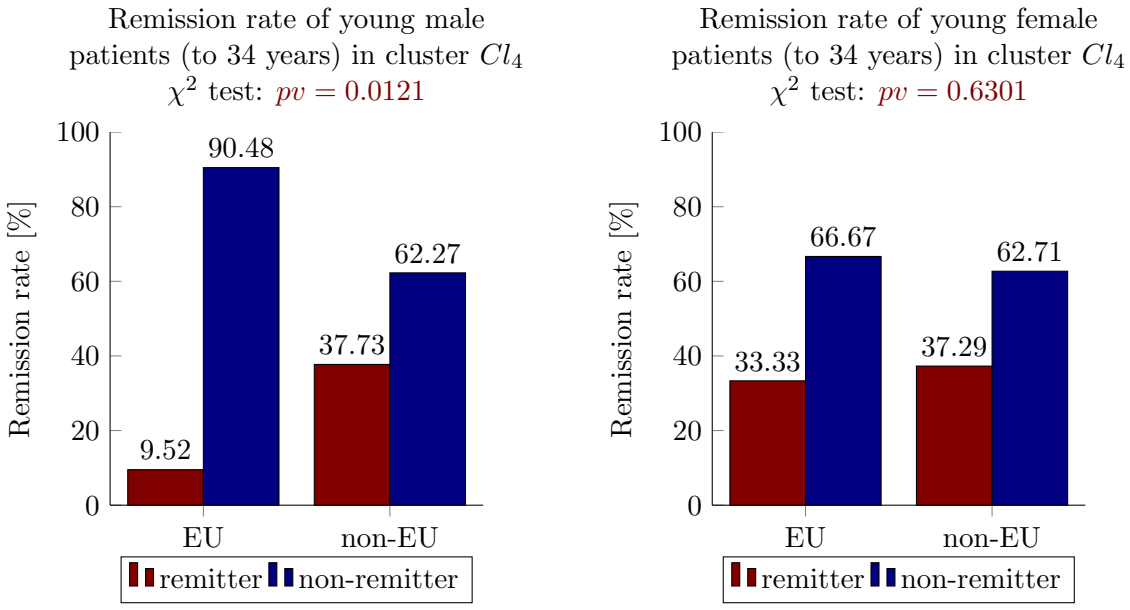

Figure 6.12.: Cluster-based identification of heterogeneity: Cluster Cl_4, remission rate in age class 1 stratified by gender and region

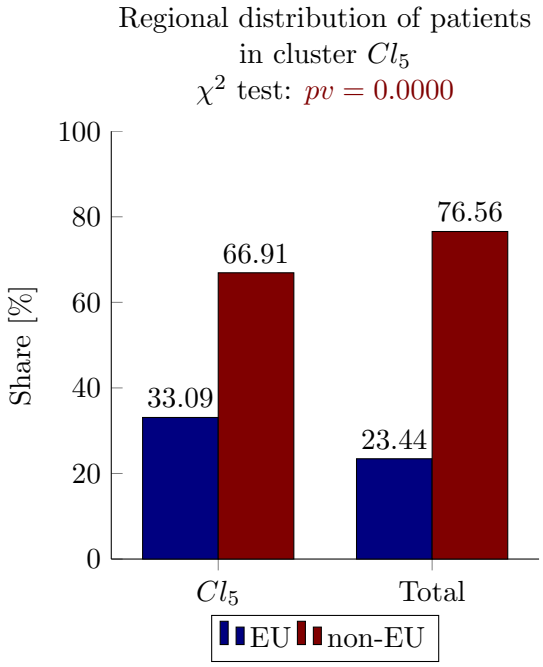

Figure 6.13.: Cluster-based identification of heterogeneity: Cluster Cl_5 vs. total population, regional distribution

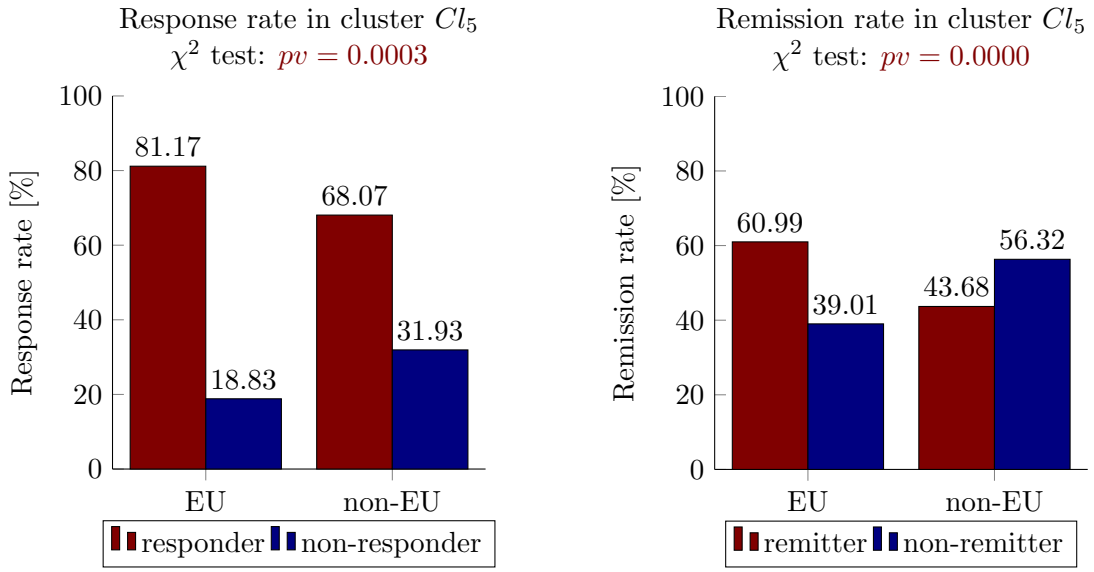

Figure 6.14.: Cluster-based identification of heterogeneity: Cluster Cl_5, response rate and remission rate stratified by region

$$\tilde{\rho}^5(\text{non-EU}) = 0.6691$$
$$\tilde{\rho}(\text{non-EU}) = 0.7656,$$

with $pv = 0.0000$. In cluster Cl_5 we also have a lower response rate in the non-EU regions,

$$\tilde{p}_1^5(\text{EU}) = 0.8117$$
$$\tilde{p}_1^5(\text{non-EU}) = 0.6807,$$

with $pv = 0.0003$, and a remission rate $\tilde{p}_2^5(\text{EU}) = 0.6099$ vs. $\tilde{p}_2^5(\text{non-EU}) = 0.4368$, with $pv = 0.0000$. These results can be seen in Figure 6.13 and Figure 6.14.

Upon the review of the socio-demographic parameters, an imbalance in the highest age class strikes in comparison to the total population: Compared to the

EU:	non-EU:
age class 1: 0.1704	age class 1: 0.1596
age class 2: 0.2422	age class 2: 0.1663
age group 3: 0.2224	age class 3: 0.1486
age class 4: 0.2466	age class 4: 0.1508
age class 5: 0.1166	age class 5: 0.3747

Table 6.33.: Cluster-based identification of heterogeneity: Cluster Cl_5, age distribution stratified by region

quite equal distributed age classes 1 to 4, th share of age class 5 is

$$\tilde{\rho}^5(\text{age class } 5) = 0.2893$$

in cluster Cl_5. The response rate in this age class is

$$\tilde{p}_1^5(\text{age class } 5) = 0.7333$$

and the remission rate is $\tilde{p}_2^5(\text{age class } 5) = 0.4941$. Overall, the age distribution shows the regional differences (EU/non-EU) presented in Table 6.33. In the EU regions, the different age classes are almost equal distributed, whereupon age class 5 seems to be under-represented. On the other hand in the non-EU group a higher rate of age class 5 can be identified. The share of elderly non-EU patients in Cluster Cl_5 tends to be increased compared to the total population

$$\begin{aligned}\tilde{\rho}^5(\text{non-EU}|\text{age class } 5) &= 0.8667\\ \tilde{\rho}(\text{non-EU}|\text{age class } 5) &= 0.8119,\end{aligned}$$

with $pv = 0,0664$, as can be seen in Figure 6.15. This could indicate that the lower percentage of responders/remitters in the non-EU group maybe due to a higher number of older patients in this group.

Cluster Cl_1 and cluster Cl_3

Cluster Cl_1 includes $\kappa_1 = 496$ and cluster Cl_3 $\kappa_3 = 2388$ patients. All patients have been treated with medication A and have gone through the study until the end. For the overall evaluation, the comparison of verum cluster Cl_1 and Cl_3

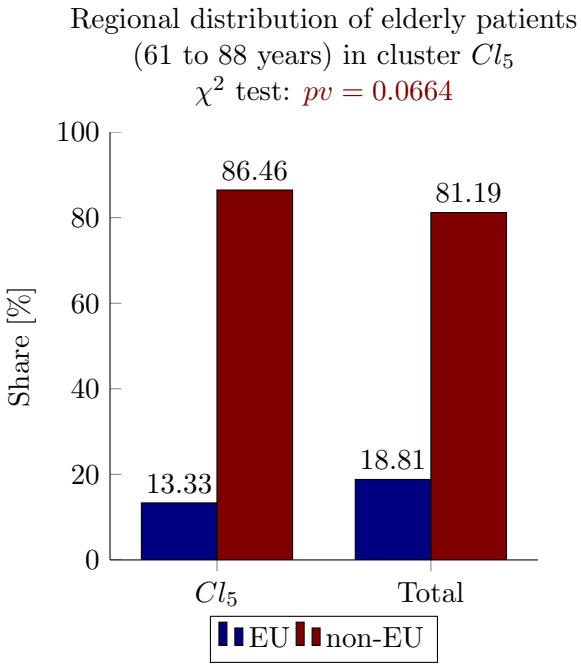

Figure 6.15.: Cluster-based identification of heterogeneity: Cluster Cl_5 vs. total population, regional distribution of elderly patients

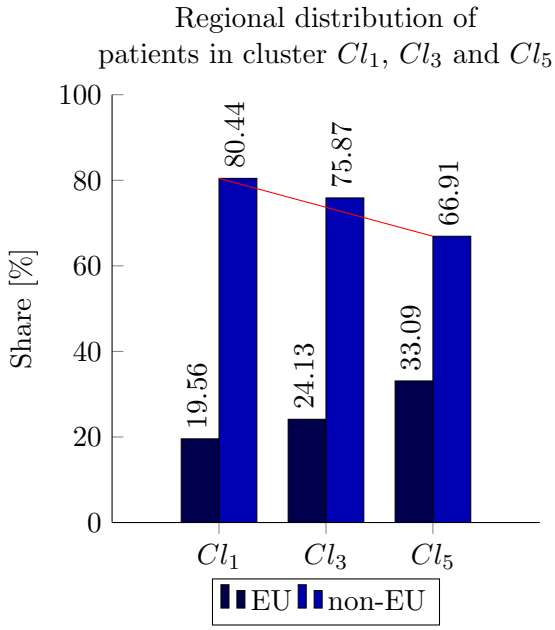

Figure 6.16.: Cluster-based identification of heterogeneity: Cluster Cl_1, Cl_3 and Cl_5, regional distribution

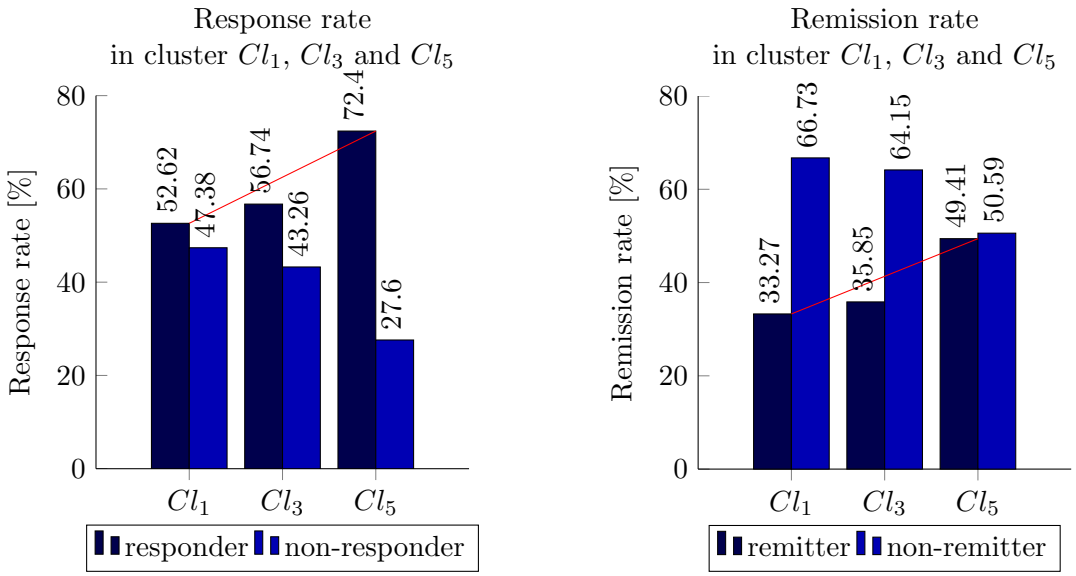

Figure 6.17.: Cluster-based identification of heterogeneity: Cluster Cl_1, Cl_3 and Cl_5, response and remission rates

(medication A) and verum cluster Cl_5 (medication C1 and C2) is of particular importance. The overall response rates in cluster Cl_1, Cl_3 and Cl_5 are

$$\tilde{p}_1^1 = 0.5262$$
$$\tilde{p}_1^3 = 0.5674$$
$$\tilde{p}_1^5 = 0.7240.$$

The remission rate in cluster Cl_1 is $\tilde{p}_2^1 = 0.3327$, in cluster Cl_3 it is $\tilde{p}_2^3 = 0.3585$ and in cluster we have Cl_5 $\tilde{p}_2^5 = 0.4941$. This contrasts with the shares of non-EU patients, which are

$$\tilde{\rho}^1 = 0.8044$$
$$\tilde{\rho}^3 = 0.7587$$
$$\tilde{\rho}^5 = 0.6691.$$

This comparison shows that the higher shares of non-EU patients seams to go hand in hand with a lower verum response rate/remission rate, see Figure 6.16 and Figure 6.17. Also within the clusters, the correspondent differences in the response rates can be found with stratification by region: in cluster Cl_3, EU

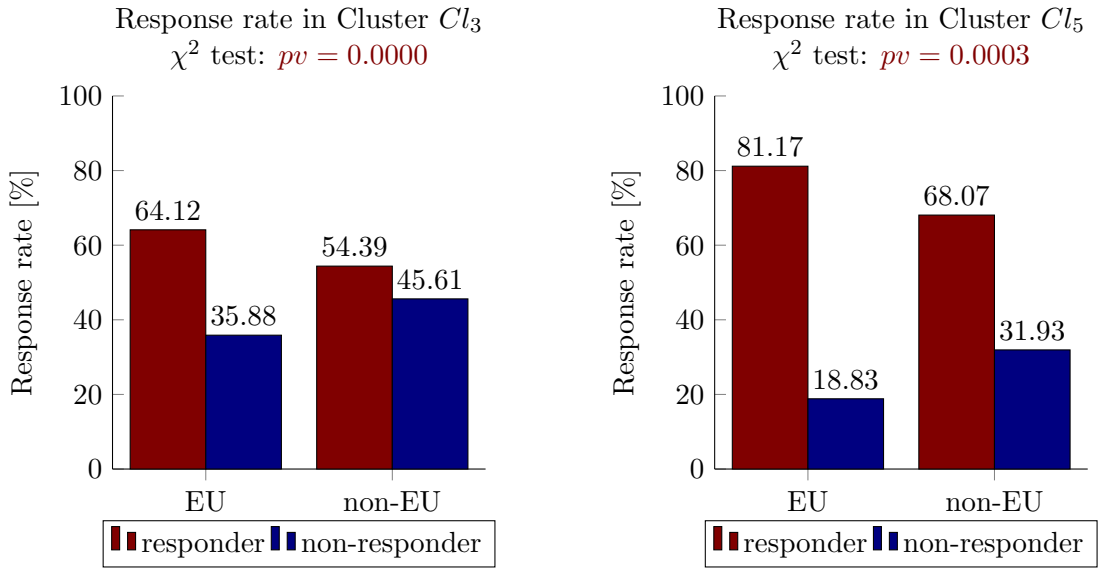

Figure 6.18.: Cluster-based identification of heterogeneity: Cluster Cl_3 and Cl_5, response rates stratified by region

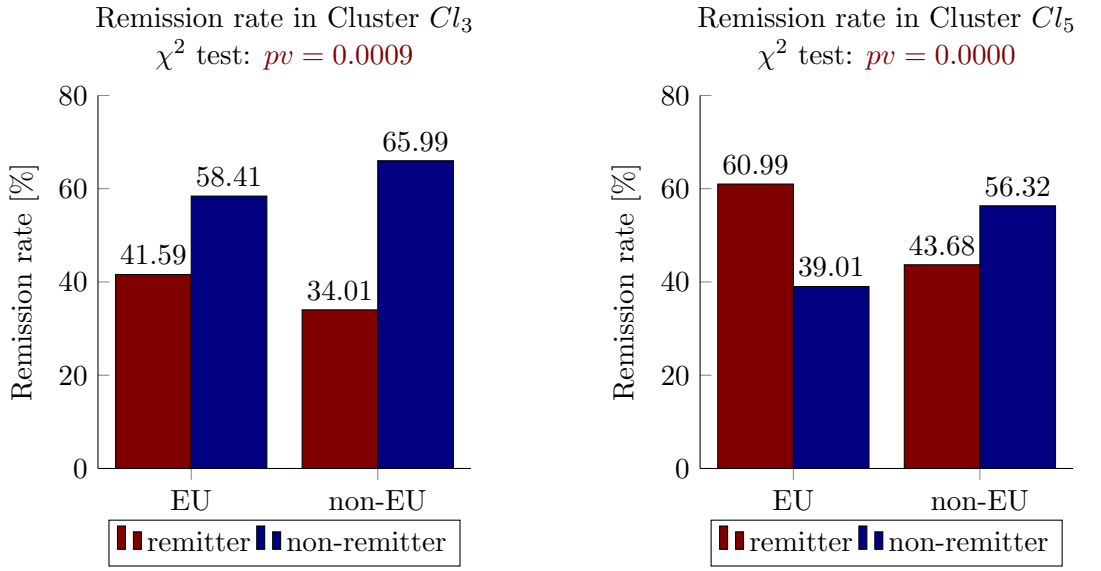

Figure 6.19.: Cluster-based identification of heterogeneity: Cluster Cl_3 and Cl_5, remission rates stratified by region

patients show a response rate of

$$\tilde{p}_1^3(\text{EU}) = 0.6412,$$

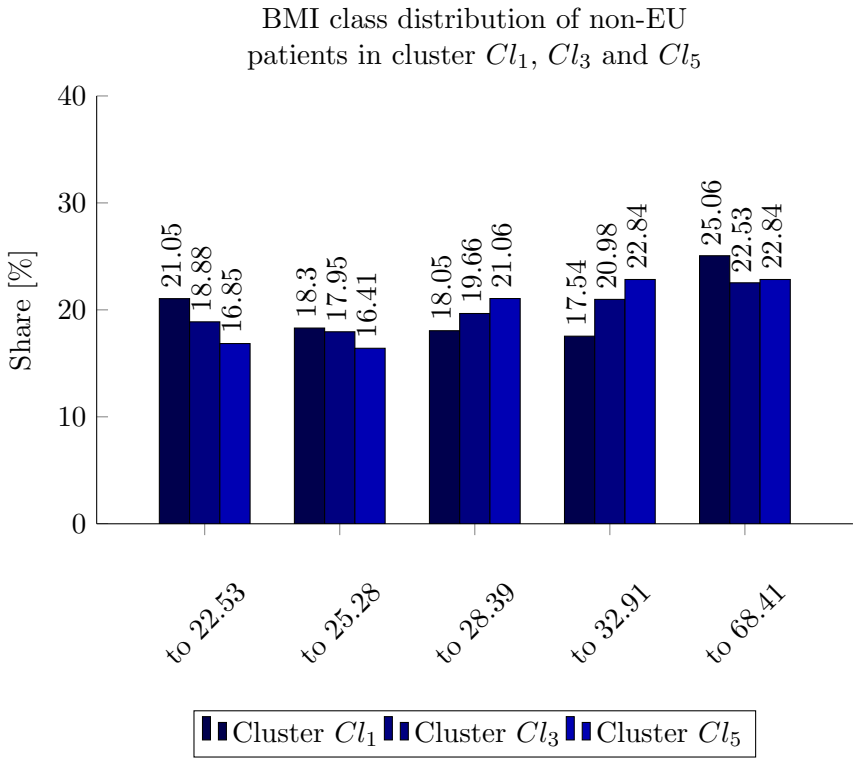

Figure 6.20.: Cluster-based identification of heterogeneity: Cluster Cl_1, Cl_3 and Cl_5, distribution of BMI classes of non-EU patients

compared to a response rate of

$$\tilde{p}_1^3(\text{non-EU}) = 0.5439$$

of non-EU patients, with $pv = 0.0000$. The response rates in Cluster Cl_5 are

$$\tilde{p}_1^5(\text{EU}) = 0.8117$$
$$\tilde{p}_1^5(\text{non-EU}) = 0.6807,$$

with $pv = 0.0003$, see Figure 6.18. With stratification by region, correspondent differences in the remission rates can be found: in cluster Cl_3 the remission rate is $\tilde{p}_2^3(\text{EU}) = 0.4159$ in the EU regions and $\tilde{p}_2^3(\text{non-EU}) = 0.3401$ in the non-EU regions, with $pv = 0.0009$. The remission rate in cluster Cl_5 is $\tilde{p}_2^5(\text{EU}) = 0.6099$ in the EU regions and $\tilde{p}_2^5(\text{non-EU}) = 0.4368$ in the non-EU regions, with $pv = 0.0000$, see Figure 6.19.

