General discussion

The overall aim of the thesis is to investigate how to optimize the design and analysis of randomized and non-randomized therapeutic studies, in order to increase the validity and reliability of causal treatment effect estimates, specifically in heterogeneous diseases.

Two specific research questions were addressed:

1) What are the benefits of more advanced statistical analyses to estimate treatment effects from RTCs in heterogeneous diseases?

a. What is the heterogeneity in acute neurological diseases with regard to baseline severity and further course of the disease?

b. What is the potential gain in efficiency of covariate adjustment and proportional odds analysis in RCTs in Guillain-Barré syndrome (GBS)?

We found substantial heterogeneity in the clinical severity and course in the acute stage and during follow-up of two well-defined acute neurological diseases (both GBS and traumatic brain injury (TBI)). Also, we found that covariate adjustment and proportional odds analysis most efficiently use available RCT data in such heterogeneous diseases and ensure balance between the treatment arms to obtain reliable and valid treatment effect estimates in RCTs in GBS.

2) What is the validity and reliability of the RD design compared to an RCT to estimate causal treatment effects?

a. What are threats to the validity of the RD design to estimate treatment effects compared to an RCT?

b. How efficient is the RD design to estimate treatment effects compared to an RCT?

c. What are the potential benefits of an alternative assignment approach in an RD design?

For the second research question we found that the RD design may provide similar but substantially less precise treatment effect estimates compared to an RCT. Most important threats to validity of the RD design include misspecification of the functional form of the relationship between the assignment variable and outcome measure in the analysis and wrong assumptions on the heterogeneity of the treatment effect over the range of the assignment variable. We found that the RD design may provide similar but substantially less precise treatment effect estimates compared to an RCT. An RD design requires at least 2.75 times as many patients compared to an RCT to estimate the same precise treatment effects. Compared to an unadjusted analysis, the efficiency of an RD design could be increased by using an assignment variable with a low correlation with the outcome of interest. However, the relative efficiency compared to an adjusted analysis of the treatment effect in an RCT, was not dependent on the correlation between the

treatment assignment variable and outcome since the adjustment affects the efficiency of an RCT as well.

In this chapter, the results of the studies are discussed with their implications. We also make recommendations and draw some overall conclusions.

Table 1. Main research findings

Question	Answer
What is the heterogeneity in acute neurological diseases with regard to baseline severity and further course of the disease?	We found substantial heterogeneity in the clinical severity and course in the acute stage and during follow-up of two well-defined acute neurological diseases (both GBS and traumatic brain injury (TBI)).
What is the potential gain in efficiency of covariate adjustment and proportional odds analysis in RCTs in Guillain-Barré syndrome (GBS)?	We found that covariate adjustment and proportional odds analysis most efficiently use available RCT data in such heterogeneous diseases and ensure balance between the treatment arms to obtain reliable and valid treatment effect estimates in RCTs in GBS.
What are threats to the validity of the RD design to estimate treatment effects compared to an RCT?	Most important threats to validity of the RD design include misspecification of the functional form of the relationship between the assignment variable and outcome measure in the analysis and wrong assumptions on the heterogeneity of the treatment effect over the range of the assignment variable.
How efficient is the RD design to estimate treatment effects compared to an RCT?	We found that the RD design may provide similar but substantially less precise treatment effect estimates compared to an RCT. An RD design requires at least 2.75 times as many patients compared to an RCT to estimate the same precise treatment effects.
What are the potential benefits of an alternative assignment approach in an RD design?	Compared to an unadjusted analysis, the efficiency of an RD design could be increased by using an assignment variable with a low correlation with the outcome of interest. However, the relative efficiency compared to an adjusted analysis of the treatment effect in an RCT, was not dependent on the correlation between the treatment assignment variable and outcome since the adjustment affects the efficiency of an RCT as well.

Randomized controlled trials

RCTs are the reference standard to study the efficacy of medical interventions. However, especially in heterogeneous and rare neurological diseases it is a challenge to include a sufficient number of patients in an RCT to reach a sufficient statistical power to be able to detect statistically significant treatment effects. Moreover, due to the heterogeneity in clinical severity and outcome, small differences in baseline risk on outcome between the treatment arms may influence the estimated treatment effect.

Heterogeneity in Guillain-Barré syndrome

In **chapter 2** we found that hospital admissions highly varied between patients with GBS, especially with regard to the number of hospital transfers and disease-related costs. GBS is a complex disorder because of the various stages in the disease course

Table 2. The pros and cons of RCTs and RD designs.

Challenge	RCT	RD	Recommendation
Selection of patients	Well-defined; still heterogeneous	Observational; focus on cut-off point	Selection in RCT based on subject knowledge; in RD based on treatment guidelines
Numbers of patients	Relatively small	Larger, but small around the point of interest	In RCTs covariate adjustment and more powerful statistical analyses; in RD using assignment variables that are feasible in clinical practice to facilitate patient inclusion
Comparability	Causal inference possible by randomization, but differences may occur in baseline risk by chance	Causal inference possible around the cut-off point; more speculative for patient further from the cut-off point	Interpret treatment effect estimates from RCTs as global estimates; interpret treatment effect estimates from RD designs primarily as local estimates
Treatment effect heterogeneity	Both treatment arms available over the full range of the population; treatment effect heterogeneity can be tested, sample size may be insufficient to detect significant treatment effect heterogeneity	Treatment groups each have data on only one side of the cut-off, the assumptions required to estimate the global treatment effect cannot be tested	Interpret treatment effect estimates from RCTs as global estimates; in RD, global treatment effect estimates from RD designs should only be presented secondary to local treatment effect estimates

