Design and Analysis of Randomized and Non-randomized Studies: Improving Validity and Reliability

Design en analyse van gerandomiseerde en niet-gerandomiseerde studies: verbeteren van validiteit en betrouwbaarheid

Nikki van Leeuwen

Layout, cover design and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

Financial support for this thesis was kindly provided by the Erasmus University Rotterdam and the Department of Public Health, Erasmus Medical Center.

Design and Analysis of Randomized and Non-randomized Studies: Improving Validity and Reliability

Design en analyse van gerandomiseerde en niet-gerandomiseerde studies: verbeteren van validiteit en betrouwbaarheid

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de rector magnificus
Prof.dr. R.C.M.E. Engels
en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 11 september 2019 om 15.30 uur

door

Nikki van Leeuwen
geboren te Rotterdam

Erasmus University Rotterdam

Promotoren

Prof.dr. E.W. Steyerberg
Prof.dr. B.C. Jacobs

Copromotor

Dr. H.F. Lingsma

Contents

I Introduction

Chapter 1

General Introduction

Over the last decades treatment guidelines have increasingly been based on the best available scientific evidence. In the early 1990s(1) the concept of Evidence Based Medicine (EBM) was introduced as a systematic approach to analyse published research as the basis of clinical decision making. The implementation of EBM was taken up worldwide by the Cochrane Collaboration. The existing medical scientific literature, with special attention to randomized trials, was summarized in tightly protocolized systematic reviews, which were then widely distributed through the Cochrane Library.(2) In this framework, meta-analysis of randomized clinical trials (RCTs) are regarded as the gold standard to provide evidence of causal effectiveness of medical interventions.(3)

However, RCTs are increasingly criticised for several reasons. First, RCTs have strict inclusion criteria limiting the generalizability for the full population of patients. Moreover, financial, ethical, and practical constraints prevent RCTs from being conducted for all clinical questions to guide clinical decision-making.(4) Also, recruitment of sufficient numbers of patients is a challenge in RCTs. Patients' treatment preferences and clinicians' lack of perceived equipoise are often cited as barriers to recruitment in RCTs.(3, 5, 6)

Recently, comparative effectiveness research (CER) gained increasing attention as a method to deliver broadly generalizable evidence on effectiveness of interventions. CER is the direct comparison of existing health care interventions to determine which work best for which patients and which pose the greatest benefits and the least harms.(7) The core question of CER is which treatment works best, for whom, and under what circumstances.(7) CER is not using data from patients with random allocation of treatments as in an RCT, but may include pragmatic RCTs or observational data that represent the current practice of treatments in 'real life'. Partly due to the ample availability of observational data, there is increasing attention for observational and quasi-experimental study designs that can be applied in such data.

The most important methodological challenge in observational data, is to determine whether the medical intervention under study is causally related to an outcome, rather than simply being correlated with another factor that is truly causally related to the outcome under study.(3) This is a particular threat as in observational studies comparison groups are different because of non-random treatment allocation. Patients are treated in accordance to the preferences of treating physicians, rather than because of a coin flip, like in randomized studies.(8, 9) These treatment choices are frequently informed by a patient's severity of illness. The treatment may be associated with outcome but could be interfered by other factors like disease severity that are causally related to outcome. Thus, observational studies assessing the causal effect of treatments are at risk of obtaining incorrect results. This type of bias is called confounding by indication.(4) It has been suggested that among non-randomized study designs, the quasi-experimental regression discontinuity (RD) design mostly resembles an RCT and overcomes confounding

by indication.(10, 11) But the methodological properties of this alternative study design are still unclear, and methods to increase the validity and efficiency need to be studied.

Thus, both randomized and non-randomized studies, like the RD design, have challenges to overcome. In this thesis, methodological challenges in both randomized and non-randomized studies are addressed. The benefits of covariate adjustment and proportional odds analysis, two different methods to optimize the validity and reliability of treatment effect estimates from RTCs in heterogeneous diseases are studied. Also, the (in)efficiency and threats to the validity of the RD design to estimate treatment effects are examined.

Randomized controlled trials

An RCT is an experimental study design in which the treatment is randomly allocated to patients. Random allocation between treatment and control group in such a study design means that patients are allocated to the groups in such a way that each participating patient has an equal chance of being allocated to either the treatment group (receiving the treatment) or the control group (not receiving the treatment).(3) All factors that can influence the outcome are on expectation equally distributed to the treatment and control group. This means when a difference in outcome between the treatment- and control group is found, this can be directly attributed to the treatment under study. The most important strength of an RCT is this controlled assignment of treatment which gives a good understanding of the assignment mechanism.(10) The treated and untreated patients in an RCT are unconditionally exchangeable.(10) This makes it possible to draw causal inference between treatment and the outcome under study in RCTs.

Nevertheless, RCTs may be difficult to set up in health care in practice for several reasons. First, the increasing complexity of regulations and logistics to conduct an RCT has raised the costs dramatically.(3, 12) Second, in part because of the high costs of RCT, an increasing proportion of studies is initiated by pharmaceutical companies that may influence the independency of the study. Third, patients may already receive a standard treatment that cannot be withheld but may interfere with the effects of a new treatment. Fourth, treating physicians may be convinced that the new treatment is better than the standard treatment and consider it unethical to withhold the new treatment even is the efficacy has not been proven.(3, 6) In addition, patient may have strong opinions on the effectiveness or risks of new treatments and not be willing to participate in a randomization. Hence, recruitment of adequate numbers of patients may be difficult in RCTs. Failure to achieve recruitment goals limits statistical precision, leads to an increase of costs, and decreases the efficiency of a RCT.(13) Even when investigators enrol a sufficient number of participants, they rarely do so on schedule.(6, 14) In addition, low recruitment rates threaten the generalizability of the findings in RCTs. A strict selection of patients enrolled in trials may poorly represent the population of interest, which limits the external validity of the results of a trial.(15, 16)

Specific challenges of RCTs in heterogeneous and rare diseases

Besides these more general limitations of RCTs, specific challenges with regard to efficiency arise when conducting RCTs in rare diseases with to the small numbers of patients(17) and in heterogeneous populations(18). In such a scenario, different approaches can be used to optimize the design and analysis in an RCT.

Random treatment allocation in RCTs ensures that observed and unobserved patient characteristics on average are similar between treatment arms.(17) However, it does not ensure full balance in small trials.(17) Differences in baseline risk on outcome other than treatment may arise between the treatment- and control group, simply due to chance.(17) In diseases with large heterogeneity in pathogenesis and natural disease course, severity and outcome, small differences in baseline risk on outcome between the treatment arms may influence the estimation of the treatment effect. In part, this effect can be compensated by increasing the number of patients included in the RCT. As indicated before, the rate of inclusion of patients is already a critical factor in most RCTs, but even more challenging in rare diseases. Small trials are also subject to a greater chance of imbalance between treatment arms than large trials.(17) Furthermore, small RCTs in rare diseases can easily fail to detect treatment benefits, due to lack of statistical power.

Covariate adjustment and ordinal outcome analysis

Two approaches to optimize the design and analysis of an RCT to increase the statistical power and to adjust for imbalances are covariate adjustment and ordinal analysis. Both approaches have been applied successfully in various acute neurological diseases such as stroke and traumatic brain injury.(19-21)

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 biased. In addition, covariate adjustment increases statistical power.(17, 18, 22) In order to adjust for covariates in RCTs, it is required to have good knowledge on the prognostic factors for outcome, as the gain in power from covariate adjustment is directly related to the predictive strength of the adjustment model.(23) Prediction research can provide information on which covariates are important to adjust for in the analysis of the treatment effect.

Ordinal analysis is an approach to analyse a full ordinal outcome scale instead of a dichotomized version. It is common in medical research to use a functional or clinical outcome scale consisting of more than two categories, but often the ordinal outcome scale is dichotomized into favorable or unfavorable outcome as primary outcome of a study. In ordinal analysis the outcome is not dichotomized but analysed as the full ordinal scale with proportional odds analysis, preventing loss of information that occurs

when dichotomizing outcome measures.(24) Both simulation studies and empirical validation studies in various fields have demonstrated that proportional odds analysis increases the statistical power of RCTs.(24-27)

Non-randomized studies

There is an increasing interest to use non-randomized and observational data to study the effectiveness of medical interventions, for example in the framework of comparative effectiveness research. However, in observational data, it is complicated to draw causal inference between treatment and outcome. The treated patients may be systematically different from the control patients. For example, physicians could treat more severely affected patients differently form less severely affected patients.(4, 28) The disease severity could influence the risk on outcome of interest and can thus be a confounder for the causal relation between treatment and outcome. When this confounder is unmeasured it is impossible to correct for it in the analysis. This can lead to bias in the treatment effect estimate. This type of bias is called confounding by indication.

Regression discontinuity design

When performing an RCT is impossible, the quasi-experimental "regression discontinuity" (RD) design is an alternative epidemiological design to study effectiveness of a medical intervention. The RD design is common in social sciences, and was introduced in public health and medicine in 1996.(29) RD has been evaluated in other fields(30-35), but the importance of studying the feasibility and robustness of this design in clinical settings has been noted.(36-38) It has been suggested that RD is the observational design that most resembles an RCT.(10, 11) In the RD design, treatment is not assigned randomly like in an RCT, but is allocated to a subset of patients, based on a cut-off of a baseline assignment variable. A subset of patients below the cut-off, not receiving a medical intervention, is considered as the control group. (Figure 1) E.g. all patients with a baseline cholesterol level 5 mmol/L may receive treatment (intervention group) and patients with a baseline cholesterol level below 5 mmol/L do not receive treatment (control group). Such treatment assignment closely resembles clinical practice especially when a standard treatment protocol is used and may thus facilitate easier recruitment of participants into a prospective, comparative study. Due to the controlled treatment assignment, an RD design achieves balance on unobserved factors between the treatment- and control group, just like in an RCT. RD may provide an opportunity to obtain unbiased causal treatment effect estimates, when an RCT is not feasible.(39) Moreover, it might be attractive to apply the RD design as a prospective study because the challenges of the randomization of patients are eluded. However, it is unclear whether the estimates from a quasi-experimental RD design might be different and substantially less efficient compared to an RCT.

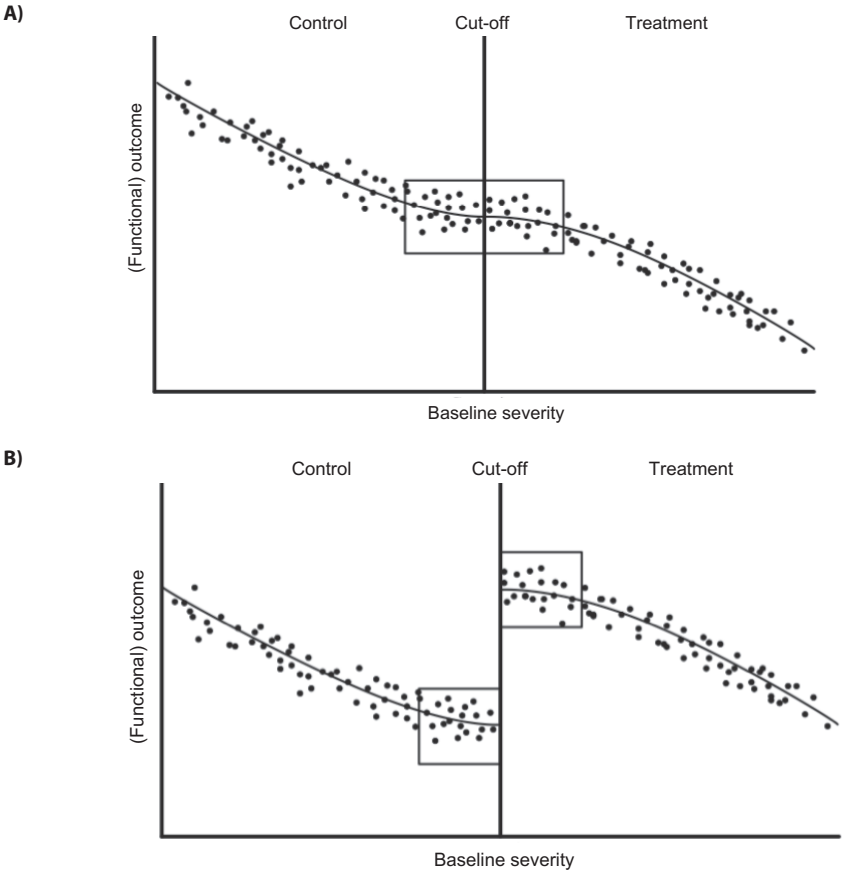


Figure 1. Graphical presentation of the regression discontinuity design in 2 studies showing no treatment effect (A) and showing a treatment effect (B).

Case-studies

The studies in this thesis test the different approaches to optimize the design and analysis of randomized and non-randomized studies in several databases on different neurological and cardio-vascular diseases.

Neurological diseases

Traumatic brain injury (TBI) is a serious public health problem with an estimated annual incidence of up to 500 cases per 100,000 population in the USA and Europe.(40-42) TBI is a major cause of death and disability, leading to great personal suffering for patients and relatives and huge direct and indirect costs to society.(40) It is defined as an injured brain as a result of an external force. TBI patients are variable with regard to causes, pathophysiology, treatment, and outcome.(40) Mild TBI patients may show full recovery, even without treatment. Severely affected TBI patients may develop serious psychologi-

cal and physical disabilities or die. A systematic literature search of the years from 1980 to 2009 revealed 27 large phase III trials in TBI; and at least further 6 unpublished trials.(43) Nevertheless, these clinical trials failed to show convincing efficacy of the treatments that were studied, mainly neuroprotective agents.(44-46) Currently the research efforts in TBI are shifting towards large observational studies to identify optimal effective treatments with CER.(47)

Guillain Barré Syndrome (GBS) is a life-threatening acute immune-mediated disorder of peripheral nerves and nerve roots (polyradiculoneuropathy)(48, 49) GBS requires early diagnosis and hospital admission for accurate monitoring, treatment and supportive care. Worldwide, the reported GBS incidence rates, vary between 0.4 and 4 per 100,000 per year, depending on age, sex, region, study methodologies and case ascertainment.(50) GBS is a heterogeneous disorder regarding pathogenesis, clinical presentation, severity and course and patients highly differ with respect to the required duration and intensity of hospital care.(51) Some patients with a mild form of GBS may show full recovery even without treatment. Other patients with a severe form of GBS may develop a full paralysis of the respiratory and limb muscles and require ventilation at an ICU for months despite treatment and may die or remain severely disabled. The current outcome of GBS is: a mortality rate of 5%, remaining unable to walk in 15% and the majority with residual complaints that interfere with daily life. In the last decade, various promising new immune-modulating treatments have been developed that may be effective in GBS as well but in this period only a very limited number of RCTs have been conducted in GBS worldwide. Because of these limitations, the treatment of GBS remained unchanged in the last 25 years.

Dementia is defined as significant loss of intellectual abilities, including memory, that is severe enough to interfere with social or occupational functioning. Increased life expectancy is associated with a steep increase of both the incidence and prevalence of dementia in the elderly. The number of 24.3 million patients that suffer from dementia is projected to almost double every 20 years to 81.1 million by the year 2040.(52, 53) Alzheimer disease is the most common cause of dementia, followed by vascular dementia.(52, 54) Treatment options for dementia are limited.(75, 76) Pharmaceutical treatment options include cholinesterase inhibitors, memantine and experimental medication. Cholinesterase inhibitors are only recommended for Alzheimer's disease and mixed dementia, not for vascular dementia or mild cognitive impairment. There is no proof of effectiveness for the other pharmaceutical options.(55) Future randomized and non-randomized studies should lead to both better prevention strategies and treatment possibilities and could help to decrease the burden of dementia.

Cardio-vascular diseases

Cardio-vascular disorders are also heterogeneous with regard to severity of symptoms, nature of clinical failure. An example of cardio-vascular diseases that is used in this thesis is acute myocardial infarction (MI). Acute MI, also known as a heart attack, is a major cause of morbidity and mortality worldwide. More than 3 million people each year are estimated to have an acute ST-elevation myocardial infarction (STEMI), with more than 4 million having a non-ST-elevation myocardial infarction (NSTEMI).(56) However, more effective treatment of patients hospitalized with acute myocardial infarction has led to a substantial decrease in deaths due to acute MI.(57) Several RCTs have established the beneficial effects and relative safety of several thrombolytic agents(58) (streptokinase(59, 60) tissue plasminogen activator(61)) and adjunctive medical therapy(62) (β-adrenergic antagonists(63), angiotensin-converting enzyme inhibitors).(64-67)

Studies used

For this thesis nine different datasets were used. An overview of the different studies, their description and in which chapters the datasets were used, is presented in Table 1.

Aim of the thesis

The 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.
The following research questions will be 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)?
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?

The thesis consists of two parts. In order to increase the validity and reliability in future RCTs in heterogeneous diseases, in part I (chapter 2, 3 and 4) the design and analysis of RCTs is studied. In chapter 2 the heterogeneity with regard to the current hospital

Table 1. Overview of datasets used in this thesis.

Abbreviation	Name	Disease	Description
PIV (68)	Pandemic Influenza & Vaccination study	GBS	The PIV study was originally designed to investigate the relation between GBS and the pandemic influenza A (H1N1) virus. Neurologists from all Dutch hospitals were requested to report patients diagnosed with GBS between November 2009 and November 2010.
PE vs IVIg trial (69)	Plasma Exchange (PE) vs Intravenous Immunoglobulin (IVIg) trial	GBS	The PE vs IVIg trial was a multicenter double-blind trial conducted between 1986 and 1989 and included 147 patients. The control group received IVIg and the treatment group received PE. The primary outcome was improvement by one or more grades on the GBS disability score after 4 weeks.
IVIg vs MP trial (70)	IVIg and placebo versus IVIg and Methyl-Prednisolone (MP) trial	GBS	In the IVIg vs MP trial, a multicenter double-blind trial, 225 patients were included between 1994 and 2000. The patients receiving IVIg and placebo were considered as control patients and the patients receiving IVIg and MP were considered as treated patients. The primary outcome was improvement by one or more grades on the GBS disability score after 4 weeks.
IMPACT (71)	International Mission on Prognosis and Clinical Trail design in TBI study	TBI	The IMPACT study combines individual patient data from 8 RCTs and three observational studies in moderate and severe TBI, mainly from the US and Europe. In Chapter 3 in this thesis we focused on the three observational studies (the European Brain Injury Consortium study (EBIC), the UK four center study (UK4), and the Traumatic Coma Databank (TCDB)). Patients were enrolled in these studies between 1984 and 1995.
CRASH (72)	Corticosteroid Randomisation After Significant Head injury trial	TBI	In the CRASH trial the effect of corticosteroids on death and disability after head injury was studied. CRASH enrolled 10,008 patients between 1999 and 2005. The primary outcome in CRASH was 14-day mortality.
TARN (73)	Trauma Audit & Research Network	TBI	TARN is a hospital based trauma registry in England and Wales including all patients with trauma resulting in immediate admission to hospital for three days or longer or death. The patients from TARN included in this study were enrolled between 1990 and 2009.
preDIVA (52)	Prevention of Dementia by Intensive Vascular Care study is	Vascular disease / dementia	An ongoing cluster-randomized trial to assess the efficacy of a multicomponent, nurse-led intervention targeting all cardiovascular risk factors in an elderly population (70-78 years). The primary outcome of this RCT is incident dementia during 6 years of follow-up. Of 3533 patients enrolled, 1894 are in the intervention and 1639 in the control group.
PROSPER (74)	PROspective Study of Pravastatin in elderly individuals at risk of vascular disease	Vascular disease	The study was conducted between December 1997 and May 1999 and enrolled 5804 patients, who were assigned to pravastatin (n=2891) or placebo (n=2913) to reduce the risk of coronary disease in elderly individuals. The outcome was a composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke at 3.2 years on average after randomization.

Table 1. (continued)

Abbreviation	Name	Disease	Description
GUSTO (61)	Global Utilization of Streptokinase and Tissue plasminogen activator for Occluded coronary arteries trial	Acute myocardial infarction	30,510 patients were entered between 1990 and 1993. 10,348 patients were assigned to treatment (accelerated tissue plasminogen activator) and 20,162 patients were used as control patients receiving streptokinase. The primary endpoint was 30-day mortality.

TBI = Traumatic Brain Injury, GBS = Guillain-Barré syndrome, RCTs = randomized controlled trials

admissions, transfers and costs in GBS is described (research question 1a). In chapter 3, also concerning research question 1a, a meta-analysis of the prognostic value of major extracranial injury in TBI patients is presented. Chapter 4 corresponds to research question 1b regarding the potential gain in efficiency of covariate adjustment and ordinal analysis in RCTs in GBS.

In part II (chapter 5, 6 and 7) the validity and reliability of the RD design compared to an RCT is addressed. Chapter 5 studies the validity and efficiency of the RD design in continuous outcomes. Similar research to chapter 5 is done in chapter 6, studying the validity and efficiency of the RD design in dichotomous outcomes. Chapter 7 focuses on the potential benefits of an alternative assignment approach to increase the efficiency of the RD design. The results of the studies in this thesis are further discussed in chapter 8, together with their implications.