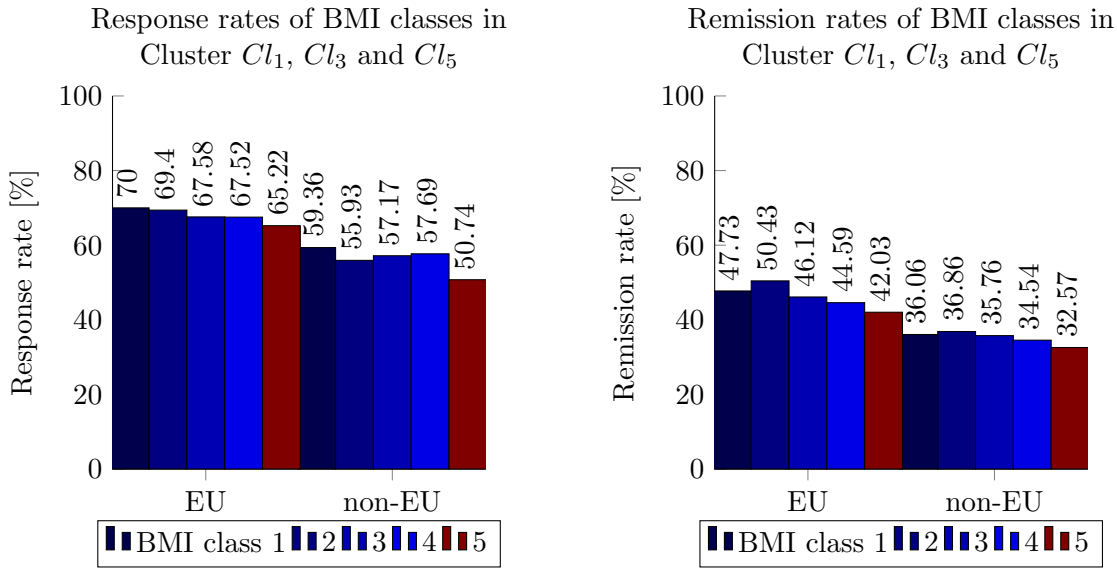

Figure 6.21.: Cluster-based identification of heterogeneity: Cluster Cl_1, Cl_3 and Cl_5 (pooled), response and remission rate stratified by BMI classes and region

As part of the evaluation of the different response and remission rates a further possible explanation comes up when considering the distribution of BMI classes. Overall, the distribution of the BMI appears to be balanced, however, in all verum cluster an increased share of non-EU patients in the higher BMI classes can be identified, especially in cluster Cl_1 and cluster Cl_3, see Figure 6.20. Altogether the response rates/remission rates of the verum clusters are lower in the higher BMI classes. With stratification by region this correlation can be found especially in the non-EU regions in cluster Cl_3, as can be seen in Figures 6.21, 6.22 and 6.23.

Summary of verum cluster

Cluster Cl_5 (medication C1 and C2) shows an overall good response rate of $\tilde{p}_1^5 = 0.7240$, whereupon the EU responder are significantly over-represented with $\tilde{p}_1^5(\text{EU}) = 0.8117$ in comparison to the non-EU responder with $\tilde{p}_1^5(\text{non-EU}) = 0.6807$. Furthermore, cluster Cl_5 shows a high remission rate of $\tilde{p}_2^5 = 0.4941$, whereupon the remission rate is $\tilde{p}_2^5(\text{EU}) = 0.6099$ for EU and $\tilde{p}_2^5(\text{non-EU}) = 0.4368$ for non-EU patients. In this cluster, a lower response/remission rate of

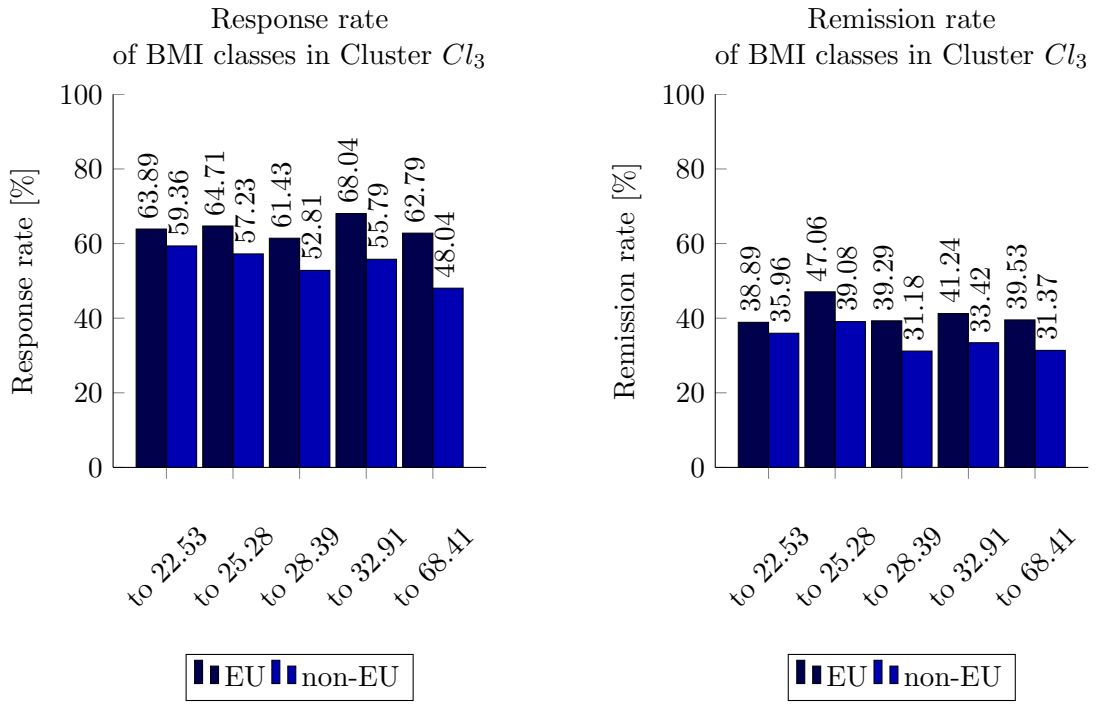

Figure 6.22.: Cluster-based identification of heterogeneity: Cluster Cl_3, response and remission rate stratified by BMI classes and region

old patients strikes who again appears to be over-represented in the non-EU group, see Figures 6.24 and 6.25.

Differences in the BMI distribution can be found between cluster Cl_1, Cl_3 and Cl_5, whereupon higher BMI classes are less represented in Cluster Cl_5 than in clusters Cl_1 and Cl_3. Stratified by region higher BMI values can be found in non-EU regions like it is known in the literature [47]. The EU group shows a consistently high response rate of at least 0.7500 and a remission rate of at least 0.5000 in all BMI classes. Overall, with stratification by BMI, there is a relationship between response rate or remission rate and BMI classes in all verum clusters. The lower the BMI the merrier the response and the remission seems to be, where this effect especially appears in the non-EU group. The comparison, stratified by BMI and response or remission, however, indicates that the region my have a greater impact on the response and remission than the BMI.

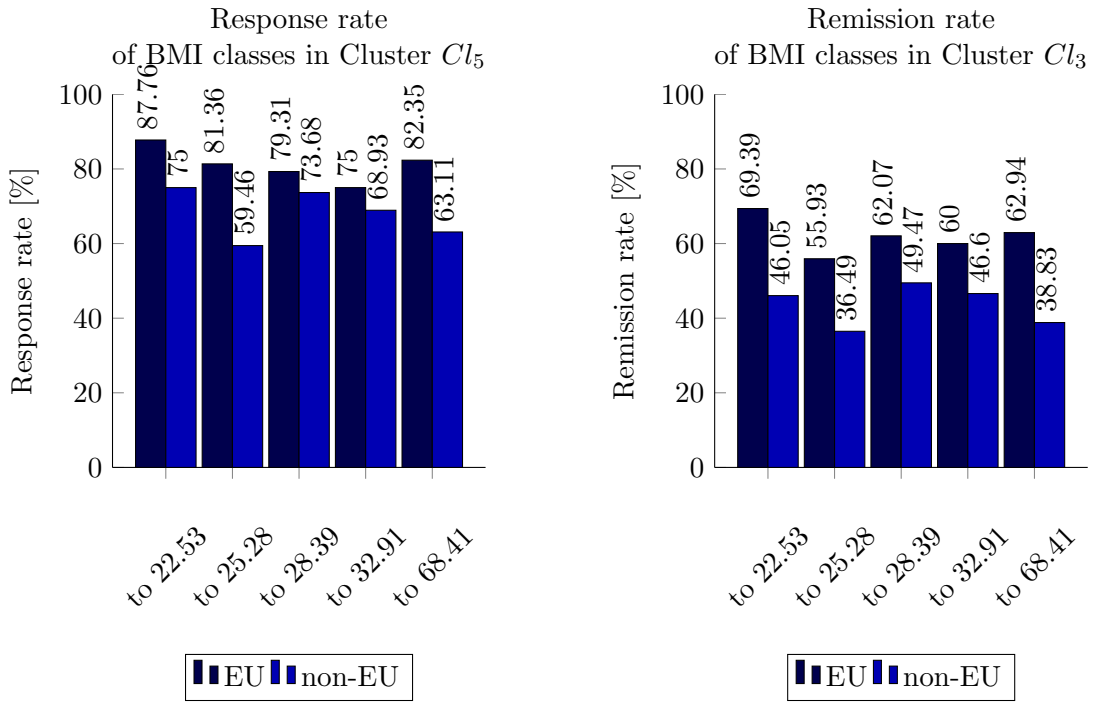

Figure 6.23.: Cluster-based identification of heterogeneity: Cluster Cl_5, response and remission rates stratified by BMI classes and region

6.5. Cluster-based prediction of treatment effects

As can be seen in the previous cluster-based analyses, it is reliable to assume that there are different treatment effects in different patient collectives with similar characteristic values combinations. In this section, the results of the endpoint-oriented clustering approach, with regard to the prediction of the treatment effects of the administered antidepressants A, C1 and C2 in the identified patient collectives, are presented. The analysis is based on the theory introduced in Section 5.9.

6.5.1. Clustering

For the prediction of the efficacy, we needed to identify homogeneous groups by the application of the endpoint-oriented clustering algorithm described in Chapter

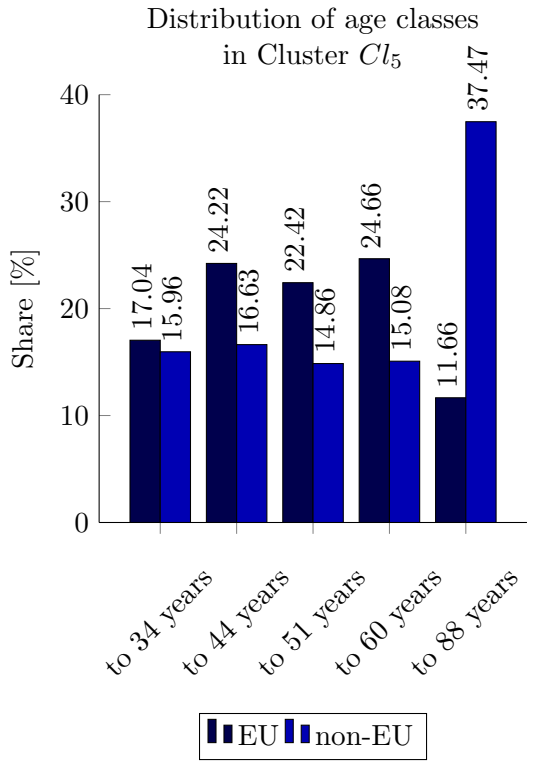

Figure 6.24.: Cluster-based identification of heterogeneity: Cluster Cl_5, age distribution stratified by region

5. Since the predictive clustering approach is supervised, in the first step, the set S^{all}, described in Section 6.1, has been divided. We used random 80% of the available patient data as training data set S and 20% as testing patient data set S^{te}. Consequently we got

$$S = \{(x_j, y_j)\}_{j=1}^{4808}$$

and

$$S^{te} = \{(x_j^{te}, y_j^{te})\}_{j=1}^{1202},$$

with $n = 4808$ and $n^{te} = 1202$. For the endpoint, we chose the binary outcome variable 'response', with the two occurrences 'yes' (= 1) and 'no' (= 0). For applying the clustering algorithm, in the next step, the characteristics with their characteristic values, which might have an influence on the response of a patient, had to be identified. All picked random variables are listed in Table 6.34. For the attributes 'BMI', 'duration' and 'age' we used the classification shown in Table 6.12. Since the characteristic values of the attributes 'treatment', 'region',

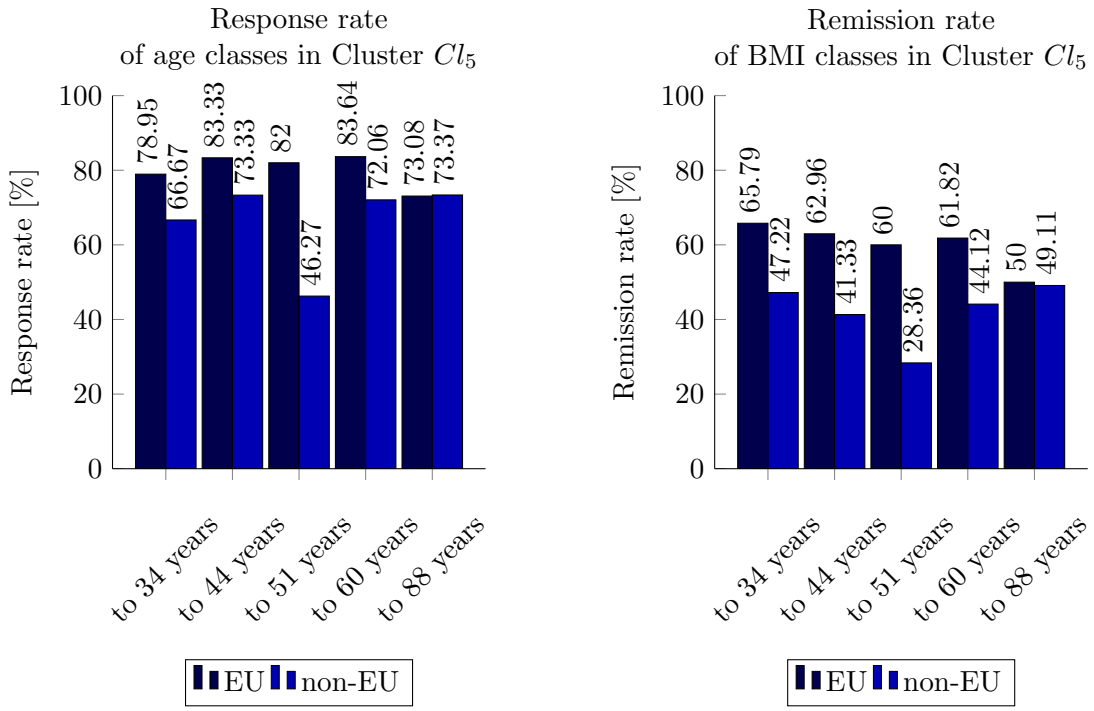

Figure 6.25.: Cluster-based identification of heterogeneity: Cluster Cl_5, response rate and remission rate stratified by region and age classes

'gender' and 'withdrawal' have a nominal level of scale, for the quantification we used the transformation technique described in Section 5.5. Then, we were able to apply the clustering algorithm to get a set of patient collectives Cl on the basis of the transformed training patient data set $\hat{S}$. The parameter setup for the clustering consisted of $k = 6$ clusters, lower bounds $l_i = 100$, for $i = 1, \ldots, 6$, and no upper bounds. Each cluster $Cl_i \in Cl$ is uniquely defined by combinations of the characteristic values of the various attributes that have been identified as similar by the algorithm.

6.5.2. Cluster-based analysis

Since the patients' outcome Y^i in the training patient data set $\hat{S}^i$ of cluster Cl_i is Bernoulli distributed, with the two occurrences 'response' (= 1) and 'no response' (= 0), we used the estimated response rate $\tilde{p}^i$ for the definition of the

A_i	Characteristic	Values
A_1	Treatment	A5mg, A10mg, A15mg, A20mg, C1, C2, PBO
A_2	Region	EU, non-EU
A_3	BMI	13 to 68.41 kg/m^2
A_4	MADRS	13 to 52
A_5	CGI-S	3 to 7
A_6	Duration	28 to 7976 days
A_7	Sex	Female, Male
A_8	Age	18 to 88 years
A_9	Withdrawal	Adverse events, Lack of efficacy, Lost to follow-up, Withdrawal of consent
	Outcome	Values
	Response	yes, no
	Remission	yes, no

Table 6.34.: Cluster-based prediction: Independent and dependent variables used for the predictive clustering approach

i	1	2	3	4	5	6	all
κ_i	1322	496	717	357	289	1627	4808
$\tilde{p}^i$	0.4085	0.5504	0.5411	0.1541	0.2042	0.6404	0.4902

Table 6.35.: Cluster-based prediction: Results with respect to 'response'; number of patients and response rate per cluster

cluster value,

$$f_i(Cl_i) = \tilde{p}^i = \frac{1}{\kappa_i} \sum_{j=1}^{\kappa_i} \mathbb{1}_{\{1\}}(y_j^i),$$

where κ_i is the number of patients in cluster Cl_i, for $i = 1, \ldots, k$. The variance is given by

$$(\tilde{\sigma}^i)^2 = \tilde{p}^i(1 - \tilde{p}^i).$$

The cluster value represents the predictive outcome or the predictive probability of responding to a medication for a patient who would be assigned to this correspondent cluster due to his characteristic values combination. For the $(1 - \alpha)$ confidence interval of the true response rate p^i in cluster Cl_i holds

$$\mathcal{I}^i = [\tilde{p}^i - z[1 - \frac{\alpha}{2}]\frac{\tilde{\sigma}^i}{\sqrt{\kappa_i}}, \tilde{p}^i + z[1 - \frac{\alpha}{2}]\frac{\tilde{\sigma}^i}{\sqrt{\kappa_i}}],$$

i	$\tilde{p}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$	
1	0.4085	0.2416	0.3862	0.4307
2	0.5504	0.2475	0.5137	0.5871
3	0.5411	0.2483	0.5105	0.5718
4	0.1541	0.1303	0.1226	0.1855
5	0.2042	0.1625	0.1652	0.2432
5	0.6404	0.2303	0.6209	0.6600

Table 6.36.: Cluster-based prediction: Response rates, variances and confidence intervals for each cluster

i	1	2	3	4	5	6	Total
.	1322	496	717			1627	4162
Adverse Event(s)				142	142		284
Lack of Efficacy				69	37		106
Lost to Follow-up				58	57		115
Withdrawal of Consent				88	53		141
Total	1322	496	717	357	289	1627	4808

Table 6.37.: Cluster-based prediction: Distribution of patients classified by reason for withdrawal

where $z[1-\frac{\alpha}{2}]$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution. In the following we set $\alpha = 0.1$.

The distribution of the $n = 4808$ patients into the patient collectives with its predictive cluster values is shown in Table 6.35. The corresponding variances and the confidence intervals can be found in Table 6.36. E.g. the predictive cluster values for clusters Cl_2 and Cl_3 are

$$\tilde{p}^2 = 0.5504 \text{ and } \tilde{p}^3 = 0.5411,$$

with overlapping confidence intervals. For the predictive clustering approach, it is reliable to combine clusters with similar cluster values. Thus, we merged cluster Cl_2 and Cl_3. As can be seen in Table 6.37, in cluster Cl_4 and Cl_5 all drop-outs are grouped with a low response rate of

$$\tilde{p}^4 = 0.1541 \text{ and } \tilde{p}^5 = 0.2042$$

i	1	2	3	4	all
κ_i	1322	1213	646	1627	4808
$\tilde{p}^i$	0.4085	0.5449	0.1765	0.6404	0.4902

Table 6.38.: Cluster-based prediction: Merged results with respect to 'response'; number of patients and response rate per cluster

i	$\tilde{p}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$	
1	0.4085	0.2416	0.3862	0.4307
2	0.5449	0.2480	0.5214	0.5684
3	0.1765	0.1453	0.1518	0.2011
4	0.6404	0.2303	0.6209	0.6600

Table 6.39.: Cluster-based prediction: Merged response rates, variances and confidence intervals for each cluster

and overlapping confidence intervals. Consequently, we also merged these two clusters. Then, the cluster numbers have been renamed and the number of clusters has been reduced to $k = 4$. The results can be found in Table 6.38. The corresponding variances and the confidence intervals are shown in Table 6.39. As we can see, there is no overlapping of the confidence intervals. Therefore, we can conclude that it is reliable to assume that the response rates are significantly different across all four clusters.