that require different health care facilities, ranging from an intensive care unit in the progressive phase and a rehabilitation unit in the recovery phase. Moreover, the clinical course and related need of these facilities highly varies between patients, ranging from short term admissions at medium care units to admissions to intensive care units and rehabilitation units for months to even years. The complexity is reflected in the high frequency of transfers between departments and hospitals, especially shortly after initial admission. Transfers within and between hospitals were frequent: 40% of the patients were transferred at least one time and half of them were transferred within two days of initial admission. Moreover, in 25% of the cases, the admission may have been suboptimal form a cost-effectiveness perspective, including admission to other than (pediatric) neurology departments or ICUs, admission of mildly affected patients to ICUs and transfers shortly after the initial admission. The related costs were highly variable between patients and mainly associated with the severity of disease (Figure 1). These findings are important with regard to designing future GBS studies. The large heterogeneity should be taken into account when designing an RCT in GBS.

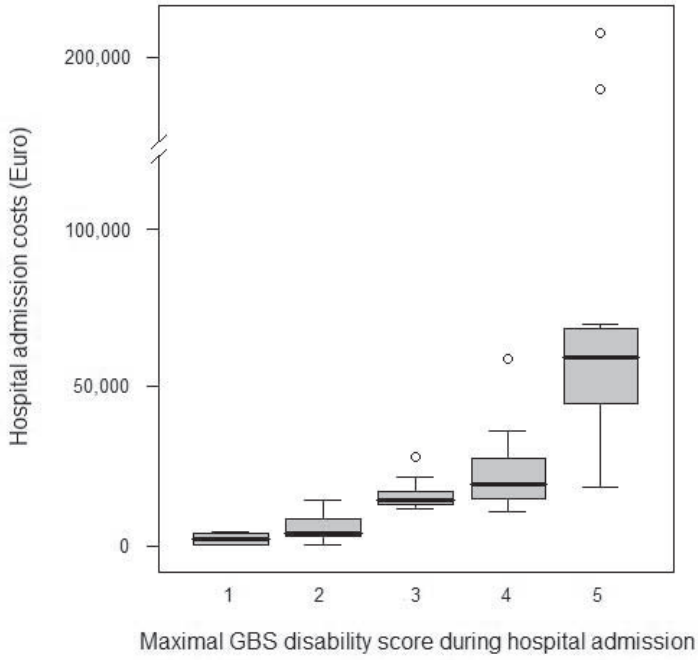

Figure 1. Interquartile ranges (grey boxes), 95% confidence intervals (whiskers) and median (dark lines in middle of the boxes) of costs of hospital admission for different maximal GBS disability scores.
Excluded was one patient who died. Circles are (extreme) outliers. Maximal GBS disability score during hospital admission: 1 = minor symptoms, 2 = able to walk 10m unassisted but unable to run, 3 = able to walk over 10m open space with help, 4 = bedridden or chair bound, 5 = needs ventilation for at least a part of the day.

Heterogeneity in traumatic brain injury

It is known that the TBI patient population is highly heterogeneous with regard to baseline severity and outcome. This hampers TBI research, especially estimation of treatment effects in RCTs. To choose the best prognostic variables to use in covariate adjustment in RCTs, studies on the prognostic value of a baseline variable on outcome should be used. We studied the prognostic value of major extracranial injury (MEI) on mortality in TBI patients. Our results in **chapter 3** show that MEI is an important prognostic factor for mortality in TBI patients. However, the prognostic effect is dependent on the population studied. First the strength of the effect is heterogeneous over the range of the brain injury severity. The prognostic effect of MEI is larger in patients with mild TBI. Moreover, we found that the effect is dependent on the time of inclusion in a study. In the registry we used in our study, MEI is strongly associated with mortality after adjustment for age, Glasgow Coma Scale motor score and pupil reactivity. In broadly selected observational studies and an RCT, including TBI patients surviving the early stage after their injury, the incremental prognostic value of MEI compared to known predictors of mortality was limited. These results are important for example to identify prognostic variables

for covariate adjustment, in the design of future TBI trials. The meta-analysis in **chapter 3** implicates specifically that MEI is an important prognostic factor to correct for when studying the effect of pre-hospital interventions, including all patients starting from the time of injury. In contrast it would be less urgent to consider MEI in studies assessing in-hospital interventions, including mainly patients with more severe brain injury and patients who survived the early phase after injury.

With regard to research question 1a we conclude in that hospital admissions for GBS patients are highly heterogeneous, with frequent transfers and higher costs for those with more severe disease. Also, MEI is an important prognostic factor for mortality in TBI patients; however, the effect varies by population.

To assess the benefits of more advanced statistical analyses to estimate treatment effects from RCTs in heterogeneous populations, we studied covariate adjustment and proportional odds analysis in GBS in **chapter 4**.