REFERENCES

1. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet. 2017;390(10092):415-23.
2. Cochrane Library [Available from: http://www.cochranelibrary.com.]
3. Bowling A. Research methods in health : investigating health and health services. 2nd ed. Buckingham ; Philadelphia: Open University Press; 2002. xiv, 486 p. p.
4. Sjoding MW, Luo K, Miller MA, Iwashyna TJ. When do confounding by indication and inadequate risk adjustment bias critical care studies? A simulation study. Crit Care. 2015;19:195.
5. Smyth RM, Jacoby A, Altman DG, Gamble C, Williamson PR. The natural history of conducting and reporting clinical trials: interviews with trialists. Trials. 2015;16:16.
6. Gotay CC. Accrual to cancer clinical trials: directions from the research literature. Soc Sci Med. 1991;33(5):569-77.
7. Greenfield S, Rich E. Welcome to the Journal of Comparative Effectiveness Research. Journal of comparative effectiveness research. 2012;1(1):1-3.
8. Miettinen OS. The need for randomization in the study of intended effects. Stat Med. 1983;2(2):267-71.
9. Walker AM. Confounding by indication. Epidemiology. 1996;7(4):335-6.
10. Labrecque JA, Kaufman JS. Commentary: Can a Quasi-experimental Design Be a Better Idea than an Experimental One? Epidemiology. 2016;27(4):500-2.
11. Lee DS, Lemieux T. Regression Discontinuity Designs in Economics. Journal of Economic Literature. 2010;48:281-355.
12. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312(7040):1215-8.
13. Vozdolska R, Sano M, Aisen P, Edland SD. The net effect of alternative allocation ratios on recruitment time and trial cost. Clin Trials. 2009;6(2):126-32.
14. Halpern SD. Prospective preference assessment: a method to enhance the ethics and efficiency of randomized controlled trials. Control Clin Trials. 2002;23(3):274-88.
15. Martinson BC, Crain AL, Sherwood NE, Hayes MG, Pronk NP, O'Connor PJ. Population reach and recruitment bias in a maintenance RCT in physically active older adults. J Phys Act Health. 2010;7(1):127-35.
16. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: who are these patients and how did they get there? Ann Intern Med. 2002;137(1):10-6.
17. 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.
18. 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.
19. Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, et al. IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury. Neurotherapeutics. 2010;7(1):127-34.
20. Optimising Analysis of Stroke Trials C, Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. Stroke. 2007;38(6):1911-5.

21. Optimising the Analysis of Stroke Trials C, Gray LJ, Bath PM, Collier T. Should stroke trials adjust functional outcome for baseline prognostic factors? Stroke. 2009;40(3):888-94.
22. Turner EL, Perel P, Clayton T, Edwards P, Hernandez AV, Roberts I, et al. Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. J Clin Epidemiol. 2012;65(5):474-81.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010;63(1):64-74.
29. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-experimental Designs for Generalized Causal Inference: Houghton Mifflin; 2002.
30. Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. Journal of Policy Analysis and Management. 2008;27(4):724-50.
31. Wong VC, Cook TD, Barnett WS, Jung K. An effectiveness-based evaluation of five state prekindergarten programs. Journal of Policy Analysis and Management. 2008;27(1):122-54.
32. Aiken LS, West SG, Schwalm DE, Carroll JL, Hsiung S. Comparison of a Randomized and Two Quasi-Experimental Designs in a Single Outcome Evaluation:Efficacy of a University-Level Remedial Writing Program. Evaluation Review. 1998;22(2):207-44.
33. Goldberger AS. Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations: University of Wisconsin--Madison; 1972.
34. Buddelmeyer H. An evaluation of the performance of regression discontinuity design on progresa: [Washington, D.C. : World Bank, 2004]; 2004.
35. Black D. Evaluating the Bias of the Regression Discontinuity Design Using Experimental Data. Working paper, University of Chicago. 2007.
36. 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.
37. 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.
38. O'Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. BMJ. 2014;349:g5293.
39. 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.

40. Maas AI, Murray GD, Roozenbeek B, Lingsma HF, Butcher I, McHugh GS, et al. Advancing care for traumatic brain injury: findings from the IMPACT studies and perspectives on future research. Lancet Neurol. 2013;12(12):1200-10.
41. Langlois JA, Rutland-Brown W, Wald MM. The epidemiology and impact of traumatic brain injury: a brief overview. J Head Trauma Rehabil. 2006;21(5):375-8.
42. Styrke J, Stalnacke BM, Sojka P, Bjornstig U. Traumatic brain injuries in a well-defined population: epidemiological aspects and severity. J Neurotrauma. 2007;24(9):1425-36.
43. Maas AI, Roozenbeek B, Manley GT. Clinical trials in traumatic brain injury: past experience and current developments. Neurotherapeutics. 2010;7(1):115-26.
44. Maas AI, Steyerberg EW, Murray GD, Bullock R, Baethmann A, Marshall LF, et al. Why have recent trials of neuroprotective agents in head injury failed to show convincing efficacy? A pragmatic analysis and theoretical considerations. Neurosurgery. 1999;44(6):1286-98.
45. Bullock MR, Lyeth BG, Muizelaar JP. Current status of neuroprotection trials for traumatic brain injury: lessons from animal models and clinical studies. Neurosurgery. 1999;45(2):207-17; discussion 17-20.
46. Narayan RK, Michel ME, Ansell B, Baethmann A, Biegon A, Bracken MB, et al. Clinical trials in head injury. J Neurotrauma. 2002;19(5):503-57.
47. Maas AI, Menon DK, Steyerberg EW, Citerio G, Lecky F, Manley GT, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. Neurosurgery. 2015;76(1):67-80.
48. Hughes RA, Cornblath DR. Guillain-Barre syndrome. Lancet. 2005;366(9497):1653-66.
49. Willison HJ et al. Guillain-Barré syndroom, Lancet. 2016 Aug 13;388(10045):717-27.
50. van der Maas NA, Kramer MA, Jacobs BC, van Soest EM, Dieleman JP, Kemmeren JM, et al. Guillain-Barre syndrome: background incidence rates in The Netherlands. J Peripher Nerv Syst. 2011;16(3):243-9.
51. Fokke C, van den Berg B, Drenthen J, Walgaard C, van Doorn PA, Jacobs BC. Diagnosis of Guillain-Barre syndrome and validation of Brighton criteria. Brain. 2014;137(Pt 1):33-43.
52. Richard E, Van den Heuvel E, Moll van Charante EP, Achthoven L, Vermeulen M, Bindels PJ, et al. Prevention of dementia by intensive vascular care (PreDIVA): a cluster-randomized trial in progress. Alzheimer Dis Assoc Disord. 2009;23(3):198-204.
53. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al. Global prevalence of dementia: a Delphi consensus study. Lancet. 2005;366(9503):2112-7.
54. Fratiglioni L, Launer LJ, Andersen K, Breteler MM, Copeland JR, Dartigues JF, et al. Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology. 2000;54(11 Suppl 5):S10-5.
55. Blom K, van den Elsen GAH, Koek HL, Sanders JB, Kruithof HC, Claassen J. [Treatment of dementia]. Ned Tijdschr Geneeskd. 2017;161(0):D1851.
56. Kasliwal RR. Acute coronary syndrome. Gurgaon, Haryana, India: Elsevier; 2009.
57. Dalen JE, Alpert JS, Goldberg RJ, Weinstein RS. The epidemic of the 20(th) century: coronary heart disease. Am J Med. 2014;127(9):807-12.
58. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Lancet. 1994;343(8893):311-22.
59. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet. 1988;2(8607):349-60.

60. Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Lancet. 1986;1(8478):397-402.
61. investigators G. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. N Engl J Med. 1993;329(10):673-82.
62. Hennekens CH, Albert CM, Godfried SL, Gaziano JM, Buring JE. Adjunctive drug therapy of acute myocardial infarction--evidence from clinical trials. N Engl J Med. 1996;335(22):1660-7.
63. Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1. First International Study of Infarct Survival Collaborative Group. Lancet. 1986;2(8498):57-66.
64. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Gruppo Italiano per lo Studio della Sopravvivenza nell'infarto Miocardico. Lancet. 1994;343(8906):1115-22.
65. Oral captopril versus placebo among 13,634 patients with suspected acute myocardial infarction: interim report from the Chinese Cardiac Study (CCS-1). Lancet. 1995;345(8951):686-7.
66. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. Lancet. 1995;345(8951):669-85.
67. 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.
68. Dieleman J, Romio S, Johansen K, Weibel D, Bonhoeffer J, Sturkenboom M, et al. Guillain-Barre syndrome and adjuvanted pandemic influenza A (H1N1) 2009 vaccine: multinational case-control study in Europe. BMJ. 2011;343:d3908.
69. van der Meche FG, Schmitz PI. A randomized trial comparing intravenous immune globulin and plasma exchange in Guillain-Barre syndrome. Dutch Guillain-Barre Study Group. N Engl J Med. 1992;326(17):1123-9.
70. van Koningsveld R, Schmitz PI, Meche FG, Visser LH, Meulstee J, van Doorn PA, et al. Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barre syndrome: randomised trial. Lancet. 2004;363(9404):192-6.
71. Marmarou A, Lu J, Butcher I, McHugh GS, Mushkudiani NA, Murray GD, et al. IMPACT database of traumatic brain injury: design and description. J Neurotrauma. 2007;24(2):239-50.
72. Edwards P, Farrell B, Lomas G, Mashru R, Ritchie N, Roberts I, et al. The MRC CRASH Trial: study design, baseline data, and outcome in 1000 randomised patients in the pilot phase. Emerg Med J. 2002;19(6):510-4.
73. Lecky F, Woodford M, Yates DW. Trends in trauma care in England and Wales 1989-97. UK Trauma Audit and Research Network. Lancet. 2000;355(9217):1771-5.
74. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet. 2002;360(9346):1623-30.
75. Birks JS, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. Cochrane Database of Systematic Reviews 2018, Issue 6. Art. No.: CD001190.
76. McGuinness B, Craig D, Bullock R, Malouf R, Passmore P. Statins for the treatment of dementia. Cochrane Database of Systematic Reviews 2014, Issue 7. Art. No.: CD007514.

II Design and analysis of randomized trials in heterogeneous neurological diseases

Chapter 2

Hospital admissions, transfers and costs of Guillian-Barré syndrome

Nikki van Leeuwen
Hester F Lingsma
Ann M Vanrolleghem
Miriam CJM Sturkenboom
Pieter A van Doorn
Ewout W Steyerberg
Bart C Jacobs

PLoS One 2016

ABSTRACT

Introduction

Guillain-Barré syndrome (GBS) has a highly variable clinical course, leading to frequent transfers within and between hospitals and high associated costs. We defined the current admissions, transfers and costs in relation to disease severity of GBS.

Methods

Dutch neurologists were requested to report patients diagnosed with GBS between November 2009 and November 2010. Information regarding clinical course and transfers was obtained via neurologists and general practitioners.

Results

87 GBS patients were included with maximal GBS disability score of 1 or 2 (28%), 3 or 4 (53%), 5 (18%) and 6 (1%). Four mildly affected GBS patients were not hospital admitted. Of the 83 hospitalized patients 68 (82%) were initially admitted at a neurology department, 4 (5%) at an ICU, 4 (5%) at pediatrics, 4 (5%) at pediatrics neurology and 3 (4%) at internal medicine. Median hospital stay was 17 days (IQR 11- 26 days, absolute range 1-133 days). Transfers between departments or hospitals occurred in 33 (40%) patients and 25 (30%) were transferred 2 times or more. From a cost-effectiveness perspective 21 (25%) of the admissions was suboptimal. Median costs for hospital admission of GBS patients were 15,060 Euro (IQR 11,226 - 23,683). Maximal GBS disability score was significantly correlated with total length of stay, number of transfers, ICU admission and costs.

Conclusions

Hospital admissions for GBS patients are highly heterogeneous, with frequent transfers and higher costs for those with more severe disease. Future research should aim to develop prediction models to early identify the most cost-effective allocation in individual patients.

INTRODUCTION

Guillain-Barré syndrome (GBS) is a life-threatening immune-mediated polyradiculoneuropathy(1, 2) which requires early diagnosis and hospital admission for accurate monitoring, treatment and supportive care.

GBS was initially treated with plasma exchange (PE) in specialized centers, but since the introduction of intravenous immunoglobulin (IVIg) in 1992, care for GBS patients was highly decentralized.(3-5) GBS is a heterogeneous disorder regarding clinical presentation and course and patients highly differ with respect to the required duration and intensity of hospital care.(6) Diagnosis may be delayed, especially in patients with atypical clinical presentation, including pain(7), and in young children.(8) After admission, patients may rapidly progress and require intensive monitoring or long-term ventilator support at an Intensive Care Unit (ICU), which may not be available in smaller hospitals. Diagnostic delay and unexpected deteriorations in 8% to 16% of the patients (4, 9), may cause more (acute) transfers between wards and ICUs or between local and academic hospitals. Previous studies showed that transfers of critically ill patients and emergency intubations in general result in longer stay at the ICU(10), and have a negative impact on patient and public health.(10, 11)

Currently it is unknown in which departments and hospitals GBS patients are admitted, how often they are transferred, and what the associated costs are. In this study we aim to evaluate the current practice of hospital admissions, transfers and costs in relation to severity of disease, with the ultimate aim to provide optimal, cost-effective care for GBS patients.

METHODS

Data collection and patient population

Data from the Pandemic Influenza & Vaccination (PIV) study were used, which was originally designed to investigate the relation between GBS and the pandemic influenza A (H1N1) virus.(12) Neurologists from all Dutch hospitals were requested to report patients diagnosed with GBS between November 2009 and November 2010. All the neurologists reported the GBS patients on a voluntary basis. Consequently, information regarding diagnostic features, clinical course and transfers was obtained via neurologists, general practitioners and discharge letters from the hospital that was specifically approved by the Medical Ethical Committee of the Erasmus Medical Center in Rotterdam. Written informed consent was not given by participants for their clinical records to be used in this study. Patient information was anonymized and de-identified prior to analysis.

Definitions

All patients included fulfilled the diagnostic criteria for GBS from the Brighton Collaboration.(13) Clinical severity was defined by the GBS disability score at nadir (0 = healthy, 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 = requiring ventilation for at least a part of the day, 6 = dead). For each patient, the number of transfers was determined. Two transfers equal three beds (e.g. patient admitted at a neurology department, transferred to an ICU and transferred back to the same neurology department). We counted both transfers between hospitals and between departments within a single hospital. Hospitals were divided in three categories; local, top clinical and academic centers. Top clinical centers are non-academic "high cure" centers, which have a high level ICU facility were prolonged mechanical ventilation is possible.(14)

Specific patterns of admission and transfer

Five specific transfer patterns were identified which might be suboptimal in terms of cost-effectiveness:

1) In adults a first admission to another department than neurology or ICU, as this may indicate misdiagnosis.
2) In children ≤ 18 year first admission to another department than pediatrics neurology as children with GBS may be misdiagnosed and require specialized neurological care.(8)
3) Relatively mildly affected patients (maximal GBS disability score ≤ 3) admitted to an academic center or an ICU, as this might implicate unnecessary high costs.
4) Inter-hospital transfers from local to academic center in the first two days of hospital admission, as such a rapid deterioration might have been anticipated on with direct admission to an academic center.
5) Mechanical ventilation at the ICU in a local smaller center, as GBS patients may require mechanical ventilation for extensive periods of time and require specialized care in larger centers (at least level 2 ICU in The Netherlands).

Costs of GBS hospital admission

We included costs of admission days in general and academic hospitals, admission days at ICU, treatment with IVIg and transfers between hospitals. Costs consist of cost prices and volumes. Cost prices are the costs of one single cost unit, e.g. one admissions day. The cost prices were obtained from standardized cost-data in "Manual for cost research". (15) Costs for medical doctors, ward doctors, nurses, other staff members, equipment, medical devices, food, standard medicines, housing and overhead costs were included in the cost prices per one admission day at the intensive care unit. The costs for mechanical ventilation are not charged separate from the costs for one admission day at the in-

tensive care unit, since the costs for the equipment and extra monitoring of the patient are already included in the cost price for an admission day at the intensive care unit. Volumes are the number of a cost unit, thus the number of admission days. We did not have specific information on the type of treatment in all individual patients with GBS. In The Netherlands the first choice treatment according to the national CBO guideline for GBS is IVIg, which is also available in all centers. According to this guideline treatment is indicated in patients with GBS disability score ≥ 3) or who are transferred to the ICU and in this study costs for treatment with one course of IVIg was allocated. Consequently, by multiplying cost prices with volumes the total costs per patient were calculated. Mean and median costs with interquartile ranges (IQR) in Euros were assessed for the total study population and for each maximal GBS disability score subgroup. To assess which patient characteristics mainly determine costs, a linear regression model was fitted with age and maximal GBS disability score as independent variables and costs as dependent variable. The total costs of all GBS hospital admissions in The Netherlands per year were determined by multiplying the incidence of GBS per year in the total Dutch population by the median hospital costs.

Statistical analyses

Patient characteristics and hospital admissions were described as medians with IQRs and absolute ranges, or as frequencies. Spearman correlation coefficients (SCC) and corresponding p-values were calculated for correlations between maximal GBS disability score and total length of stay, frequency of transfers, ICU admission and days to first transfer. Similarly, correlation coefficients were calculated for maximal GBS disability score with days between hospital admission and transfer to the ICU, length of stay at ICU and total length of stay in patients admitted to an ICU during hospital stay.

All analyses were performed with SPSS 20.0 (SPSS Inc, Chicago, Illinois), figures were made with Graphpad Prism 6.01 (Graphpad Software Inc) and R statistical software 2.15.3 (R Foundation for Statistical Computation, Vienna, Austria).

RESULTS

Patient population and characteristics

The study population consisted of 87 GBS patients from a representative combination of 41 different hospitals in The Netherlands (13% academic, 33% top clinical and 58% local centers in our cohort compared to 9%, 31% and 60% in The Netherlands). The maximal GBS disability scores were: 1 or 2 (28%), 3 or 4 (53%), 5 (18%) and 6 (1%) (Table 1) and was representative for the general population of GBS patients as described in a previous Dutch observational GBS study.(7) Four (5%) patients had a relatively mild variant of

Table 1. Characteristics of 83 hospitalized patients with GBS.

Characteristics		**Missing (%)**
Age, median (IQR)	49 (30 – 64)	0 (0)
Sex, male (%)	49 (56)	1 (1)
Severity at nadir (maximal GBS disability score)*		0 (0)
1 (%)	4 (5)	
2 (%)	16 (19)	
3 (%)	25 (30)	
4 (%)	21 (25)	
5 (%)	16 (19)	
6 (%)	1 (1)	
Preceding diarrhoea (%)	13 (26)	33 (40)
Facial and/or bulbar weakness (%)	30 (36)	4 (5)
Days between onset weakness and admission, median (IQR)	2 (0 – 5)	6 (8)
Length of stay in hospital		2 (2)
median (IQR)	17 (11 – 26)	
absolute range	1 - 133	
1st hospital		0 (0)
Academic center (%)	12 (15)	
Top clinical center (%)	33 (40)	
Local center (%)	38 (46)	
Departments during hospital stay**		
Neurology (%)	74 (89)	
ICU (%)	26 (31)	
Medium Care/ Neurology- ICU (%)	4 (5)	
Internal Medicine (%)	3 (4)	
Paediatric Neurology (%)	6 (7)	
Paediatrics (%)	4 (5)	
Transfers during hospital stay		0 (0)
0 (%)	50 (60)	
1 (%)	7 (8)	
2 (%)	18 (22)	
3 (%)	6 (7)	
4 (%)	2 (2)	
Days between 1st and 2nd bed, median (IQR)***	2 (1 – 4)	5 (15)
ICU admission		
mechanical ventilation (%)	17 (65)	
Discharge direction		2 (2)
Home	41 (49)	
Rehabilitation center	37 (45)	
Nursing home	3 (4)	

Data are presented as numbers (percentages) or medians (interquartile ranges), excluding patients with missing data.

** 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, 6 = dead.*

*** These figures indicate in which departments the GBS patients were admitted at some time point during hospital stay. A proportion of patients was admitted at various departments, explaining the total number exceeds 83.*

**** Calculated for patients with at least one transfer (n=33).*

GBS, not reaching a GBS disability score >3. They were not hospitalized and excluded from further analyses. The hospitalized patients had a median age of 49 (IQR 30 – 63), with 11 (13%) children (≤ 18 years old) and 48 (59%) males.

Hospital admissions

Of 83 hospitalized patients, 12 (15%) were initially referred to an academic center, 33 (40%) to a top clinical center and 38 (46%) to a local center. The patients were initially referred to various departments: 68 (82%) to a neurology department, 4 (5%) to an ICU, 3 (4%) to internal medicine, 4 (5%) to pediatrics, and 4 (5%) to pediatric neurology. The median hospital stay was 17 days (IQR 11-26 days; absolute range 1-133 days). A higher maximal GBS disability score was significantly correlated with a longer total length of stay (SCC 0.59, $p < 0.001$) (Table 2). Of the 83 admitted patients, 33 (40%) had at least one transfer to another department or hospital, and more than 50% of patients were transferred within 2 days after admission. Moreover, 26 (31%) patients were transferred 2 times or more of which 2 (2%) were transferred 4 times. A higher maximal GBS disability score was significantly correlated with more transfers (SCC 0.62, $p < 0.001$). One patient had died in the hospital. More detailed information regarding the hospital admission is presented in Table 1. The course of hospital admission for all patients is presented in Figure 1.

Table 2. Spearman correlations with maximal GBS disability score.