6.5.3. Statistical evaluation

For the evaluation of the reliability of the predictive response rates in the identified patient collectives, we used the hypothesis test procedure introduced in Section 5.9.3. In the first step, we put the cluster values in the right order and renumbered the clusters,

$$f_1(Cl_1) = 0.1765,\ f_2(Cl_2) = 0.4085,\ f_3(Cl_3) = 0.5449,\ f_4(Cl_4) = 0.6404.$$

Thus, for the cluster values holds

$$f_1(Cl_1) \leq f_2(Cl_2) \leq f_3(Cl_3) \leq f_4(Cl_4).$$

In the next step, we set the system of hypotheses. The right tailed hypotheses tests are given by

$$H_{0,l}^i : p^i \leq f_i(Cl_i)\delta_l^i =: \tilde{p}_l^i$$

which indicates the assumption that the true response rate p^i is at most $\tilde{p}_l^i$, for $i = 2, \ldots, 4$. The null hypotheses of the left tailed hypotheses tests are formulated by

$$H_{0,u}^1 : p^i \geq f_i(Cl_i)\delta_u^i =: \tilde{p}_u^i$$

which stands for the assumption that p^i is at least $\tilde{p}_r^i$, for $i = 1, \ldots, 3$. The corresponding elements of the parameter set Δ are specified by

$$\delta_u^i = 1 - 0.4 + \frac{0.4 f_{i+1}(Cl_{i+1})}{f_i(Cl_i)}, \text{ for } i = 1, \ldots, 3$$
$$\delta_l^i = 1 - 0.4 + \frac{0.4 f_{i-1}(Cl_{i-1})}{f_i(Cl_i)}, \text{ for } i = 2, \ldots, 4.$$

In the next step, we assigned the patients in the transformed testing data set $\hat{S}^{te}$ to the corresponding clusters, due to their characteristic values combination. Then, we were able to calculate the response rate of the resulting set $\hat{S}_{Cl_i}^{te}$ in cluster Cl_i,

$$f_i^{te}(Cl_i) = \tilde{p}^{i,te} = \frac{\sum_{j=1}^{\kappa_i^{te}} \mathbb{1}_{\{1\}}(y_j^{i,te})}{\kappa_i^{te}},$$

for the comparison with the response rate of the patients in the training data set $\hat{S}^i$ of cluster Cl_i. Then, all requirements were met for the formulation of the realization of the test statistics,

$$t_u^i := \frac{f_i^{te}(Cl_i) - \tilde{p}_u^i}{\sqrt{\frac{\tilde{p}_u^i(1-\tilde{p}_u^i)}{\kappa_i^{te}}}}, \text{ for } i = 1, \ldots, 3$$
$$t_l^i := \frac{f_i^{te}(Cl_i) - \tilde{p}_l^i}{\sqrt{\frac{\tilde{p}_l^i(1-\tilde{p}_l^i)}{\kappa_i^{te}}}}, \text{ for } i = 2, \ldots, 4.$$

For the evaluation of the results of the test statistics, we used the p-values,

$$pv_u^i := P(T_u^i < t_u^i | H_{0,u}^i) = F_{\mathcal{N}}(t_u^i), \text{ for } i = 1, \ldots, 3$$
$$pv_l^i := P(T_l^i > t_l^i | H_{0,l}^i) = 1 - F_{\mathcal{N}}(t_l^i), \text{ for } i = 2, \ldots, 4,$$

where $F_{\mathcal{N}}$ denotes the cumulative distribution function of the standard normal distribution. In case of a p-value smaller than the predefined $\alpha = 0.15$ we rejected the null hypothesis.

On the basis of the underlying patient data set, for the right tailed hypotheses tests we got

$$H_{0,l}^2 : p^2 \leq 0.3157$$
$$H_{0,l}^3 : p^3 \leq 0.4903$$
$$H_{0,l}^4 : p^4 \leq 0.6022$$

and the null hypotheses of the left tailed hypothesis test were formulated by

$$H_{0,u}^1 : p^1 \geq 0.2693$$
$$H_{0,u}^2 : p^2 \geq 0.4631$$
$$H_{0,u}^3 : p^3 \geq 0.5831.$$

In Table 6.40 the distribution of the $n^{te} = 1202$ patients into the four clusters and the corresponding response rates are presented. The comparison of the predictive values in the training and the testing data set is shown in Figure 6.26. Here a very accurate prediction is visible. The evaluation of the reliability of the predicted response rates are based on the realizations of the statistics T_l^i and T_u^i shown in Table 6.41. For the decision, if a hypothesis test can be rejected we used the p-values, which can also be found in Table 6.41. All null hypotheses could be rejected due to a p-value smaller than $\alpha = 0.15$, expect null hypothesis $H_{0,l}^4$. E.g. in cluster Cl_1, the null hypothesis $H_{0,u}^1$ could be rejected due to the p-value of $pv_u^1 = 0.0002$. This implies that the probability to get this result or a higher value for the realization of the test statistic T_u^1 is only 0.2% under the assumption that the true response rate p^1 is greater than $\tilde{p}_u^1 = 0.2693$.

In these results, we identified a trade-off between the accuracy and reliability of the predicted response rates in the patient collectives identified by the cluster-

i	1	2	3	4	all
κ_i	646	1322	1213	1627	4808
$\tilde{p}^i$	0.1765	0.4085	0.5449	0.6404	0.4902
κ_i^{te}	165	317	310	410	1202
$\tilde{p}^{i,te}$	0.1455	0.3912	0.5226	0.5854	0.4576

Table 6.40.: Cluster-based prediction: Merged results with respect to 'response'; number of patients and response rate per cluster

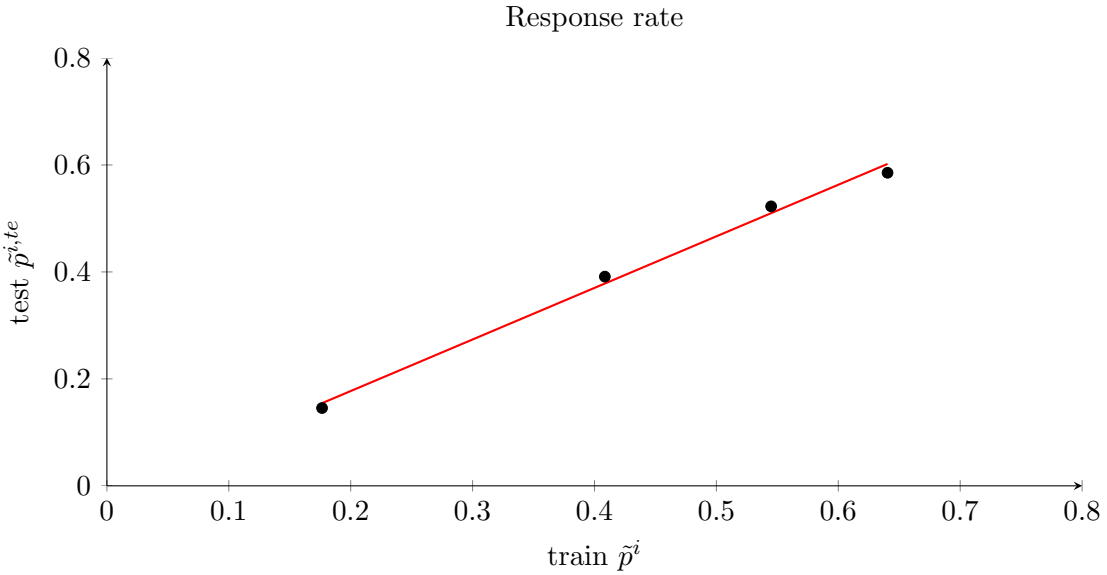

Figure 6.26.: Cluster-based prediction: Comparison of response rates of training and testing sets

ing approach. The accuracy of the prediction is high, but the reliability of the response rate in cluster Cl_4 has to be questioned and the prediction has to be done conditionally under the consideration of the p-value.

Cluster i	1		2		3		4	
Bounds		u	l	u	l	u	l	
$\tilde{p}^i$	0.1765		0.4085		0.5449			0.6404
$\tilde{p}^i_{l,u}$		0.2693	0.3157	0.4631	0.4903	0.5831	0.6022	
$\tilde{p}^{i,te}$	0.1455		0.3912		0.5226			0.5854
$t^i_{l,u}$		-3.5855	2.8920	-2.5669	1.1353	-2.1624	-0.6980	
$pvl^i_{l,u}$		0.0002	0.0019	0.0051	0.1281	0.0153	0.7574	

Table 6.41.: Cluster-based prediction: hypotheses test procedure results

7. Conclusion

In this thesis, we introduced new cluster-based statistical methods for the analysis of big patient data to improve the quality of healthcare delivery in terms of health economic evaluation and evidence-based medicine. These new approaches deal with the evaluation and prediction of the efficacy of medical interventions and the identification of clinical heterogeneity in the treatment effects for different patient collectives. For the determination of these sufficiently large and homogeneous collectives, by means of similar combinations of the characteristic values, an innovative endpoint-oriented clustering technology, supervised and unsupervised, has to be applied for all introduced statistical methods. For determining the corresponding clusters, the underlying clustering algorithm uses geometrical proximity after the transformation of the patient data into its one-dimensional conditional expected values. With the use of the transformation technique, it is possible to consider also nominal patient characteristics as independent variable if it is assumed that this characteristic has an influence on the outcome of a patient or the measured endpoint. The clustering theory is based on the work of Brieden and Gritzmann and a short overview was given in this thesis in Chapter 5.

For the application of the new invented cluster-based theories, as data base we used patient data derived by RCTs on three different antidepressants in the acute treatment of depression. Thereby, the goal was to analyze the measured endpoints if there is an additional benefit for patients treated with a new invented drug, here called A, in comparison to the standard therapies C1 and C2.

7.1. Cluster-based meta-analysis

In Section 5.7, we introduced the new invented cluster-based meta-analysis for patient collectives which have been identified by the unsupervised endpoint-oriented clustering algorithm. Since the collectives are determined by patients with similar combinations of their characteristic values, the main assumption is that for the patients of a cluster the efficacy of an intervention is equal and the treatment effects across cluster vary. Thus, for the evaluation of the efficacy of a medication in a patient collective, we adjusted the common meta-analytic approaches presented in Chapter 3 and introduced the cluster-based fixed-effects and random-effects model. With these approaches, clinical and especially regional heterogeneity in the treatment effects, which is a crucial problem in meta-analysis, can be considered. For this purpose, we developed new methods for the assessment and evaluation of the heterogeneity within a cluster and across clusters on the basis of the statistical theory presented in Chapter 4.

The empirical results of this new approach were compared to those of the common meta-analysis. The new invented drug A, with its dosages 5mg, 10mg, 15mg and 20mg, showed no additional benefit compared to C1 and C2 when applying the meta-analytic methods. But with the use of the cluster-based meta-analysis, on specific patient collectives medication A was more efficient than medication C1 and C2. This implies that the assumption of meta-analysis, that there is one true treatment effect for a drug, should be regarded cautiously. Moreover, with the new invented statistical assessment method of heterogeneity across clusters, it was shown that it is reliable to assume that there are different treatment effects for patients with different combinations of their characteristic values.

7.2. Cluster-based identification of heterogeneity

In the presented cluster-based identification of heterogeneity of Section 5.8, the discovery of hidden structures, in terms of region-specific patterns of clinical or socio-demographic parameters of patient data derived by RCTs, was introduced.

The identification methods are applied on the patient collectives identified with the help of the unsupervised endpoint-oriented clustering algorithm. Then, the differences in the conditional mean outcomes or shares of patients with different combinations of their characteristic values can be evaluated with the help of common statistical tests, like the χ^2 test for independence. The calculation of the conditional mean outcome is thereby dependent on the level of scale of the patients outcome.

In the underlying patient data, the cluster-based identification of the heterogeneity in the treatment effects was applied to response and remission data. Thereby, the focus has been on the differences in the centers of EU vs. non-EU regions and in the clinical and demographic parameters. Both, in the response analysis as well as in the ensuring analysis of remitting patients, in the correspondent clusters, consistent indications could be found that can also attribute divergent clinical endpoint data to region-specific constellations. Using multivariate cluster analysis, significant differences were found in the gender composition, age cohorts, the drop-out rates, the placebo response rates and in clinical parameters, e.g. BMI. The described structures, which can be attributed to a different composition of the base population as well as to the recruitment behaviour, may have contributed to the differences in the endpoints in the regions EU vs. non-EU of the medication A studies. Comparable factors are known from literature as therapeutic effect influencing parameters. [70, 47, 66, 39]. Impacts on verum-placebo differences, which can be associated with clinical and demographic parameters, appear to be of particular importance. The placebo response seems to be higher in non-EU regions due to different listed factors (gender, age, region). The effects shown here fit in the known scientific literature [43, 53]. The reported results, which point to correspondent regional differences in the analyzed clusters, provide a mathematical founded and also clinical reasonable explanation for high heterogeneity identified in the meta-analyses of 6.2.

7.3. Cluster-based prediction of treatment effects

In Section 5.9, we introduced how the efficacy of medical interventions can be predicted for patient collectives identified by the supervised endpoint-oriented clustering approach. Thereby, a training patient data is used for determining the different clusters. Then, dependent on the level of scale of the outcome data of the participating patients, a mean outcome or a general treatment effect for each collective can be calculated as predictive value. For the assessment of the reliability of the given predictive values, we presented statistical tests based on the comparison of the cluster values of the training and the testing data set. For the correspondent test statistics, the patients of the testing data set are assigned to the corresponding clusters and the mean outcome or treatment effects of these testing patients in the correspondent clusters are calculated.

In the practical application, after the identification of the patient collectives by the endpoint-oriented clustering approach, we had clusters with similar predictive values. Since the goal is to predict the efficacy of a medical intervention, it was reliable to merge clusters with similar predictions. The empirical results of the cluster-based prediction of the treatment effects then showed a very high accuracy in the predictive values for the different patient collectives. Also the results of the statistical tests showed a high reliability of the calculated predictive cluster values.

7.4. Outlook

The new introduced cluster-based theories are based on the innovative endpoint-oriented clustering approach, which allows adjustments in many ways. The clustering algorithm of the individual patient data can be divided into two components. The data transformation, and the clustering itself, including the cell decomposition. Like it is stated in [48], in each of these steps there are parameters and procedures that can be modified. For the data transformation, the one-dimensional conditional expected values could be extended to multidimensional conditional expectations. For the computation of the clustering, there are

numerous adjustable parameters to be investigated, like lower and upper bounds, the approximation of the ellipsoidal norm and last but not least the number of clusters. Additionally, the calculation of good initial sites or improvements of the iterative procedure, like steepest descent, could be analyzed.

The cluster-based meta-analysis, which is conducted on the patient collectives, could take into account the cluster-based identified heterogeneity by calculating the treatment effects for the corresponding patient groups with different clinical or socio-demographic parameters in a cluster, instead of differentiating between the single trials. The heterogeneity across trials in a cluster, caused by the random-effect, could be taken into account by using the trial information as influencing independent variable for the clustering approach. If it is necessary, the summary treatment effect for a patient collective could then be calculated by the weighted aggregation of the single treatment effect of the patient groups by considering the heterogeneity with a specific regional-effect. With such an approach the heterogeneity could be eliminated as well as possible.

For the cluster-based identification of heterogeneity, beside the statistical tests, new indices for the assessment of heterogeneity could be developed. Also the analysis if clinical heterogeneity could be completely considered when increasing the cluster number is an interesting topic of future work.

The assessment of the reliability of the cluster-based prediction of the treatment effects for the clusters is based on a statistical hypothesis testing approach. The setting of the underlying hypotheses and the detailed analysis of the trade-off between the accuracy and the reliability of the prediction could be investigated in the future. We introduced the comparing distance as a first parameter scheme for the statistical hypotheses setting. Additional analysis of different parameter settings in combination with different cluster numbers could also be an interesting extension of this part.

Appendices

A. Meta-analysis

Study $j = 1$ (T11)

	s	f	n
A5mg	72	35	107
PBO	47	55	102
	119	90	209

Study $j = 2$ (T12)

	s	f	n
A5mg	86	64	150
PBO	67	78	145
	153	142	295

Study $j = 3$ (T13)

	s	f	n
A5mg	88	62	150
PBO	53	89	142
	141	151	292

Study $j = 4$ (T21)

	s	f	n
A5mg	138	144	282
PBO	131	144	275
	269	288	557

Study $j = 5$ (T22)

	s	f	n
A5mg	66	84	150
PBO	49	90	139
	115	174	289

Study $j = 6$ (T23)

	s	f	n
A5mg	62	75	137
PBO	33	104	137
	95	179	274

Study $j = 7$ (T27)

	s	f	n
A5mg	70	71	141
PBO	58	88	146
	128	159	287

Table A.1.: Meta-analysis: 2×2-table of patients treated with A5mg

Study $j = 1$ (T11)

	s	f	n
A10mg	67	30	97
PBO	47	55	102
	114	85	199

Study $j = 2$ (T12)

	s	f	n
A10mg	84	63	147
PBO	67	78	145
	151	141	292

Study $j = 3$ (T15)

	s	f	n
A10mg	92	95	187
PBO	58	131	189
	150	226	376

Study $j = 4$ (T23)

	s	f	n
A10mg	68	69	137
PBO	33	104	137
	101	173	274

Study $j = 5$ (T25)

	s	f	n
A10mg	50	99	149
PBO	45	107	152
	95	206	301

Study $j = 6$ (T26)

	s	f	n
A10mg	54	88	142
PBO	49	95	144
	103	183	286

Study $j = 7$ (T27)

	s	f	n
A10mg	80	67	147
PBO	58	88	146
	138	155	293

Table A.2.: Meta-analysis: 2×2-table of patients treated with A10mg

Study $j = 1$ (T14)

	s	f	n
A15mg	82	59	141
PBO	51	102	153
	133	161	294

Study $j = 2$ (T24)

	s	f	n
A15mg	64	75	139
PBO	59	91	150
	123	166	289

Study $j = 3$ (T26)

	s	f	n
A15mg	51	86	137
PBO	49	95	144
	100	181	281

Table A.3.: Meta-analysis: 2×2-table of patients treated with A15mg

Study $j = 1$ (T14)

	s	f	n
A 20mg	91	55	146
PBO	51	102	153
	142	157	299

Study $j = 2$ (T15)

	s	f	n
A 20mg	120	80	200
PBO	58	131	189
	178	211	389

Study $j = 3$ (T24)

	s	f	n
A 20mg	63	77	140
PBO	59	91	150
	122	168	290

Study $j = 4$ (T25)

	s	f	n
A 20mg	58	83	141
PBO	45	107	152
	103	190	293

Study $j = 5$ (T27)

	s	f	n
A 20mg	75	72	147
PBO	58	88	146
	133	160	293

Table A.4.: Meta-analysis: 2×2-table of patients treated with A20mg

Study $j = 1$ (T12)

	s	f	n
C1	85	58	143
PBO	67	78	145
	152	136	288

Study $j = 2$ (T13)

	s	f	n
C1	99	45	144
PBO	53	89	142
	152	134	286

Study $j = 3$ (T14)

	s	f	n
C1	107	35	142
PBO	51	102	153
	158	137	295

Study $j = 4$ (T22)

	s	f	n
C1	75	66	141
PBO	49	90	139
	124	156	280

Study $j = 5$ (T24)

	s	f	n
C1	80	62	142
PBO	59	91	150
	139	153	292

Table A.5.: Meta-analysis: 2×2-table of patients treated with C1

Study $j = 1$ (T11)

	s	f	n
C2	80	62	142
PBO	59	91	150
	139	153	292

Table A.6.: Meta-analysis: 2×2-table of patients treated with C2

Treatment group: A5mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.6729	-1.3565	0.0819	-1.8272	-0.8857
T12	2	0.4621	0.5733	-1.2560	0.0550	-1.6418	-0.8703
T13	3	0.3732	0.5867	-0.8985	0.0576	-1.2932	-0.5037
T21	4	0.4764	0.4894	-1.2921	0.0288	-1.5711	-1.0131
T22	5	0.3525	0.4400	-0.7928	0.0586	-1.1909	-0.3947
T23	6	0.2409	0.4526	-0.2475	0.0694	-0.6807	0.1858
T27	7	0.3973	0.4965	-0.9695	0.0570	-1.3621	-0.5768

Table A.7.: Meta-analysis: Odds Ratio, antidepressant A5mg

Treatment group: A10mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.6907	-1.3863	0.0877	-1.8735	-0.8992
T12	2	0.4621	0.5714	-1.2549	0.0555	-1.6425	-0.8673
T15	3	0.3069	0.4920	-0.5718	0.0463	-0.9256	-0.2180
T23	4	0.2409	0.4964	-0.2385	0.0691	-0.6709	0.1940
T25	5	0.2961	0.3356	-0.6346	0.0617	-1.0431	-0.2261
T26	6	0.3403	0.3803	-0.7833	0.0608	-1.1889	-0.3776
T27	7	0.3973	0.5442	-0.9773	0.0560	-1.3666	-0.5879

Table A.8.: Meta-analysis: Odds Ratio, antidepressant A10mg

Treatment group: A15mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T14	1	0.3333	0.5816	-0.7201	0.0586	-1.1181	-0.3221
T24	2	0.3933	0.4604	-0.9593	0.0569	-1.3516	-0.5669
T26	3	0.3403	0.3723	-0.7917	0.0622	-1.2019	-0.3816

Table A.9.: Meta-analysis: Odds Ratio, antidepressant A15mg

Treatment group: A20mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T14	1	0.3333	0.6233	-0.7559	0.0586	-1.1540	-0.3578
T15	2	0.3069	0.6000	-0.6124	0.0457	-0.9640	-0.2607
T24	3	0.3933	0.4500	-0.9630	0.0568	-1.3550	-0.5710
T25	4	0.2961	0.4113	-0.5521	0.0609	-0.9578	-0.1463
T27	5	0.3973	0.5102	-0.9698	0.0558	-1.3585	-0.5812