Covariate adjustment

Covariate adjustment is a statistical method that adjusts the treatment effect for baseline risk on poor outcome in the treatment and control arms. When the treatment arms are unbalanced, the unadjusted estimate of the treatment effect may be different than when treatment arms are fully balanced. Also, when there are no differences in baseline risk, the adjusted estimates will be more extreme than the unadjusted estimates.(1) On expectation, covariate adjustment leads to more extreme treatment effect estimates (further away from $\beta = 0$ *or odds ratio* $= 1$) and larger standard errors for non-linear regression models.(2) Although the standard error is larger when covariate adjustment is applied, the statistical power increases.(3, 4) The p-values are a function of the treatment effect estimates and standard error. The increase in treatment effect estimate will outweigh increased in standard error and the p-values will be lower compared to unadjusted analysis.(2)

Indeed, in **chapter 4**, we found increased standard errors in all adjusted analyses compared to the unadjusted analyses. The better prognosis in the treatment group decreased the treatment effect estimate β after covariate adjustment in the Plasma Exchange (PE) vs Intravenous Immunoglobulin (IVIg) (PE vs IVIg) trial in patients with GBS. In the IVIg and placebo versus IVIg and Methyl-Prednisolone (MP) (IVIg vs MP) trial in patients with GBS, the treatment group had a lower probability of favorable outcome. Therefore, in the IVIg vs MP trial covariate adjustment led to a larger β and a smaller p value.

When investigating the effectiveness of a medical intervention in rare and heterogeneous neurological diseases, such as GBS, one has to deal with limited sample sizes. In GBS trials, the outcome 'minimal one grade improvement' on the GBS disability score, is often used as primary endpoint and implicitly involves a form of covariate adjust-

ment. The baseline disease severity of the patient is taken into account in the analysis by estimating improvement for each patient from his or her own starting position at admission. This principle of a measure of change between baseline and follow up seems attractive to control for baseline imbalance. However, analyzing change does not control for baseline imbalance caused by regression to the mean(5, 6); baseline values are negatively correlated with change because patients with high scores (more severely affected patients) at baseline generally improve more than those with low scores.(7) Therefore covariate adjustment with the absolute baseline value is still preferable over implicitly taking into account baseline severity in the outcome measure 'improvement' (Table 3). Moreover, disease severity at baseline is not the only relevant covariate. For example, age will be an important covariate in most diseases.

When designing a trial, the analysis plan should be precisely pre-specified, including the covariates that will be used for adjustment. Previous studies showed that the stronger the effect of the covariates on outcome, the larger the increase in statistical power with covariate adjustment will be.(8-10) In GBS, predictors of outcome are relatively well known(11, 12) and therefore pre-specifying important baseline variables for covariate adjustment is possible in GBS trials.

Proportional odds analysis

Another, more advanced statistical method for analyses of outcome in RCTs is proportional odds analysis. Proportional odds analysis optimally exploits the ordinal nature of outcome scales, which are frequently used as primary outcome measures in RCTs. The proportional odds analysis estimates the treatment effect on each cut-off of the ordinal outcome scale, instead of estimating the treatment effect on the difference between the averages scores in the treatment arms, as in linear regression. The proportional odds model results in a common OR, which is interpretable as a pooled or overall OR for the different cut-offs. The common OR can be interpreted as the average shift over the total ordinal outcome scale caused by the treatment under study.(13-16) Because the ordinal analysis uses the full ordinal outcome scale instead of one dichotomy, the variance will be smaller compared to binary analysis. This was confirmed in our study in **chapter 4**, where the proportional odds resulted in lower standard errors compared to the binary approaches.

In the PE vs IVIg trial in patients with GBS, the ORs for each cut-off were very similar and as a result the common OR was also similar. Thus, with a smaller SE, the p-value was lower. In contrast, in the IVIg vs IVIg+MP trial in patients with GBS, the ORs for each cut-off were more scattered. One explanation is chance: the ORs for the different cut-offs are uncertain, especially at the tails of the outcome scale where numbers are usually small. However, almost all binary ORs have confidence intervals that overlap. Another explanation is that the treatment effect is truly different for different cut-offs, although

this is considered unlikely for a disorder like GBS. In hindsight, the cut-off chosen in the reference approach (more than the other possible cut-offs) improvement appeared to be the optimal cut-off from a statistical perspective, since it was the only cut-off resulting in a significant treatment effect.

Table 3. Characteristics of different methods of treatment effect analysis in GBS trials. Approach in BOLD is the recommended approach.

	Takes into account baseline imbalance	**Takes into account ordinal nature of the outcome measure**
Unadjusted binary logistic regression on cutoff for GBS disability score	NO	NO
Adjusted binary logistic regression on cutoff for GBS disability score	YES	NO
Unadjusted binary logistic regression on ≥ 1 grade improvement on GBS disability score	PARTLY*	NO
Adjusted binary logistic regression on ≥ 1 grade improvement on GBS disability score	YES	NO
Unadjusted proportional odds logistic regression on GBS disability score	NO	YES
Adjusted proportional odds logistic regression on GBS disability score	**YES**	**YES**
Unadjusted proportional odds logistic regression on Δ GBS disability score	PARTLY*	YES
Adjusted proportional odds logistic regression on Δ GBS disability score	YES	YES