Total population (n=83)	Correlation coefficient	P-value
Total length of hospital stay	0.59	<0.001
Frequency of transfers	0.62	<0.001
ICU admission*	0.67	<0.001
Days to first transfer**	0.15	0.44
ICU admissions (n=26)*		
Days between hospital admission and transfer to ICU	0.20	0.38
Length of stay at ICU	0.46	0.03
Total length of stay	0.37	0.08

** This correlation coefficient is based on the total GBS cohort. The correlation coefficient for patients with a maximal GBS disability score ≤ 4 is 0.28, p = 0.03.*

*** This correlation coefficient is based on GBS patients with at least one transfer (n=33).*

****All correlation coefficients below are calculated only for patients admitted at an ICU (n=26).*

ICU admissions

26 (31%) patients stayed at an ICU at some time during follow-up, of which 17 (65%) were ventilated. In patients with a GBS disability score of 4 or lower (i.e. not by definition admitted to the ICU) a higher maximal GBS disability score was significantly correlated to ICU admission (SCC 0.28, $p = 0.025$) (Table 2). Median time between onset of weak-

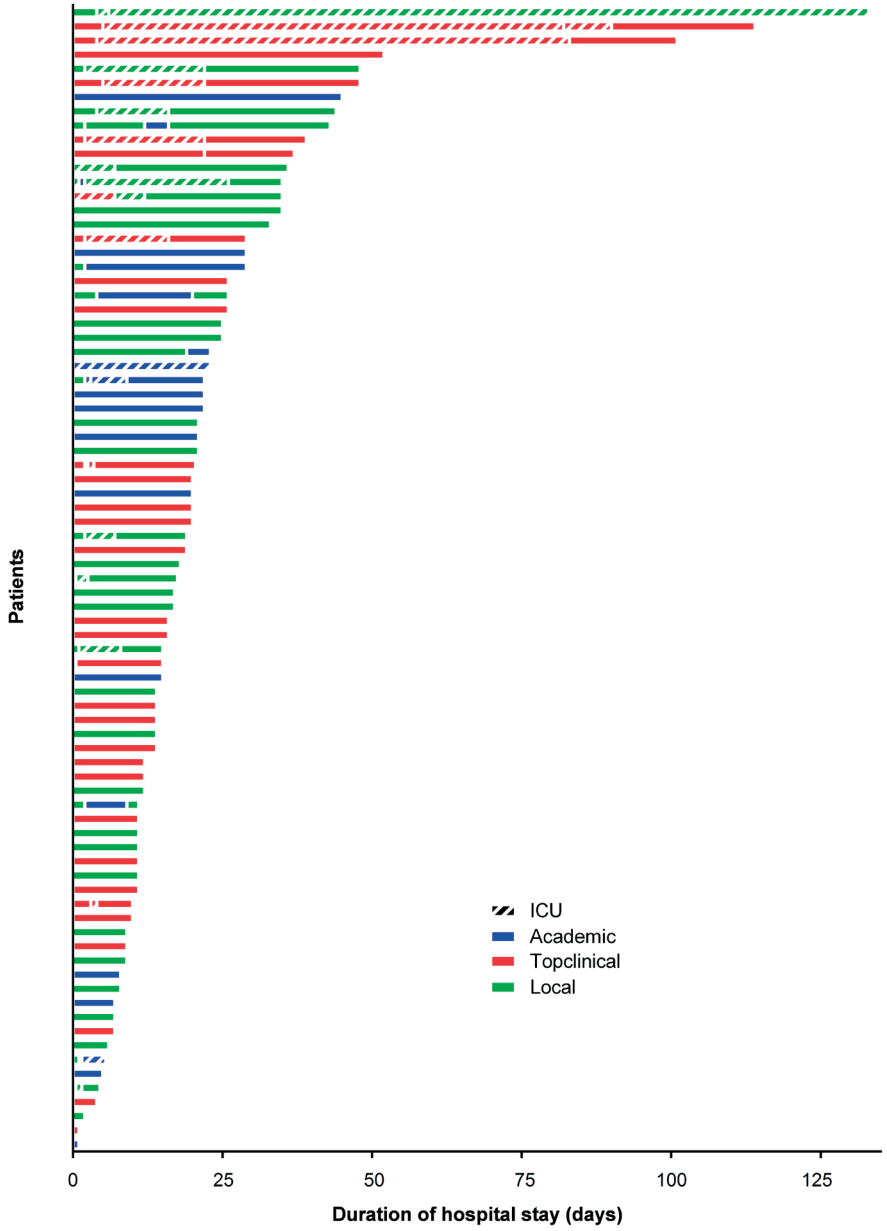


Figure 1. Overview hospital stay of 87 GBS patients.
7 patients not included in figure since limited data were available on exact days of admission and department(s) of admission.
Hatched bars are ICU admissions, non-hatched bars are admissions in any other department.

ness and admission to the ICU was 4 days (IQR 2 – 7 days; absolute range 1 – 14 days). Median length of stay at the ICU was 12 days (IQR 2-20 days). A higher maximal GBS disability score was significantly correlated to a longer stay at the ICU (SCC 0.46, p = 0.03). Median duration of ICU admission was 4 days (IQR 1 – 10 days) for patients with a maximal GBS disability score of 4 and 17 days (IQR 10 – 24 days) for patients with a maximal GBS disability score of 5. Main (documented) reasons for transfer to an ICU were (risk for) mechanical ventilation or short-term admission for observation.

Patients transferred between different types of hospitals

Nine (11%) patients were transferred between different types of centers, including 7 patients from a local center to an academic center (median time of transfer after admission was 2 days; range 1-19 days). Two patients were transferred from a top clinical center to an academic center (after 1 day) or local center (after 7 days). No patients were transferred from a local to a top clinical center, and no patients initially admitted to an academic center were transferred to another type of hospital. 78% of the inter-hospital transfers occurred in the first week of admission and in 56% of the patients within two days. Five (56%) patients were children (≤ 18 years) and 3 (33%) patients had a maximal GBS disability score 5. (S1 Table)

Specific patterns of admission and transfer

In 21 (25%) patients, the admission and transfers might be classified as suboptimal from a cost-effectiveness perspective.

1) Three (4%) patients were initially admitted to an internal medicine department.
2) Four (5%) children were initially admitted to a general pediatric department, and transferred to pediatric neurology department or ICU. Three of them were transferred from a local to an academic center.
3) Six (7%) relatively mildly affected patients (maximal GBS disability score ≤ 3) were admitted to an academic center (3; 4%) or ICU (3; 4%).
4) Four (5%) patients were transferred within 2 days of admission from a local to an academic center.
5) Seven (8%) patients were mechanically ventilated in a local center.

Costs of GBS hospital admissions

Seven patients were excluded from cost analyses because of lack of data. Median costs of the remaining 80 patients were 15,060 Euro (IQR 11,226 – 23,683 Euro), with an absolute range of 575 – 208,018 Euro. These costs were composed of admission days in a general hospital (435 Euro per day), academic hospital (575 Euro per day), or ICU (2183 Euro per day), frequency of inter-hospital transfers (262 Euro per transfer) and treatment with one course of IVIg (8,100 Euro per course). The estimated total costs for

all GBS hospital admissions in the Netherlands per year were 4,832,000 Euro (estimated frequency in total Dutch population of 200 patients multiplied by the median hospital costs of 24,160 Euro).

Median costs were highly associated with disease severity (expressed as maximal GSB disability score), ranging from 2,428 Euro (IQR 796 – 3,806 Euro) for patients with a score of 1, to 59,167 Euro (IQR 45,031 – 68,369 Euro) for patients with a score of 5 (Table 3). The correlation between GBS disability score and costs was observed in both children and adults (Figure 2).

Table 3. (Minimum) Costs* of hospital admission in GBS patients.**

Maximal GBS disability score	N	Mean costs in Euros	Median costs in Euros (IQR)
1	4	2,428	2,175 (796-3,806)
2	16	5,558	4,258 (3,045-8,644)
3	23	15,866	14,625 (13,320-17,018)
4	20	22,715	19,296 (15,219-26,384)
5	12	75,066	59,167 (45,031-68,369)
6	1	17,529	17,529
Total	**76**	**24,160**	**15,060 (11,226-23,683)**

**These costs include costs for nursing days, treatment and transfers. Costs of diagnostic tests, physiotherapy and mechanical ventilation were not included in our calculations.*
***Excluded were 7 patients from the cost analyses since limited data were available on exact days of admission and department(s) of admission.*

The beta for the effect of maximal GBS disability score on costs (adjusted for age) was 16,442 (95% CI 10,939 – 21,945). This means, for example, that the costs for a patient with a maximal GBS disability score of 4 are is on average 16,442 Euro higher than those for a patient with a maximal GBS disability score of 3.

DISCUSSION

In this study the current practice of hospital admissions of patients with GBS was evaluated in terms of location, duration, transfers and costs, in a representative cohort of GBS patients in The Netherlands. Transfers within and between hospitals were frequent: 40% of the patients were transferred at least one time and half of them were transferred within 2 days of admission. Moreover, in 25% the admission may have been suboptimal form a cost-effectiveness perspective, including admission to other than (paediatric) 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. These findings may suggest

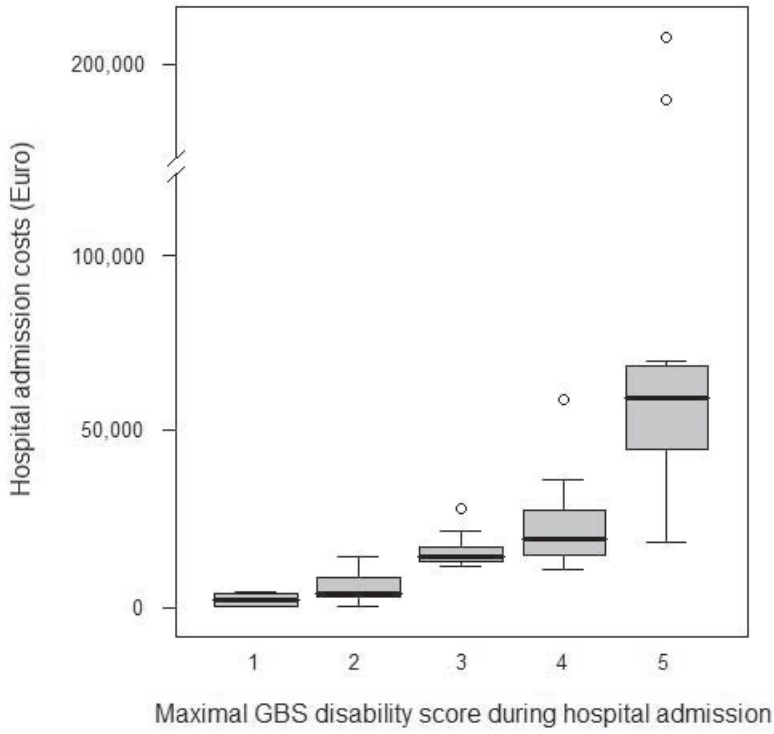


Figure 2. Interquartile ranges (grey boxes), 95% confidence intervals (whiskers) and median (dark lines in middle of the boxes) of costs of hospital admission for different maximal GBS disability scores.
Excluded was one patient who died.
Circles are (extreme) outliers.
Maximal GBS disability score during 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.

that the care of GBS patients in The Netherlands can be improved by developing more cost-effective referral strategies based on early diagnosis and prediction of clinical course and outcome.

Strengths and limitations

Only very few studies have described the practice of current hospital admission of patients with GBS. Most studies on the clinical course of GBS are based on data from therapeutic trials, which may be biased to severe cases. Although reporting of GBS cases by the neurologists was voluntary in this study, we had a representative cohort of GBS patients. In the Netherlands, all care for all patients with GBS is primarily coordinated by neurologists. Therefor it is highly unlikely that GBS cases were missed because treatment was coordinated at another department. The types of hospital (academic, top clinical

and local) were similarly distributed as the total number of hospitals in the Netherlands. The distribution of age, disease severity at nadir, proportion of ventilated patients was similar to previous studies on GBS patients in The Netherlands. Previous studies were performed in the United States, which has a different health care system than European countries, and focused largely on indirect costs.(16) Other studies only measured the costs of specific treatments for GBS(17-19) or analyzed costs of a specific subgroup of GBS.(20) In the current study we aimed to determine the current costs of hospital admissions across the full spectrum of this heterogeneous disorder.

Optimal and cost-effective care for GBS

GBS is a complex disorder for cost-effective care because of the various stages in the clinical course and diversity in clinical course between patients. The complexity is reflected in the high frequency of transfers between departments and hospitals, especially shortly after initial admission. Patients initially admitted at the internal medicine department may result in delayed specialized treatment and monitoring, and an extra transfer. From a costs point of view, ideally mildly affected patients are admitted to a local or top clinical center with good general care for GBS but relatively low costs. More severely affected patients with a higher chance of respiratory failure and complications may benefit from admission in a top-clinical or academic center. Four patients were transferred from a local to an academic center within two days of admission and ideally these patients would have been admitted directly to a specialized center. Adequate assessment of prognosis could aid decision making at the time of admission. Prognostic models have been developed to support this assessment, including the externally validated modified Erasmus GBS Outcome Score (mEGOS) to predict disability outcome in GBS patients at the time of admission.(21)

Seven patients were initially mechanically ventilated in a local center, which could have been prevented when earlier transferred to a top clinical or academic center. The Erasmus GBS Respiratory Insufficiency Score (ERGIS) (22) was developed to predict respiratory insufficiency at time of admission. When a patient has a high chance of respiratory insufficiency, careful monitoring can potentially avoid an unexpected emergency intubation and acute transfer to the ICU. The ERGIS could help clinicians to decide to admit or transfer a patient to an academic center before the critical stage of disease. Direct admission to a top clinical or academic is preferred above transfer since inter-hospital transfers have negative impact on patient outcome.(10, 11) ERGIS could also help avoiding unnecessary ICU admissions of mild GBS patients to save costs. In this study, 9 of the ICU admitted patients had no need for mechanical ventilation. We cannot exclude the possibility that these patients were admitted to an ICU because of autonomic dysfunction, although this sole indication for admission to an ICU may be relatively rare.

Children with GBS

Almost half of the children in this cohort were initially admitted to a general pediatrics department. Considering the challenging neurologic examination, monitoring and treatment of children with GBS, they should preferably be seen by a pediatric neurologist and be admitted to a center with a pediatric ICU. All children initially admitted to a pediatric department were later transferred to a pediatric neurology department or ICU. This referral pattern may indicate a delay in diagnosis of GBS in young children compared to adults or problems with monitoring children during the progressive state.(8) In one child, admitted to a pediatric department in a local center, the delayed diagnosis and insufficient monitoring resulted in death due to hypoxia after emergency intubation. (8, 23)

Costs of GBS hospital admission

We found that the costs of hospitals admission in GBS are highly variable and mainly depend on maximal GBS disability score. These are the minimal costs of GBS hospital admissions, since costs of diagnostic tests, physiotherapy and mechanical ventilation were not included. Moreover, one course with IVIg for each patient was assigned, although some patients may have received more (or no) course(s) with IVIg due to treatment related fluctuations or received other treatment like PE.

Length of stay is the main driver for high costs in GBS hospital admission, especially (long) admission to an ICU. This also explains the strong correlation between costs and GBS disability score. Compared to other costs during hospital admission, costs for inter hospital transfers are relatively low. Also, a course with IVIg (8,100 Euro), although considered to be an expensive treatment, has relatively low costs compared to ICU admission (2,183 Euro per day).

Conclusion

In conclusion, substantial heterogeneity in admission and transfer patters of GBS patients and associated costs was found. As this study lacks outcome data, no definite conclusions can be drawn, but we suggested several possibilities for improving to cost-effectiveness of care for GBS patients. Future research should focus on identifying subgroups of patients who benefit most from specialized care in an academic center, e.g. based on prognostic models, and subsequently on developing admission guidelines to provide optimal, cost-effective care for GBS patients.

REFERENCES

1. Hughes RA, Cornblath DR. Guilain-Barre syndrome. Lancet. 2005 Nov;366(9497):1653-66.
2. Hughes RA, Rees JH. Clinical and epidemiologic features of Guillain-Barre syndrome. J Infect Dis. 1997 Dec;176 Suppl 2:S92-8.
3. Van der Meché FG, Schmitz PI. A randomized trial comparing intravenous immune globulin and plasma exchange in Guillain-Barré syndrome. Dutch Guillain-Barré Study Group. N Engl J Med. 1992 Apr;326(17):1123-9.
4. Hughes RA, Swan AV, Raphaël JC, Annane D, Van Koningsveld R, Van Doorn PA. Immunotherapy for Guillain-Barré syndrome: a systematic review. Brain. 2007 Sep;130(Pt 9):2245-57.
5. Van Koningsveld R, Van Doorn PA, Schmitz PI, Van der Meché FG. Changes in referral pattern and its effect on outcome in patients with Guillain-Barré syndrome. Neurolgy. 2001 Feb;56(4):564-6.
6. Fokke C, Van den Berg B, Drenthen J, Walgaard C, Van Doorn PA, Jacobs BC. Diagnosis of Guillain-Barré syndrome and validation of Brighton criteria. Brain. 2014 Jan;137(Pt 1):33-43.
7. Ruts L, Drenthen J, Jongen JL, Hop WC, Visser GH, Jacobs BC, et al. ; Dutch GBS Study Group. Pain in Guillain-Barré syndrome: a long-term follow-up study. Neurology. 2010 Oct;75(16):1439-47.
8. Roodbol J, De Wit MC, Walgaard C, De Hoog M, Catsman-Berrevoets CE, Jacobs BC. Recognizing Guillain-Barre syndrome in preschool children. Neurology. 2011 Mar;76(9):907-10.
9. Ruts L, Van Koningsveld R, Van Doorn PA. Distinguishing acute-onset CIDP from Guillain-Barré syndrome with treatment related fluctuations. Neurology. 2005 Jul;65(1):138-40.
10. Barrat H, Harrison DA, Rowan KM, Raine R. Effect of non-clinical inter-hospital critical care unit to unit transfer of critically ill patients: a propensity-matched cohort analysis. Crit Care. 2012 Oct;16(5):R179.
11. Ligtenberg JJ, Arnold LG, Stienstra Y, Van der Werf TS, Meertens JH, Tulleken JE, et al. Quality of interhospital transport of critically ill patients: a prospective audit. Crit Care. 2005 Aug;9(4):R446-51.
12. Dieleman JP, Romio S, De Vries CS, Sammon C, Andrews N, Hviid A, et al. Guillain-Barre syndrome and adjuvanted pandemic influenza A (H1N1) 2009 vaccine: multinational case-control study in Europe. BMJ. 2011 Jul;12(343):d3908.
13. Sejvar JJ, Kohl KS, Gidudu J, Amato A, Bakshi N, Baxter R, et al. Guillain-Barré syndrome and Fisher syndrome: case definitions and guidelines for collection, analysis and presentation of immunization safety data. Vaccine. 2011 Jan;29(3):599-612.
14. STZ [homepage on the Internet]. Bureau Samenwerkende Topklinische opleidingsZiekenhuizen; [cited 5 February 2013]. Avaiable from: http://www.stz.nl/
15. Handleiding voor kostenonderzoek, methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg [homepage on the Internet]. College voor zorgverzekeringen; [updated January 2010.; cited 5 February 2013]. Avaiable from: http://www.cvz.nl/binaries/live/cvzinternet/hst_content/nl/documenten/losse-publicaties/handleiding-kostenonderzoek-2010.pdf
16. Frenzen PD. Economic cost of Guillain-Barré syndrome in the United States. Neurology. 2008 Jul;71(1):21-7.
17. Winters JL, Brown D, Hazard E, Chainani A, Andrzejewski C Jr. Cost-minimization analysis of the direct costs of TPE and IVIg in the treatment of Guillain-Barré syndrome. BMC Health Serv Res. 2011 May;11:101.
18. Mehndiratta MM, Chowdhury D, Goel V. Efficacy and cost effectiveness of current therapies in Guillain Barre syndrome. J Assoc Physicians India. 2001 Apr;49:459-69.

19. Nagpal S, Benstead T, Shumak K, Rock G, Brown M, Anderson DR. Treatment of Guillain-Barré syndrome: a cost-effectiveness analysis. J Clin Apher. 1999;14(3):107-13.
20. Havelaar AH, De Wit MA, Van Koningsveld R, Van Kempen E. Health burden in the Netherlands due to infection with thermophilic Campylobacter. Epidemiol Infect. 2000 Dec;125(3):505-22.
21. Walgaard C, Lingsma HF, Ruts L, Van Doorn PA, Steyerberg EW, Jacobs BC. Early recognition of poor prognosis in Guillain-Barré syndrome. Neurology. 2011 Mar;76(11):968-75.
22. Walgaard C, Lingsma HF, Ruts L, Drenthen J, Van Koningsveld R, Garssen MJ, et al. Prediction of respiratory insufficiency in Guillain-Barré syndrome. Ann Neurol. 2010 Jun;67(6):781-7.
23. De Wit MC, Roodbol J, De Hoog M, Catsman-Berrevoets CE, Jacobs BC. Imminent respiratory insufficiency in childeren resulting from Guillain-Barré syndrome. Ned Tijdschr Geneeskd. 2011;155(42):A3808.

S1 Table. Transfers between different types of hospitals (n=9).

Age	Maximal GBS disability score	Length of stay in 1st hospital	Department admission 1st hospital	Reason for transfer	Length of stay in 2nd hospital	Department admission 2nd hospital	Transferred back to 1st hospital	Total length of stay in hospital
From top clinical to academic center (n=1)*								
10	5	1	Pediatrics Neurology	Unknown	19	ICU	Yes	N/A
From top clinical to local center (n=1)*								
21	5	7	ICU	Parents of patient live closer to the local center on the other side of the county	28	ICU	No	35
From local to academic center (n=7)*								
28	4	19	Neurology	Unknown	4	Neurology	No	23
5	4	2	Pediatrics	Unknown	27	Pediatric Neurology	No	29
11	3	4	Pediatrics	Unknown	16	Pediatric Neurology	Yes	26
36	4	2	Neurology	Unknown	7	Neurology	Yes	11
6	4	2	Pediatrics	Unknown	17	Neurology	Yes	22
4	6	1	Pediatrics	Unknown	4	ICU	-	5
59	3	12	Neurology	Risk for mechanical ventilation	4	Neurology	Yes	43

** There were no transfers from an academic center to another type of hospital and no transfers from a local center to a top clinical center.*

Chapter **3**

Prognostic value of major extracranial injury in traumatic brain injury: an individual patient data meta-analysis in 39,274 patients

Nikki van Leeuwen
Hester F Lingsma
Pablo Perel
Fiona Lecky
Bob Roozenbeek
Juan Lu
Haleema Shakur
James Weir
Ewout W Steyerberg
Andrew I R Maas

On behalf of the IMPACT study group, the CRASH trial collaborators and the Trauma Audit and Research Network

Neurosurgery 2012

ABSTRACT

Introduction

Major extracranial injury (MEI) is common in Traumatic Brain Injury (TBI) patients, but the effect on outcome is controversial.