Table A.10.: Meta-analysis: Odds Ratio, antidepressants A20mg

Treatment group: C1

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T12	1	0.4621	0.5944	-1.2706	0.0568	-1.6624	-0.8787
T13	2	0.3732	0.6875	-1.0195	0.0624	-1.4305	-0.6085
T14	3	0.3333	0.7535	-0.9903	0.0673	-1.4171	-0.5635
T22	4	0.3525	0.5319	-0.7824	0.0600	-1.1853	-0.3795
T24	5	0.3933	0.5634	-0.9692	0.0566	-1.3604	-0.5780

Table A.11.: Meta-analysis: Odds Ratio, antidepressant C1

Treatment group: C2

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}$	
T11	1	0.4608	0.7273	-1.4606	0.0853	-1.9409	-0.9802

Table A.12.: Meta-analysis: Odds Ratio, antidepressant C2

Treatment group: A5mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.6729	0.2121	0.0045	0.1019	0.3224
T12	2	0.4621	0.5733	0.1113	0.0033	0.0161	0.2064
T13	3	0.3732	0.5867	0.2134	0.0033	0.1195	0.3074
T21	4	0.4764	0.4894	0.0130	0.0018	-0.0567	0.0827
T22	5	0.3525	0.4400	0.0875	0.0033	-0.0068	0.1818
T23	6	0.2409	0.4526	0.2117	0.0031	0.1195	0.3039
T27	7	0.3973	0.4965	0.0992	0.0034	0.0031	0.1953

Table A.13.: Meta-analysis: Risk Difference, antidepressant A5mg

Treatment group: A10mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.6907	0.2299	0.0046	0.1179	0.3420
T12	2	0.4621	0.5714	0.1094	0.0034	0.0137	0.2050
T15	3	0.3069	0.4920	0.1851	0.0025	0.1035	0.2667
T23	4	0.2409	0.4964	0.2555	0.0032	0.1630	0.3479
T25	5	0.2961	0.3356	0.0395	0.0029	-0.0486	0.1276
T26	6	0.3403	0.3803	0.0400	0.0032	-0.0533	0.1333
T27	7	0.3973	0.5442	0.1470	0.0033	0.0521	0.2418

Table A.14.: Meta-analysis: Risk Difference, antidepressant A10mg

Treatment group: A15mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T14	1	0.3333	0.5816	0.2482	0.0032	0.1555	0.3410
T24	2	0.3933	0.4604	0.0671	0.0034	-0.0285	0.1627
T26	3	0.3403	0.3723	0.0320	0.0033	-0.0620	0.1260

Table A.15.: Meta-analysis: Risk Difference, antidepressant A15mg

Treatment group: A20mg

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T14	1	0.3333	0.6233	0.2900	0.0031	0.1990	0.3810
T15	2	0.3069	0.6000	0.2931	0.0023	0.2138	0.3724
T24	3	0.3933	0.4500	0.0567	0.0034	-0.0387	0.1520
T25	4	0.2961	0.4113	0.1153	0.0031	0.0239	0.2067
T27	5	0.3973	0.5102	0.1129	0.0033	0.0179	0.2080

Table A.16.: Meta-analysis: Risk Difference, antidepressants A20mg

Treatment group: C1

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T12	1	0.4621	0.5944	0.1323	0.0034	0.0364	0.2282
T13	2	0.3732	0.6875	0.3143	0.0031	0.2221	0.4064
T14	3	0.3333	0.7535	0.4202	0.0028	0.3338	0.5066
T22	4	0.3525	0.5319	0.1794	0.0034	0.0834	0.2754
T24	5	0.3933	0.5634	0.1700	0.0033	0.0752	0.2649

Table A.17.: Meta-analysis: Risk Difference antidepressant C1

Treatment group: C2

Study	j	$\tilde{p}_{C_j}$	$\tilde{p}_{T_j}$	$\tilde{\theta}_j$	$\tilde{\sigma}_j^2$	$\mathcal{I}_j$	
T11	1	0.4608	0.7273	0.2665	0.0042	0.1594	0.3736

Table A.18.: Meta-analysis: Risk Difference antidepressant C2

Fixed-effects model

Drug	$\hat{\theta}$	var($\hat{\Theta}$)	$\mathcal{I}$		q	pv	l^2
A5mg	-1.0188	0.0076	-1.1622	-0.8754	14.7534	0.0223	59.33%
A10mg	-0.8175	0.0086	-0.9702	-0.6648	14.3076	0.0264	58.06%
A15mg	-0.8257	0.0197	-1.0566	-0.5947	0.5224	0.7701	0.00%
A20mg	-0.7668	0.0110	-0.9393	-0.5942	2.6977	0.6096	0.00%
C1	-1.0092	0.0121	-1.1900	-0.8285	2.0965	0.7180	0.00%
C2	-1.4606	0.0853	-1.9409	-0.9802	0.0000		

Table A.19.: Meta-analysis: Odds Ratio estimated by the fixed-effects model

Fixed-effects model

Drug	$\hat{\theta}$	var($\hat{\Theta}$)	$\mathcal{I}$		q	pv	l^2
A5mg	0.1200	0.0004	0.0857	0.1543	14.0933	0.0286	57.43%
A10mg	0.1405	0.0005	0.1054	0.1757	13.7101	0.0330	56.24%
A15mg	0.1175	0.0011	0.0632	0.1718	8.3689	0.0152	76.10%
A20mg	0.1841	0.0006	0.1439	0.2242	16.6561	0.0023	75.98%
C1	0.2517	0.0006	0.2102	0.2933	19.2615	0.0007	79.23%
C2	0.2665	0.0042	0.1594	0.3736	0.0000		

Table A.20.: Meta-analysis: Risk Difference estimated by the fixed-effects model

Random-effects model

Drug	$\hat{\theta}$	var($\hat{\Theta}$)	$\mathcal{I}$	
A5mg	-1.0115	0.0086	-1.1641	-0.8590
A10mg	-0.8189	0.0097	-0.9806	-0.6573
A15mg	-0.8257	0.0197	-1.0566	-0.5947
A20mg	-0.7668	0.0110	-0.9393	-0.5942
C1	-1.0092	0.0121	-1.1900	-0.8285
C2	-1.4606	0.0853	-1.9409	-0.9802

Table A.21.: Meta-analysis: Odds Ratio estimated by the random-effects model

Random-effects model

drug	$\hat{\theta}$	var($\hat{\Theta}$)	$\mathcal{I}$	
A5mg	0.1201	0.0004	0.0857	0.1545
A10mg	0.1405	0.0005	0.1053	0.1758
A15mg	0.1175	0.0011	0.0622	0.1727
A20mg	0.1837	0.0006	0.1430	0.2245
C1	0.2513	0.0007	0.2089	0.2938
C2	0.2665	0.0042	0.1594	0.3736

Table A.22.: Meta-analysis: Risk Difference estimated by the random-effects model

B. Clustering

i	1		2		3		4		5		6	
Study $j = 1$ (T11)												
A5mg	67	26	0	7	0	0	0	0	1	1	4	1
PBO	46	36	1	9	0	0	0	0	0	5	0	5
Study $j = 2$ (T12)												
A5mg	0	0	6	11	17	13	53	21	1	11	9	8
PBO	0	0	2	13	17	18	34	30	1	6	13	11
Study $j = 3$ (T13)												
A5mg	0	0	1	9	20	21	56	25	1	3	10	4
PBO	0	0	0	11	13	26	33	39	0	3	7	10
Study $j = 6$ (T21)												
A5mg	0	0	2	24	47	44	71	54	1	11	17	11
PBO	0	0	6	20	44	45	72	58	2	11	7	10
Study $j = 7$ (T22)												
A5mg	0	0	3	14	17	28	33	24	3	8	10	10
PBO	0	0	1	8	19	22	20	41	2	8	7	11
Study $j = 8$ (T23)												
A5mg	53	50	0	5	0	0	0	0	1	2	8	18
PBO	31	75	0	8	0	0	0	0	0	2	2	19
Study $j = 12$ (T27)												
A5mg	0	0	1	7	24	28	33	25	1	5	11	6
PBO	0	0	0	4	22	27	24	32	0	7	12	18

Table B.1.: Cluster-based meta-analysis: 2×2-table of patients treated with A5mg

i	1		2		3		4		5		6	
Study $j = 1$ (T11)												
A10mg	62	18	2	8	0	0	0	0	1	4	2	0
PBO	46	36	1	9	0	0	0	0	0	5	0	5
Study $j = 2$ (T12)												
A10mg	0	0	3	23	11	12	58	19	0	4	12	5
PBO	0	0	2	13	17	18	34	30	1	6	13	11
Study $j = 3$ (T15)												
A10mg	0	0	3	8	15	17	60	54	1	2	13	14
PBO	0	0	4	15	18	15	31	78	0	7	5	16
Study $j = 4$ (T23)												
A10mg	53	52	2	9	0	0	0	0	0	4	13	4
PBO	31	75	0	8	0	0	0	0	0	2	2	19
Study $j = 5$ (T25)												
A10mg	0	0	1	8	19	33	19	37	5	11	6	10
PBO	0	0	2	3	14	40	22	32	3	5	4	27
Study $j = 6$ (T26)												
A10mg	0	0	0	9	21	31	19	19	3	8	11	21
PBO	0	0	1	5	17	44	22	28	4	7	5	11
Study $j = 7$ (T27)												
A10mg	0	0	0	7	28	22	34	23	1	7	17	8
PBO	0	0	0	4	22	27	24	32	0	7	12	18

Table B.2.: Cluster-based meta-analysis: 2×2-table of patients treated with A10mg

i	1		2		3		4		5		6	
Study $j = 4$ (T14)												
A15mg	66	33	1	13	0	0	0	0	0	10	15	3
PBO	42	70	1	11	0	0	0	0	2	6	6	15
Study $j = 9$ (T24)												
A15mg	0	0	2	7	23	33	26	14	4	13	9	8
PBO	0	0	0	11	32	34	21	25	1	9	5	12
Study $j = 11$ (T26)												
A15mg	0	0	2	8	21	40	17	17	3	10	8	11
PBO	0	0	1	5	17	44	22	28	4	7	5	11

Table B.3.: Cluster-based meta-analysis: 2 × 2-table of patients treated with A15mg

i	1		2		3		4		5		6	
Study $j = 4$ (T14)												
A20mg	79	34	1	10	0	0	0	0	0	10	11	1
PBO	42	70	1	11	0	0	0	0	2	6	6	15
Study $j = 5$ (T15)												
A20mg	79	34	1	10	0	0	0	0	0	10	11	1
PBO	42	70	1	11	0	0	0	0	2	6	6	15
Study $j = 9$ (T24)												
A20mg	0	0	4	7	28	22	18	23	3	13	10	12
PBO	0	0	0	11	32	34	21	25	1	9	5	12
Study $j = 10$ (T25)												
A20mg	0	0	2	6	24	33	25	22	3	8	4	14
PBO	0	0	2	3	14	40	22	32	3	5	4	27
Study $j = 12$ (T27)												
A20mg	0	0	1	10	28	16	30	27	0	4	16	15
PBO	0	0	0	4	22	27	24	32	0	7	12	18

Table B.4.: Cluster-based meta-analysis: 2 × 2-table of patients treated with A20mg

i	1		2		3		4		5		6	
Study $j = 1$ (T11)												
C1	0	0	1	10	28	16	30	27	0	4	16	15
PBO	0	0	0	4	22	27	24	32	0	7	12	18
Study $j = 2$ (T12)												
C1	0	0	4	17	14	8	51	22	1	8	15	3
PBO	0	0	2	13	17	18	34	30	1	6	13	11
Study $j = 3$ (T13)												
C1	0	0	2	6	30	9	56	11	2	6	9	13
PBO	0	0	0	11	13	26	33	39	0	3	7	10
Study $j = 7$ (T22)												
C1	0	0	5	13	26	11	32	24	2	11	10	7
PBO	0	0	1	8	19	22	20	41	2	8	7	11
Study $j = 9$ (T24)												
C1	0	0	6	7	40	20	23	16	4	10	7	9
PBO	0	0	0	11	32	34	21	25	1	9	5	12

Table B.5.: Cluster-based meta-analysis: 2×2-table of patients treated with C1

i	1		2		3		4		5		6	
Study $j = 1$ (T11)												
C2	75	13	2	10	0	0	0	0	0	6	3	1
PBO	46	36	1	9	0	0	0	0	0	5	0	5

Table B.6.: Cluster-based meta-analysis: 2×2-table of patients treated with C2

i	Study	j	A5mg	A10mg	A15mg	A20mg	C1	C2	PBO	Total
1	T11	1	93	80				88	82	343
	T14	4			99	113	115		112	439
	T23	8	103	105					106	314
2	T11	1	7	10				12	10	39
	T12	2	17	26			21		15	79
	T13	3	10				8		11	29
	T14	4			14	11	7		12	44
	T15	5		11		14			19	44
	T21	6	26						26	52
	T22	7	17				18		9	44
	T23	8	5	11					8	24
	T24	9			9	11	13		11	44
	T25	10		9		8			5	22
	T26	11		9	10				6	25
	T27	12	8	7		11			4	30
3	T12	2	30	23			22		35	110
	T13	3	41				39		39	119
	T15	5		32		39			33	104
	T21	6	91						89	180
	T22	7	45				37		41	123
	T24	9			56	50	60		66	232
	T25	10		52		57			54	163
	T26	11		52	61				61	174
	T27	12	52	50		44			49	195
4	T12	2	74	77			73		64	288
	T13	3	81				67		72	220
	T15	5		114		115			109	338
	T21	6	125						130	255
	T22	7	57				56		61	174
	T24	9			40	41	39		46	166
	T25	10		56		47			54	157
	T26	11		38	34				50	122
	T27	12	58	57		57			56	228
5	T11	1	2	5				6	5	18
	T12	2	12	4			9		7	32
	T13	3	4				8		3	15
	T14	4			10	10	4		8	32
	T15	5		3		8			7	18
	T21	6	12						13	25
	T22	7	11				13		10	34
	T23	8	3	4					2	9
	T24	9			17	16	14		10	57
	T25	10		16		11			8	35
	T26	11		11	13				11	35
	T27	12	6	8		4			7	25
6	T11	1	5	2				4	5	16
	T12	2	17	17			18		24	76
	T13	3	14				22		17	53
	T14	4			18	12	16		21	67
	T15	5		27		24			21	72
	T21	6	28						17	45
	T22	7	20				17		18	55
	T23	8	26	17					21	64
	T24	9			17	22	16		17	72
	T25	10		16		18			31	65
	T26	11		32	19				16	67
	T27	12	17	25		31			30	103
	Total		1117	1006	417	774	712	110	1874	6010

Table B.7.: Cluster-based meta-analysis: Distribution of patients in study and treatment

i / Treatment	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
A10mg	0.6216	0.1688	0.1325	0.1744	0.4498	0.1493	0.5556	0.1756	0.2157	0.1522	0.5441	0.1801	0.4920	0.1674
A15mg	0.6667	0.0903	0.1515	0.0693	0.3761	0.0836	0.5811	0.0380	0.1750	0.1194	0.5926	0.0715	0.4724	0.0694
A20mg	0.6991	0.1031	0.1818	0.1155	0.5421	0.1357	0.6000	0.1335	0.1429	0.1463	0.4860	0.1417	0.5258	0.1288
A5mg	0.6122	0.1788	0.1444	0.1891	0.4826	0.1850	0.6228	0.2028	0.1800	0.1493	0.5433	0.1682	0.5210	0.1859
C1	0.8087	0.1049	0.2836	0.1408	0.6962	0.1129	0.6894	0.1206	0.2292	0.1433	0.5730	0.1179	0.6264	0.1185
PBO	0.3967	0.2737	0.1324	0.2857	0.4197	0.3336	0.4346	0.3296	0.1648	0.2716	0.3067	0.3152	0.3735	0.3118
C2	0.8523	0.0803	0.1667	0.0252					0.0000	0.0179	0.7500	0.0053	0.7273	0.0183
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.8.: Cluster-based meta analysis: Response rate in clusters classified by treatment

i / Region	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
eu	0.6086		0.1121	0.2248					0.1186	0.1761	0.5034	0.1947	0.5394	0.2344
non eu			0.1789	0.7752	0.4800		0.5524		0.1920	0.8239	0.4605	0.8053	0.4666	0.7656
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.9.: Cluster-based meta analysis: Response rate in clusters classified by region

i / BMI	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
$]0, 22.53]$	0.6136	0.2409	0.1525	0.2479	0.4843	0.1593	0.6079	0.1951	0.1471	0.2030	0.4771	0.2026	0.4992	0.2007
$]22.53, 25.28]$	0.6436	0.2509	0.1759	0.2269	0.4435	0.1643	0.5671	0.1874	0.1212	0.1970	0.4808	0.2066	0.4900	0.1997
$]25.53, 28.39]$	0.5683	0.2473	0.1705	0.1849	0.4848	0.1886	0.5303	0.1946	0.1600	0.1493	0.5101	0.1974	0.4846	0.1998
$]28.39, 32.91]$	0.6256	0.1779	0.1277	0.1975	0.5270	0.2250	0.5377	0.1976	0.2500	0.2149	0.4643	0.1854	0.4913	0.1998
$]32.91, \infty]$	0.5714	0.0830	0.2059	0.1429	0.4565	0.2629	0.5239	0.2254	0.2025	0.2358	0.4140	0.2079	0.4534	0.2000
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.10.: Cluster-based meta analysis: Response rate in clusters classified by BMI

i / MADRS	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
22			1.0000	0.0021					0.0000	0.0030	0.3750	0.0106	0.4000	0.0017
23			0.5000	0.0042	0.7143	0.0050	0.6154	0.0067	0.5000	0.0060			0.6250	0.0040
24			0.2500	0.0084	0.5000	0.0029	0.6667	0.0062					0.5500	0.0033
25	0.0000	0.0009	0.3333	0.0063	0.6000	0.0071	0.6667	0.0046	0.5000	0.0060			0.5600	0.0042
26	0.4595	0.0338	0.1250	0.0504	0.5122	0.0586	0.5922	0.0529	0.2857	0.0627			0.4831	0.0444
27	0.5362	0.0630	0.2083	0.0504	0.4706	0.0607	0.5789	0.0878	0.0000	0.0478			0.4959	0.0607
28	0.5258	0.0885	0.2927	0.0861	0.5643	0.1000	0.5444	0.0924	0.2857	0.0627			0.5136	0.0797
29			0.1212	0.0693					0.0400	0.0746	0.4614	0.5828	0.4177	0.0829
30	0.5821	0.1223	0.1509	0.1113	0.4865	0.1057	0.5405	0.1330	0.1429	0.0836			0.4855	0.1035
31	0.6433	0.1432	0.0952	0.0882	0.4277	0.1236	0.5336	0.1299	0.2821	0.1164			0.4895	0.1105
32	0.6232	0.1259	0.1579	0.0798	0.4324	0.1057	0.5833	0.1109	0.1429	0.0836			0.5035	0.0945
33	0.6742	0.1204	0.1538	0.1092	0.4336	0.1021	0.5138	0.0929	0.1667	0.0896			0.4777	0.0895
34	0.6023	0.0803	0.2353	0.0714	0.4472	0.0879	0.5935	0.0796	0.0800	0.0746			0.4941	0.0707
35	0.5859	0.0903	0.1212	0.0693	0.5327	0.0764	0.5177	0.0724	0.2778	0.0537			0.4950	0.0662
36	0.6724	0.0529	0.2222	0.0567	0.4789	0.0507	0.5577	0.0534	0.1905	0.0627			0.5018	0.0468
37	0.6800	0.0456	0.1250	0.0504	0.4500	0.0571	0.4915	0.0303	0.2222	0.0537	0.5625	0.0212	0.4656	0.0411
38			0.0714	0.0294					0.0000	0.0328	0.4931	0.1907	0.4260	0.0281
39	0.5333	0.0137	0.1111	0.0189	0.5333	0.0321	0.5306	0.0252	0.2727	0.0328			0.4806	0.0215
40			0.0000	0.0126					0.0000	0.0090	0.4688	0.1272	0.4286	0.0175
41	0.9000	0.0091	0.0000	0.0021	0.5000	0.0143	0.5000	0.0113	0.2500	0.0119			0.5439	0.0095
42	0.5000	0.0073	0.0000	0.0084	0.5000	0.0014			0.5000	0.0060	0.4848	0.0437	0.4490	0.0082
43	1.0000	0.0018	0.0000	0.0063	0.7143	0.0050	0.6471	0.0087	0.0000	0.0119			0.5455	0.0055
44			0.5000	0.0042					0.0000	0.0090	0.5714	0.0093	0.4167	0.0020
45	1.0000	0.0009			1.0000	0.0007	0.6667	0.0015					0.8000	0.0008
46			0.0000	0.0021							0.3333	0.0079	0.2857	0.0012
47					0.7500	0.0029	0.0000	0.0005	1.0000	0.0030			0.6667	0.0010
48			0.0000	0.0021							0.0000	0.0013	0.0000	0.0003
49									0.0000	0.0030	0.3333	0.0040	0.2500	0.0007
52											0.0000	0.0013	0.0000	0.0002
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.11.: Cluster-based meta analysis: Response rate in clusters classified by MADRS

i / CGI-S	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
3	0.6667	0.0027	0.0000	0.0021			0.6000	0.0026	0.0000	0.0030	0.0000	0.0013	0.4545	0.0018
4	0.5879	0.2856	0.1768	0.4160	0.4758	0.4279	0.5527	0.4579	0.2086	0.4149	0.4508	0.3907	0.4758	0.4053
5	0.5929	0.4909	0.1818	0.4853	0.4798	0.4957	0.5508	0.4594	0.1523	0.4507	0.4632	0.4318	0.4801	0.4717
6	0.6708	0.2190	0.0222	0.0945	0.5047	0.0764	0.5467	0.0770	0.1818	0.1313	0.5194	0.1709	0.5217	0.1190
7	0.5000	0.0018	0.0000	0.0021			0.8333	0.0031			0.7500	0.0053	0.6923	0.0022
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.12.: Cluster-based meta analysis: Response rate in clusters classified by CGI-Score