*Only baseline GBS disability score, no other covariates

Proportional odds assumption

The common OR from a proportional odds analysis is formally valid if the ORs for each cut-off are the same. This is called the proportional odds assumption. We can, however, interpret the common OR as a summary measure of the treatment effect, even if the ORs differ per cut-off.(13, 17) In a recent RCT on decompressive craniectomy for traumatic intracranial hypertension, the common OR from the proportional odds model was not presented because the proportional odds assumption was violated; surgery strongly reduced mortality but at the cost of more vegetative state and severe disability.(18) Instead, the authors reported a descriptive analysis, ignoring the ordering in the outcome. The overall trial result was difficult to interpret. However, it is not the violation of the proportional odds assumption that complicates the interpretation of a proportional odds ratio, but the lack of consensus on the value judgment on the ordering of dead, vegetative state and severe disability in the ordinal scale. If there is agreement that each score on a certain scale is more favorable than a one point lower score, statistical testing of the proportional odds assumption is redundant.(19) Proportional odds analysis

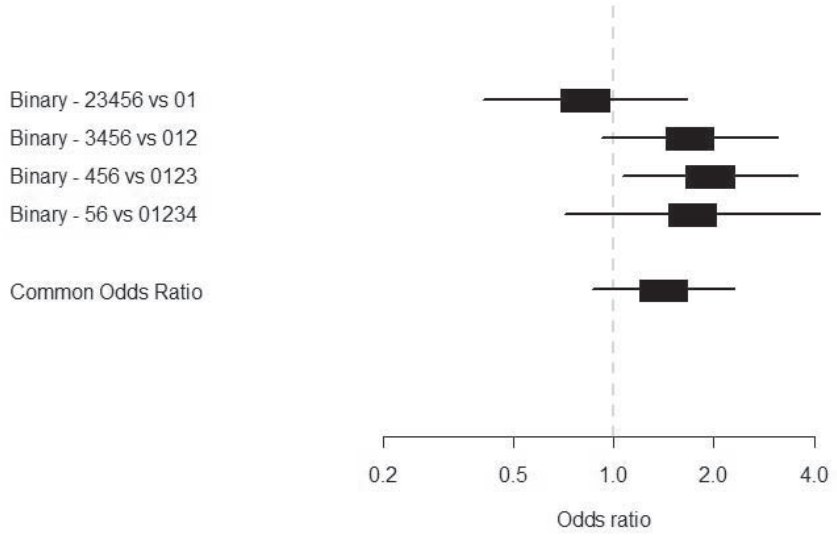

Figure 2. Treatment effect analysis: forest plots of the adjusted binary and proportional odds logistic regression in the IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial

allows sample sizes to be reduced substantially, even when the proportional odds assumption is not met.(15) We encourage the use of proportional odds analysis for the primary analysis of treatment effect in RCTs with an ordinal outcome. For transparency, the binary odds ratios for each cut-off of the ordinal outcome should be presented, as in Figure 2 and **chapter 4**. If there is consensus on the ordering, the common OR can be presented and interpreted as a summary estimate of the treatment effect, regardless of violation of the proportional odds assumption.

In summary, covariate adjustment and proportional odds analysis most efficiently use the available RCT data and ensure balance between the treatment arms to obtain reliable and valid treatment effect estimates. These approaches merit application in future trials in rare and heterogeneous neurological diseases like GBS. For GBS, covariate adjustment should be applied with known predictors for (functional) clinical outcome, specifically age at diagnosis, presence of preceding diarrhea, GBS disability score and MRC sum score.(11, 12) Although covariate adjustment and proportional odds analysis increase statistical power, it is not advised to lower the sample size of the study, since in practice most trials are underpowered.

Regression discontinuity design

In some situations, an RCT might be complicated to perform, due to regulatory requirements, patients' treatment preferences or (perceived) lack of equipoise. In such situation, data from observational studies may be used to estimate a treatment effect by comparing the clinical course in subgroups of patients receiving different treatments.

A major challenge in such observational studies of the effectiveness of treatment is to correct for unmeasured confounders. Estimating the causal relation between treatment and outcome is often hampered by confounding by indication. It is stated that the quasi-experimental RD design is a promising design to assess the causal inference between a medical intervention and outcome.(20) The second part of this thesis focused on the validity and reliability of this alternative study design.

Causality in a regression discontinuity design

The controlled allocation of treatment is the most important advantage of a prospective RD design over an observational study. This characteristic of the design is similar to an RCT. In both an RCT as in a (prospective) RD design, we have good understanding of the mechanism of assignment of treatment.(21) In RCTs, treatment allocation is at random and in RD the assignment of treatment is based on a baseline assignment variable. Treatment effect estimates from an RCT can be interpreted as a causal relation between treatment and the outcome, because the treated and the control patients are exchangeable. In an RD design the treated and the control patients are not exchangeable over the complete range of the assignment variable since they have a systematically different baseline value. In RD the treated and control patients are only replaceable around the cut-off of the assignment variable.(21, 22) Therefore, in an RD design, causal inference can only be made around the cut-off. This assumption can be tested, by showing a histogram of the treatment assignment variable, like is presented in the supplementary figures of **chapter 6**. Hahn et al.(23) shows that without this area of overlap, continuity in the assignment variable near the cut-off is sufficient to obtain unbiased estimates of the treatment effect. Visual inspection of the data can confirm that the assignment variable is continuous at the cut-off.(24)