Objective

To assess the prognostic value of MEI on mortality after TBI in an individual patient data meta-analysis of three observational TBI studies (IMPACT), a randomized controlled trial (CRASH), and a trauma registry (TARN).

Methods

MEI (extracranial injury with an AIS ≥ 3 or "requiring hospital admission") was related to mortality with logistic regression analysis, adjusted for age, GCS motor score and pupil reactivity, stratified by TBI severity. We pooled odds ratios (ORs) with random effects meta-analysis.

Results

We included 39,274 patients. Mortality was 25% and 32% had MEI. MEI was a strong predictor for mortality in TARN, with adjusted ORs and 95% confidence intervals (95%CI) of 2.81 (2.44-3.23) in mild, 2.18 (1.80-2.65) in moderate and 2.14 (1.95-2.35) in severe TBI patients. The prognostic effect was smaller in IMPACT and CRASH with pooled adjusted ORs and 95%CIs of 2.14 (0.93-4.91) in mild, 1.46 (1.14-1.85) in moderate and 1.18 (1.03-1.55) in severe TBI. When patients who died within 6 hours after injury were excluded from TARN, the effect of MEI was comparable with IMPACT and CRASH.

Conclusion

MEI is an important prognostic factor for mortality in TBI patients. However, the effect varies by population, which explains the controversy in the literature. The strength of the effect is smaller in patients with more severe brain injury, and depends on time of inclusion in a study.

INTRODUCTION

Major extracranial injury (MEI) is frequently present in patients with traumatic brain injury (TBI). The prevalence differs from 23%(1) to 41%(2) dependent on study population and definition of MEI. Relatively few studies have however focused on the effect of MEI on mortality after TBI. Most studies concerning TBI and MEI have investigated patients with extracranial trauma, with or without TBI. These studies show that the coexistence of traumatic brain injury with extracranial injury is associated with both increased mortality and morbidity.(3-6)

In contrast, there is no consensus on the degree to which the presence of MEI worsens outcome in TBI patients. Some studies demonstrate that outcome mainly depends on the severity of the primary cerebral damage and is not worsened by the presence of extracranial injuries.(2, 7) Other studies suggest that the presence of MEI carries a poorer outcome in TBI patients.(1, 8-10) Differences between studies might be due to patient population, setting and study design. Determining the importance of MEI in outcome after TBI has relevance for understanding and potentially improving the patient pathway, and for improving prognostic models that might be used to benchmark care(6), or to inform relatives and medical decisions.

We report a collaborative analysis on a large number of TBI patients with and without documented MEI, including data from the International Mission on Prognosis and Clinical Trial design in TBI (IMPACT) study, the Medical Research Council Corticosteroid Randomization after Significant Head Injury (MRC CRASH) trial, and the Trauma Audit & Research Network (TARN) registry. Our aim was to determine the role of MEI as a prognostic factor for mortality after TBI and to solve the current disagreement in the literature. We hypothesize that the presence of MEI is associated with higher mortality in patients with TBI.

METHODS

Patient population and data collection

We included individual patient data from the International Mission on Prognosis and Clinical Trail design in TBI (IMPACT) study, the Medical Research Council Corticosteroid Randomization after Significant Head Injury (MRC CRASH) trial, and the Trauma Audit & Research Network (TARN).

IMPACT combines individual patient data from randomized controlled trials (RCTs) and three observational studies in moderate and severe TBI, mainly from the US and Europe. Here we focused on the three observational studies (the European Brain Injury Consortium core data survey (EBIC), the UK four centre study (UK4), and the Traumatic

Coma Databank (TCDB)), as the presence of MEI was not an exclusion criterium for these studies. Patients were enrolled in these studies between 1984 and 1995.

The CRASH trial is a trial with broad inclusion criteria studying the effect of corticosteroids on death and disability after head injury. CRASH was conducted in both high and low/middle income countries. In CRASH we analyzed low/middle income countries and high income countries separately, as trauma organizations may be different.(1) CRASH enrolled 10,008 patients between 1999 and 2005, of which 9554 had complete outcome data.

TARN is a hospital based trauma registry in England and Wales including all patients with trauma resulting in immediate admission to hospital for three days or longer or death. From these, we selected TBI patients defined as having an Abbreviated Injury Scale for the Head Region of 3 or higher, which was not resulting from scalp laceration, scalp avulsion or penetrating injury. The patients from TARN included in this study were enrolled between 1990 and 2009.

Detailed descriptions of all the studies and data collection and management can be found in previous publications.(11-13)

Outcome and major extracranial injury

The primary outcome examined in this analysis was mortality at six months in IMPACT and CRASH and discharge mortality in TARN. In IMPACT, six-month mortality was missing in 3 patients who were excluded. CRASH had also 14 day mortality available. Major Extracranial Injury (MEI) was defined as "Abbreviated Injury Scale (AIS) ≥ 3" or "an injury requiring hospital admission on its own".

Statistical analyses

The strength of the association between MEI and mortality was analyzed univariably and multivariably using binary logistic regression models. We adjusted for core prognostic parameters: age, GCS motor score (1=makes no movements, 2=extension to painful stimuli, 3=abnormal flexion to painful stimuli, 4 =flexion/withdrawal to painful stimuli, 5=localizes painful stimuli, 6=obeys commands) and pupil reactivity (1= both responsive, 2=one responsive, 3=both unresponsive) at admission. We also adjusted for hypotension (prior to hospital admission) to better understand the pathway of the prognostic effect of MEI. When IMPACT was analyzed as a single study (in mild and moderate TBI), we additionally adjusted for study, since IMPACT actually consists of three studies. In CRASH we also adjusted for treatment by adding the treatment variable to the multivariable regression model, since there was a significant treatment effect. Since the patients in TARN and IMPACT were included in a wide time range, we tested for interaction between MEI and year of injury.

Results were expressed as odds ratio for mortality with MEI compared to absent MEI, with 95% confidence intervals. An overall summary measure was derived using random effects meta-analysis (Der Simonian-Laird pooling). We assessed the heterogeneity between the studies based on the between-study variance τ^2 and its p-value to test for heterogeneity.

TARN was not included in the pooled analysis because of the different nature of the study and the different time point of the outcome. Forest plots were used to display consistency of findings across the datasets. We calculated partial R^2 statistics to indicate the amount of variance explained by MEI, both univariable and multivariable. In CRASH and IMPACT we corrected the univariable and multivariable R^2s for the variance explained by study and treatment.

Absolute risks of patients with and without MEI were calculated from the models by taking the mean of the probabilities predicted by the multivariable models, stratified for brain injury severity.

Missing data is common in medical scientific research. One distinguishes three types of mechanisms leading to missing values. Missing completely at random (MCAR) are missing values due to for example administrative errors or accidents. Missingness related to known patient characteristics, time or place is called missing at random (MAR). The third mechanism, missing not at random (MNAR), is a problematic situation in which missingness is related to unknown predictors. In epidemiology, it is generally acknowledged that imputation is preferable over complete case analysis in case of missing values. (14-17) Estimating associations using complete case analysis is less efficient, since part of the data is not used. The simplest approach for imputation ('simple imputation') is imputing a fixed value for all patients with a missing value for a particular variable, e.g. the mean or the most common category. Such simple methods ignore the correlation between variables and are hence suboptimal. In 'single imputation', multivariable regression models are used to predict the missing value based on associations with other variables. In multiple imputation this procedure is repeated several times resulting in multiple datasets, all with slightly different imputed values. Subsequent analyses are performed on each dataset separately and summarized to obtain more precise standard errors and P-values.(15) The assumption underlying single and multiple imputation is that missing values are MAR.

In our study, missing data were imputed for the motor score of the Glasgow Coma Scale (GCS), pupil reactivity and MEI with single imputation using all relevant prognostic factors and outcome. We thus assume MAR. Imputations were done separately for TARN, CRASH and IMPACT, using the *AregImpute* function in R statistical software.

Analyses were performed with R statistical software 2.7.1 (R Foundation for Statistical Computation, Vienna) using packages *Rmeta*, *Hmisc* and *Design*, and SPSS 15.0 (SPSS Inc, Chicago).

Sensitivity analyses

In preliminary analysis we found a large difference between IMPACT and CRASH versus TARN in terms of the effect of MEI on outcome. We hypothesized that this might be due to the different setting (TBI studies versus a trauma registry), the different distribution of TBI severity across the studies (only moderate and severe TBI in IMPACT, many mild TBI patients in TARN), or the different time point of outcome assessment (discharge versus 6 month). We tested these hypotheses by three approaches.

1) We tested for interaction between MEI and brain injury severity (GCS), by adding an interaction term between MEI and GCS to the binary logistic regression model containing age, GCS motor score, pupil reactivity, MEI and GCS as main effects. We assessed the p-value of the interaction term and subsequently stratified the analyses for brain injury severity, defining mild TBI as Glasgow Coma Scale (GCS) 13-15, moderate TBI as GCS 9-12 and severe traumatic brain injury as GCS 3-8.
2) We excluded the patients from TARN who died within 6 hours after injury since the majority of these patients is not likely to be included in IMPACT or CRASH.
3) We analyzed in CRASH both 14 day and 6 month mortality.

RESULTS

Patient population

We included 2,216 patients from IMPACT (791 from UK4, 603 from TCDB, and 824 from EBIC), 9,554 from CRASH (7,205 from low/middle income countries, and 2,349 from high income countries), and 27,504 from TARN. This resulted in 39,274 patients for the analysis. For all variables missing was less than 10%, except for TARN where 90% of the pupil reactivity data was missing since this variable was only recorded from 2005 onwards.

Patient characteristics

The majority of the patients (17,136, 44%) had severe TBI. A total of 7,229 (18%) had moderate and 14,909 (38%) had mild TBI. The IMPACT studies included mainly severe TBI patients (81%) and TARN mainly mild (43%) and severe (42%) TBI patients. In CRASH the distribution of brain injury severity was more equal (30% mild, 30% moderate, 40% severe). In IMPACT, mortality was 41%, compared to 24% in CRASH and 28% in TARN. In IMPACT, 41% of the patients had MEI, in CRASH this was 23% and in TARN 34%. MEI was observed more frequently in patients with severe TBI (30-46%), than in those with mild TBI (14-41%). (Table 1)

Table 1. Patient Characteristics of 11 Studies in the IMPACT database, the CRASH trial and the TARN registry.

	Age	**GCS score**	**Motor score**	**Pupillary reactivity**	**Major extracranial injury**	**Mortality**
	median (25^{th}-75^{th} percentile)	*Mild - GCS 13-15* *Moderate - GCS 9-12* *Severe - GCS 3-8*	*none extension* *abnormal flexion* *normal flexion* *localize/obeys* *untestable/ missing*	*both responsive* *one responsive* *both unresponsive*	*yes*	*dead*
UK4 (n=791)	36 (22-55)	24 (3%) 83 (11%) 684 (87%)	113 (14%) 85 (11%) 37 (5%) 141 (18%) 221 (28%) 194 (26%)	434 (55%) 113 (14%) 244 (31%)	303 (38%)	359 (45%)
TCDB (n=603)	26 (21-40)	22 (4%) 45 (8%) 536 (89%)	136 (23%) 107 (18%) 74 (12%) 121 (20%) 134 (22%) 31 (5%)	299 (50%) 55 (9%) 249 (41%)	280 (46%)	264 (44%)
EBIC (n=822)	37.5 (24-59)	73 (9%) 168 (20%) 581 (71%)	150 (18.2%) 80 (10%) 55 (7%) 113 (14%) 281 (34%) 143 (17%)	532 (65%) 80 (10%) 210 (26%)	316 (38%)	281 (34%)
CRASH LOW/ MIDDLE INCOME (n=7,205)	32 (24-45)	2108 (29%) 2331 (32%) 2766 (38%)	356 (5%) 403 (6%) 531 (7%) 891 (12%) 5024 (70%) 0 (0%)	6135 (85%) 450 (6%) 620 (9%)	1694 (23%)	1854 (26%)
CRASH HIGH INCOME (n=2,349)	37 (24-54)	760 (32%) 551 (24%) 1038 (44%)	429 (18%) 112 (5%) 128 (5%) 290 (12%) 1390 (59%) 0 (0%)	1965 (84%) 147 (6%) 237 (10%)	522 (23%)	469 (20%)
TARN (n=27,504)	39 (24-60)	11922 (43%) 4051 (15%) 11531 (42%)	4117 (15%) 838 (3%) 973 (4%) 1449 (5%) 11892 (43%) 8235 (30%)	21548 (78%) 1630 (6%) 4326 (16%)	9452 (34%)	7673 (28%)

Major Extracranial Injury and mortality

We found a moderate prognostic effect of MEI in IMPACT and CRASH with pooled adjusted ORs and 95% confidence intervals (95%CIs) of 2.14 (0.93-4.91) in mild, 1.46 (1.14-1.85) in moderate and 1.18 (1.03-1.55) in severe TBI patients. The between-study variances τ^2 and p-values for heterogeneity were 0.39 (p=0.02) for the mild, 0.11 (p=0.10) for the moderate and 0.0 (p=0.98) for the severe TBI studies. In TARN MEI was a strong prognostic factor for mortality, with adjusted odds ratios (OR) and 95%CIs of 2.81 (2.44-3.23) in mild, 2.18 (1.80-2.65) in moderate and 2.14 (1.95-2.35) in severe TBI patients (Figure 1 and Table 2). The unadjusted ORs were all smaller than adjusted ORs, indicating that the effect of MEI on mortality was independent of other predictors of mortality.

Adjusting the effect of extracranial injury for hypotension led to a small decrease of the prognostic effect (ORs decreasing by 0.1-0.4) of MEI, indicating that hypotension indeed explains part of the relationship between extracranial injury and outcome. Hypotension itself was a strong prognostic factor for mortality, independent of MEI (adjusted ORs 2.9 to 3.6).

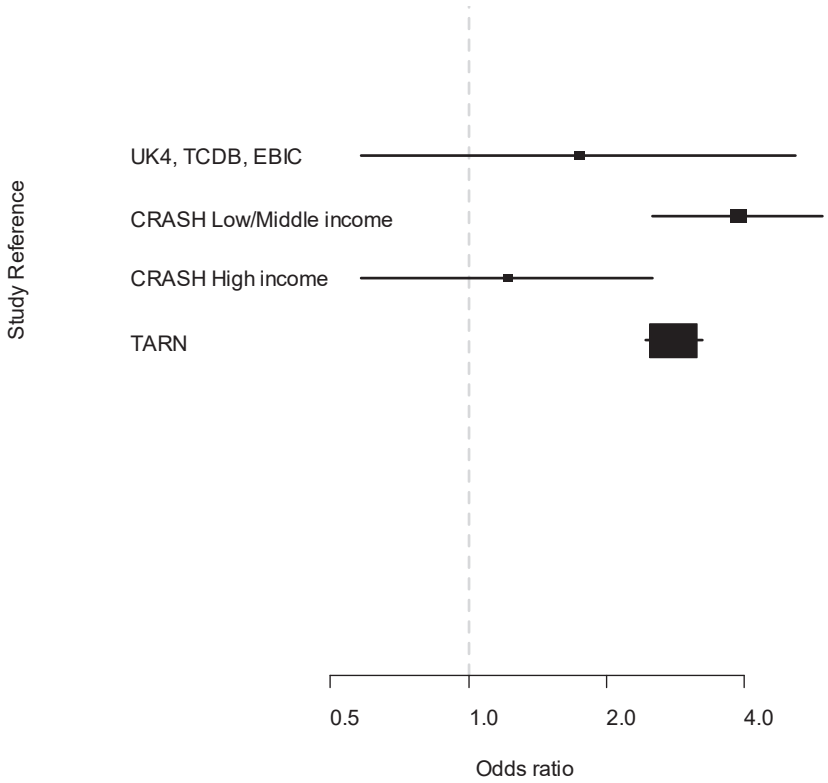


Figure 1. Forest plots showing the strength of the adjusted association between major extracranial injury and mortality in mild (left), moderate (middle) and severe (right) TBI patients

Table 2. Associations between major extracranial injury (versus no and minor extracranial injury) and mortality in the IMPACT, CRASH and TARN data in mild, moderate and severe TBI patients.

	Mild TBI (GCS 13-15)				**Moderate TBI** (GCS 9-12)				**Severe TBI** (GCS 3-8)			
	N Mortality (%)	*N Major extracranial injury (%)*	*Unadjusted*	*Adjusted*	*N Mortality (%)*	*N Major extracranial injury (%)*	*Unadjusted*	*Adjusted*	*N Mortality (%)*	*N Major extracranial injury (%)*	*Unadjusted*	*Adjusted*
UK4	10 (42)	9 (38)	1.14 (0.46-2.86)	1.75 (0.56-5.46)	34 (41)	33 (40)	0.70 (0.39-1.26)	1.15 (0.55-2.41)	315 (46)	261 (38)	0.90 (0.66-1.23)	1.21 (0.81-1.80)
TCDB	6 (27)	9 (41)	-	-	9 (20)	24 (53)	-	-	249 (47)	247 (46)	0.98 (0.70-1.38)	1.13 (0.74-1.73)
EBIC	11 (15)	19 (26)	-	-	26 (16)	51 (30)	-	-	244 (42)	246 (42)	0.79 (0.56-1.10)	1.12 (0.73-1.71)
CRASH L-M income	112 (5)	309 (14)	3.96 (2.64-5.94)	3.86 (2.52-5.91)	348 (15)	530 (22)	1.48 (1.15-1.92)	1.43 (1.09-1.88)	1393 (50)	852 (30)	1.16 (0.99-1.37)	1.24 (1.03-1.49)
CRASH High income	64 (8)	115 (15)	1.39 (0.73-2.64)	1.21 (0.58-2.52)	81 (15)	111 (20)	1.16 (0.66-2.04)	2.04 (1.01-4.12)	324 (31)	296 (30)	0.98 (0.73-1.31)	1.04 (0.75-1.44)
Overall	203 (7)	453 (15)	1.96 (0.84-4.59)	2.14 (0.93-4.91)	498 (16)	737 (23)	1.13 (0.73-1.75)	1.46 (1.14-1.85)	2525 (45)	1904 (34)	1.00 (0.86-1.15)	1.18 (1.03-1.55)
TARN	1132 (10)	3147 (26)	2.24 (1.98-2.54)	2.81 (2.44-3.23)	764 (19)	1178 (29)	1.68 (1.42-1.80)	2.18 (1.80-2.65)	5777 (50)	5127 (45)	1.92 (1.78-2.07)	2.14 (1.95-2.35)

In table: odds ratio (95% confidence intervals)

GCS = Glasgow Coma Scale

Adjusted analyses – adjusted for age, pupil reactivity and motor-score. In IMPACT and CRASH also adjusted for respectively study and treatment.

In mild and moderate TBI the logistic regression analyses were done together for UK4, TCDB and EBIC. The ORs are reported in the UK4 row.

In IMPACT there was no significant interaction between MEI and year of injury (p=0.618). In TARN there was a significant interaction between MEI and year of injury (p=0.000), with outcomes slightly improving over time and the effect of MEI slightly decreasing.

The prognostic value of MEI in terms of univariable R^2 (Figure 2) varied from 0.0% (in severe patients in IMPACT and CRASH) to 3.4% (in severe patients in TARN), and was considerably smaller than the prognostic value of core predictors as age, GCS motor score and pupil reactivity.

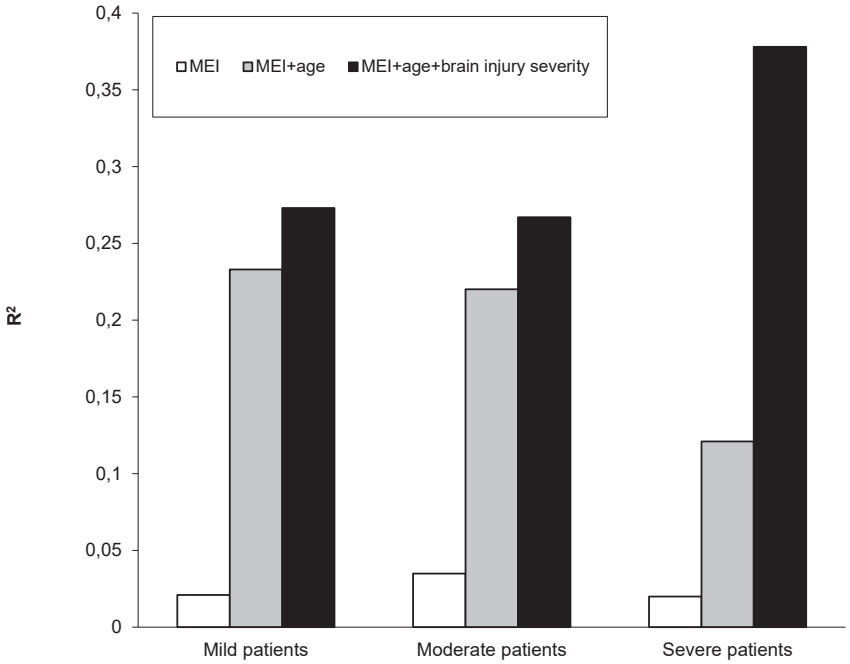


Figure 2. The prognostic value of major extracranial injury (MEI), univariable and in combination with age and brain injury severity (GCS motor score and pupil reactivity), expressed in percentage explained variance (R^2)

Absolute risks

In CRASH and IMPACT, the increase in absolute risk on mortality associated with MEI was 8% (6% vs. 14%) in mild, 4% (15% vs. 19%) in moderate and 1% (45% vs. 46%) in severe TBI patients. The prevalence of MEI in TBI patients was larger in TARN for all brain injury severities than in IMPACT and CRASH, as was the increase in absolute risks on mortality. The increase in absolute risk on mortality associated with MEI was 8% (7% vs. 15%) in mild, 9% (16% vs. 25%) in moderate and 16% (43% vs. 59%) in severe TBI patients in TARN (Table 3).