i / Duration (in days)	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
$]0, 113]$	0.5811	0.2026	0.1829	0.3445			0.5850	0.3835			0.5038	0.1762	0.5237	0.2106
$]113, 142]$	0.5897	0.2646	0.1259	0.2836			0.5276	0.2793			0.4655	0.2305	0.4864	0.1902
$]142, 197]$	0.6643	0.2609	0.1751	0.3718			0.5358	0.3373			0.4600	0.1987	0.5055	0.2113
$]197, 323]$	0.5777	0.1880			0.4943	0.4364			0.1481	0.4836	0.4620	0.2093	0.4556	0.1892
$]323, \infty]$	0.6304	0.0839			0.4689	0.5636			0.2081	0.5164	0.4571	0.1854	0.4422	0.1987
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.13.: Cluster-based meta analysis: Response rate in clusters classified by duration of current episode

i / Age (in years)	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
$]0, 34]$	0.6029	0.1861	0.1984	0.2647	0.5083	0.2164	0.6317	0.2105	0.2692	0.2328	0.5217	0.1828	0.5195	0.2095
$]34, 44]$	0.6272	0.2546	0.1400	0.2101	0.4682	0.1907	0.5250	0.1848	0.1692	0.1940	0.5380	0.2093	0.4874	0.2045
$]44, 51]$	0.5741	0.1971	0.1294	0.1786	0.4847	0.1871	0.5112	0.1838	0.1806	0.2149	0.3553	0.2013	0.4472	0.1905
$]51, 60]$	0.5975	0.2199	0.1609	0.1828	0.4479	0.2264	0.5074	0.2084	0.1429	0.1672	0.5029	0.2318	0.4696	0.2133
$]60, \infty]$	0.6474	0.1423	0.1795	0.1639	0.4940	0.1793	0.5773	0.2125	0.1094	0.1910	0.4167	0.1748	0.4932	0.1822
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.14.: Cluster-based meta analysis: Response rate in clusters classified by age

i / Gender	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
F	0.6245	0.6268	0.1486	0.6786	0.5022	0.6557	0.5641	0.6689	0.1875	0.6209	0.4857	0.6954	0.4962	0.6596
M	0.5819	0.3732	0.1961	0.3214	0.4378	0.3443	0.5287	0.3311	0.1654	0.3791	0.4304	0.3046	0.4594	0.3404
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.15.: Cluster-based meta analysis: Response rate in clusters classified by gender

i / Reason for withdrawal	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
	0.6086				0.4800		0.5524				0.4689		0.5326	0.8651
Adverse Event(s)			0.1606	0.4055					0.1509	0.4746			0.1563	0.0586
Lack of Efficacy			0.0000	0.1912					0.0213	0.1403			0.0072	0.0230
Lost to Follow-up			0.2911	0.1660					0.3014	0.2179			0.2961	0.0253
Withdrawal of Consent			0.2124	0.2374					0.2321	0.1672			0.2189	0.0281
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.16.: Cluster-based meta analysis: Response rate in clusters classified by reason for withdrawal

i / Study	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
T11	0.7289	0.3130	0.1282	0.0819					0.1111	0.0537	0.5625	0.0212	0.6394	0.0692
T12			0.1899	0.1660	0.5364	0.0786	0.6806	0.1478	0.0938	0.0955	0.6447	0.1007	0.5504	0.0973
T13			0.1034	0.0609	0.5294	0.0850	0.6591	0.1129	0.2000	0.0448	0.4906	0.0702	0.5505	0.0725
T14	0.6378	0.4005	0.1136	0.0924					0.1250	0.0955	0.6269	0.0887	0.5687	0.0968
T15			0.2045	0.0924	0.5385	0.0743	0.5148	0.1735	0.1111	0.0537	0.4028	0.0954	0.4688	0.0958
T21			0.1538	0.1092	0.5056	0.1286	0.5608	0.1309	0.1200	0.0746	0.5333	0.0596	0.4829	0.0927
T22			0.2045	0.0924	0.5041	0.0879	0.4885	0.0893	0.2059	0.1015	0.4909	0.0728	0.4419	0.0715
T23	0.4363	0.2865	0.0833	0.0504					0.1111	0.0269	0.3594	0.0848	0.3966	0.0684
T24			0.2727	0.0924	0.5302	0.1657	0.5301	0.0852	0.2105	0.1701	0.4306	0.0954	0.4658	0.0950
T25			0.2273	0.0462	0.3497	0.1164	0.4204	0.0806	0.3143	0.1045	0.2154	0.0861	0.3462	0.0735
T26			0.1200	0.0525	0.3391	0.1243	0.4754	0.0626	0.2857	0.1045	0.3582	0.0887	0.3641	0.0704
T27			0.0667	0.0630	0.5231	0.1393	0.5307	0.1170	0.0800	0.0746	0.5437	0.1364	0.4871	0.0967
Total	0.6086		0.1639		0.4800		0.5524		0.1791		0.4689		0.4837	

Table B.17.: Cluster-based meta analysis: Response rate in clusters classified by study protocol

Treatment group: A5mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.7204	-1.8475	0.1029	-2.3752	-1.3199
	T23	8	0.2925	0.5146	-0.5036	0.0845	-0.9817	-0.0256
2	T12	1	0.1333	0.3529	0.3950	0.8345	-1.1076	1.8976
	T21	6	0.2308	0.0769	-1.4410	0.7583	-2.8734	-0.0086
	T22	7	0.1111	0.1765	0.1507	1.5298	-1.8837	2.1851
3	T12	2	0.4857	0.5667	-1.3471	0.2501	-2.1697	-0.5244
	T13	3	0.3333	0.4878	-0.6937	0.2130	-1.4529	0.0654
	T21	6	0.4944	0.5165	-1.3649	0.0890	-1.8555	-0.8743
	T22	7	0.4634	0.3778	-1.3013	0.1926	-2.0232	-0.5794
	T27	12	0.4490	0.4615	-1.1874	0.1599	-1.8451	-0.5298
4	T12	2	0.5313	0.7162	-1.7185	0.1292	-2.3098	-1.1272
	T13	3	0.4583	0.6914	-1.3776	0.1138	-1.9325	-0.8227
	T21	6	0.5538	0.5680	-1.6212	0.0637	-2.0364	-1.2059
	T22	7	0.3279	0.5789	-0.6937	0.1464	-1.3230	-0.0644
	T27	12	0.4286	0.5690	-1.1178	0.1432	-1.7403	-0.4953
5	T12	2	0.1429	0.0833	-0.7802	2.2576	-3.2516	1.6913
	T21	6	0.1538	0.0833	-0.8672	1.6818	-3.0003	1.2660
	T22	7	0.2000	0.2727	-0.2314	1.0833	-1.9435	1.4806
6	T12	2	0.5417	0.5294	-1.5568	0.4039	-2.6022	-0.5114
	T13	3	0.4118	0.7143	-1.2326	0.5929	-2.4991	0.0339
	T21	6	0.4118	0.6071	-1.0766	0.3926	-2.1072	-0.0460
	T22	7	0.3889	0.5000	-0.9343	0.4338	-2.0176	0.1490
	T23	8	0.0952	0.3077	0.7049	0.7332	-0.7035	2.1133
	T27	12	0.4000	0.6471	-1.0713	0.3965	-2.1070	-0.0356

Table B.18.: Cluster-based meta-analysis: Odds Ratio, antidepressant A5mg

Treatment group: A10mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.7750	-1.9917	0.1212	-2.5643	-1.4190
	T23	3	0.2925	0.5048	-0.5029	0.0837	-0.9787	-0.0270
2	T11	1	0.1000	0.2000	0.3646	1.7361	-1.8026	2.5319
	T12	2	0.1333	0.1154	-0.4103	0.9537	-2.0166	1.1961
	T15	5	0.2105	0.2727	-0.2960	0.7750	-1.7440	1.1521
	T25	10	0.4000	0.1111	-1.9095	1.9583	-4.2114	0.3923
3	T12	2	0.4857	0.4783	-1.3310	0.2886	-2.2147	-0.4474
	T15	5	0.5455	0.4688	-1.5725	0.2477	-2.3912	-0.7539
	T25	10	0.2593	0.3654	-0.4117	0.1794	-1.1083	0.2849
	T26	11	0.2787	0.4038	-0.4730	0.1614	-1.1339	0.1879
	T27	12	0.4490	0.5600	-1.1960	0.1637	-1.8614	-0.5306
4	T12	2	0.5313	0.7532	-1.8079	0.1326	-2.4069	-1.2089
	T15	5	0.2844	0.5263	-0.4663	0.0803	-0.9323	-0.0003
	T25	10	0.4074	0.3393	-1.1207	0.1564	-1.7711	-0.4702
	T26	11	0.4400	0.5000	-1.1451	0.1864	-1.8553	-0.4349
	T27	12	0.4286	0.5965	-1.1366	0.1458	-1.7646	-0.5085
5	T25	10	0.3750	0.3125	-1.0270	0.8242	-2.5203	0.4663
	T26	11	0.3636	0.2727	-1.0581	0.8512	-2.5757	0.4594
6	T12	2	0.5417	0.7059	-1.7391	0.4512	-2.8440	-0.6343
	T15	5	0.2381	0.4815	-0.2245	0.4109	-1.2788	0.8298
	T23	8	0.0952	0.7647	0.5361	0.8796	-1.0065	2.0787
	T25	10	0.1290	0.3750	0.4587	0.5537	-0.7652	1.6827
	T26	11	0.3125	0.3438	-0.7006	0.4294	-1.7785	0.3773
	T27	12	0.4000	0.6800	-1.1196	0.3227	-2.0540	-0.1852

Table B.19.: Cluster-based meta-analysis: Odds Ratio, antidepressant A10mg

Treatment group: A15mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.6667	-0.9933	0.0835	-1.4687	-0.5178
2	T14	4	0.0833	0.0714	-0.3153	2.1678	-2.7371	2.1065
	T26	11	0.1667	0.2000	-0.2231	1.8250	-2.4452	1.9989
3	T24	9	0.4848	0.4107	-1.3581	0.1344	-1.9612	-0.7550
	T26	11	0.2787	0.3443	-0.5374	0.1542	-1.1832	0.1085
4	T24	9	0.4565	0.6500	-1.3063	0.1975	-2.0373	-0.5752
	T26	11	0.4400	0.5000	-1.1451	0.1988	-1.8786	-0.4117
5	T24	9	0.1000	0.2353	0.4820	1.4380	-1.4904	2.4545
	T26	11	0.3636	0.2308	-1.1691	0.8262	-2.6642	0.3260
6	T14	4	0.2857	0.8333	-1.0578	0.6333	-2.3668	0.2512
	T24	9	0.2941	0.5294	-0.5143	0.5194	-1.6998	0.6712
	T26	11	0.3125	0.4211	-0.6231	0.5068	-1.7941	0.5479

Table B.20.: Cluster-based meta-analysis: Odds Ratio, antidepressant A15mg

Treatment group: A20mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	1	0.3750	0.6991	-1.0481	0.0802	-1.5139	-0.5824
2	T14	4	0.0833	0.0909	-0.0953	2.1909	-2.5300	2.3394
	T15	5	0.2105	0.1429	-0.7783	0.9000	-2.3388	0.7821
	T25	10	0.4000	0.2500	-1.2685	1.5000	-3.2830	0.7460
3	T15	5	0.5455	0.5897	-1.6014	0.2282	-2.3871	-0.8156
	T24	9	0.4848	0.5600	-1.3402	0.1418	-1.9596	-0.7207
	T25	10	0.2593	0.4211	-0.3617	0.1684	-1.0367	0.3133
	T27	12	0.4490	0.6364	-1.2588	0.1807	-1.9580	-0.5596
4	T15	5	0.2844	0.7217	-0.6826	0.0884	-1.1716	-0.1936
	T24	9	0.4565	0.4390	-1.2269	0.1867	-1.9376	-0.5163
	T25	10	0.4074	0.5319	-1.0157	0.1622	-1.6780	-0.3533
	T27	12	0.4286	0.5263	-1.1014	0.1433	-1.7240	-0.4788
5	T24	9	0.1000	0.1875	0.3156	1.5214	-1.7132	2.3444
	T25	10	0.3750	0.2727	-1.1069	0.9917	-2.7449	0.5311
6	T14	4	0.2857	0.9167	-1.6556	1.3242	-3.5485	0.2372
	T15	5	0.2381	0.4583	-0.2301	0.4303	-1.3091	0.8489
	T24	9	0.2941	0.4545	-0.5191	0.4667	-1.6428	0.6045
	T25	10	0.1290	0.2222	0.1542	0.6085	-1.1289	1.4372
	T27	12	0.4000	0.5161	-0.9819	0.2681	-1.8335	-0.1303

Table B.21.: Cluster-based meta-analysis: Odds Ratio, antidepressant A20mg

Treatment group: C1

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.8087	-1.3554	0.0943	-1.8605	-0.8503
2	T12	2	0.1333	0.1905	0.0023	0.8857	-1.5458	1.5503
	T14	4	0.0833	0.2857	0.8087	1.7909	-1.3926	3.0099
	T22	7	0.1111	0.2778	0.4731	1.4019	-1.4745	2.4206
3	T12	2	0.4857	0.6364	-1.4064	0.3108	-2.3234	-0.4894
	T13	3	0.3333	0.7692	-1.0356	0.2598	-1.8740	-0.1971
	T22	7	0.4634	0.7027	-1.4192	0.2275	-2.2037	-0.6348
	T24	9	0.4848	0.6667	-1.4435	0.1357	-2.0493	-0.8376
4	T12	2	0.5313	0.6986	-1.6832	0.1278	-2.2713	-1.0952
	T13	3	0.4583	0.8358	-1.8191	0.1647	-2.4866	-1.1515
	T22	7	0.3279	0.5714	-0.6891	0.1473	-1.3204	-0.0578
	T24	9	0.4565	0.5897	-1.2447	0.1936	-1.9684	-0.5210
5	T12	2	0.1429	0.1111	-0.5232	2.2917	-3.0133	1.9668
	T14	4	0.2500	0.5000	-0.2877	1.6667	-2.4112	1.8358
	T22	7	0.2000	0.1538	-0.6526	1.2159	-2.4663	1.1612
	T24	9	0.1000	0.2857	0.6080	1.4611	-1.3803	2.5962
6	T12	2	0.5417	0.8333	-2.1411	0.5678	-3.3806	-0.9017
	T13	3	0.4118	0.4091	-1.0632	0.4309	-2.1430	0.0165
	T14	4	0.2857	0.6250	-0.5345	0.5000	-1.6976	0.6285
	T22	7	0.3889	0.5882	-0.9659	0.4766	-2.1015	0.1696
	T24	9	0.2941	0.4375	-0.5266	0.5373	-1.7323	0.6791

Table B.22.: Cluster-based meta-analysis: Odds Ratio, antidepressant C1

Treatment group: C2

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.8523	-2.3174	0.1398	-2.9323	-1.7024
2	T11	1	0.1000	0.1667	0.2231	1.7111	-1.9285	2.3748

Table B.23.: Cluster-based meta-analysis: Odds Ratio, antidepressant C2

Treatment group: A5mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.7204	0.1595	0.0052	0.0412	0.2777
	T23	8	0.2925	0.5146	0.2221	0.0044	0.1133	0.3309
2	T11	1	0.1000	0.0000	-0.1000	0.0090	-0.2560	0.0560
	T12	2	0.1333	0.3529	0.2196	0.0211	-0.0195	0.4587
	T13	3	0.0000	0.1000	0.1000	0.0090	-0.0560	0.2560
	T21	6	0.2308	0.0769	-0.1538	0.0096	-0.3147	0.0070
	T22	7	0.1111	0.1765	0.0654	0.0195	-0.1645	0.2952
	T27	12	0.0000	0.1250	0.1250	0.0137	-0.0673	0.3173
3	T12	2	0.4857	0.5667	0.0810	0.0153	-0.1227	0.2846
	T13	3	0.3333	0.4878	0.1545	0.0118	-0.0241	0.3331
	T21	6	0.4944	0.5165	0.0221	0.0056	-0.1005	0.1447
	T22	7	0.4634	0.3778	-0.0856	0.0113	-0.2604	0.0891
	T27	12	0.4490	0.4615	0.0126	0.0098	-0.1505	0.1756
4	T12	2	0.5313	0.7162	0.1850	0.0066	0.0510	0.3190
	T13	3	0.4583	0.6914	0.2330	0.0061	0.1047	0.3613
	T21	6	0.5538	0.5680	0.0142	0.0039	-0.0881	0.1164
	T22	7	0.3279	0.5789	0.2511	0.0079	0.1050	0.3972
	T27	12	0.4286	0.5690	0.1404	0.0086	-0.0122	0.2929
5	T11	1	0.0000	0.5000	0.5000	0.1250	-0.0815	1.0815
	T12	2	0.1429	0.0833	-0.0595	0.0239	-0.3136	0.1945
	T13	3	0.0000	0.2500	0.2500	0.0469	-0.1061	0.6061
	T21	6	0.1538	0.0833	-0.0705	0.0164	-0.2810	0.1400
	T22	7	0.2000	0.2727	0.0727	0.0340	-0.2307	0.3762
	T23	8	0.0000	0.3333	0.3333	0.0741	-0.1143	0.7810
	T27	12	0.0000	0.1667	0.1667	0.0231	-0.0836	0.4169
6	T11	1	0.0000	0.8000	0.8000	0.0320	0.5058	1.0942
	T12	2	0.5417	0.5294	-0.0123	0.0250	-0.2723	0.2478
	T13	3	0.4118	0.7143	0.3025	0.0288	0.0233	0.5818
	T21	6	0.4118	0.6071	0.1954	0.0228	-0.0528	0.4436
	T22	7	0.3889	0.5000	0.1111	0.0257	-0.1526	0.3748
	T23	8	0.0952	0.3077	0.2125	0.0123	0.0301	0.3948
	T27	12	0.4000	0.6471	0.2471	0.0214	0.0062	0.4879

Table B.24.: Cluster-based meta-analysis: Risk Difference, antidepressant A5mg

Treatment group: A10mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.7750	0.2140	0.0052	0.0956	0.3324
	T23	8	0.2925	0.5048	0.2123	0.0043	0.1040	0.3206
2	T11	1	0.1000	0.2000	0.1000	0.0250	-0.1601	0.3601
	T12	2	0.1333	0.1154	-0.0179	0.0116	-0.1953	0.1594
	T15	5	0.2105	0.2727	0.0622	0.0268	-0.2070	0.3314
	T23	8	0.0000	0.1818	0.1818	0.0135	-0.0095	0.3731
	T25	10	0.4000	0.1111	-0.2889	0.0590	-0.6883	0.1106
	T26	11	0.1667	0.0000	-0.1667	0.0231	-0.4169	0.0836
3	T12	2	0.4857	0.4783	-0.0075	0.0180	-0.2280	0.2131
	T15	5	0.5455	0.4688	-0.0767	0.0153	-0.2801	0.1267
	T25	10	0.2593	0.3654	0.1061	0.0080	-0.0411	0.2534
	T26	11	0.2787	0.4038	0.1252	0.0079	-0.0213	0.2716
	T27	12	0.4490	0.5600	0.1110	0.0100	-0.0533	0.2753
4	T12	2	0.5313	0.7532	0.2220	0.0063	0.0914	0.3526
	T15	5	0.2844	0.5263	0.2419	0.0041	0.1372	0.3466
	T25	10	0.4074	0.3393	-0.0681	0.0085	-0.2195	0.0833
	T26	11	0.4400	0.5000	0.0600	0.0115	-0.1164	0.2364
	T27	12	0.4286	0.5965	0.1679	0.0086	0.0154	0.3204
5	T11	1	0.0000	0.2000	0.2000	0.0320	-0.0942	0.4942
	T12	2	0.1429	0.0000	-0.1429	0.0175	-0.3604	0.0747
	T15	5	0.0000	0.3333	0.3333	0.0741	-0.1143	0.7810
	T25	10	0.3750	0.3125	-0.0625	0.0427	-0.4025	0.2775
	T26	11	0.3636	0.2727	-0.0909	0.0391	-0.4160	0.2342
	T27	12	0.0000	0.1250	0.1250	0.0137	-0.0673	0.3173
6	T11	1	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000
	T12	2	0.5417	0.7059	0.1642	0.0226	-0.0828	0.4113
	T15	5	0.2381	0.4815	0.2434	0.0179	0.0234	0.4634
	T23	8	0.0952	0.7647	0.6695	0.0147	0.4701	0.8688
	T25	10	0.1290	0.3750	0.2460	0.0183	0.0236	0.4683
	T26	11	0.3125	0.3438	0.0313	0.0205	-0.2041	0.2666
	T27	12	0.4000	0.6800	0.2800	0.0167	0.0674	0.4926