Global vs. local treatment effect estimates

RD may provide similar estimates of treatment effects to RCT estimates, but it requires the assumption of a global treatment effect over the full range of the assignment variable. However, the causal treatment effect estimated in RD should be primarily interpreted as a local treatment effect estimate, around the cut-off. Even with comparable RCT and RD data, it might not be completely straightforward to compare estimates from an RCT and an RD design.(21) The overall RCT estimate is the average treatment effect in the whole RCT population.(2, 8, 25, 26) An RD estimate is a local treatment effect among patients at the cut-off and may vary dependent on the cut-off for treatment assignment.(27) Only when the treatment effect is constant over the full range of the assignment variable, the treatment effect estimate from an RD design can be interpreted as a global treatment effect estimate, and is comparable to the global RCT estimate.(21) In order to estimate a global treatment effect estimate in RD, one would have to feel confident modeling

the relationship between the assignment variable and the outcome even where it is not observed in the data.(21, 28, 29) In a prospective RD design, it is not possible to assess whether there is heterogeneity of the treatment effect over the range of the baseline assignment variable, since the treatment groups each have data on only one side of the cut-off. So, the assumptions required to estimate the global treatment effect cannot be tested in a prospective RD design. Therefore, we suggest that global treatment effect estimates from RD designs should only be presented secondary to local treatment effect estimates and not as the primary parameter of interest.

This thesis shows that when there is no interaction between the assignment variable and treatment – and thus a global treatment effect can be estimated – the results from the RCS or polynomial adjusted analyses and local logistic regression are more similar to each other than when there is treatment effect heterogeneity over the assignment variable. For example, in **chapter 5 and 6**, we found no interaction between treatment and the assignment variable in one of the validation studies and the results from both logistic regression with RCS adjustment and local logistic regression were similar in this example. In the other two validation studies in **chapter 6**, non-linear restricted cubic spline functions of the interaction of the intervention effects over the assignment variables showed interaction between the assignment variable and treatment, and the results from the analysis with local logistic regression and the RCS adjusted analyses were less similar.

In conclusion, RD may provide similar estimates of treatment effects to RCT estimates but requires the assumption of a global treatment effect over the full range of the assignment variable. This assumption is not verifiable within the RD design.

Efficiency of the RD design compared to an RCT

The RD estimates appeared to be substantially less efficient than RCT estimates. In **chapter 5 and 6**, we assessed the difference in efficiency of RD compared to an RCT for both continuous and dichotomous outcome parameters. For continuous outcomes, in terms of statistical precision, the RD with RCS adjustment was 1 to 4 times less efficient than an RCT for the local effects estimated. An RD design analyzed with adjusted logistic regression using RCS adjustment implies that 7 to 12 times more patients need to be included in the study compared to an RCT design. If one would analyze the RD design with local logistic regression, this study would need about 3 times more patients than an RCT. So, the local regression approach was more efficient compared to the adjusted logistic regression. In terms of efficiency, local logistic regression would be preferred to analyze an RD design.

In summary, the RD design provides substantially less precise treatment effect estimates compared to an RCT. When considering a prospective RD design, researchers need to weigh better recruitment against the substantial loss in precision.

Efficient assignment approach in RD

In **chapter 7** we assessed the potential efficiency of an alternative treatment assignment strategy. When assignment in RD was close to at random, or based on a variable that poorly correlates with outcome, estimates were more efficient than RD based on a variable highly correlating with outcome. These comparisons were made with the unadjusted treatment effect estimate from a similarly-sized RCT. However, compared to an adjusted treatment effect estimate from an RCT, the (in)efficiency of the RD design is independent of the correlation between assignment variable and outcome measure. In the case study, RD estimates from assignment based on a random variable or variable poorly correlating with outcome were more similar to the global RCT estimates than the RD estimates from assignment based on a variable highly correlating with outcome. These findings show that the relative efficiency of the RD design is not dependent on the correlation between the treatment assignment variable and outcome. We recommend researchers to use assignment variables that are feasible in clinical practice but do not necessarily have a high correlation with outcome, to facilitate patient inclusion and optimize efficiency in a prospective RD design.

Fuzzy RD

So far, we have discussed a sharp RD; an RD design with full adherence to the cut-off for treatment assignment. There could be cases in which assignment to treatment does not adhere fully to the cut-off. This could especially be the case in settings where retrospective data would be available to estimate treatment effectiveness with an RD design. This may result in what is called a fuzzy RD.(30) If the threshold is fuzzy(31), this means that other considerations to allocate treatment came into play that leads to the suspicion of confounding by indication.(20) If the range of miss-assignment is confined around the threshold score to a narrow range, then patients within that range can be excluded. This solution may work well only if the range being excluded is narrow, otherwise it will be difficult to accurately model the regression line near the threshold.(30) Fuzzy RD shows similarities with instrumental variable (IV) analysis; some say fuzzy RD is a form of IV.(35) In IV analysis an instrument is used to mimic randomization. In fuzzy RD the adherence of treatment assignment according to the cut-off can be used as an instrument; the analysis of the treatment effect would in this case be similar to IV analysis in which two-stage least squares (2SLS) regression analysis.