Table 3. Absolute risks of major extracranial injury and no and minor extracranial injury in different TBI severity groups on mortality in IMPACT& CRASH vs. TARN.

		Mild TBI patients	Moderate TBI patients	Severe TBI patients
IIMPACT & CRASH	**No major extracranial injury**	5.5 (5.2-5.8)	14.8 (14.2-15.3)	44.8 (44.1-45.6)
	Major extracranial injury	13.9 (12.6-15.2)	18.7 (17.7-19.8)	45.5 (44.5-46.6)
TARN	**No major extracranial injury**	7.4 (7.2-7.4)	16.4 (15.8-17.1)	42.9 (42.2-43.6)
	Major extracranial injury	15.3 (14.7-15.8)	24.8 (23.5-26.0)	59.1 (58.3-59.8)

Differences between CRASH, IMPACT and TARN

There was a significant interaction between MEI and brain injury severity in CRASH ($p<0.001$) and TARN ($p=0.029$) but not in IMPACT.

Since we found, also after stratification, a considerable difference in the prognostic effect of MEI between IMPACT-CRASH and TARN across all TBI severities, we excluded 912 patients from TARN who died within 6 hours after injury since the majority of these patients would not have been included in IMPACT or CRASH. This resulted in decreased ORs of MEI for mortality: 2.4 in mild, 1.8 in moderate and 1.6 in severe TBI (IMPACT and CRASH: 2.1 in mild, 1.6 in moderate and 1.2 in severe TBI).

To assess the difference between IMPACT-CRASH and TARN further, we analyzed 14 day mortality in CRASH. In low/middle income countries MEI was less strongly related to 14 day mortality than to 6 month mortality (ORs 0.1-1 point lower for 14 day mortality). In high income countries however, effects were opposite (ORs 0.1 to 0.4 points higher for 14 day mortality).

We performed all analyses also in a subset of the TARN collected after 2005 (n=6078) and found similar results.

DISCUSSION

Our study shows that MEI is a prognostic factor in patients with TBI. However, the effect varies by the population studied in two ways, which explains the disagreement in the literature. First the strength of the effect interacts with brain injury severity, with larger effects in milder TBI patient populations. Second the effect is dependent on the time of inclusion in a study. In TARN (a registry including all TBI patients form the time of injury) MEI is strongly associated with mortality after adjustment for age, GCS motor score and pupil reactivity. In IMPACT and CRASH (broadly selected observational studies and an RCT, including TBI patients surviving the early stage) the incremental prognostic value of MEI compared to known predictors of mortality is limited.

We found a large difference in prognostic effect between TARN and IMPACT / CRASH. The larger effect in TARN was largely explained by inclusion of patients who died before or shortly after admission. The ORs in IMPACT and CRASH thus could be interpreted as the effect of MEI when a TBI patient survives the early stage (first hours) after trauma. The effect in TARN could be interpreted as the effect of MEI in the unselected TBI population. For example: a victim of a road traffic accident with severe TBI and MEI has an odds for mortality 2.14 fold that of a similar patient without MEI. When this patient survives the early stage, the prognostic effect of MEI is reduced to a 1.18 fold increased risk.

Our study shows thus that the magnitude of the effect of MEI on mortality depends on the study design. This is also an explanation for the disagreement in the literature about the prognostic effect of MEI. Studies demonstrating that outcome is not worsened by MEI only included (often severe) patients admitted to an intensive-care unit.(2, 7) These studies are mostly comparable to IMPACT and CRASH with regard to study population and results. The studies showing an effect of MEI in TBI patients, obtained the data from a Trauma Registry like TARN.(8-10)

This means that prognostic effect of MEI is also dependent on the application of a prognosis in a clinical setting. For counseling of relatives of severe TBI patients in the hospital for example, MEI is more likely to be a highly relevant prognostic factor in the Emergency Department than a few hours later if the patient has survived the immediate risk of death from haemorrhage caused by major extracranial injury and has been admitted to intensive care. Thus, this study demonstrates that it is important not only to formulate a clear research question but also to define the specific patient population, which is often not done in prognostic research. To interpret results of a prognostic study and to determine applicability to a particular setting it is important to be aware of the study population and design.

We reported absolute risks in the different studies and the different strata of patients, which further provide some relevant clinical insights. For example, patients with mild TBI & MEI have a similar risk on mortality to one with moderate TBI and no MEI. Absolute risks on mortality were higher in TARN than in IMPACT and CRASH across all TBI severities. This is probably partly due to the previously mentioned difference in patient population. Further, differences in mortality between the studies might be caused by differences in health care system and resources (low/middle income countries in CRASH). Also, the time of data collection varied between the studies (1984 for TCDB and 2009 for the most recent patients in TARN), which might be considered a limitation, but we found that the effect of MEI was constant over time in IMPACT and slightly decreasing in TARN.

It might be expected that MEI is more associated with early mortality than with late mortality. This is supported by our finding that ORs decrease when excluding early deaths in TARN. In CRASH we analyzed both 14 day and 6 month mortality, with inconsistent results. In high income countries the ORs for 14 day mortality were indeed higher

than those for 6 month mortality, in low/middle income countries it was the other way round. An explanation might be that within high income countries trauma deaths after 14 days are rare, while lack of resources and also a greater level of underlying comorbidity make late trauma deaths more prevalent in low/middle income countries. MEI will have an impact there because it will often cause immobility, resulting from e.g. limb and pelvic fractures, which may cause mortality in less resourced settings. In general, the prognostic effect of MEI was larger in low/middle income countries, which might be partly explained by structure and processes of care (e.g. longer times to admissions, less resources). These findings illustrate the necessity to take resources and post acute facilities into account when including patients in TBI studies from regions where resources may be more limited. This is particularly important as a tendency has been noted for pharmaceutical companies and researchers to involve centers from other regions of the world in TBI studies, because of higher patient potential and lower cost.(18)

The unadjusted ORs were all smaller than adjusted ORs. This means that the effect of MEI on mortality was not explained by other predictors of mortality. Adjusting only for brain injury severity led to a small decrease in the effect of MEI, since patients with MEI have more severe brain injury, which is also related to mortality. Adjusting for age led to an increase of the effect of MEI since patients with MEI are younger on average, which is related to less mortality.

Hypotension explained a small part of the association between MEI and mortality. This was expected since systemic injuries can cause major bleedings and thus hypotension. The finding that the ORs of MEI change only very little after adjustment for hypotension and that hypotension is also a strong predictor of mortality independent of MEI suggests that the threshold values for defining hypotension may be too restrictive, or that other mechanisms, such as inflammatory response to multiple injuries, play a role in the relationship between extracranial injury and mortality.

Previous studies have shown that TBI increases the risk of both mortality and morbidity in the general trauma population.(3-5) We find that the presence of MEI is also associated with increased mortality in patients with TBI. Whether this effect may be greater or smaller than in the general trauma population cannot be answered from our study, since we only included patients with TBI. Within the TARN registry work is currently ongoing to analyse the effect of TBI in the general trauma population. It is however an artificial distinction between patients with TBI and patients with MEI. In clinical practice there are patients with trauma and they have often multiple injuries, both extracranial and intracranial. Based on our results and findings from previous studies we would provisionally conclude that both MEI and TBI carry a high risk of mortality, and that a combination of both further increases this risk. The relation is however multidimensional and interaction effects exist with the severity of brain injury.

We used a very simple definition of MEI, since extracranial injury severity was reported differently in each dataset. Analysis of the prognostic value of the full AIS or Injury Severity Scale (ISS), and the different body parts in which the extracranial injury occurred may provide additional insights in the mechanism of effect. This would be of high interest, but unfortunately these data were not available in sufficient numbers to permit meaningful analysis. This represents a limitation of our study. On the other hand, the definition of AIS ≥ 3 we use is quite common, easy to use in practice and showed to discriminate well.

A second limitation of our study is the presence of missing values. We preformed imputation, which is better than deleting missing variables.(14-17) In TARN, where pupil reactivity was imputed in the majority of patients, we performed all analyses also in a subset of the TARN collected after 2005 (n=6078) with complete pupillary reactivity and found similar results.

It could be argued that another limitation is the heterogeneity between the three studies used in the meta-analysis, concerning patient population (enrollment criteria), setting and timing of outcome. However, this heterogeneity allowed us to disentangle the effects of MEI on mortality and to explain to some extent the conflicting results in the current literature. A strength of this study is obviously the many patients included in the study. Also, the meta-analysis is based on individual patient data.

In conclusion, this meta-analysis demonstrates that MEI is a prognostic factor for increasing mortality in patients with TBI. However, the strength of the effect is smaller in patients with more severe brain injury. Also, the strength of the effect decreases when only considering patients who survive the early phase after injury, instead of considering all patients, starting from the time of injury.

REFERENCES

1. Perel P, Arango M, Clayton T, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ. 2008; 336(7642):425-429.
2. Heinzelmann M, Platz A, Imhof HG. Outcome after acute extradural haematoma, influence of additional injuries and neurological complications in the ICU. Injury. 1996;27(5):345-349.
3. McMahon CG , Yates DW, Campbell FM, Hollis S, Woodford M. Unexpected contribution of moderate traumatic brain injury to death after major trauma. J Trauma. 1999;47(5):891-895.
4. Gennarelli TA , Champion HR, Sacco WJ, Copes WS, Alves WM. Mortality of patients with head injury and extracranial injury treated in trauma centers. J Trauma. 1989;29(9):1193-1202.
5. Gennarelli TA, Champion HR, Copes WS, Sacco WJ. Comparison of mortality, morbidity, and severity of 59,713 head injured patients with 114,447 patients with extracranial injuries. J Trauma. 1994;37(6):962-968.
6. Patel HC, Bouamra O, Woodford M, King AT, Yates DW, Lecky FE. Trends in head injury outcome from 1889 to 2003 and the effect of neurosurgical care: an observational study. Lancet. 2005;366(9496):1538-1544.
7. Sarrafzadeh AS, Peltonen EE, Kaisers U, Kuchler I, Lanksch WR, Unterberg AW. Secondary insults in severe head injury - do multiply injured patients do worse? Crit Care Med. 2001;29(6):1116-1123.
8. Lefering R, Paffrath T, Linker R, Bouillon B, Neugebauer EAM, Head injury and outcome – What influence do concomitant injuries have? J Trauma. 2008;65(5):1036-1044.
9. Jacobs B, Beems T, Stulemeijer M, et al. Outcome prediction in mild traumatic brain injury: age and clinical variables are stronger predictors than CT abnormalities. J Neurotrauma. 2010;27(4): 655-668.
10. Ho KM, Burrell M, Rao S, et al. Extracranial injuries are important in determining mortality of neurotrauma. Crit Care Med. 2010;38(7):1562-1568.
11. Marmarou A, Lu J, Butcher I, et al. IMPACT database of traumatic brain injury: design and description, J Neurotrauma 2007;24(2):239-250.
12. Edwards P, Farrell B, Lomas G, et al. The MRC CRASH Trial: study design, baseline data, and outcome in 1000 randomised patients in the pilot phase. Emerg Med J. 2002:19(6):510-514.
13. Lecky F, Woodford M, Yates DW. Trends in trauma care in England and Wales 1989-97, UK Trauma Audit and Research Network. Lancet. 2000;355(9217):1771-1775.
14. Steyerberg EW, Van Veen M. Imputation is beneficial for handling missing data in predictive models. J Clin Epidemiol. 2007; 60(9): 979.
15. Donders AR, Van der Heijden GJ, Stijnen T, et al. Review: A gentle introduction to imputation of missing values. J Clin Epidemiol. 2006; 59(10):1087-91.
16. Van der Heijden GJ, Donders AR, Stijnen T, et al. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. J Clin Epidemiol. 2006; 59(10):1102-9.
17. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008 5;5(8):e165.
18. Maas AIR, Roozenbeek B, Manley GT. Clinical trials in traumatic brain injury: past experience and current developments, Neurotherapeutics. 2010;7(1):115-126.

Chapter **4**

Efficient design and analysis of randomized controlled trials in rare neurological diseases: an example in Guillain-Barré syndrome

Nikki van Leeuwen
Christa Walgaard
Pieter A van Doorn
Bart C Jacobs
Ewout W Steyerberg
Hester F Lingsma

PLoS One 2019

ABSTRACT

Introduction

Randomized controlled trials (RCTs) pose specific challenges in rare and heterogeneous neurological diseases due to the small numbers of patients and heterogeneity in disease course. Two analytical approaches have been proposed to optimally handle these issues in RCTs: covariate adjustment and ordinal analysis. We investigated the potential gain in efficiency of these approaches in rare and heterogeneous neurological diseases, using Guillain-Barré syndrome (GBS) as an example.

Methods

We analyzed two published GBS trials with primary outcome 'at least one grade improvement' on the GBS disability scale. We estimated the treatment effect using logistic regression models with and without adjustment for prognostic factors. The difference between the unadjusted and adjusted estimates was disentangled in imbalance (random differences in baseline covariates between treatment arms) and stratification (change of the estimate due to covariate adjustment). Second, we applied proportional odds regression, which exploits the ordinal nature of the GBS disability score. The standard error of the estimated treatment effect indicated the statistical efficiency.

Results

Both trials were slightly imbalanced with respect to baseline characteristics, which was corrected in the adjusted analysis. Covariate adjustment increased the estimated treatment effect in the two trials by 8% and 18% respectively. Proportional odds analysis resulted in lower standard errors indicating more statistical power.

Conclusion

Covariate adjustment and proportional odds analysis most efficiently use the available 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.

INTRODUCTION

RCTs are the standard to investigate the effectiveness of medical interventions. However, RCTs are challenging in rare heterogeneous diseases. The randomization process in RCTs ensures that observed and unobserved patient characteristics on average are similar between treatment arms.(1) However, it does not ensure full balance.(1) Different baseline risks for outcome can arise between treatment arms, simply due to chance.(1) In diseases with large between-patient differences in natural disease course, severity and outcome, small imbalances in covariates between the treatment arms may, positively or negatively, affect the estimated treatment effect.

Sample sizes in RCTs in rare diseases are usually small. Small trials are a subject to a greater chance of imbalance than large trials.(1)(Moreover, small RCTs can easily fail to detect treatment benefits, due to lack of statistical power. In rare neurological disorders, such as inflammatory neuropathies like Guillain-Barré syndrome (GBS), Chronic Inflammatory Demyelinating Polyneuropathy (CIDP) and Multifocal Motor Neuropathy (MMN), this heterogeneity and rarity is a major challenge for conducting RCTs.

Two approaches to optimize RCT design and analysis that have been successfully applied in other acute neurological diseases such as stroke and traumatic brain injury are covariate adjustment and ordinal analysis.(2-4) (Table S1) Covariate adjustment is a statistical method that adjusts the treatment effect for baseline risk on poor outcome in the treatment arms. When the treatment arms are imbalanced, an unadjusted analysis is suboptimal to estimate the treatment effect. In addition, previous studies found that covariate adjustment could increase statistical power.(1, 5-9) Ordinal analysis is an approach to analyze a full ordinal outcome scale instead of a dichotomized version. Although these techniques already have been successfully applied in stroke and traumatic brain injury, it is still relevant to study this in other diseases like GBS, since the effect of the different approaches can work out differently in different study settings. The most commonly used outcome in GBS is the ordinal GBS disability score, consisting of seven categories. Usually the scale is dichotomized into favorable or unfavorable outcome, or the improvement on the GBS disability score from admission calculated and dichotomized as minimal one grade improvement. In ordinal analysis the outcome is not dichotomized but analyzed as the full ordinal scale with proportional odds analysis, preventing loss of information.(10) Simulation studies and empirical validation studies in other fields have demonstrated that proportional odds analysis increases statistical power in RCTs.(10-13)

To test the applicability and value of these approaches in rare and heterogeneous neurological diseases, we use Guillain-Barré syndrome (GBS) as an example. GBS is a life-threatening acute immune-mediated polyradiculoneuropathy(14, 15), which requires early diagnosis and hospital admission for accurate monitoring, treatment and

supportive care. Some patients may show spontaneous and full recovery, while others require ventilation at an ICU for months and remain severely disabled. Several RCTs have successfully been conducted in GBS.(16-18)

We aimed to explore the potential benefit of covariate adjustment and proportional odds analysis in rare and heterogeneous neurological diseases, compared to the conventional statistical approaches. We hereto re-analyzed two RCTs in GBS.

METHODS

Patient Population

We analyzed data from two RCTs in GBS, the Plasma Exchange (PE) vs Intravenous Immunoglobulin (IVIg) (PE vs IVIg) trial(17) and the IVIg and placebo versus IVIg and Methyl-Prednisolone (MP) (IVIg vs MP) trial(18), conducted between 1986 and 2000. In the PE vs IVIg trial, the control group received IVIg and the treatment group received PE. In the IVIg vs MP trial, the patients receiving IVIg and placebo were considered as control patients and the patients receiving IVIg and MP were considered as treated patients. The primary outcome in both trials was improvement (corresponding to lower GBS disability scores) by one or more grades on the GBS disability score after 4 weeks. The GBS disability score is an ordinal scale ranging from 0 = healthy to 6 = dead. However, in order to estimate treatment effects for a positive outcome for all the analyses, we used the reversed GBS disability score at 4 weeks, to keep the estimates easy to compare. For all the regression models used in this paper, higher numbers (in outcome) mean better health outcomes.

Statistical analysis

The predicted probabilities for one grade improvement on the GBS disability score were calculated and used as a measure for baseline risk to indicate potential unbalance between the treatment arms in baseline characteristics.

To estimate treatment effects, we used two commonly used primary (dichotomous) outcomes in GBS trials as reference; (1) favorable outcome (0-2) on the GBS disability scale at 4 weeks as outcome and (2) minimal one grade improvement on the GBS disability score between the moment of randomization and 4 weeks as outcome, both analyzed with binary logistic regression without covariate adjustment. Consequently, these references were compared with the two approaches under study: covariate adjustment and ordinal analysis.

Covariate adjustment

With covariate adjustment, conditional treatment effects are estimated with regression models. Adjusting for GBS disability score at admission results in an estimated treatment effect for a patient with a given GBS disability score, while unadjusted analysis results in an average estimated treatment effect over all patients, irrespective of the GBS disability score. Unadjusted analysis is expressed by the following formula:

$$log\ odds\ (improvement) = \alpha + \beta * treatment,$$

where improvement is by one or more grades on the GBS disability score, and treatment is an indicator for the randomization arm. The coefficients α and β indicate the intercept and regression coefficient for treatment. In logistic regression, exp(β) indicates the odds ratio (OR).

For adjusted analysis, we used three well-known predictors of outcome(19, 20): age, preceding diarrhea and GBS disability score at admission. The covariate adjusted model is expressed by the following formula:

$$log\ odds\ (improvement) = \alpha + \beta * treatment + \beta 1 * age + \beta 2 * preceding\ diarrhea + \beta 3 * GBS\ disability\ score\ at\ admission.$$

This results in an adjusted regression coefficient β for the estimated treatment effect. In the trial analysis, the observed difference of the unadjusted and adjusted regression coefficient for the treatment variable is a result of imbalance and stratification.(8) We hereto calculated the linear predictor based on age, diarrhea and GBS disability score at admission. We then calculated the difference in treatment effect that was attributable to imbalance as the difference between the mean value of the linear predictor between the treatment arms.(8) The remaining part of the difference between the unadjusted and the adjusted treatment effect was attributed to stratified estimation, i.e. conditioning on covariates.(8)

Proportional odds analysis

For ordinal analysis we used proportional odds logistic regression to exploit the ordinal nature of the GBS disability score. A proportional odds logistic regression model was fitted with the GBS disability score collapsed to a 5-point scale. We combined both healthy (0) and minor symptoms (1), as well as needs ventilation at least a part of the day (5) and dead (6) because of small numbers in these extreme categories. We used the reversed GBS disability scale to estimate treatment effects on a positive outcome, and to keep these estimates comparable to the estimates of the other logistic regression models on positive dichotomous outcomes (improvement and favorable outcome). The proportional odds model uses an ordinal outcome variable with more than two possible categories. It estimates a common OR over all possible cut-offs of the outcome scale. Next, we used the difference between the GBS disability score at admission and the GBS

disability score at four weeks as outcome. A proportional odds logistic regression model was used to analyse the difference in GBS disability score.

Treatment effect estimates

The coefficient β of the treatment effect and the corresponding standard error (SE) were calculated for the four approaches to analyse outcome, with and without covariate adjustment. The SE of the treatment effect indicates the precision of the calculated treatment effect. The SEs in the proportional odds regression models are expected to be smaller than those in the logistic models. Both trials were analysed with complete case analysis, ignoring 1 and 4 patients with incomplete baseline data. Statistical analyses were performed in R Statistical Software version 2.15.3 using the rms package (R Foundation for Statistical Computation, Vienna, Austria).

RESULTS

Patient population and reference strategies

We analysed data from 146 patients in the PE vs IVIg trial and 221 patients in the IVIg vs IVIg+MP trial. Both trials were slightly imbalanced with regard to the baseline characteristics. In the IVIg vs IVIg+MP trial the treatment group (with MP) had a probability of 0.60 to improve at least one grade on the GBS disability score compared to a predicted probability of 0.64 in the control group (without MP). So without any treatment, the prognosis of the treatment arm was slightly better. An opposite distribution of baseline covariates between treatment arms is shown in the PE vs IVIg trial. The treatment group (PE) has a higher predicted probability (0.45) to improve at least one grade on the GBS disability score compared to the control group (IVIg; predicted probability 0.41, Table 1).

Regarding the actual outcome, 63 (57%) control patients treated with IVIg and placebo and 74 (67%) patients treated with IVIg and methylprednisolone improved minimal one grade on the GBS disability score after 4 weeks. In the other trial, 25 (34%) control patients treated with IVIg and 38 (52%) patients receiving PE improved minimal one grade on the GBS disability score after 4 weeks (Appendix 1).