Table B.25.: Cluster-based meta-analysis: Risk Difference, antidepressant A10mg

Treatment group: A15mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.6667	0.2917	0.0043	0.1833	0.4000
2	T14	4	0.0833	0.0714	-0.0119	0.0111	-0.1852	0.1614
	T24	9	0.0000	0.2222	0.2222	0.0192	-0.0057	0.4502
	T26	11	0.1667	0.2000	0.0333	0.0391	-0.2921	0.3588
3	T24	9	0.4848	0.4107	-0.0741	0.0081	-0.2222	0.0740
	T26	11	0.2787	0.3443	0.0656	0.0070	-0.0720	0.2032
4	T24	9	0.4565	0.6500	0.1935	0.0111	0.0203	0.3666
	T26	11	0.4400	0.5000	0.0600	0.0123	-0.1223	0.2423
5	T14	4	0.2500	0.0000	-0.2500	0.0234	-0.5018	0.0018
	T24	9	0.1000	0.2353	0.1353	0.0196	-0.0949	0.3655
	T26	11	0.3636	0.2308	-0.1329	0.0347	-0.4392	0.1735
6	T14	4	0.2857	0.8333	0.5476	0.0174	0.3304	0.7648
	T24	9	0.2941	0.5294	0.2353	0.0269	-0.0343	0.5049
	T26	11	0.3125	0.4211	0.1086	0.0263	-0.1580	0.3751

Table B.26.: Cluster-based meta-analysis: Risk Difference, antidepressant A15mg

Treatment group: A20mg

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.6991	0.3241	0.0040	0.2207	0.4275
2	T14	4	0.0833	0.0909	0.0076	0.0139	-0.1862	0.2014
	T15	5	0.2105	0.1429	-0.0677	0.0175	-0.2852	0.1499
	T24	9	0.0000	0.3636	0.3636	0.0210	0.1251	0.6022
	T25	10	0.4000	0.2500	-0.1500	0.0714	-0.5896	0.2896
	T27	12	0.0000	0.0909	0.0909	0.0075	-0.0517	0.2335
3	T15	5	0.5455	0.5897	0.0443	0.0137	-0.1484	0.2369
	T24	9	0.4848	0.5600	0.0752	0.0087	-0.0784	0.2287
	T25	10	0.2593	0.4211	0.1618	0.0078	0.0162	0.3074
	T27	12	0.4490	0.6364	0.1874	0.0103	0.0204	0.3544
4	T15	5	0.2844	0.7217	0.4373	0.0036	0.3385	0.5362
	T24	9	0.4565	0.4390	-0.0175	0.0114	-0.1931	0.1581
	T25	10	0.4074	0.5319	0.1245	0.0098	-0.0381	0.2871
	T26	11	0.4400	0.0000	0.0000	0.0000	0.0000	0.0000
	T27	12	0.4286	0.5263	0.0977	0.0087	-0.0561	0.2516
5	T14	4	0.2500	0.0000	-0.2500	0.0234	-0.5018	0.0018
	T15	5	0.0000	0.1250	0.1250	0.0137	-0.0673	0.3173
	T24	9	0.1000	0.1875	0.0875	0.0185	-0.1364	0.3114
	T25	10	0.3750	0.2727	-0.1023	0.0473	-0.4601	0.2556
6	T14	4	0.2857	0.9167	0.6310	0.0161	0.4223	0.8396
	T15	5	0.2381	0.4583	0.2202	0.0190	-0.0064	0.4469
	T24	9	0.2941	0.4545	0.1604	0.0235	-0.0916	0.4125
	T25	10	0.1290	0.2222	0.0932	0.0132	-0.0960	0.2824
	T27	12	0.4000	0.5161	0.1161	0.0161	-0.0923	0.3246

Table B.27.: Cluster-based meta-analysis: Risk Difference, antidepressant A20mg

Treatment group: C1

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T14	4	0.3750	0.8087	0.4337	0.0034	0.3373	0.5301
2	T12	2	0.1333	0.1905	0.0571	0.0150	-0.1446	0.2589
	T13	3	0.0000	0.2500	0.2500	0.0234	-0.0018	0.5018
	T14	4	0.0833	0.2857	0.2024	0.0355	-0.1076	0.5124
	T22	7	0.1111	0.2778	0.1667	0.0221	-0.0780	0.4113
	T24	9	0.0000	0.4615	0.4615	0.0191	0.2341	0.6890
3	T12	2	0.4857	0.6364	0.1506	0.0177	-0.0679	0.3692
	T13	3	0.3333	0.7692	0.4359	0.0102	0.2694	0.6024
	T22	7	0.4634	0.7027	0.2393	0.0117	0.0613	0.4173
	T24	9	0.4848	0.6667	0.1818	0.0075	0.0395	0.3242
4	T12	2	0.5313	0.6986	0.1674	0.0068	0.0320	0.3028
	T13	3	0.4583	0.8358	0.3775	0.0055	0.2555	0.4994
	T22	7	0.3279	0.5714	0.2436	0.0080	0.0966	0.3905
	T24	9	0.4565	0.5897	0.1332	0.0116	-0.0439	0.3104
5	T12	2	0.1429	0.1111	-0.0317	0.0285	-0.3093	0.2458
	T13	3	0.0000	0.2500	0.2500	0.0234	-0.0018	0.5018
	T14	4	0.2500	0.5000	0.2500	0.0859	-0.2322	0.7322
	T22	7	0.2000	0.1538	-0.0462	0.0260	-0.3114	0.2191
	T24	9	0.1000	0.2857	0.1857	0.0236	-0.0669	0.4383
6	T12	2	0.5417	0.8333	0.2917	0.0181	0.0706	0.5127
	T13	3	0.4118	0.4091	-0.0027	0.0252	-0.2640	0.2586
	T14	4	0.2857	0.6250	0.3393	0.0244	0.0825	0.5960
	T22	7	0.3889	0.5882	0.1993	0.0275	-0.0732	0.4719
	T24	9	0.2941	0.4375	0.1434	0.0276	-0.1298	0.4166

Table B.28.: Cluster-based meta-analysis: Risk Difference, antidepressant C1

Treatment group: C2

i	Study	j	$\tilde{p}^i_{C_j}$	$\tilde{p}^i_{T_j}$	$\tilde{\theta}^i_j$	$(\tilde{\sigma}^i_j)^2$	$\mathcal{I}^i_j$	
1	T11	1	0.5610	0.8523	0.2913	0.0044	0.1818	0.4008
2	T11	1	0.1000	0.1667	0.0667	0.0206	-0.1693	0.3026
6	T11	1	0.0000	0.7500	0.7500	0.0469	0.3939	1.1061

Table B.29.: Cluster-based meta-analysis: Risk Difference, antidepressant C2

Cluster-based fixed-effects model

OR	i	$\tilde{\theta}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$		q^i	pv^i	l^2
A5mg	1	-1.1094	0.0464	-1.4637	-0.7552	9.6393	0.0019	52.47%
	2	-0.4190	0.3154	-1.3427	0.5048	2.3836	0.3037	49.45%
	3	-1.2163	0.0319	-1.5100	-0.9227	1.6413	0.8013	0.00%
	4	-1.3770	0.0217	-1.6195	-1.1346	5.4970	0.2400	46.30%
	5	-0.5482	0.5100	-1.7229	0.6265	0.1769	0.9153	0.00%
	6	-0.9743	0.0775	-1.4323	-0.5164	4.8525	0.4341	0.00%
A10mg	1	-1.1110	0.0495	-1.4770	-0.7450	10.8178	0.0010	8.48%
	2	-0.4604	0.2919	-1.3491	0.4283	1.5019	0.6818	0.00%
	3	-0.9256	0.0394	-1.2521	-0.5991	5.4471	0.2444	0.00%
	4	-1.0500	0.0259	-1.3147	-0.7853	8.7072	0.0688	46.70%
	5	-1.0423	0.4187	-2.1067	0.0221	0.0006	0.9808	0.00%
	6	-0.6193	0.0768	-1.0752	-0.1635	7.5666	0.1818	38.63%
A15mg	1	-0.9933	0.0835	-1.4687	-0.5178	0.0000		
	2	-0.2653	0.9908	-1.9026	1.3721	0.0021	0.9632	0.00%
	3	-0.9758	0.0718	-1.4166	-0.5350	2.3338	0.1266	18.68%
	4	-1.2260	0.0991	-1.7437	-0.7082	0.0655	0.7980	0.00%
	5	-0.5666	0.5247	-1.7581	0.6249	1.2040	0.2725	12.26%
	6	-0.7102	0.1826	-1.4130	-0.0073	0.2796	0.8695	0.00%
A20mg	1	-1.0481	0.0802	-1.5139	-0.5824	0.0000		
	2	-0.7850	0.4476	-1.8855	0.3154	0.3730	0.8299	0.00%
	3	-1.1168	0.0437	-1.4605	-0.7731	4.8780	0.1810	0.00%
	4	-0.9473	0.0335	-1.2485	-0.6461	1.4065	0.7040	64.92%
	5	-0.5456	0.6003	-1.8200	0.7289	0.8052	0.3695	0.00%
	6	-0.5952	0.0944	-1.1005	-0.0899	2.6519	0.6177	3.91%
C1	1	-1.3554	0.0943	-1.8605	-0.8503	0.0000		
	2	0.3297	0.4165	-0.7319	1.3913	0.2638	0.8764	0.00%
	3	-1.3481	0.0531	-1.7271	-0.9691	0.4762	0.9241	20.03%
	4	-1.3664	0.0387	-1.6899	-1.0429	5.2204	0.1564	20.27%
	5	-0.2051	0.3932	-1.2365	0.8264	0.6654	0.8813	0.00%
	6	-1.0272	0.0996	-1.5463	-0.5081	3.1480	0.5334	0.00%
C2	1	-2.3174	0.1398	-2.9323	-1.7024	0.0000		
	2	0.2231	1.7111	-1.9285	2.3748	0.0000		

Table B.30.: Cluster-based meta-analysis: Odds Ratio estimated by the cluster-based fixed-effects model

Cluster-based fixed-effects model

RD	i	$\tilde{\theta}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$		q^i	pv^i	l^2
A5mg	1	0.1934	0.0024	0.1133	0.2735	0.4112	0.5213	52.47%
	2	0.0136	0.0020	-0.0601	0.0873	8.2494	0.0162	49.45%
	3	0.0308	0.0019	-0.0413	0.1029	2.7097	0.6075	0.00%
	4	0.1444	0.0012	0.0869	0.2020	7.3741	0.1174	46.30%
	5	0.0750	0.0046	-0.0363	0.1863	5.4136	0.0667	0.00%
	6	0.2422	0.0032	0.1498	0.3346	13.2773	0.0209	0.00%
A10mg	1	0.2131	0.0024	0.1332	0.2930	0.0003	0.9860	8.48%
	2	0.0206	0.0034	-0.0748	0.1160	5.5056	0.1383	0.00%
	3	0.0736	0.0021	-0.0021	0.1493	2.4500	0.6536	0.00%
	4	0.1536	0.0014	0.0926	0.2146	9.2524	0.0551	46.70%
	5	0.0354	0.0045	-0.0745	0.1453	5.0814	0.0242	0.00%
	6	0.2961	0.0030	0.2058	0.3864	13.9964	0.0156	38.63%
A15mg	1	0.2917	0.0043	0.1833	0.4000	0.0000		
	2	0.0677	0.0060	-0.0593	0.1947	1.8442	0.1745	0.00%
	3	0.0009	0.0038	-0.0999	0.1017	1.2924	0.2556	18.68%
	4	0.1302	0.0058	0.0046	0.2557	0.7626	0.3825	0.00%
	5	-0.0619	0.0082	-0.2105	0.0867	3.6403	0.0564	12.26%
	6	0.3339	0.0075	0.1911	0.4768	4.9158	0.0856	0.00%
A20mg	1	0.3241	0.0040	0.2207	0.4275	0.0000		
	2	0.0740	0.0031	-0.0174	0.1654	6.1933	0.0452	0.00%
	3	0.1229	0.0024	0.0419	0.2039	1.3087	0.7271	0.00%
	4	0.2468	0.0017	0.1785	0.3150	20.2449	0.0002	64.92%
	5	0.0054	0.0052	-0.1136	0.1245	4.4382	0.0351	0.00%
	6	0.2436	0.0034	0.1479	0.3393	12.3744	0.0148	3.91%
C1	1	0.4337	0.0034	0.3373	0.5301	0.0000		
	2	0.2208	0.0043	0.1134	0.3281	4.9901	0.0825	0.00%
	3	0.2567	0.0027	0.1715	0.3418	4.5447	0.2083	20.03%
	4	0.2502	0.0018	0.1795	0.3209	5.1457	0.1614	20.27%
	5	0.1090	0.0059	-0.0170	0.2351	2.9505	0.3993	0.00%
	6	0.2033	0.0048	0.0894	0.3171	3.0031	0.5573	0.00%
C2	1	0.2913	0.0044	0.1818	0.4008	0.0000		
	2	0.0667	0.0206	-0.1693	0.3026	0.0000		

Table B.31.: Cluster-based meta-analysis: Risk Difference estimated by the cluster-based fixed-effects model

Cluster-based random-effects model

OR	i	$\tilde{\theta}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$	
A5mg	1	-1.1673	0.3743	-2.1736	-0.1610
	2	-0.4190	0.3154	-1.3427	0.5048
	3	-1.2163	0.0319	-1.5100	-0.9227
	4	-1.3757	0.0221	-1.6203	-1.1310
	5	-0.5482	0.5100	-1.7229	0.6265
	6	-0.9743	0.0775	-1.4323	-0.5164
A10mg	1	-1.2347	0.5569	-2.4622	-0.0073
	2	-0.4604	0.2919	-1.3491	0.4283
	3	-0.9272	0.0405	-1.2581	-0.5962
	4	-1.0688	0.0312	-1.3593	-0.7783
	5	-1.0423	0.4187	-2.1067	0.0221
	6	-0.1292	0.0170	-0.3435	0.0851
A15mg	1	-0.9933	0.0835	-1.4687	-0.5178
	2	-0.2653	0.9908	-1.9026	1.3721
	3	-0.9700	0.0904	-1.4646	-0.4755
	4	-1.2260	0.0991	-1.7437	-0.7082
	5	-0.5666	0.5247	-1.7581	0.6249
	6	-0.7102	0.1826	-1.4130	-0.0073
A20mg	1	-1.0481	0.0802	-1.5139	-0.5824
	2	-0.7850	0.4476	-1.8855	0.3154
	3	-1.1178	0.0468	-1.4735	-0.7620
	4	-0.9473	0.0335	-1.2485	-0.6461
	5	-0.5456	0.6003	-1.8200	0.7289
	6	-0.5952	0.0944	-1.1005	-0.0899
C1	1	-1.3554	0.0943	-1.8605	-0.8503
	2	0.3297	0.4165	-0.7319	1.3913
	3	-1.3481	0.0531	-1.7271	-0.9691
	4	-1.3657	0.0421	-1.7031	-1.0282
	5	-0.2051	0.3932	-1.2365	0.8264
	6	-1.0272	0.0996	-1.5463	-0.5081
C2	1	-2.3174	0.1398	-2.9323	-1.7024
	2	0.2231	1.7111	-1.9285	2.3748

Table B.32.: Cluster-based meta-analysis: Odds Ratio estimated by the cluster-based random-effects model

Cluster-based random-effects model

RD	i	$\tilde{\theta}^i$	$(\tilde{\sigma}^i)^2$	$\mathcal{I}^i$	
A5mg	1	0.1934	0.0024	0.1133	0.2735
	2	0.0002	0.0000	-0.0088	0.0092
	3	0.0308	0.0019	-0.0413	0.1029
	4	0.1446	0.0012	0.0869	0.2023
	5	0.0750	0.0046	-0.0363	0.1863
	6	0.2428	0.0033	0.1487	0.3369
A10mg	1	0.2131	0.0024	0.1332	0.2930
	2	0.0206	0.0034	-0.0748	0.1160
	3	0.0736	0.0021	-0.0021	0.1493
	4	0.1532	0.0014	0.0917	0.2147
	5	0.0354	0.0045	-0.0745	0.1453
	6	0.8514	0.0007	0.8090	0.8938
A15mg	1	0.2917	0.0043	0.1833	0.4000
	2	0.0677	0.0060	-0.0593	0.1947
	3	0.0009	0.0038	-0.1000	0.1017
	4	0.1302	0.0058	0.0046	0.2557
	5	-0.0623	0.0083	-0.2122	0.0876
	6	0.3320	0.0079	0.1855	0.4785
A20mg	1	0.3241	0.0040	0.2207	0.4275
	2	0.0740	0.0031	-0.0178	0.1657
	3	0.1229	0.0024	0.0419	0.2039
	4	0.2243	0.0023	0.1456	0.3031
	5	0.0000	0.0000	-0.0053	0.0053
	6	0.2439	0.0037	0.1445	0.3433
C1	1	0.4337	0.0034	0.3373	0.5301
	2	0.2208	0.0043	0.1134	0.3282
	3	0.2567	0.0027	0.1714	0.3420
	4	0.2502	0.0019	0.1793	0.3210
	5	0.1090	0.0059	-0.0170	0.2351
	6	0.2033	0.0048	0.0894	0.3171
C2	1	0.2913	0.0044	0.1818	0.4008
	2	0.0667	0.0206	-0.1693	0.3026

Table B.33.: Cluster-based meta-analysis: Risk Difference estimated by the cluster-based random-effects model

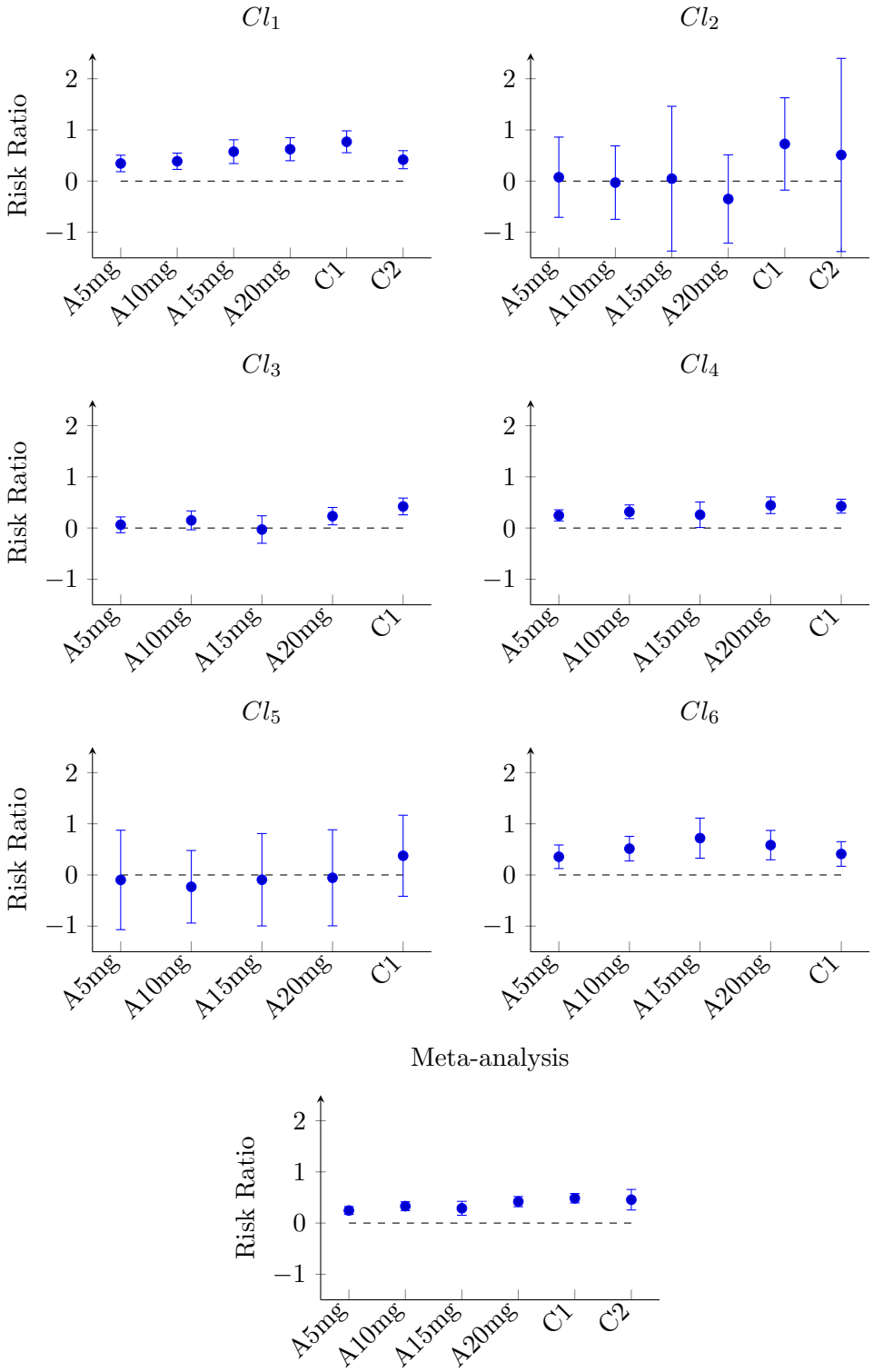

Figure B.1.: Cluster-based meta-analysis: Confidence intervals of Risk Ratio estimated by the fixed-effects model and the cluster-based fixed-effects model

Drug	i	$\tilde{\theta}^i$(drug,C1)	$(\tilde{\sigma}^i)^2$(drug,C1)
A5mg	1	-0.4238	0.0265
	2	-0.6504	0.5278
	3	-0.3583	0.0187
	4	-0.1800	0.0112
	5	-0.4726	0.5806
	6	-0.0548	0.0409
A10mg	1	-0.3806	0.0263
	2	-0.7555	0.4909
	3	-0.2715	0.0225
	4	-0.1099	0.0134
	5	-0.6065	0.4176
	6	**0.1034**	0.0427
A15mg	1	-0.1931	0.0369
	2	-0.6786	1.0399
	3	-0.4510	0.0365
	4	-0.1681	0.0297
	5	-0.4687	0.5327
	6	**0.3095**	0.0781
A20mg	1	-0.1456	0.0356
	2	-1.0758	0.5764
	3	-0.1892	0.0206
	4	**0.0173**	0.0166
	5	-0.4306	0.5572
	6	**0.1732**	0.0517