Potential applications of RD

Although RD with treatment assignment based on poorly correlating values with outcome could result in more valid and efficient effect estimates, one can debate about the feasibility of such a prospective RD design. In clinical practice there would be more support to assign treatment to the patients with a high risk on poor outcome, because

these patients would have more absolute benefit of being treated, when the relative benefit is similar over the whole range of the assignment variable. However, an RD design with treatment assignment for high risk patients would be inefficient. Moreover, application of a prospective RD design on a single baseline measurement, like blood pressure or age, which would have in general a lower correlation with outcome than a complete prognostic model, could be more practical. In clinical practice, it is common that treatment is assigned based on a single baseline measurement, and this highly resembles the RD design. A few examples (Table 4) could be thought of and are described in literature. For example eligibility of medical interventions that are assigned based on a low birth weight (babies weighing less than 1,500 g) cut-off.(24, 32) In TBI, it is recommended to treat patients with more aggressive therapy when intracranial pressure rises above 22 mmHg.(33) In HIV patients, a CD4 count threshold rule is used to determine treatment assignment for immediate vs. deferred antiretroviral therapy.(24, 31) Another example could be treatment assignment based on time, like is included in stroke guidelines. Patients with an onset-to-door time below six hours are treated with intravenous thrombolysis. Patients outside this timeframe are refrained from treatment. These are examples of treatment assignment in daily clinical practice that resemble a 'natural' application of the RD design. Observational data of these examples could be used to assess the (local) effectiveness of treatment with a retrospective application of the RD design. Based on the studies in this thesis the recommendation would be to select an assignment variable that resembles clinical practice, but not to strive for a high correlation of the assignment variable with outcome by combining multiple variables in an assignment model.

There is also potential for RD to be used in public health.(20) Often public health interventions are applied below or above a certain threshold. For example, public health interventions could be applied to a population below or above a certain age or income level. The effectiveness of such public health interventions could be assessed using an RD design.

Table 4. Examples of potential applications of the RD design.

Disease / condition	Assignment variable	Cut-off for treatment
Babies with low birth weight	Birth weight	< 1,500 g
TBI	Intracranial pressure	> 22 mmHg
HIV	CD4 count	< 350 cells/mm^3
Acute ischemic stroke	Onset-to-door time	< 6 hours
High blood pressure	Systolic blood pressure	> 140 mmHg

An example of application of the RD design in pediatric oncology care

The potential application of the RD design can furthermore be illustrated in the following example/application. We aimed to assess the effectiveness of treatment in a specialized pediatric oncology care compared to treatment in a regular hospital in pediatric oncology patients using an RD approach. Since 2018, all Dutch pediatric oncology patients are treated in one specialized pediatric oncology center in the country.(34) However, there is little evidence on whether treatment of pediatric oncology patients in specialized pediatric oncology centers is beneficial compared to treatment in a regular hospital. An RD approach was used to estimate the causal effect of being treated in a pediatric oncology center (treatment) on mortality compared to being treated with regular hospital care (control). Observational data between 2004 and 2013 of all Dutch leukemia patients and patients with an astrocytoma with age at diagnosis between 0 and 24 years was available in the nationwide Netherlands Cancer Registry. Baseline age was used as assignment variable. A (fuzzy) cut-off value of an age at diagnosis below 17 years was used for treatment assignment in a specialized pediatric oncology center (Figure 3). The treatment effect on mortality in this RD design was analyzed using Cox regression with RCS adjustment for age. A sensitivity analysis using two-stage least squares (2SLS) regression analysis was performed correcting the fuzzy treatment assignment. Preliminary results showed a significant beneficial effect of being treated in a pediatric oncology center compared to being treated with regular hospital care (Figure 4). A hazard ratio (HR) of 0.54 (95% Confidence Interval (CI): 0.34-0.88) for treatment on mortality was estimated, with RCS adjustment for age. 2SLS Cox regression showed an HR for treatment on mortality of 0.50 (95% CI: 0.29-0.86). Although this study using an RD design does not provide definite evidence on the effectiveness of treatment in specialized pediatric oncology centers, we can conclude that treatment of pediatric oncology patients in specialized pediatric oncology centers might be beneficial on mortality compared to regular hospital care.

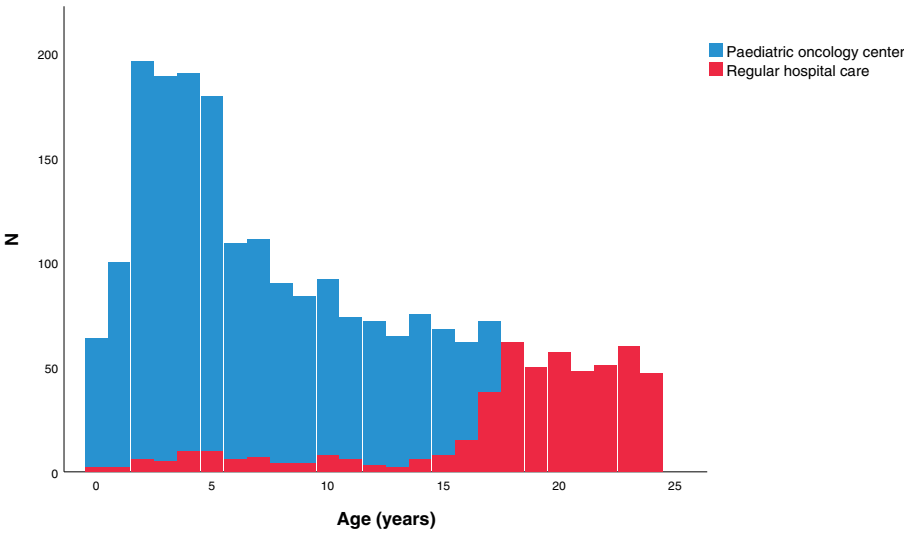

Figure 3. Histogram of the distribution of baseline age of patients treated in either pediatric oncology center or with regular hospital care.