The treatment under study in both trials had a positive effect on health outcomes. With the reference strategy of logistic regression on a favorable GBS disability scale (0 – 2) at 4 weeks as outcome, the estimated treatment OR was 1.80 (95% confidence interval (CI) 0.84 – 3.85, SE 0.39, p = 0.13) in the PE vs IVIg trial and 1.69 (95% CI 0.93 – 3.08, SE 0.31, p = 0.09) in the IVIg vs IVIg+MP trial. The treatment effect estimates on one grade improvement were slightly larger (Table 2).

Table 1. Distribution of baseline predictors and outcome distribution in two randomized controlled trials in GBS.

	PE vs IVIg trial			IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial		
	Total (n = 146)	**Control (PE) (n =73)**	**Treatment (IVIg) (n = 73)**	**Total (n = 221)**	**Control (IVIg) (n = 111)**	**Treatment (IVIg+MP) (n = 110)**
Age (Median, Interquartile Range 25th-75th Percentile)	49 (32 – 63)	51 (33 – 66)	47 (32 – 61)	55 (35 - 67)	52 (35 – 67)	57 (34 – 68)
Preceding diarrhea	27 (19%)	16 (22%)	11 (15%)	60 (27%)	30 (27%)	30 (27%)
GBS disability score at admission						
Able to walk over 10m open space with help	29 (20%)	16 (22%)	13 (18%)	58 (26%)	32 (30%)	26 (24%)
Bedridden or chair bound	92 (63%)	44 (60%)	48 (66%)	153 (49%)	78 (70%)	75 (68%)
Needs ventilation for at least a part of the day	25 (17%)	13 (18%)	12 (16%)	10 (5%)	1 (1%)	9 (8%)
Predicted probability of one or more grades improvement on the GBS disability score after 4 weeks	0.43	0.41	0.45	0.62	0.64	0.60
One or more grades improvement on the GBS disability score after 4 weeks	63 (43%)	25 (34%)	38 (52%)	137 (62%)	63 (57%)	74 (67%)
GBS disability score after 4 weeks						
0 = Healthy	0 (0%)	0 (0%)	0 (0%)	5 (2%)	0 (0%)	5 (5%)
1 = Minor symptoms	16 (11%)	6 (8%)	10 (14%)	37 (17%)	24 (22%)	13 (12%)
2 = Able to walk 10m unassisted but not able to run	30 (21%)	12 (16%)	18 (25%)	74 (34%)	31 (28%)	43 (39%)
3 = Able to walk over 10m open space with help	19 (13%)	9 (12%)	10 (14%)	22 (10%)	10 (9%)	12 (11%)
4 = Bedridden or chair bound	48 (33%)	27 (37%)	21 (29%)	54 (24%)	31 (28%)	23 (21%)
5 = Needs ventilation for at least a part of the day	31 (21%)	17 (23%)	14 (19%)	26 (12%)	14 (13%)	12 (11%)
6 = Dead	2 (1%)	2 (3%)	0 (0%)	3 (1%)	1 (1%)	2 (2%)

Table 2. Treatment effect analysis: unadjusted and adjusted binary and proportional odds logistic regression.

		PE vs IVIg trial (n = 146)		IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial (n = 221)	
		Unadjusted	Adjusted*	Unadjusted	Adjusted*
Binary logistic regression – GBS disability 3-6 vs 0-2 **	OR (95% CI)	1.90 (0.93 – 3.87)	1.80 (0.84 – 3.85)	1.27 (0.75 – 2.15)	1.69 (0.93 – 3.08)
	SE	0.36	0.39	0.27	0.31
	P-value	0.08	0.13	0.38	0.09
Binary logistic regression – improvement on GBS disability score	OR (95% CI)	2.08 (1.07 – 4.06)	1.95 (0.96 – 4.00)	1.57 (0.91 – 2.71)	1.96 (1.08 – 3.56)
	SE	0.34	0.36	0.28	0.31
	P-value	0.03	0.06	0.11	0.03
Proportional odds logistic regression – reversed GBS disability score at 4 weeks**	OR (95% CI)	1.76 (0.98 – 3.19)	1.76 (0.98 – 3.19)	1.12 (0.70 – 1.80)	1.41 (0.87 – 2.28)
	SE	0.30	0.30	0.24	0.25
	P-value	0.06	0.06	0.63	0.17
Proportional odds logistic regression – Δ GBS disability score (grades improvement between admission and 4 weeks)	OR (95% CI)	1.93 (1.07 – 3.49)	1.80 (0.99 – 3.27)	1.43 (0.89 – 2.30)	1.34 (0.89 – 2.32)
	SE	0.30	0.30	0.24	0.25
	P-value	0.03	0.05	0.14	0.14

**Adjustment for age, preceding diarrhea and GBS disability score at admission. ** 0 = Healthy / 1 = Minor symptoms / 2 = Able to walk 10m unassisted but not able 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 / 6 = Dead*

*** In order to estimate the treatment effect for a positive outcome, we used the reversed GBS disability score at 4 weeks*

Covariate adjustment

With covariate adjustment, the estimated treatment effect was larger in the IVIg vs IVIg+MP trial, partly as a result of adjustment, which makes the estimates more extreme, and partly because of the imbalance at baseline. Poorer prognosis at baseline for the intervention (IVIg + MP) group implied a +31% increase in the adjusted treatment effect (Table 3). The stratification effect of adjustment was an additional 18% increase in the treatment effect (OR = 1.96). In contrast, the treatment effect was smaller with adjustment for baseline characteristics in the PE vs IVIg trial. The stratification effect increased the treatment effect with 8%, but the better prognosis in the intervention (IVIg) group at baseline reduced the estimated treatment effect by -24%. The net effect was a difference in treatment effect of -16%. These results were similar for all binary and ordinal outcome analyses (Table 2).

Table 3. Results of unadjusted and adjusted binary logistic regression analysis of the effect of treatment versus control on GBS disability score at four weeks in both PE vs IVIg trial (n = 146) and the IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial (n = 221).

	OR	Coefficient	Absolute difference in treatment effect between adjusted and unadjusted	Imbalance between treatment arms	Relative difference in treatment effect between adjusted and unadjusted due to imbalance	Relative difference in treatment effect between adjusted and unadjusted due to stratification
			PE vs IVIg trial			
Unadjusted	2.08	0.73				
Adjusted for age, preceding diarrhea and GBS disability score at admission	1.95	0.67	- 0.06^	-0.12	-16%*	8%#
			IVIg vs MP trial			
Unadjusted	1.57	0.45				
Adjusted for age, preceding diarrhea and GBS disability score at admission	1.96	0.67	0.22^	0.14	31%*	18%#

^ *Adjusted coefficient – Unadjusted coefficient*

* *Imbalance between treatment arms / Unadjusted coefficient*

 # (Absolute difference in treatment effect between adjusted and unadjusted - Imbalance between treatment arms) / Unadjusted coefficient

Proportional odds analysis

For illustration of the proportional odds analyses we calculated the treatment effect estimates (ORs) for each cut-off of the reversed ordinal scale. The common OR can be interpreted as the pooled estimate of these binary ORs. The treatment under study in both trials had a positive effect on health outcomes in all the ordinal analyses. In the PE vs IVIg trial the ORs over each cut-off were relatively similar (Figure 1c and 1d). The common OR was similar as well, but the SE and CI were smaller. In the IVIg vs IVIg+MP trial, the ORs were more variable (Figure 1a and 1b). The common OR was less extreme compared to ORs for the cut-off used in the reference approach (0-2 vs. 3-6 and minimal one grade improvement vs. no improvement). But again, the SE and CI were smaller. This can also be seen in table 2; in all analyses, the proportional odds analysis on the GBS disability score after four weeks and on the improvement on the GBS disability score resulted in lower SEs of the treatment effect compared to the binary approaches.

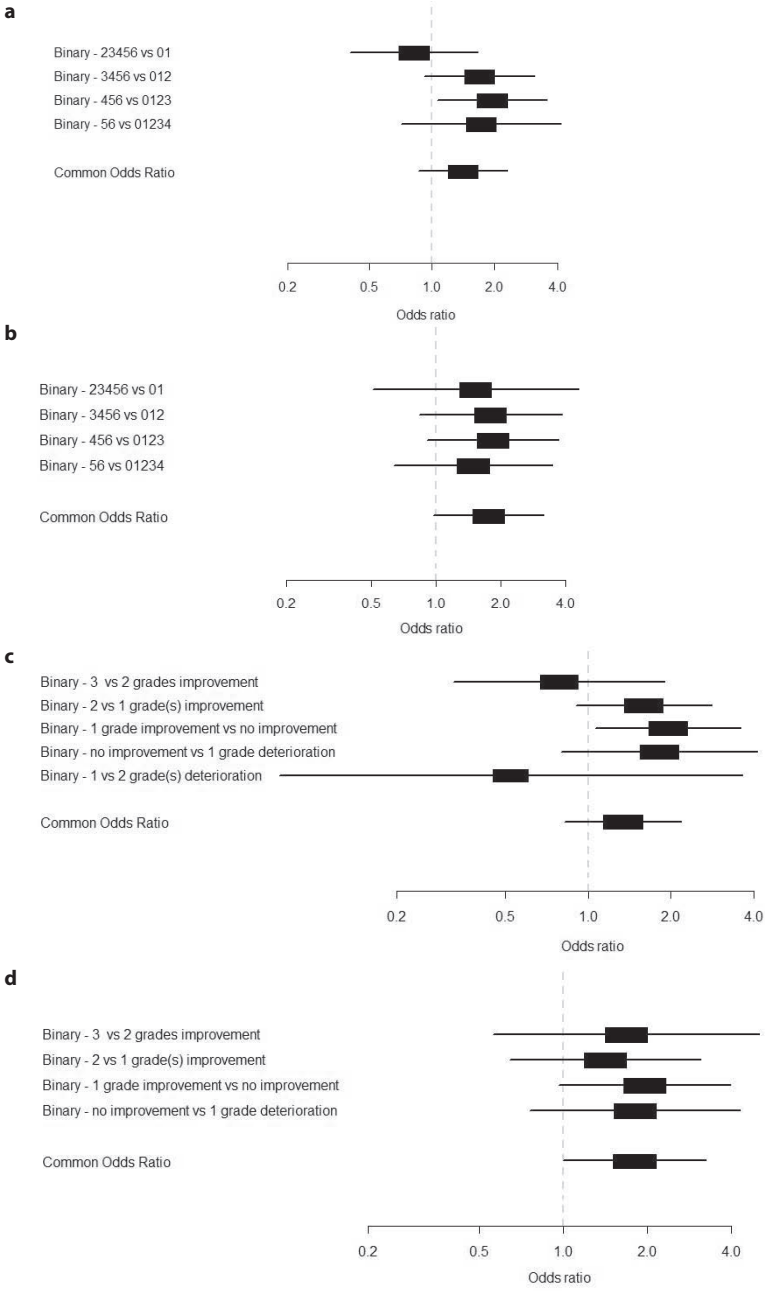


Figure 1. Treatment effect analysis: forest plots of the adjusted binary and proportional odds logistic regression in the IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial (a and b) and PE vs IVIg trial (c and d) show smaller confidence intervals for the common odds ratio compared to the binary estimates.

Table 4. Characteristics of four 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

DISCUSSION

In this study we assessed the potential benefit of the use of covariate adjustment and proportional odds analysis in RCTs compared to the conventional method, by reanalyzing two GBS trials. We found that covariate adjustment increased the estimated treatment effect in one trial, and decreased the estimated treatment effect in the other trial, due to imbalances in baseline characteristics between the treatment arms. Although such imbalances are fully due to chance if a proper randomization procedure is followed, our results illustrate that their impact on interpretability of treatment effect estimates can be substantial and can be different in several study settings. We found that the proportional odds analysis resulted in lower standard errors and thus smaller confidence intervals of the treatment effect estimate compared to the conventional method of logistic regression on dichotomized outcome measures. Thus, dichotomization of ordinal outcome measures does not merit application. In future trials in rare and heterogeneous neurological diseases like GBS both covariate adjustment and proportional odds analysis are advised.

Covariate adjustment

On expectation, covariate adjustment leads to more extreme treatment effect estimates and larger standard errors for non-linear regression models.(21) The p values are a function of the treatment effect estimates and standard error. With covariate adjustment the increase in treatment effect estimate will outweigh increased in standard error and the p values will be lower compared to unadjusted analysis.(21)

Indeed, 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 PE vs IVIg trial. In the IVIg vs MP trial, 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.

Covariate adjustment increases statistical power, despite the larger standard error.(1, 7) When there are no baseline imbalances, the adjusted conditional estimates will be more extreme than the unadjusted marginal estimates.(22) However, the size and the direction of the difference between the unadjusted and adjusted estimates are dependent on the strength of the prognostic factors and the imbalance in baseline risk between the treatment- and control group in the specific trial and this is shown in our study. When investigating the effectiveness of a medical intervention in rare and heterogeneous neurological diseases, such as GBS, one has to deal with small sample sizes. We therefore recommend performing covariate adjustment in future trials in rare and heterogeneous neurological diseases. For GBS this covariate adjustment should be applied with known predictors for (functional) outcome, specifically age, preceding diarrhea, GBS disability score and MRC sum score.(19, 20)

The outcome 'minimal one grade improvement' implicitly involves a form of covariate adjustment. 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 (Table 4). 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 because of regression to the mean;(23, 24) baseline values are negatively correlated with change because patients with high scores at baseline generally improve more than those with low scores.(25) Therefore covariate adjustment with the absolute baseline value is still preferable over implicitly taking into account baseline severity in the outcome measure 'improvement'. Moreover, disease severity at baseline is not the only covariate we could adjust for. Especially, the age of the patient will be an important covariate in most neurological diseases.

Thus, in general, ignoring baseline imbalance between treatment arms in trials may cause invalid conclusions on both the magnitude and significance of the treatment effect estimate compared to analysis using covariate adjustment. The impact on interpretability of treatment effect estimates can be substantial and can be different in several study

settings. When designing a trial, the analysis plan should be precisely pre-specified. Also, the covariates that will be used for adjustment should be pre-specified. Previous studies have shown that the stronger the relation of the covariates with outcome, the larger the increase in statistical power with covariate adjustment will be.(5, 26, 27) In GBS, predictors of outcome are relatively well known(19, 20) and therefore pre-specifying important baseline variables for covariate adjustment is possible in GBS trials.

Proportional odds analysis

It is evident that the GBS disability scale is not a linear scale. For example, improvement from "needs ventilation for at least a part of the day" to "bedridden or chair bound" is not the same improvement as the improvement from "able to walk over 10m open space with help" to "able to walk 10m unassisted but not able to run". However, whether or not the ordinal outcome under study is a linear scale is not relevant for the validity of the proportional odds analysis. Proportional odds analysis merely requires ordering of outcomes. The proportional odds analysis estimates the treatment effect on each cut-off of the scale, instead of estimating the treatment effect on the difference between the averages scores in the treatment arms, as linear regression. The proportional odds model results in a common OR, which is interpretable as a pooled OR over all ORs for the different cut-offs. The common OR is formally valid if the ORs for each cut-off are the same (the proportional odds assumption). We can, however, interpret the common OR as a summary measure of the treatment effect, even if the ORs differs slightly per cut-off.(12, 28) The common OR can also be interpreted as the average shift over the total ordinal outcome scale caused by the treatment under study.(10-13) Moreover, simulation studies have shown that ordinal analysis is more efficient than binary analysis, even if the proportional odds assumption is violated.(11) Because the ordinal analysis uses the full ordinal outcome scale instead of one dichotomy, the variability will be smaller compared to binary analysis. This was confirmed in our study, where the proportional odds resulted in lower standard errors compared to the binary approaches. Although the importance of applying proportional odds analysis already has been assessed in other diseases, it is still relevant to study this for specific cases like GBS. For example it is important to have more insight in the effect of treatment on the different cut-offs for the specific ordinal outcome measure, in this case the GBS disability score, and see if the proportional odds assumption holds.

In the PE vs IVIg trial, the ORs for each cut-off were very similar and as a result the common OR was also similar. Thus, with the smaller SE, the p value was lower. In contrast, in the IVIg vs IVIg+MP trial, the ORs 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 effect is truly different for different

cut-offs, although this is clinically unlikely. The cut-off chosen in the reference approach in the analysis of improvement appeared to be the most optimal cut-off from a statistical perspective, since it was the only cut-off resulting in a significant treatment effect.

However, if we assume a relatively constant treatment effect across the different cut-offs of an ordinal outcome scale, it is unpredictable which cut-off will show the strongest effect. Therefore, the ordinal analysis is a 'safe' choice and the common OR is a fair representation of the effect of treatment on the ordinal outcome compared to the binary approach, because it takes into account improvement over all levels of the GBS disability score. Since it is also more efficient, we recommend the use of the full ordinal outcome scale in future trials in rare and heterogeneous neurological diseases. In observational studies, ordinal analyses could be combined with propensity score methods to maximize statistical power.

Limitations

Patients with missing covariate data were excluded from the analyses. Data from 367 patients were analyzed rather than 372 patients in the original analyses. We did not assess heterogeneous treatment effects according to baseline risk, which could influence the ability of covariate adjustment to improve the statistical power in an RCT. In this study we only investigated GBS which may not fully be representative for other neurological disorders, although covariate adjustment and proportional odds analysis have shown advantages in other fields, such as stroke and traumatic brain injury.(3, 4, 7, 12)

Conclusion and implications

Covariate adjustment corrects for baseline imbalance and increases power. Proportional odds analysis optimally exploits the ordinal nature of outcome scales. A combined approach is advised for reliable and efficient estimation of treatment effects in small RCTs in rare and heterogeneous diseases like GBS.

REFERENCES

1. 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.
2. Maas AI, Steyerberg EW, Marmarou A, McHugh GS, Lingsma HF, Butcher I, et al. IMPACT recommendations for improving the design and analysis of clinical trials in moderate to severe traumatic brain injury. Neurotherapeutics. 2010;7(1):127-34.
3. Optimising Analysis of Stroke Trials C, Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcomes in stroke trials. Stroke. 2007;38(6):1911-5.
4. Optimising the Analysis of Stroke Trials C, Gray LJ, Bath PM, Collier T. Should stroke trials adjust functional outcome for baseline prognostic factors? Stroke. 2009;40(3):888-94.
5. 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.
6. 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.
7. 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.
8. 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.
9. Turner EL, Perel P, Clayton T, Edwards P, Hernandez AV, Roberts I, et al. Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. J Clin Epidemiol. 2012;65(5):474-81.
10. 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.
11. 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.
12. 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.
13. 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.
14. Hughes RA, Cornblath DR. Guillain-Barre syndrome. Lancet. 2005;366(9497):1653-66.
15. Hughes RA, Rees JH. Clinical and epidemiologic features of Guillain-Barre syndrome. J Infect Dis. 1997;176 Suppl 2:S92-8.
16. Garssen MP, van Koningsveld R, van Doorn PA, Merkies IS, Scheltens-de Boer M, van Leusden JA, et al. Treatment of Guillain-Barre syndrome with mycophenolate mofetil: a pilot study. J Neurol Neurosurg Psychiatry. 2007;78(9):1012-3.
17. van der Meche FG, Schmitz PI. A randomized trial comparing intravenous immune globulin and plasma exchange in Guillain-Barre syndrome. Dutch Guillain-Barre Study Group. N Engl J Med. 1992;326(17):1123-9.

18. van Koningsveld R, Schmitz PI, Meche FG, Visser LH, Meulstee J, van Doorn PA, et al. Effect of methylprednisolone when added to standard treatment with intravenous immunoglobulin for Guillain-Barre syndrome: randomised trial. Lancet. 2004;363(9404):192-6.
19. 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.
20. 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.
21. 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.
22. Hauck WW, Neuhaus JM, Kalbfleisch JD, Anderson S. A consequence of omitted covariates when estimating odds ratios. J Clin Epidemiol. 1991;44(1):77-81.
23. Bland JM, Altman DG. Regression towards the mean. BMJ. 1994;308(6942):1499.
24. Bland JM, Altman DG. Some examples of regression towards the mean. BMJ. 1994;309(6957):780.
25. Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. BMJ. 2001;323(7321):1123-4.
26. 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.
27. 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.
28. Senn S. A random effects model for ordinal responses from a crossover trial. Stat Med. 1993;12(22):2147-51.
29. Hernandez AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. Neurosurgery. 2005;57(6):1244-53; discussion -53.
30. Ciolino J, Zhao W, Martin R, Palesch Y. Quantifying the cost in power of ignoring continuous covariate imbalances in clinical trial randomization. Contemp Clin Trials. 2011;32(2):250-9.
31. Ciolino JD, Martin RH, Zhao W, Jauch EC, Hill MD, Palesch YY. Covariate imbalance and adjustment for logistic regression analysis of clinical trial data. J Biopharm Stat. 2013;23(6):1383-402.
32. Garofolo KM, Yeatts SD, Ramakrishnan V, Jauch EC, Johnston KC, Durkalski VL. The effect of covariate adjustment for baseline severity in acute stroke clinical trials with responder analysis outcomes. Trials. 2013;14:98.
33. Senn S, Julious S. Measurement in clinical trials: a neglected issue for statisticians? Stat Med. 2009;28(26):3189-209.
34. Diaz I, Colantuoni E, Rosenblum M. Enhanced precision in the analysis of randomized trials with ordinal outcomes. Biometrics. 2016;72(2):422-31.

Table S1. Overview of a selection of methodological studies considering covariate adjustment and ordinal analysis in RCTs.