Table B.34.: Cluster-based meta-analysis fixed-effects model: Indirect comparison of the Risk Ratios of medication A5mg, A10mg, A15mg and A20mg with C1

Drug	$\tilde{\theta}$(drug,C1)	$\tilde{\sigma}^2$(drug,C1)
A5mg	-0.2396	0.0052
A10mg	-0.1549	0.0058
A15mg	-0.1960	0.0099
A20mg	-0.0656	0.0067

Table B.35.: Meta-analysis fixed-effects model: Indirect comparison of the treatment effects of medication A5mg, A10mg, A15mg and A20mg with C1

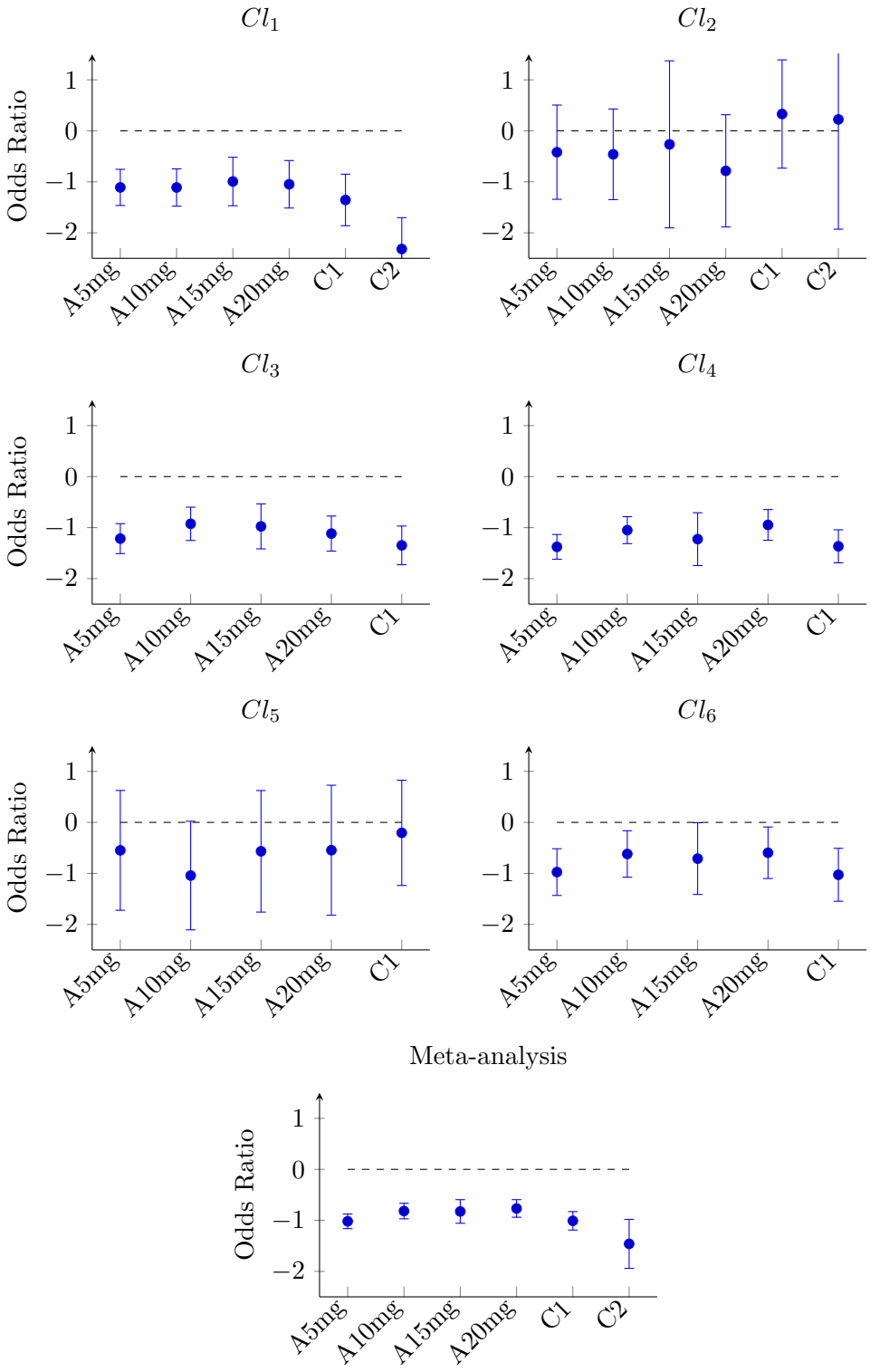

Figure B.2.: Confidence intervals of Odds Ratio estimated by the fixed-effects model and the cluster-based fixed-effects model

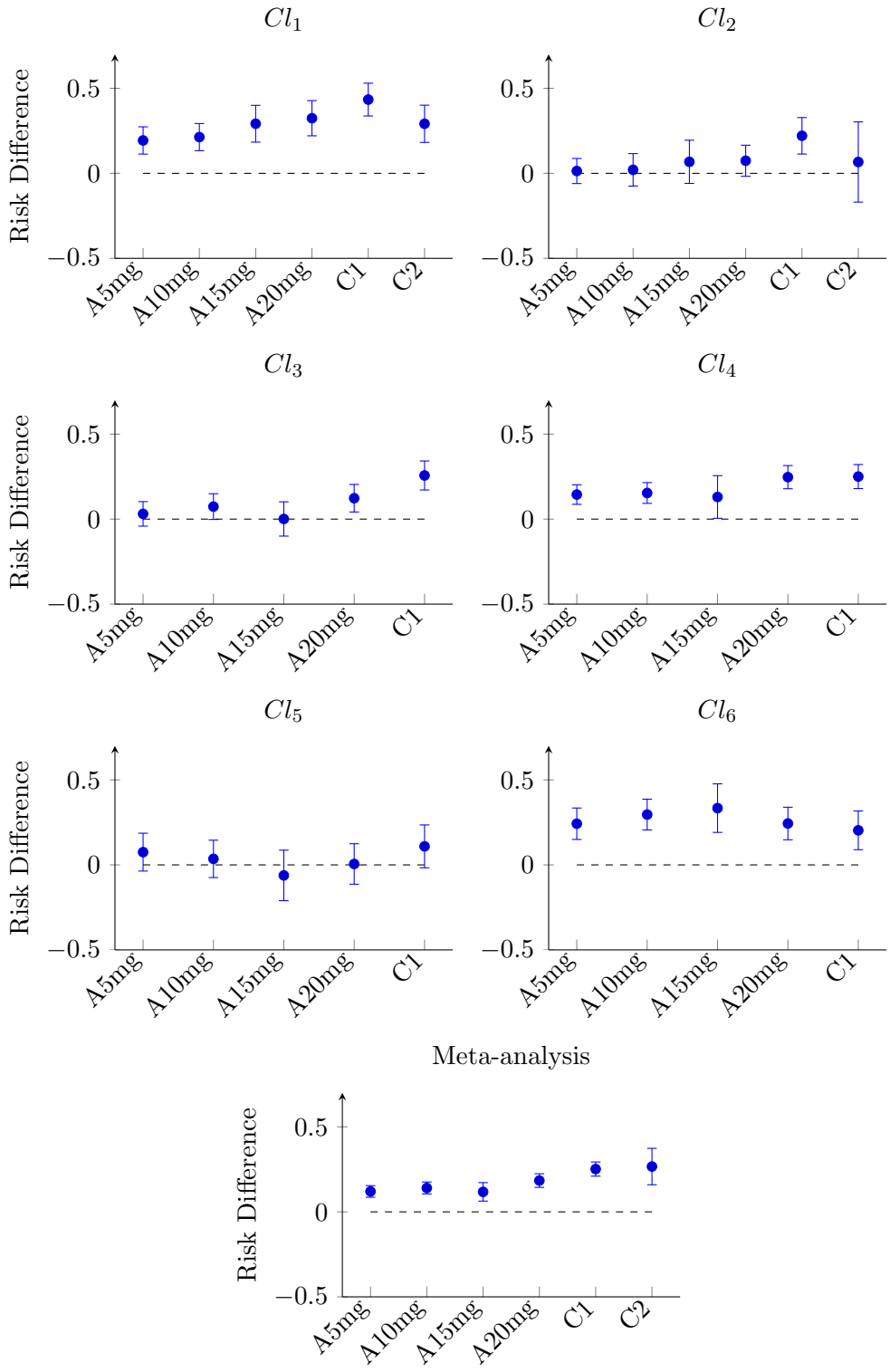

Figure B.3.: Confidence intervals of Risk Difference estimated by the fixed-effects model and the cluster-based fixed-effects model

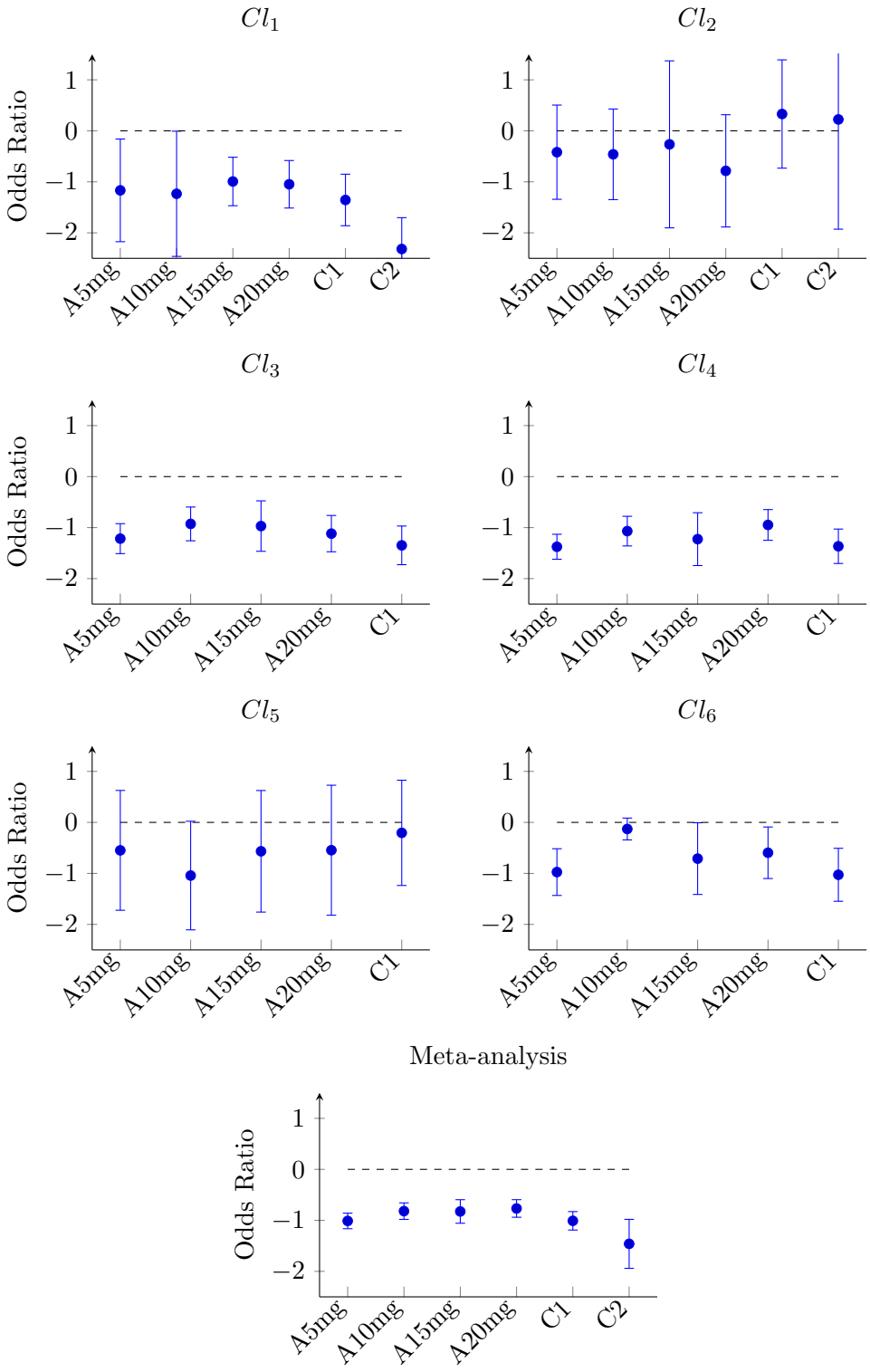

Figure B.4.: Confidence intervals of Odds Ratio estimated by the random-effects model and the cluster-based random-effects model

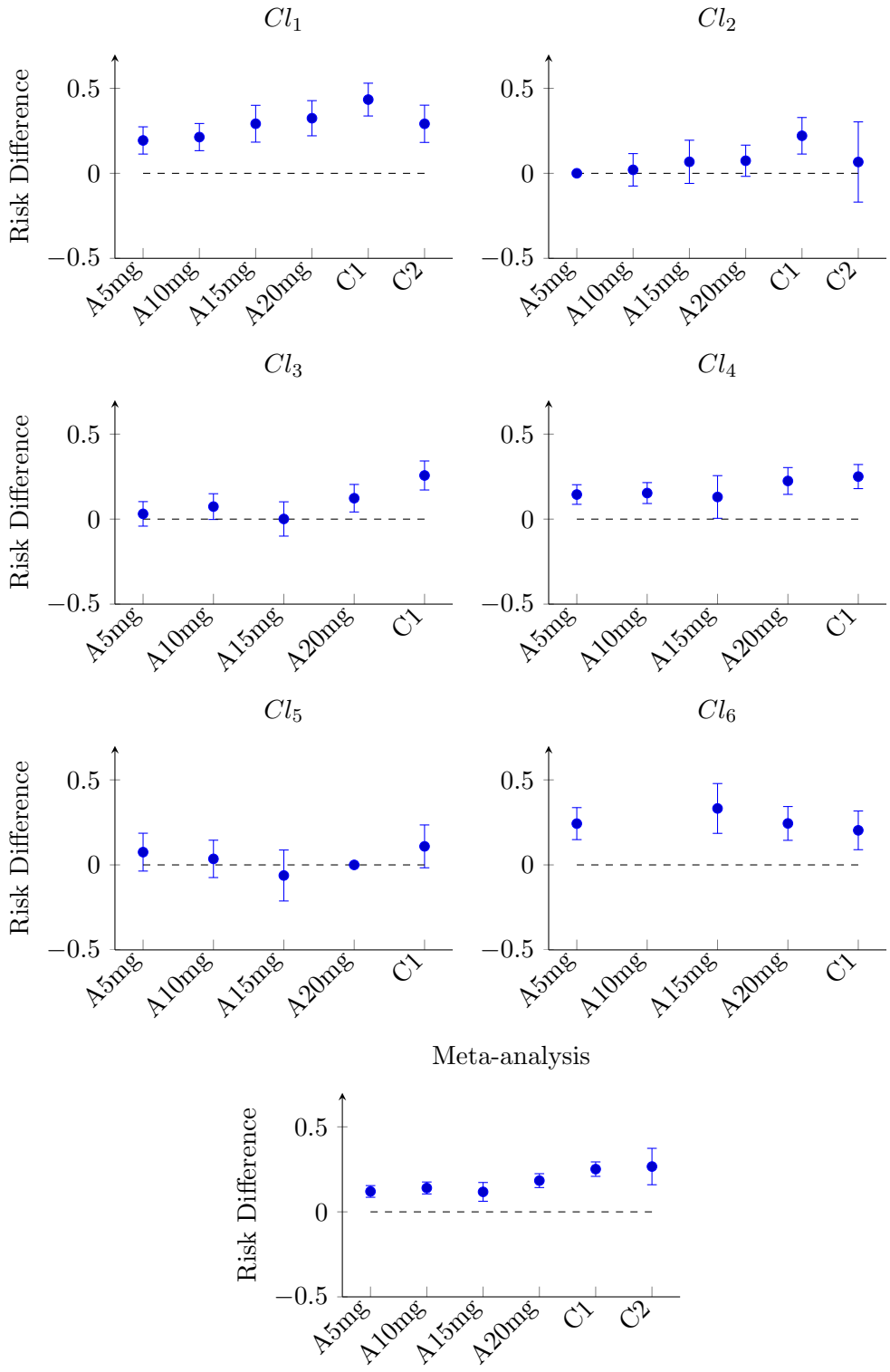

Figure B.5.: Confidence intervals of Risk Difference estimated by the random-effects model and the cluster-based random-effects model

i / Treatment	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
A10mg	0.5434	0.3488	0.1642	0.2295	0.5422	0.2927							0.4920	0.1674
A15mg	0.5714	0.1411	0.1644	0.1250	0.5292	0.1147							0.4724	0.0694
A20mg	0.4766	0.2157	0.1635	0.1781	0.6021	0.2358							0.5258	0.1288
A5mg	0.5476	0.2540	0.1571	0.2397	0.5770	0.3564							0.5210	0.1859
C1			0.2273	0.1507					0.7178	0.7151			0.6421	0.0948
PBO	0.3333	0.0121					0.4052				0.1454		0.3735	0.3118
C2			0.1111	0.0308					0.8478	0.1365			0.7273	0.0183
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.36.: Cluster-based identification of heterogeneity: Response rate in clusters classified by treatment

Region	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
eu	0.6495	0.1956	0.1240	0.2072	0.6412	0.2416	0.3671	0.2108	0.8117	0.3309	0.0889	0.1982	0.5394	0.2344
non eu	0.4962	0.8044	0.1944	0.7928	0.5439	0.7584	0.4154	0.7892	0.6807	0.6691	0.1593	0.8018	0.4666	
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.37.: Cluster-based identification of heterogeneity: Response rate in clusters classified by region

BMI	1 $\tilde{p}^i$	1 $\tilde{\rho}^i$	2 $\tilde{p}^i$	2 $\tilde{\rho}^i$	3 $\tilde{p}^i$	3 $\tilde{\rho}^i$	4 $\tilde{p}^i$	4 $\tilde{\rho}^i$	5 $\tilde{p}^i$	5 $\tilde{\rho}^i$	6 $\tilde{p}^i$	6 $\tilde{\rho}^i$	Total $\tilde{p}^i$	Total $\tilde{\rho}^i$
$]0, 22.53]$	0.5135	0.2238	0.1439	0.2380	0.6070	0.2035	0.4094	0.1816	0.8000	0.1855	0.1702	0.2070	0.4992	0.2007
$]22.53, 25.28]$	0.5161	0.1875	0.1846	0.2226	0.5962	0.2002	0.4224	0.1962	0.6917	0.1973	0.0682	0.1938	0.4900	0.1997
$]25.53, 28.39]$	0.6129	0.1875	0.1818	0.1695	0.5524	0.2077	0.3489	0.1956	0.7582	0.2270	0.1282	0.1718	0.4846	0.1998
$]28.39, 32.91]$	0.5111	0.1815	0.1852	0.1849	0.5828	0.1997	0.4154	0.1980	0.7063	0.2122	0.1724	0.2555	0.4913	0.1998
$]32.91, \infty]$	0.4862	0.2198	0.2130	0.1849	0.4945	0.1889	0.4267	0.2285	0.6583	0.1780	0.1795	0.1718	0.4534	0.2000
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.38.: Cluster-based identification of heterogeneity: Response rate in clusters classified by BMI