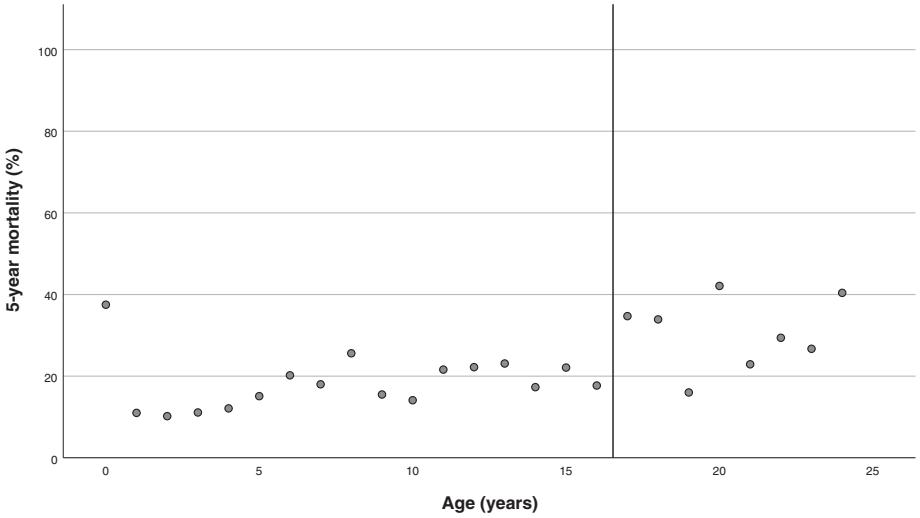

Figure 4. Scatterplot of the probability on 5-year mortality (ignoring censoring) per age.

Implications, recommendations and practical guidelines

In summary, when it is feasible to randomize (enough) patients, a randomized design is preferred over a non-randomized design, to study the effectiveness of a medical intervention. Based on this thesis, implications and specific recommendations can be made when designing a future RCT in a heterogeneous disease.

- Covariate adjustment and proportional odds analysis most efficiently use the available trial data and ensure balance between the treatment and control group to obtain reliable and valid treatment effect estimates. Both covariate adjustment and proportional odds analysis merit application in future trials in rare and heterogeneous neurological diseases like GBS.
- To apply covariate adjustment in future trials good knowledge of the prognostic value of baseline characteristics is crucial to pre-specify the covariate adjustment. These variables can be identified based on clinical experience and past literature on the prognostic value of baseline characteristics.
- The common OR from a proportional odds analysis is a fair representation of the overall effect of treatment on the (ordinal) outcome. Moreover, this approach is more efficient compared to the binary approach. Therefore, we recommend the use of the full ordinal outcome scale in future trials in rare and heterogeneous neurological diseases. The binary odds ratios for each cut-off of the ordinal outcome should be reported as well. The common OR can be presented and interpreted as a summary estimate of the treatment effect, regardless of violation of the proportional odds assumption, when there is consensus on the ordering of the outcome scale.

However, when an RCT is impossible, an RD design can be considered and is (when applicable) preferred over an observational design to assess effectiveness of a medical intervention. Based on this thesis the following implications and recommendations for the use of RD in both epidemiologic and clinical research can be made:

- In an RD design we have full understanding of the allocation of treatment, in contrast to other observational studies. The treated and control patients are exchangeable around the cut-off of the assignment variable and this enables local causal inference.
- The RD design may result in similar treatment effect estimates compared to an RCT but showed to be substantially less efficient than the RCT estimates. The assumption, of exchangeability of both treatment arms around the cut-off, can be tested, by showing a histogram of the treatment assignment variable. Without an area of overlap, continuity in the assignment variable near the cut-off is sufficient to obtain unbiased local estimates of the treatment effect.
- If it is possible to design a prospective RD design, we need sample sizes far larger than achievable in RCTs. Otherwise, large observational registry data should be available to apply a retrospective RD. Observational data of treatment assignment strategies in daily clinical practice that resemble a 'natural' application of the RD design could be used to assess the (local) effectiveness of treatment.
- With an RD design, cautious conclusions should be drawn with respect to treatment effectiveness. RD estimates should primarily be interpreted as local treatment effects since causal inference can most reasonably be drawn at the cut-off for treatment assignment. Global treatment effect estimates from RD designs should only be presented secondary to local treatment effect estimates and not as the primary parameter of interest.
- The relative efficiency compared to an adjusted analysis of the treatment effect in an RCT, was not dependent on the correlation between the treatment assignment variable and outcome since the adjustment affects the efficiency of an RCT as well. When designing a prospective RD study, we recommend researchers to use assignment variables that are feasible in clinical practice but do not necessarily have a high correlation with outcome, to facilitate patient inclusion and optimize efficiency in a prospective RD design.

In conclusion, neurologic diseases are highly heterogeneous with regard to pathogenesis and natural disease course, severity and outcome. Both heterogeneity and small sample sizes can cause insufficient statistical power to detect true treatment effect in RCTs. Covariate adjustment and proportional odds analysis are solutions for these challenges.