First author	Year	Field	Key findings and conclusions
Covariate adjustment			
Robinson(21)	1991	-	In classic linear regression, the adjustment for a non-confounding predictive covariate, results in improved precision, whereas such adjustment in logistic regression results in a loss of precision. However, when testing for a treatment effect in randomized studies, it is always more efficient to adjust for predictive covariates when logistic models are used, and thus in this regard the behavior of logistic regression is the same as that of a classic linear regression.
Hauck(22)	1991	-	In the epidemiologic literature, one finds three criteria for confounding, which we will call the classical (marginal), operational (change-in-estimate) and conditional criteria. We define mavericks to be covariates that satisfy the operational criterion, but not the classical criterion. We present what is known about the problems of mavericks for estimating odds ratios and clarify the interpretation of oddsratios. Key results are: (1) omitting mavericks biases odds ratios towards 1; (2) omitting mavericks cannot artificially introduce an effect in contrast to omitting classical confounders; (3) the operational criterion for confounding corresponds to the conditional criterion when estimating odds ratios, but for relative risks, there are no mavericks (i.e. the classical and operational criterion correspond); and (4) the interpretation of odds ratios obtained from standard methods is that of comparing proportions, not of individual risk.
Pocock(27)	2002	-	When reporting the trial's findings baseline data can be used for i.a. covariate-adjusted analyses which aim to refine the analysis of the overall treatment difference by taking account of the fact that some baseline characteristics are related to outcome and may be unbalanced between treatment groups and baseline comparisons which compare the baseline characteristics of patients in each treatment group for any possible differences. This paper examines how these issues are currently tackled in the medical journals, based on a recent survey of 50 trialreports in four major journals. Key issues include: inconsistencies in the use of covariate-adjustment; the lack of clear guidelines on covariate selection; the overuse of baseline comparisons in some studies; the misuses of significance tests for baseline comparability, and the need for trials to have a predefined statistical analysis plan for all these uses of baseline data.
Hernandez(6)	2004	-	Logistic regression analysis was applied to simulated data sets (n=360) with different treatment effects, covariate effects, outcome incidences, and covariate prevalence. Treatment effects were estimated with or without adjustment for a single dichotomous covariate. The power was highest with prespecified adjustment. The potential reduction in sample size was higher with stronger covariate effects (from 3 to 46%, at 50% outcome incidence and covariate prevalence) and independent of the treatment effect. At lower outcome incidences and/ or covariate prevalence, the reduction was lower.

Table S1. Overview of a selection of methodological studies considering covariate adjustment and ordinal analysis in RCTs. (continued)

First author	Year	Field	Key findings and conclusions
Hernandez(29)	2005	Traumatic brain injury	18 RCTs (n = 6439) were identified in a systematic review of therapeutic phase III RCTs, including adult patients with acute, moderate-to-severe TBI to assess actual reporting of covariate adjustment according to the Consolidated Standards of Reporting Trials (CONSORT) recommendations. Five RCTs reported covariate adjustment. The number of covariates was limited (<=5), most frequently including age. Many covariates were outcome predictors. Four RCTs reported only adjusted treatment effects as the main efficacy parameter. The reported covariate adjustment in TBI trials had several methodological shortcomings. Appropriate performance and reporting of covariate adjustment and subgroup analysis should be considerably improved in future TBI trials because interpretation of treatment benefits may be misleading otherwise.
Hernandez(5)	2006	Traumatic brain injury	Individual patient data from seven therapeutic phase III randomized clinical trials (RCTs; n = 6166) in moderate or severe TBI, and three TBI surveys (n = 2238) were used to calculate the potential sample size reduction obtained by adjustment of a hypothetical treatment effect for one to seven predictors with logistic regression models. The distribution of predictors was more heterogeneous in surveys than in trials. Adjustment of the treatment effect for the strongest predictors (age, motor score, and pupillary reactivity) yielded a reduction in sample size of 16-23% in RCTs and 28-35% in surveys. Adjustment for seven predictors yielded a reduction of about 25% in most studies: 20-28% in RCTs and 32-39% in surveys.
Optimizing the Analysis of Stroke Trials (OAST) Collaboration(5)	2009	Acute stroke	Data from 23 stroke trials (N = 25 674) assessing functional outcome were included. Unadjusted and adjusted ordinal logistic regression models were compared using simulated data from each trial (10 000 simulations per trial). Three levels of treatment effect were assessed with ORs of 0.95, 0.74, and 0.57. Adjusting for prognostic factors in stroke trials can reduce sample size by at least 20% to 30% (the lower interquartile range) for a given power and is similar across all 3 treatment effects
Roozenbeek(7)	2009	Traumatic brain injury	Statistical modeling studies in three surveys and six randomized controlled trials were performed using individual patient data from the IMPACT database. Covariate adjustment reduced sample sizes by 30% in surveys and 16% in RCTs. Covariate adjusted analysis in a broadly selected group of patients is advisable if a uniform treatment effect is assumed, since there is no decrease in recruitment.
Steyerberg(8)	2010	Acute myocardial infarction	The effects of adjustment (correction for imbalance and stratification) were studied with logistic regression analysis in the GUSTO-I trial. When adjusted for 17 characteristics, the odds ratio was 0.820, an increase of 25% compared to the unadjusted odds ratio. The increase in effect estimate was largely explained by the stratification effect and only partly by imbalance of predictors. Adjustment for predictive baseline characteristics, even when largely balanced, may lead to clearly different estimates of the treatment effect on mortality rates.

Table S1. Overview of a selection of methodological studies considering covariate adjustment and ordinal analysis in RCTs. (continued)

First author	Year	Field	Key findings and conclusions
Ciolino(30)	2011	Acute ischemic stroke	Based on data from a clinical trial of acute ischemic stroke treatment, computer simulations were used to create scenarios varying from the best possible baseline covariate balance to the worst possible imbalance, with multiple balance levels between the two extremes. Our simulation results show that the worst possible imbalance is highly unlikely, but it can still occur under simple random allocation. Also, power loss could be nontrivial if balancing distributions of important continuous covariates were ignored even if adjustment is made in the analysis for important covariates. This situation, although unlikely, is more serious for trials with a small sample size and for covariates with large influence on primary outcome.
Turner(9)	2012	Traumatic brain injury	14-day mortality was analyzed in 9,497 TBI patients in the CRASH trial of corticosteroid vs.placebo. Adjustment was made using logistic regression for baseline covariates of two validated risk models (IMPACT and CRASH) derived from external data. The relative sample size (RESS) measure, defined as the ratio of the sample size required by an adjusted analysis to attain the same power as the unadjusted reference analysis, was used to assess the impact of adjustment. RESS of 0.79 and 0.73 were obtained by adjustment using the IMPACT and CRASH models, respectively, which, for example, implies an increase from 80% to 88% and 91% power, respectively.
Ciolino(31)	2013	Acute stroke	This article uses simulation to quantify the benefit of covariate adjustment in logistic regression. Results suggest that if adjustment is not possible or unplanned in a logistic setting, balance in continuous covariates can alleviate some (but never all) of the shortcomings of unadjusted analyses.
Garofolo(32)	2013	Acute stroke	Using a current stroke clinical trial and its pilot studies to guide simulation parameters, 1,000 clinical trials were simulated at varying sample sizes under several treatment effects to assess power and type I error. Covariate-adjusted and unadjusted logistic regressions were used to estimate the treatment effect under each scenario. Under various treatment effect settings, the operating characteristics of the unadjusted and adjusted analyses do not substantially differ. Power and type I error are preserved for both the unadjusted and adjusted analyses.
Thompson(1)	2015	Stroke and acute myocardial infarction	In two large trial data sets GUSTO-I (N = 30,510) and IST (N = 18,372) random samples (500,000 times) of sizes 300 and 5,000 per arm were repeatedly drawn, and simulated each primary outcome using the control arms. The power gained from a covariate adjusted analysis for small and large samples was between 5% and 6%. Similar proportions of discordance with respect to statistical significance were noted irrespective of the sample size in both the GUSTO-I and the IST data sets.
Ordinal outcome analysis			

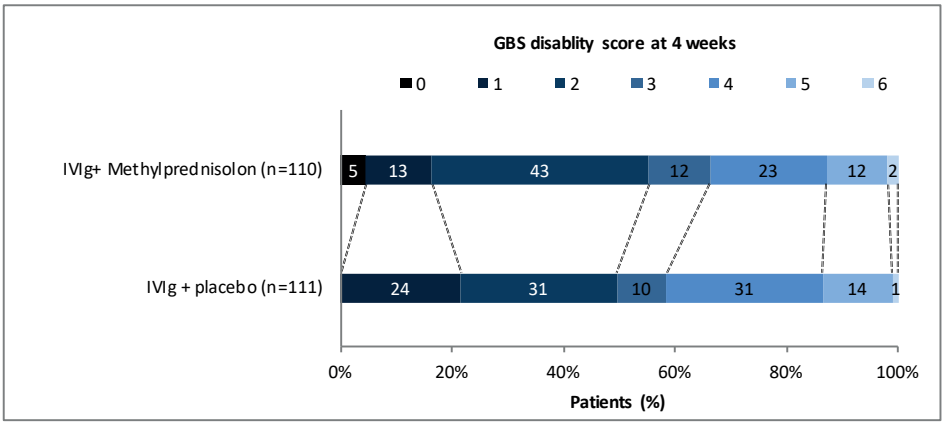
Table S1. Overview of a selection of methodological studies considering covariate adjustment and ordinal analysis in RCTs. (continued)

First author	Year	Field	Key findings and conclusions
Valenta(10)	2006	-	In this article conceptual and methodological aspects of employing proportional odds logistic regression for a three level ordinal factor as a suitable alternative to ordinary logistic regression when dealing with limited uncertainty in classifying clinical outcome as a binary variable are reviewed. Classifying a measurable clinical outcome as a dichotomous variable often involves difficulty with borderline cases that could fairly be assigned either of the two binary class memberships. In such situations the indicated class membership is often highly subjective and subject to, for instance, a measurement error. In other situations the intermediate level of a three-level ordinal factor may sometimes be explicitly reserved for cases which could likely belong to either of the two binary classes.
Optimizing the Analysis of Stroke Trials (OAST) Collaboration(3)	2007	Acute stroke	Data from 55 datasets (47 trials, 54,173 patients) from acute, rehabilitation and stroke unit trials studying the effects of interventions were used to asses which statistical approaches are most efficient in analyzing outcomes from stroke trials. The test results differed substantially so that approaches which use the ordered nature of functional outcome data (ordinal logistic regression, t test, robust ranks test, bootstrapping the difference in mean rank) were more efficient statistically than those which collapse the data into 2 groups (chi(2); ANOVA, $P<0.001$). The findings were consistent across different types and sizes of trial and for the different measures of functional outcome.
Saver(13)	2007	Acute stroke	Dichotomized, global statistic, responder, and shift analyses each offer distinctive benefits and drawbacks. Choice of primary end point analytic technique should be tailored to the study population, expected treatment response, and study purpose. Shift analysis generally provides the most comprehensive index of a treatment's clinical impact. Shift analysis gauges change in outcome distributions over the full range of ascertained outcomes, incorporating benefit and harm at all health state transitions valued by patients and clinicians, and often increasing study power.
Senn(33)	2009	-	Biostatisticians have frequently uncritically accepted the measurements provided by their medical colleagues engaged in clinical research, which often involve considerable loss of information. Particularly, unfortunate is the widespread use of the so-called 'responder analysis', which may involve not only a loss of information through dichotomization, but also extravagant and unjustified causal inference regarding individual treatment effects at the patient level, and, increasingly, the use of the so-called number needed to treat scale of measurement. Other problems involve inefficient use of baseline measurements, the use of covariates measured after the start of treatment, the interpretation of titrations and composite response measures. Statisticians should pay more attention to this aspect of their work.

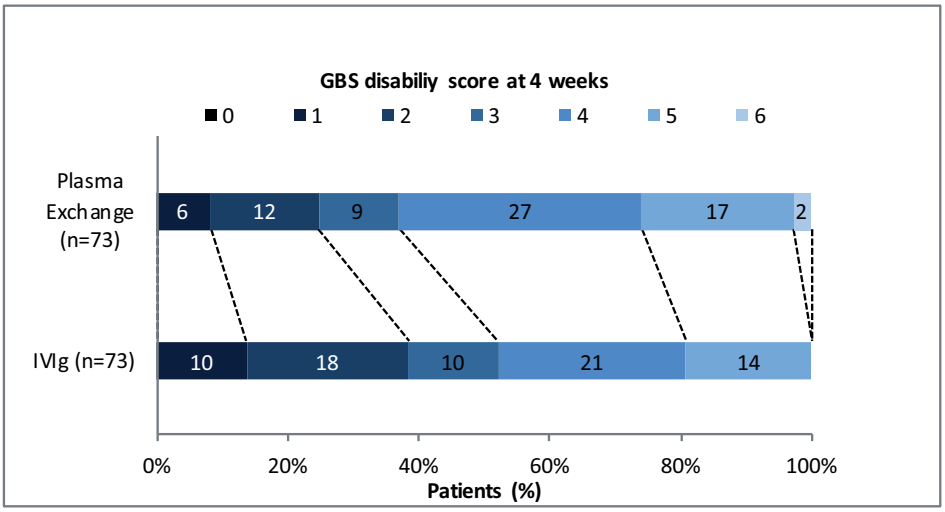
Table S1. Overview of a selection of methodological studies considering covariate adjustment and ordinal analysis in RCTs. (continued)

First author	Year	Field	Key findings and conclusions
McHugh(11)	2010	Traumatic brain injury	This study was based on simulations, which were built around a database of patient-level data extracted from eight Phase III trials and three observational studies in traumatic brain injury. Two different putative treatment effects were explored, one which followed the proportional odds model, and the other which assumed that the effect of the intervention was to reduce the risk of death without changing the distribution of outcomes within survivors. The simulation results show substantial efficiency gains. Use of the sliding dichotomy allows sample sizes to be reduced by up to 40% without loss of statistical power. The proportional odds model gives modest additional gains over and above the gains achieved by use of the sliding dichotomy.
Roozenbeek(12)	2011	Traumatic brain injury	Two techniques for ordinal analysis were applied using data from the CRASH trial (n = 9,554): proportional odds analysis and the sliding dichotomy approach, where the GOS is dichotomized at different cut-offs according to baseline prognostic risk. These approaches were compared to dichotomous analysis. Ordinal analysis with proportional odds regression or sliding dichotomy showed highly statistically significant treatment effects, with 2.05-fold and 2.56-fold higher information density compared to the dichotomous approach respectively. Analysis of the CRASH trial data confirmed that ordinal analysis of outcome substantially increases statistical power.
Diaz(34)	2016	Acute stroke	A general method for estimating the effect of a treatment on an ordinal outcome in randomized trials is presented. The method is robust in that it does not rely on the proportional odds assumption. Our estimator leverages information in prognostic baseline variables, and has all of the following properties: (i) it is consistent; (ii) it is locally efficient; (iii) it is guaranteed to have equal or b;etter asymptotic precision than both the inverse probability-weighted and the unadjusted estimators. The estimator is demonstrated in simulations based on resampling from a completed randomized clinical trial of a new treatment for stroke; we show potential gains of up to 39% in relative efficiency compared to the unadjusted estimator.

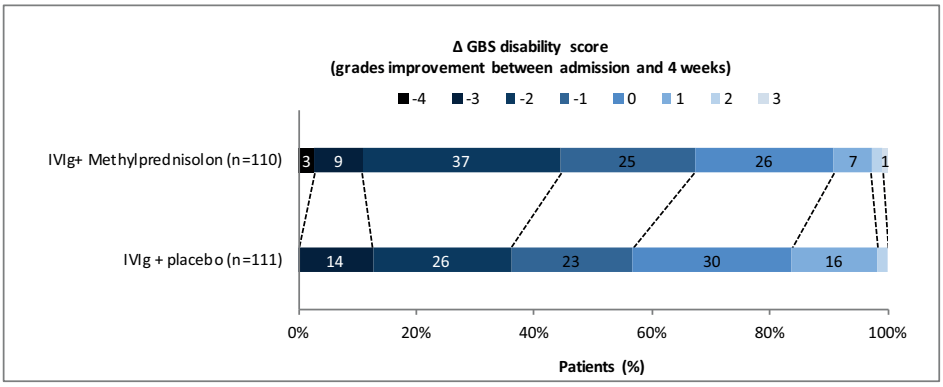
a)


GBS disablity score at 4 weeks
0
1
2
3
4
5
6
IVIg+ Methylprednisolon (n=110)
5
13
43
12
23
12
2
IVIg + placebo (n=111)
24
31
10
31
14
1
0%
20%
40%
60%
80%
100%
Patients (%)

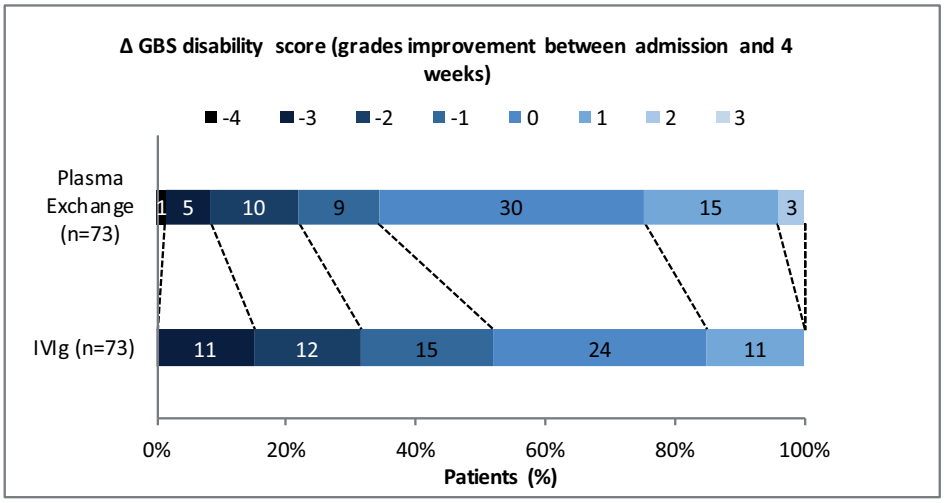
b)


GBS disabiliy score at 4 weeks
0
1
2
3
4
5
6
Plasma Exchange (n=73)
6
12
9
27
17
2
IVIg (n=73)
10
18
10
21
14
0%
20%
40%
60%
80%
100%
Patients (%)

c)


Δ GBS disability score
(grades improvement between admission and 4 weeks)
-4
-3
-2
-1
0
1
2
3
IVIg+ Methylprednisolon (n=110)
3
9
37
25
26
7
1
IVIg + placebo (n=111)
14
26
23
30
16
0%
20%
40%
60%
80%
100%
Patients (%)

d)



Appendix 1. Distribution of the GBS disability score at four weeks and improvement on the GBS disability score after four weeks in the IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial (a and c) and PE vs IVIg trial (b and d).

III Design and analysis of non-randomized studies: the regression discontinuity design

Chapter **5**

Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes

Nikki van Leeuwen
Hester F Lingsma
Anton J de Craen
Daan Nieboer
Simon Mooijaart
Edo Richard
Ewout W Steyerberg

Epidemiology 2016

ABSTRACT

Introduction

In epidemiology the regression discontinuity design has received increasing attention recently and might be an alternative to a randomized controlled trial (RCT) to evaluate treatment effects. In RD treatment is assigned above a certain threshold of an assignment variable for which the treatment effect is adjusted in the analysis.

Methods

We performed simulations and a validation study in which we used treatment effect estimations from an RCT as the reference for a prospectively performed regression discontinuity study. We estimated the treatment effect using linear regression with adjustment for the assignment variable both as linear terms and restricted cubic spline and using local linear regression models.

Results

In the first validation study the estimated treatment effect from a cardiovascular RCT was -4.0 mmHg blood pressure (95%CI -5.4;-2.6) at two years after inclusion. The estimated effect in regression discontinuity was -5.9 mmHg (95%CI -10.8; -1.0) with restricted cubic spline adjustment. Regression discontinuity showed different, local effects when analyzed with local linear regression. In the second RCT, regression discontinuity treatment effect estimates on total cholesterol level at three months after inclusion were similar to RCT estimates, but at least 6 times less precise.

Conclusion

Concluding, regression discontinuity may provide similar estimates of treatment effects to RCT estimates, but requires the assumption of a global treatment effect over the range of the assignment variable. In addition to a risk of bias due to wrong assumptions, researchers need to weigh better recruitment against the substantial loss in precision when considering a study with regression discontinuity versus RCT design.

INTRODUCTION

Randomized controlled trials (RCTs) are the reference standard to demonstrate the efficacy of medical interventions. However, recruitment of sufficient numbers of patients is a challenge in RCTs, and becomes increasingly difficult due to regulatory requirements. Including participants is particularly challenging when the effect of an intervention on an outcome is of interest, but withholding treatment is considered unethical. Also, patients may not want to be randomized(1, 2) or physicians are not willing to include patients.(3) It is estimated that between only 3% and 5% of all eligible adult cancer patients in the United Stated and the United Kingdom enroll in RCTs, partly due to dislike of the concept of trials and the idea of randomization of both patient and clinician.(4) Low recruitment rates in trials are also common in other fields, especially in surgery(5-6) and elderly(7, 8) research. Failure to achieve recruitment goals limits statistical precision, leads to an increase of costs, and decreases the efficiency of a RCT.(9) Even when investigators enroll an adequate number of participants, they rarely do so on schedule.(3, 10, 11)

Second, low recruitment rates threaten the generalizability of the findings in RCTs. Patients enrolled in trials may poorly represent the population of interest.(8, 12) Mostly women and elderly are underrepresented in RCTs.(8, 13)

An alternative epidemiologic design to assess effectiveness of medical treatment is the quasi-experimental "regression discontinuity" design. This design is common in the social sciences, and was introduced in public health and medicine in 1996.(14) Although in other fields regression discontinuity has been evaluated(15-20), recently Vandenbroucke et al.(21), Bor et al.(22) and O'Keeffe et al.(23) noted the importance of studying the feasibility and robustness of this design in clinical settings. In the regression discontinuity design, treatment is not allocated randomly, but is assigned to a subset of patients, based on a threshold of a baseline characteristic. The control group consists of a subset of patients below the threshold, not receiving treatment. The threshold variable does not necessarily have to be prognostic for the outcome measure assessed in the study. E.g. all patients with a baseline blood pressure (BP) over 140 mmHg may receive treatment (intervention group) and patients with a baseline BP below 140 mmHg do not receive treatment (control group). Such treatment allocation closely resembles clinical practice and may thus facilitate easier recruitment of participants into a prospective, comparative study. Due to the assignment rule, regression discontinuity designs can achieve balance on unobserved factors, just like in an RCT. When estimating the treatment effect, a regression analysis compares treated to control patients, while adjusting for the assignment variable, in this example baseline BP. Regression discontinuity provides a possible opportunity for obtaining unbiased causal effect estimates, when experiments are not feasible or when we want to evaluate interventions under "real-life" circumstances.(24)

The regression discontinuity design as a prospective quasi-experimental study might be attractive because the challenges of the randomization process are avoided. However, the estimates from a quasi-experimental regression discontinuity design might be different and substantially less efficient compared to an RCT. We aimed to assess validity of this design as a prospective quasi-experimental design compared to a traditional RCT, since this has not been discussed in detail in the epidemiologic literature. In this study we perform simulation studies and analyze data from two large cardiovascular RCTs as validation studies.