MADRS	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
22.0000	0.5000	0.0081	1.0000	0.0017			0.0000	0.0012	0.5000	0.0030	0.0000	0.0044	0.4000	0.0017
23.0000			0.6667	0.0051	0.3333	0.0013	0.6250	0.0049	0.7778	0.0134	0.0000	0.0044	0.6250	0.0040
24.0000			0.2500	0.0068	0.6667	0.0025	0.6667	0.0037	0.5000	0.0059			0.5500	0.0033
25.0000			0.6667	0.0051	0.6250	0.0034	0.6000	0.0030	0.5714	0.0104	0.0000	0.0088	0.5600	0.0042
26.0000			0.1515	0.0565	0.5495	0.0465	0.4225	0.0433	0.7250	0.0593	0.3333	0.0529	0.4831	0.0444
27.0000			0.1333	0.0514	0.5575	0.0729	0.3925	0.0652	0.8409	0.0653	0.1000	0.0441	0.4959	0.0607
28.0000			0.3478	0.0788	0.5819	0.0972	0.4088	0.0835	0.7708	0.0712	0.1250	0.0705	0.5136	0.0797
29.0000	0.5574	0.4738	0.1064	0.0805			0.2617	0.0908	0.5893	0.0831	0.0000	0.0485	0.4177	0.0829
30.0000			0.1406	0.1096	0.5623	0.1244	0.4162	0.1054	0.7183	0.1053	0.1765	0.0749	0.4855	0.1035
31.0000			0.2295	0.1045	0.5478	0.1315	0.4409	0.1133	0.6747	0.1231	0.0500	0.0881	0.4895	0.1105
32.0000			0.1875	0.0822	0.5714	0.1202	0.4329	0.0999	0.8039	0.0757	0.0556	0.0793	0.5035	0.0945
33.0000			0.1509	0.0908	0.5524	0.1039	0.4173	0.0847	0.7101	0.1024	0.1724	0.1278	0.4777	0.0895
34.0000			0.1389	0.0616	0.5659	0.0858	0.4444	0.0713	0.7273	0.0653	0.2174	0.1013	0.4941	0.0707
35.0000			0.2121	0.0565	0.6230	0.0800	0.3153	0.0676	0.7556	0.0668	0.1111	0.0793	0.4950	0.0662
36.0000			0.2286	0.0599	0.5489	0.0557	0.5205	0.0445	0.7407	0.0401	0.1538	0.0573	0.5018	0.0468
37.0000	0.5424	0.1190	0.1071	0.0479	0.5077	0.0272	0.3846	0.0317	0.7931	0.0430	0.2857	0.0617	0.4656	0.0411
38.0000	0.5169	0.1794	0.0556	0.0308			0.3953	0.0262	0.6667	0.0178	0.0000	0.0308	0.4260	0.0281
39.0000			0.2667	0.0257	0.5806	0.0260	0.4615	0.0238	0.5000	0.0119	0.0000	0.0220	0.4806	0.0215
40.0000	0.4783	0.1391	0.0000	0.0103			0.3913	0.0140	0.7500	0.0059	0.0000	0.0132	0.4286	0.0175
41.0000			0.0000	0.0034	0.6000	0.0126	0.4615	0.0079	0.6667	0.0134	0.3333	0.0132	0.5439	0.0095
42.0000	0.4800	0.0504	0.0000	0.0051			0.2727	0.0067	0.8571	0.0104	0.3333	0.0132	0.4490	0.0082
43.0000			0.0000	0.0120	0.7647	0.0071	0.3333	0.0037	1.0000	0.0045			0.5455	0.0055
44.0000	0.5000	0.0081	0.0000	0.0068			0.5000	0.0012	1.0000	0.0015	1.0000	0.0044	0.4167	0.0020
45.0000					1.0000	0.0008	0.5000	0.0012	1.0000	0.0015			0.8000	0.0008
46.0000	0.3333	0.0121	0.0000	0.0017									0.2857	0.0012
47.0000			1.0000	0.0017	0.6667	0.0013	0.5000	0.0012					0.6667	0.0010
48.0000	0.0000	0.0020	0.0000	0.0017									0.0000	0.0003
49.0000	0.3333	0.0060	0.0000	0.0017									0.2500	0.0007
52.0000	0.0000	0.0020											0.0000	0.0002
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.39.: Cluster-based identification of heterogeneity: Response rate in clusters classified by MADRS

CGI-S	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
3.0000	0.0000	0.0020	0.0000	0.0034	0.7500	0.0017	0.6667	0.0018	0.0000	0.0015			0.4545	0.0018
4.0000	0.5244	0.3306	0.2103	0.4315	0.5530	0.4112	0.4026	0.4223	0.7192	0.3858	0.1294	0.3744	0.4758	0.4053
5.0000	0.4958	0.4758	0.1741	0.4623	0.5504	0.4862	0.4324	0.4552	0.7023	0.4585	0.1607	0.4934	0.4801	0.4717
6.0000	0.6044	0.1835	0.0847	0.1010	0.7029	0.1001	0.3061	0.1194	0.8100	0.1484	0.1333	0.1322	0.5217	0.1190
7.0000	0.7500	0.0081	0.0000	0.0017	1.0000	0.0008	0.5000	0.0012	0.7500	0.0059			0.6923	0.0022
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.40.: Cluster-based identification of heterogeneity: Response rate in clusters classified by CGI-Score

Duration (in days)	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
$]0, 113]$	0.5647	0.1714	0.2222	0.1849	0.6275	0.2136	0.4423	0.2163	0.7105	0.2255	0.1071	0.2467	0.5237	0.2106
$]113, 142]$	0.5446	0.2036	0.1250	0.1781	0.5885	0.2035	0.3600	0.1828	0.7438	0.1795	0.1290	0.1366	0.4864	0.1902
$]142, 197]$	0.5773	0.1956	0.1797	0.2192	0.6000	0.2094	0.4140	0.2090	0.7386	0.2270	0.1633	0.2159	0.5055	0.2113
$]197, 323]$	0.5133	0.2278	0.1441	0.2021	0.5161	0.1817	0.3808	0.1840	0.7698	0.1869	0.1591	0.1938	0.4556	0.1892
$]323, \infty]$	0.4400	0.2016	0.2222	0.2158	0.4913	0.1918	0.4194	0.2078	0.6557	0.1810	0.1702	0.2070	0.4422	0.1987
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.41.: Cluster-based identification of heterogeneity: Response rate in clusters classified by duration of current episode

Age (in years)	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
$]0, 34]$	0.5714	0.2117	0.2125	0.2740	0.6091	0.2035	0.4915	0.2157	0.7091	0.1632	0.2727	0.1938	0.5195	0.2095
$]34, 44]$	0.6117	0.2077	0.1552	0.1986	0.5612	0.2052	0.3977	0.2084	0.7752	0.1914	0.1429	0.2159	0.4874	0.2045
$]44, 51]$	0.4286	0.1976	0.1650	0.1764	0.5662	0.1930	0.3622	0.1901	0.6154	0.1736	0.1296	0.2379	0.4472	0.1905
$]51, 60]$	0.5081	0.2500	0.1700	0.1712	0.5380	0.2312	0.3676	0.2072	0.7724	0.1825	0.1163	0.1894	0.4696	0.2133
$]60, \infty]$	0.5000	0.1331	0.1810	0.1798	0.5664	0.1671	0.3993	0.1785	0.7333	0.2893	0.0541	0.1630	0.4932	0.1822
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.42.: Cluster-based identification of heterogeneity: Response rate in clusters classified by age

Gender	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
Female	0.5543	0.7056	0.1641	0.6678	0.5809	0.6575	0.4153	0.6441	0.7346	0.6766	0.1631	0.6211	0.4962	0.6596
Male	0.4589	0.2944	0.2113	0.3322	0.5416	0.3425	0.3870	0.3559	0.7018	0.3234	0.1163	0.3789	0.4594	0.3404
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.43.: Cluster-based identification of heterogeneity: Response rate in clusters classified by gender

Reasons for withdrawal	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
.	0.5262				0.5674		0.4052		0.7240				0.5326	0.8651
Adverse Event(s)			0.1588	0.4743							0.1467	0.3304	0.1563	0.0586
Lack of Efficacy			0.0000	0.1267							0.0156	0.2819	0.0072	0.0230
Lost to Follow-up			0.3000	0.1884							0.2857	0.1850	0.2961	0.0253
Withdrawal of Consent			0.2276	0.2106							0.1957	0.2026	0.2189	0.0281
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.44.: Cluster-based identification of heterogeneity: Response rate in clusters classified by reason for withdrawal

Study	1		2		3		4		5		6		Total	
	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$	$\tilde{p}^i$	$\tilde{\rho}^i$
T11	0.7857	0.0282	0.1429	0.0719	0.7470	0.0695	0.5287	0.0530	0.8478	0.1365	0.0667	0.0661	0.6394	0.0692
T12	0.5833	0.0726	0.1685	0.1524	0.6881	0.0846	0.5203	0.0750	0.7080	0.1677	0.1364	0.0969	0.5504	0.0973
T13	0.7143	0.0282	0.2000	0.0514	0.6230	0.0511	0.4141	0.0780	0.7422	0.1899	0.0000	0.0617	0.5505	0.0725
T14	0.8056	0.0726	0.1071	0.0959	0.6860	0.0867	0.3636	0.0804	0.7863	0.1944	0.1500	0.0881	0.5687	0.0968
T15	0.4630	0.1089	0.1944	0.0616	0.6074	0.1248	0.3272	0.0987			0.1538	0.1145	0.4688	0.0958
T21	0.6333	0.0605	0.0789	0.0651	0.5392	0.0909	0.5236	0.1420			0.2051	0.1718	0.4829	0.0927
T22	0.5000	0.0403	0.2203	0.1010	0.4902	0.0427	0.3833	0.0731	0.6182	0.1632	0.1579	0.0837	0.4419	0.0715
T23	0.4894	0.0948	0.1304	0.0394	0.5098	0.0854	0.2598	0.0774			0.0000	0.0441	0.3966	0.0684
T24	0.4667	0.1210	0.2875	0.1370	0.5083	0.0758	0.4496	0.0786	0.6400	0.1484	0.0476	0.0925	0.4658	0.0950
T25	0.3171	0.0827	0.2500	0.0753	0.4098	0.0858	0.2878	0.0847			0.3846	0.0573	0.3462	0.0735
T26	0.3788	0.1331	0.1860	0.0736	0.4211	0.0716	0.3492	0.0768			0.2941	0.0749	0.3641	0.0704
T27	0.6026	0.1573	0.0909	0.0753	0.5559	0.1311	0.4296	0.0823			0.0000	0.0485	0.4871	0.0967
Total	0.5262		0.1798		0.5674		0.4052		0.7240		0.1454		0.4837	

Table B.45.: Cluster-based identification of heterogeneity: Response rate in clusters classified by study protocol

Bibliography

[1] Aurenhammer, F. Power diagrams:properties, algorithms and applications. *SIAM Journal on Computing 16*, 1 (1998), 61–76.

[2] Backhaus, K., Erichson, B., Plinke, W., and Weiber, R. *Multivariate Analysemethoden: Eine anwendungsorientierte EinfÃ̃¼hrung*. Springer, Berlin, 2008.

[3] Barnes, E., Hoffman, A., and Rothblum, U. Optimal partitions having disjoint convex and conic hulls. *Mathematical Programming 54*, 1-3 (1992), 69–86.

[4] Bodlaender, H., Gritzmann, P., Klee, V., and Leeuwen, J. Computational complexity of norm-maximization. *Combinatorica 10*, 2 (1990), 203–225.

[5] Borenstein, M., Hedges, L., Higgins, J. P., and Rothsteinh, H. *Introduction to Meta-Analysis*. John Wiley & Sons, 2009.

[6] Borgwardt, S. *A Combinatorial Optimization Approach to Constrained Clustering*. Dissertation. Technische Universität München, 2010.

[7] Borgwardt, S., Brieden, A., and Gritzmann, P. Constrained minimum-k-star clustering and its application to the consolidation of farmland. *International Journal of Operational Research 1101* (2011), 1–17.

[8] Borgwardt, S., Brieden, A., and Gritzmann, P. Mathematics in agriculture and forestry: Geometric clustering for land consolidation. *IFORS News 7*, 4 (2013), 10–11.

[9] Borgwardt, S., Brieden, A., and Gritzmann, P. Geometric clustering for the consolidation of farmland and woodland. *Math. Intelligencer 26* (2014), 37–44.

[10] Borgwardt, S., Brieden, A., and Gritzmann, P. A balanced k-means algorithm for weighted point sets. 2015.

[11] Breyer, F., Zweifel, P., and Kifmann, M. *Gesundheitsökonomik*. Springer-Lehrbuch. Springer, 2004.

[12] Brieden, A. Geometric optimization problems likely not contained in apx. *Discrete and Computational Geometry 28* (2002), 201–209.

[13] Brieden, A., and Gritzmann, P. On the inapproximability of polynomial-programmining, the geometry of stable sets, and the power of relaxation. *Discrete and Computational Geometry* (2003), 301–311.

[14] Brieden, A., and Gritzmann, P. A quadratic optimization model for the consolidation of farmland by means of lend-lease agreements. *Proceedings of Operations Research 2003* (2004), 324–331.

[15] Brieden, A., and Gritzmann, P. On clustering bodies: Geometry and polyhedral approximation. *Discrete and Computational Geometry 44*, 3 (2010), 508–534.

[16] Brieden, A., and Gritzmann, P. On optimal weighted balanced clusterings: gravity bodies and power diagrams. *SIAM Journal of Discrete Mathematics 26*, 2 (2012), 415–434.

[17] Brieden, A., Gritzmann, P., Kannan, R., Klee, V., LovÃ¡sz, L., and Simonovits, M. Approximation of diameters: Randomization does not help. *Proceeding of the 39th Annual Symposium of the FOCS* (1998), 244–251.

[18] Brieden, A., Gritzmann, P., Kannan, R., Klee, V., LovÃ¡sz, L., and Simonovits, M. Deterministic and randomized oracle-polynomial-time approximation of radii. *Mathematika 48* (2001), 63–105.

[19] Brieden, A., and Hinnenthal, M. Cluster-based meta-analysis. *Manuscript*.

[20] Brieden, A., Hinnenthal, M., and Öllinger, M. Probabilistic hypotheses in geometric clusterings. *Manuscript*.

[21] Brieden, A., Hinnenthal, M., and Pogarell, O. Regional heterogeneity of multi-centric studies explanations with the help of endpoint-oriented cluster optimization.

[22] Cohn, L., and Becker, B. How meta-analysis increases statistical power. *Psychological Methods 8*, 3 (2003), 243–253.

[23] DerSimonian, R., and Kacker, R. Random-effects model for meta-analysis of clinical trials, an update. *Contemporary Clinical Trials 28* (2007), 105–114.

[24] DerSimonian, R., and Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials 7* (1986), 177–188.

[25] Dickersin, K., and Berlin, J. Meta-analysis: State-of-the-science. *Epidemiologic Reviews 14* (1992).

[26] Fleiss, J. The statistical basis of meta-analysis. *Statistical methods in medical research 2* (1993), 121–145.

[27] Gagnier, J., Moher, D., Boon, H., Beyene, J., and Bombardier, C. Investigating clinical heterogeneity in systematic reviews, a methodologic review of guidance in the literature. *BMC Medical Research Methodology 12*, 111 (2012).

[28] Hardy, R., and Thompson, S. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine 17*, 8 (1998), 841–856.

[29] Hedges, L., and Pigott, T. The power of statistical tests in meta-analysis. *Psychological Methods 6*, 3 (2001), 203–217.

[30] Higgins, J., and Green, S. *Cochrane Collaboration Cochrane Handbook for Systematic Reviews of Interventions*, vol. Version 5.1.0. 2011.

[31] Higgins, J., and Thompson, S. Quantifying heterogeneity in meta-analysis. *Statistics in medicine 21* (2002), 1539–1558.

[32] Higgins, J., Thompson, S., Deeks, J., and Altman, D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy 7*, 1 (2002), 51–61.

[33] Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., and Botella, J. Assessing heterogeneity in meta-analysis: Q statistic or i2 index? *Center for Health, Intervention, and Prevention (CHIP)* (2006).

[34] Hwang, F., Onn, S., and Rothblum, U. Representations and characterizations of vertices of bounded-shaped partition polytopes. *Linear Algebra and its Applications 278*, 1-3 (1998), 263–284.

[35] Iovieno, N., and Papakostas, G. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. *The Journal of clinical psychiatry 73*, 10 (2012), 1300–1306.

[36] IQWiG. General methods, 2013.

[37] IQWiG. Methodik für die bewertung von verhältnissen zwischen nutzen und kosten im system der deutschen gesetzlichen krankenversicherung version, 2015.

[38] Jackson, D. The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine 25*, 15 (2005), 2688–2699.

[39] Khin, N., Chen, Y., Yang, Y., Yang, P., and Laughren, T. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the us food and drug administration in support of new drug applications. *The Journal of clinical psychiatry 72*, 4 (2011), 464–472.

[40] Kravitz, R., Duan, N., and Braslow, J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quaterly 82*, 4 (2004), 661–687.

[41] Lindsey, J. Statistical inference based on the likelihood. adelchi azzalini. *Statistics in Medicine 19*, 9 (2000), 1260–1261.

[42] Mittlbock, M., and Heinzl, H. A simulation study comparing properties of heterogeneity measures in meta-analyses. *Statistics in Medicine 25*, 24 (2006), 4321–4333.

[43] Moeller, H.-J. Das problem der heterogenitaet zwischen in den usa und nicht in den usa durchgefuehrten antidepressiva-studien. *Psychopharmakotherapie 21* (2014), 211–218.

[44] Mukhtar, A. *Beruecksichtigung von Heterogenitaet in Meta-Analyse von Randomisierten Kontrollierten Studien.* Dissertation. Universität Bremen, 2008.

[45] Murphy, K. Machine learning: a probabilistic perspective. *Adaptive Computation and Machine Learning series. The MIT Press* (2012).

[46] Newby, D., and Hill, S. Use of pharmacoeconomics in prescribing research. prat 2: Cost-minimization analysis - when are two therapies equal? *Journal of Clinical Pharmacy and Therapeutics 28*, 2 (2003), 145–150.

[47] Niklson, I., and Reimitz, P. Baseline characteristics of major depressive disorder patients in clinical trials in europe and united states: is there a transatlantic difference? *Journal of psychiatric research 35*, 2 (2001), 71–81.

[48] Öllinger, M. *Geometric clustering and its applications in binary classification problems.* Dissertation. Universität der Bundeswehr, 2014.

[49] Pacchiarotti, I., Pappdopulos, E., Mandel, F., Lombardo, I., Loebel, A., and Vieta, E. Differential response to treatment across countries in a randomized clinical trial in mania. *European Neuropsychopharmacology 20* (2010), 93–104.

[50] Pacchiarotti, I., Valenti, M., Bonnin, C., Rosa, A., Murru, A., and Kotzalidis, G. Factors associated with initial treatment response with antidepressants in bipolar disorder. *European Neuropsychopharmacology 21*, 5 (2011), 362–369.

[51] Pal, K. What is the influence of big data in medicine? *KDnuggets* (2016).

[52] Panityakul, T., Bumrungsup, C., and Knapp, G. On estimating residual heterogeneity in random-effects meta-regression, a comparative study. *Journal of Statistical Theory and Applications 12*, 3 (2013), 253–265.

[53] Pogarell, O. Regionale heterogenitaet und deren einfluss auf die ergebnissicherheit klinischer studien, systematischer reviews und metaanalysen, 2014.

[54] Raghupathi, W., and Raghupathi, V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems 2*, 3 (2014).

[55] Rennie, D., and Dickersin, K. Registering clinical trials. *JAMA 290*, 4 (2003).

[56] Rücker, G., Schwarzer, G., Carpenter, J., and Schumacher1, M. Undue reliance on i^2 in assessing heterogeneity may mislead. *BMC Med Res Methodol 8*, 79 (2008).

[57] Sackett, D., Rosenberg, W., Gray, J., and R.B. Haynes, W. R. Evidence based medicine: what it is and what it isn't. *BMJ 312*, 7023 (1996), 71–72.

[58] Schöffski, O., and Schulenburg, J. *Gesundheitsökonomisch Evaluation*. Springer, 2012.

[59] Sidik, K., and Jonkman, J. A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in medicine 26* (2007), 1964–1981.

[60] Smith, G., and M.Egger. Incommunicable knowledge, interpreting and applying the results of clinical trials and meta-analyses. *J Clin Epidemiol 51*, 4 (1998), 289–295.

[61] Snedecor, G., and Cochran, W. *Statistical Methods*. Iowa State University Press., 1989.

[62] Song, F., Sheldon, T., Sutton, A., Abrams, K., and Jones, D. Methods for exploring heterogeneity in meta-analysis. *Evaluation and the health professions 24*, 2 (2001), 126–151.

[63] Sprott, D. Normal likelihoods and their relation to large sample theory of estimation. *Biometrika 60*, 3 (1973), 457–465.

[64] Stanley, T., and Jarrell, S. Meat-regression analysis, a quantitative method of literature surveys. *Journal of Economic Surveys 3*, 2 (1989), 161–169.

[65] Stassen, H., Angst, J., and Delini-Stula, A. Severity at baseline and onset of improvement in depression. meta-analysis of imipramine and moclobemide versus placebo. *European psychiatry : the journal of the Association of European Psychiatrists 9* (1994), 129–136.

[66] Stein, D., Baldwin, D., Dolberg, O., Despiegel, N., and Bandelow, B. Which factors predict placebo response in anxiety disorders and major depression? an analysis of placebo-controlled studies of escitalopram. *The Journal of clinical psychiatry 67*, 11 (2006), 1741–1746.

[67] Sullivan, F. . Drowning in big data? reducing information technology complexities and costs for healthcare organizations. *50 Years of Growth, Innovation and Leadership*.

[68] Sutton, A., Abrams, K., Jones, D., Sheldon, T., and Song, F. Systematic reviews of trials and other studiessystematic reviews of trials and other studies. *Health Technology Assessment 2*, 19 (1998).

[69] Takkouche, B., Cadarso-Suarez, C., and Spiegelman, D. Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *American Journal of Epidemiology 150*, 2 (1999), 206–215.

[70] Thase, M., Last, C., Hersen, M., Bellack, A., and Himmelhoch, J. Symptomatic volunteers in depression research: a closer look. *The Journal of clinical psychiatry 11*, 1 (1984), 25–33.

[71] Thompson, S., and Pocock, S. Can meta-analyses be trusted? *Controlled Clinical Trials 338* (1991), 1127–1130.

[72] Thompson, S., and Sharp, S. Explaining heterogeneity in meta-analysis, a comparison of methods. *Statistics in medicine 18* (1999), 2693–2708.

[73] Thompson, S., Smith, T., and Sharp, S. Investigating unferlying risk as source of heterogeneity in meta-analysis. *Statistics in medicine 16* (1997), 2741–2758.

[74] Uman, L. Systematic reviews and meta-analyses. *J Can Acad Child Psychiatry 20*, 1 (2011), 57–59.

[75] van der Beek, K., and van der Beek, G. *Gesundheitsökonomik: Einführung*. Oldenbourg Wissenschaftsverlag, 2011.

[76] Walter, E., and Zehetmayr, S. Guidelines zur gesundheitsökonomischen evaluation, 2006.

[77] Whitehead, A. *Meta-Analysis of controlled clinical trials.* John Wiley & Sons, 2002.

[78] Yusuf, S., Wittes, J., Probstfield, J., and Tyroler, H. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA 266*, 1 (1991), 93–98.