Based on our findings it is recommended to consider an RD design only when it is infeasible to design randomized studies to assess the effect of treatment. The RD design may be a valid alternative to estimate local treatment effects, although this design is substantially less efficient than an RCT and only cautious conclusions can be drawn.

REFERENCES

1. Hauck WW, Neuhaus JM, Kalbfleisch JD, Anderson S. A consequence of omitted covariates when estimating odds ratios. Journal of clinical epidemiology. 1991;44(1):77-81.
2. Robinson LD, Jewell NP. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. International Statistical Review / Revue Internationale de Statistique. 1991;59(2):227-40.
3. Thompson DD, Lingsma HF, Whiteley WN, Murray GD, Steyerberg EW. Covariate adjustment had similar benefits in small and large randomized controlled trials. J Clin Epidemiol. 2015;68(9):1068-75.
4. Roozenbeek B, Maas AI, Lingsma HF, Butcher I, Lu J, Marmarou A, et al. Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment? Crit Care Med. 2009;37(10):2683-90.
5. Bland JM, Altman DG. Some examples of regression towards the mean. BMJ. 1994;309(6957):780.
6. Bland JM, Altman DG. Regression towards the mean. BMJ. 1994;308(6942):1499.
7. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. BMJ. 2001;323(7321):1123-4.
8. Hernandez AV, Steyerberg EW, Butcher I, Mushkudiani N, Taylor GS, Murray GD, et al. Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. J Neurotrauma. 2006;23(9):1295-303.
9. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med. 2002;21(19):2917-30.
10. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet. 2000;355(9209):1064-9.
11. Walgaard C, Lingsma HF, Ruts L, van Doorn PA, Steyerberg EW, Jacobs BC. Early recognition of poor prognosis in Guillain-Barre syndrome. Neurology. 2011;76(11):968-75.
12. van Koningsveld R, Steyerberg EW, Hughes RA, Swan AV, van Doorn PA, Jacobs BC. A clinical prognostic scoring system for Guillain-Barre syndrome. Lancet Neurol. 2007;6(7):589-94.
13. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. Crit Care. 2011;15(3):R127.
14. Valenta Z, Pitha J, Poledne R. Proportional odds logistic regression--effective means of dealing with limited uncertainty in dichotomizing clinical outcomes. Stat Med. 2006;25(24):4227-34.
15. McHugh GS, Butcher I, Steyerberg EW, Marmarou A, Lu J, Lingsma HF, et al. A simulation study evaluating approaches to the analysis of ordinal outcome data in randomized controlled trials in traumatic brain injury: results from the IMPACT Project. Clin Trials. 2010;7(1):44-57.
16. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. Stroke. 2007;38(11):3055-62.
17. Senn S. A random effects model for ordinal responses from a crossover trial. Stat Med. 1993;12(22):2147-51.
18. Hutchinson PJ, Kolias AG, Timofeev IS, Corteen EA, Czosnyka M, Timothy J, et al. Trial of Decompressive Craniectomy for Traumatic Intracranial Hypertension. N Engl J Med. 2016;375(12):1119-30.
19. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? Stat Med. 2009;28(26):3189-209.

20. Vandenbroucke JP, le Cessie S. Commentary: regression discontinuity design: let's give it a try to evaluate medical and public health interventions. Epidemiology. 2014;25(5):738-41.
21. Labrecque JA, Kaufman JS. Commentary: Can a Quasi-experimental Design Be a Better Idea than an Experimental One? Epidemiology. 2016;27(4):500-2.
22. Lee DS, Lemieux T. Regression Discontinuity Designs in Economics. Journal of Economic Literature. 2010;48:281-355.
23. Hahn J, Todd P, Van der Klaauw W. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. Econometrica. 2001;69(1):201-9.
24. Moscoe E, Bor J, Barnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol. 2015;68(2):122-33.
25. Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J Clin Epidemiol. 2004;57(5):454-60.
26. Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? Am Heart J. 2000;139(5):745-51.
27. van Leeuwen N, Lingsma HF, de Craen AJ, Nieboer D, Mooijaart SP, Richard E, et al. Regression Discontinuity Design: Simulation and Application in Two Cardiovascular Trials with Continuous Outcomes. Epidemiology. 2016;27(4):503-11.
28. Bor J, Moscoe E, Barnighausen T. Three approaches to causal inference in regression discontinuity designs. Epidemiology. 2015;26(2):e28-30; discussion e.
29. Rubin DB. Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational Statistics. 1977;2(1):1-26.
30. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-experimental Designs for Generalized Causal Inference: Houghton Mifflin; 2002.
31. Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25(5):729-37.
32. Almond D, Doyle JJ, Kowalski AE, Williams H. Estimating Marginal Returns to Medical Care: Evidence from at-Risk Newborns. Q J Econ. 2010;125(2):591-634.
33. Carney N, Totten AM, O'Reilly C, Ullman JS, Hawryluk GW, Bell MJ, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. Neurosurgery. 2017;80(1):6-15.
34. van Goudoever H. Concentrating childhood cancer treatment in the Netherlands. Padiatr Padol. 2015;50(Suppl 2):38-41.
35 Lee DS, Lemieux T. Regression Discontinuity Designs in Economics. Journal of Economic Literature. 2010;48:281-355.