METHODS

Monte Carlo Simulations

We used Monte Carlo simulations to compare regression discontinuity and RCT. One hundred patients were simulated with a hypothetical mean prognostic measurement of 10 and a standard deviation (SD) of 2. The mean outcome was 90 (SD 20) and was normally distributed. A treatment effect of -10 was simulated. Linear correlations of the prognostic measurement with outcome were varied (R^2 0.0, 0.2, 0.5 and 0.8) to assess the importance of the prognostic strength of the adjustment model in the regression discontinuity design.

For the RCT, treatment was allocated at random, and sample size was 100 patients in total (50 per arm). In the regression discontinuity setting, treatment was assigned to 50 patients with a prognostic measurement above 10; 50 patients with a prognostic measurement below 10 were used as controls. For both the RCT and regression discontinuity settings, linear regression models were used to estimate the treatment effect, adjusted for the baseline variable both in a linear term and a restricted cubic spline term. The regression discontinuity setting was also analyzed with local linear regression analysis. In local linear regression, only patients around the threshold are used in the analysis to estimate a *local* treatment effect while normal regression uses all patients, resulting in a *global,* or average, treatment effect estimate. Treatment effect estimates were compared in terms of bias (expressed as mean estimated treatment effect) and precision (expressed as mean squared error of the treatment effect estimate). The simulation code is provided in the Appendix.

Validation study

Two cardiovascular trials were used to validate the regression discontinuity design in empirical data. The "Prevention of Dementia by Intensive Vascular Care" study (preDIVA) is an ongoing cluster-randomized trial to assess the efficacy of a multicomponent, nurse-led intervention targeting all cardiovascular risk factors in an elderly population (70-78

years). The study protocol has been approved by the medical ethical committee of the Academic Medical Centre.(25) The primary outcome of this RCT is incident dementia during 6 years of follow-up. Of 3533 patients enrolled, 1894 are in the intervention and 1639 in the control group. As this RCT is ongoing we randomly sampled 3000 patients from the enrolled patients.

The second RCT was the "PROspective Study of Pravastatin in elderly (70 to 82 years) individuals at risk of vascular disease" (PROSPER). It was conducted between December 1997 and May 1999 and enrolled 5804 patients, who were assigned to pravastatin (n=2891) or placebo (n=2913) to reduce the risk of coronary disease in elderly individuals.(26)

To validate the regression discontinuity design we used continuous measures collected during follow-up, which were not the primary endpoints of the trials, as the outcome variable. To evaluate the influence of the choice of the assignment variable on the estimates we did multiple analyses using two different baseline measures. For preDIVA both age and the blood pressure (BP) at baseline were used as the assignment variable and the BP at 2-year intermediate follow-up as outcome. Both BP measures were calculated as the mean of two systolic blood pressure measurements during a visit (expressed in mmHg). BP data at 2 years were unavailable for part of the 3000 randomly selected patients, mostly because they reached a clinical endpoint before 2 years, or missed one study visit. We could hence include 2346 patients for analysis. For PROSPER, we considered total cholesterol level measured three months after inclusion (expressed in mmol/L) as the outcome and both age and total cholesterol at baseline as the treatment assignment variables. After exclusion of patients with missing 3-month total cholesterol we were able to analyze 5581 patients from PROSPER.

Statistical analysis

Baseline characteristics were described with standard descriptive statistics; median and interquartile range for continuous variables and frequencies and percentages for categorical variables.

To analyze the data as an RD design, we selected those patients with a value of the assignment variable above a certain threshold treated as the intervention group, and those with a value below that threshold and not treated as control group. This thus led to a sample of approximately half the trial population. In both trials we used a threshold of the assignment variable known to be used in clinical practice(27, 28) (e.g. BP > 140 mmHg), or, if not available, a hypothetical threshold (e.g. age > 72 years). Histograms of the assignment variables in both preDIVA and PROSPER are shown in the eAppendix 1. In this patient selection the treatment effect was estimated using a linear regression model, with adjustment for the assignment variable (Y ~ Tx + Baseline assignment variable). We further analyzed a hypothetical different cut-off to assess the robustness of the

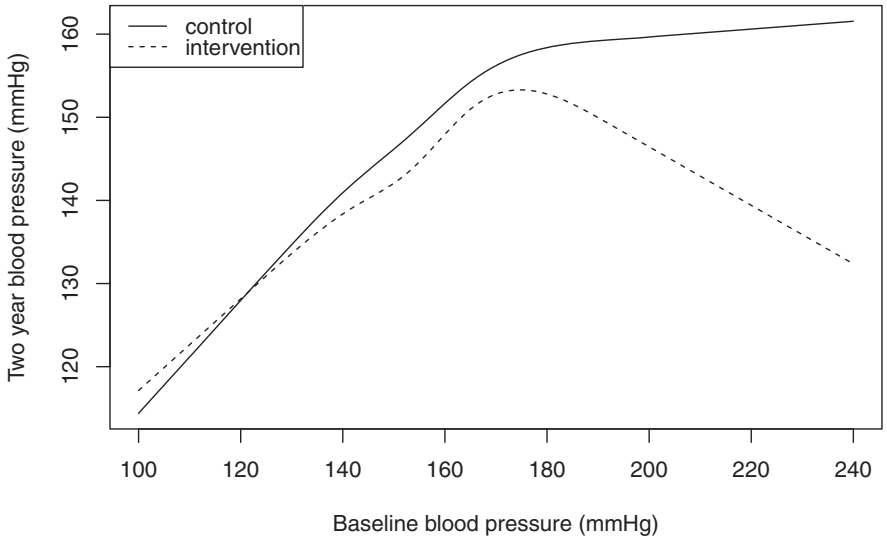
treatment effect estimate to the chosen cut-off. In other words, whether the treatment effect was global or local. Usually this would not be possible in regression discontinuity as the cut-off is determined in advance.

Analysis on these sets of patients was compared to a random sample (repeated 5,000 times) of half of the of the RCT patients. To compare relative efficiency in terms of required sample size between the different approaches we used standard errors (SEs) of the estimated treatment effects in the following formula: $(SE_{RD} / SE_{RCT})^2$. This is the ratio of variances in both designs.

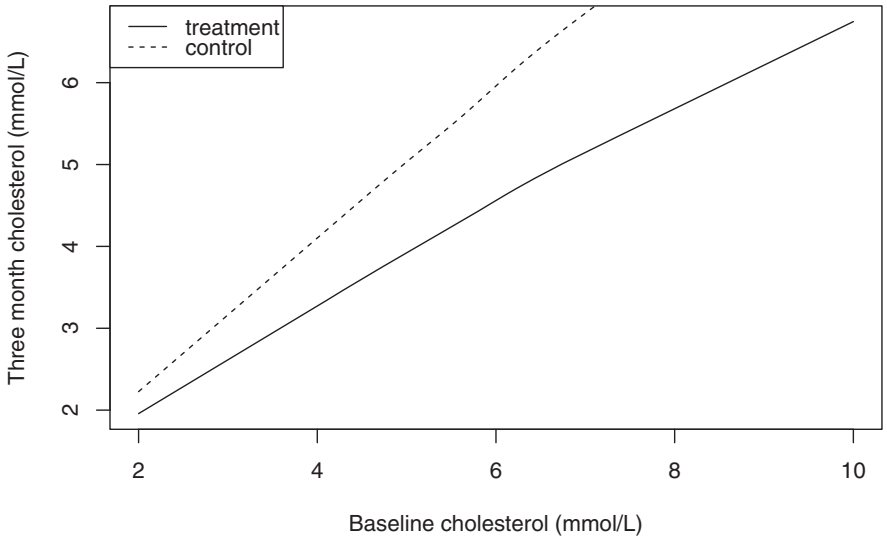
As the validity of the regression discontinuity design is dependent on proper adjustment for the assignment variable we explored the relation between the assignment variable and outcome in detail. We assessed non-linearity with non-linear restricted cubic spline functions and interaction between the baseline assignment variable and the treatment effect. Both were presented graphically (Figure 1a and 1b). A restricted cubic spline function is a smooth function that consists of pieced-together cubic splines that are restricted to be linear in the tails.(29) We used the default setting for flexibility of the model with five knots.(30) Consequently we used the restricted cubic spline of the assignment variable in the adjustment model to obtain the optimal model fit for adjustment. This method was compared to local linear regression models for the adjustment of the baseline variable. All analyses (RCT and regression discontinuity) were adjusted for the assignment measurement that was used to assign treatment; both age and baseline BP in preDIVA and both age and baseline cholesterol in the PROSPER trial. R^2 statistics were calculated to indicate the explained variance of the assignment model.

We further explored different assumptions on interaction between the assignment variable and treatment. We assumed no interaction between age and treatment in both studies. Therefore, we considered the treatment effect estimates in the regression discontinuity studies in which treatment was assigned on age, global treatment effects. We compared these treatment effect estimates to the global effect estimated in the RCT. The treatment effect estimates from the regression discontinuity studies, where treatment was assigned on baseline BP and baseline total cholesterol level, could be considered as local treatment effects, since we assumed interaction between treatment over both baseline BP and baseline cholesterol level. Therefore we compared these estimates to the local effects in the RCT, estimated with restricted cubic spline adjustment. These estimates are the differences between the treated and untreated lines in figures 1a and 1b. Treatment effects were estimated using linear regression and expressed as regression coefficients with 95% confidence Intervals (95% CIs) for blood pressure or cholesterol level in the treatment group compared to the controls.

All statistical analyses were performed in R statistical software version 2.15.3 (R Foundation for Statistical Computation, Vienna, Austria) using the rdd package and Harrell's rms package.

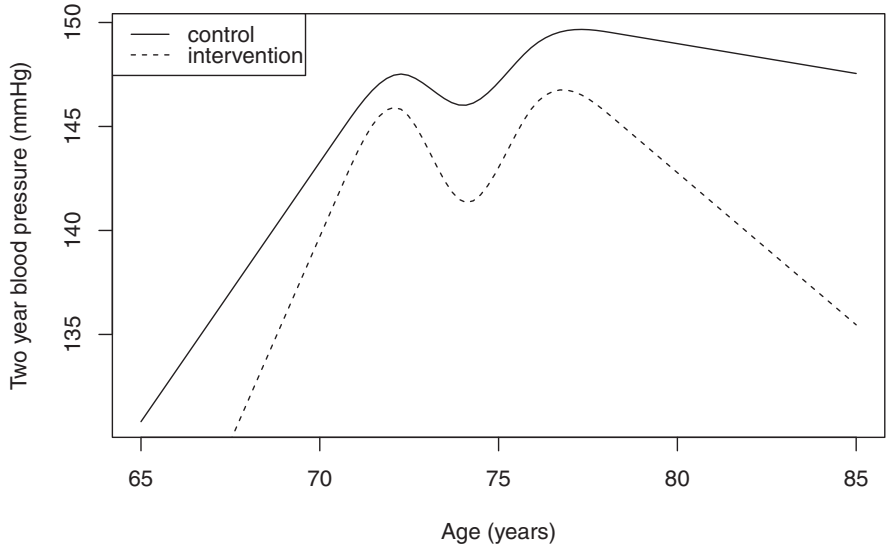


a) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the intervention effect over baseline blood pressure in the preDIVA study. ** The function fitted is: Two year blood pressure ~ Intervention * rcs(Baseline blood pressure).*

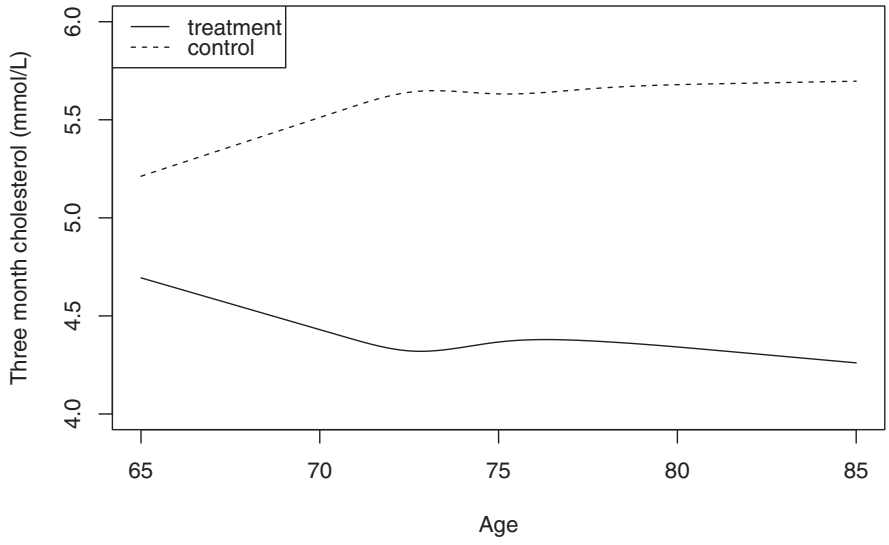


b) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the treatment effect over baseline total cholesterol in the PROSPER trial. ** The function fitted is: Three month cholesterol ~ Treatment * rcs(Baseline cholesterol).*

Figure 1. Non-linear restricted cubic spline functions of the interaction of the intervention effects over the assignment variables in both the preDIVA study (n = 2346) and the PROSPER trial (n = 5581).



c) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the intervention effect over baseline age in the preDIVA study. ** The function fitted is: Two year blood pressure ~ Intervention * rcs(Baseline age).*



d) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the treatment effect over baseline age in the PROSPER trial. ** The function fitted is: Three month cholesterol ~ Treatment * rcs(Baseline age).*

Figure 1. Non-linear restricted cubic spline functions of the interaction of the intervention effects over the assignment variables in both the preDIVA study (n = 2346) and the PROSPER trial (n = 5581).

RESULTS

Simulations

Simulations showed that under the ideal situation of a linear relation between the assignment variable and the outcome, no unmeasured confounders, and no treatment effect interaction, regression discontinuity provided unbiased treatment effect estimates all scenarios (Table 1). However, RD with linear adjustment resulted in substantially less precise effect estimates compared to the RCT. For instance, in the hypothetical setting with an R^2 of 20% for the correlation of the assignment measurement with outcome, the mean squared error of the estimated treatment effect estimate in an RCT was 3.2 compared to 9.0 in the RD design. In this example this means that when the regression discontinuity design is used and the analysis matches the underlying true model, triple the number of patients is required to obtain the same statistical precision as when using an RCT. This magnitude of (in)efficiency was consistent for all simulated correlations between the assignment measurement and outcome. Regression discontinuity estimates adjusted with restricted cubic splines were 7 times less efficient than the RCT estimates analyzed with restricted cubic splines. When analyzing regression discontinuity with local linear regression, the estimated were on average 1.4 times less efficient than analyzing regression discontinuity with restricted cub spline adjustment (Table 1).

Table 1. Simulations treatment effect over baseline in randomized control trial and regression discontinuity design, analyzed with linear regression, restricted cubic spline functions and local linear regression.

		Randomized controlled trial				Regression discontinuity design			
	R^2 (%)	0	20	50	80	0	20	50	80
Linear regression	**Mean squared error**	4.2	3.2	2.0	0.8	11.1	9.0	5.6	2.3
Restricted cubic spline adjustment	**Mean squared error**	4.3	3.3	2.1	0.8	29.5	23.3	14.9	6.0
Local linear regression	**Mean squared error**	NA	NA	NA	NA	39.8	32.6	20.7	7.9

Validation study

In the validation study we assessed blood pressure in 2346 patients from the preDIVA trial. The median age was 74 years (IQR: 72, 76) and the median BP at baseline was 153 mmHg (IQR: 140, 168). We analyzed 5581 patients from the PROSPER trial, who had a median age of 74 years (IQR: 72 - 77 years) and a median total cholesterol level of 5.6 mmol/L (IQR: 5.0, 6.3) at baseline (Table 2).

In the RCT design, the treatment effect estimate on BP at two years adjusted for baseline BP was -4.0 mmHg (95% CI -5.4; -2.6) (Table 3).

Table 2. Patient characteristics of preDIVA (n=2346) and PROSPER (n=5581). Numbers are the median (IQR) or frequency (%)

Characteristic	preDIVA	PROSPER
Age, years	74 (72, 76)	74 (72, 77)
Sex, male	1071 (46)	2708 (49)
Baseline blood pressure in mmHg	153 (140, 168)	-
2-year blood pressure in mmHg	148 (136, 162)	-
Baseline cholesterol in mmol/L	-	5.6 (5.0, 6.3)
Three month cholesterol in mmol/L	-	4.9 (4.2, 5.7)

The adjusted estimated effect was -5.9 mmHg (95% CI -10.8; -1.0) with the hypothetical regression discontinuity design, adjusted for BP as restricted cubic spline variable. Here, patients with a BP over a baseline BP of 140 mmHg would receive treatment and patients with a lower BP would not receive treatment. An additional analysis of a hypothetical RD design in which patients with a BP of 160 mmHg would receive treatment and the higher baseline BP patients not, adjusted with a restricted cubic spline showed a treatment effect estimate of -5.9 mmHg (95% CI -11.4; -0.3) (Table 3).

In the preDIVA trial, there appeared to be a different treatment effect for patients with relative low baseline BP compared to high baseline BP, indicating statistical interaction (Figure 1a). This explains the different effect estimates in the regression discontinuity setting with treated high-risk patients and low-risk controls compared to the RCT estimate. When the treatment effect was analyzed with local linear regression, the intervention effect estimates in the two regression discontinuity designs were more different than the RCT estimate: -9.3 mmHg (95% CI -18.5; -0.1) for the setting with the cut-off at 140 mmHg and -10.2 mmHg (95% CI -21.0; 0.6) for the setting with the cut-off at 160 mmHg (Table 3).

In Figure 1c shows interaction between treatment and age. A global effect of treatment over the age range was assumed. The estimates from the regression discontinuity design analyzed with restricted cubic splines showed more similar estimates for the different cut-offs and these were closer to the estimate from the RCT (-0.66 (-6.44; 5.12) -2.71 (-7.16; 1.74)).

In the PROSPER trial, the estimated treatment effect on total cholesterol level at three months in the RCT, adjusted for baseline total cholesterol level was -1.31 mmol/L (95% CI -1.35; -1.27). In a hypothetical regression discontinuity, we used a threshold of 5.0 mmol/L for treatment assignment. The treatment effect estimate in PROSPER was beneficial over the whole range of baseline total cholesterol level, but differed in magnitude (Figure 1b). Analysis with local linear regression showed different treatment effect estimates: -1.04 mmol/L (95% CI -1.16; -0.93) with the cut-off at baseline BP of 5.0 mmol/L and -1.29 mmol/L (95% CI -1.40; -1.18) with the cut-off at baseline BP of 5.5 mmol/L, Table 4).

Table 3. RCT and RD analyses in the preDIVA study (n = 2346)

Analysis	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for two year blood pressure	Standard error (SE) of treatment effect estimate	Design and patients	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for two year blood pressure	Standard error (SE) of treatment effect estimate
		RCT						**RCT**			
Linear adjustment	1173	-	25	-4.0; (-5.4, -2.6)	0.7	*Linear adjustment*	1173	-	1	-3.2 (-5.5, -0.9)	1.2
		RD – assignment on blood pressure[1]						**RD – assignment on age[3]**			
Linear adjustment	1218	965 / 253	21	3.3 (0.1, 6.5)	1.6	*Linear adjustment*	1246	278 / 968	0	-2.4 (-6.0, 1.2)	1.8
RCS adjustment	1218	965 / 253	23	-5.9 (-10.8, -1.0)	2.5	*RCS adjustment*	1246	278 / 968	0	-0.7 (-6.4, 5.1)	3.0
Local linear regression	1218	965 / 253	NA	-9.3 (-18.5, -0.1)	4.7	*Local linear regression*	1246	278 / 968	NA	2.4 (-3.7, 8.5)	3.1
		RD – assignment on blood pressure[2]						**RD – assignment on age[4]**			
Linear adjustment	1201	505 / 696	18	-3.3 (-6.7, 0.2)	1.8	*Linear adjustment*	1176	547 / 629	0	-2.7 (-7.2, 1.7)	2.3
RCS adjustment	1201	505 / 696	22	-5.9 (-11.4, -0.3)	2.8	*RCS adjustment*	1176	547 / 629	0	-2.5 (-9.2, 4.2)	3.4
Local linear regression	1201	505 / 696	NA	-10.2 (-21.0, 0.6)	5.5	*Local linear regression*	1176	547 / 629	NA	-2.8 (-8.2, 2.7)	2.8

Patient selections:

[1] BP ≤ 140 receiving no treatment and BP > 140 receiving treatment

[2] BP ≤ 160 receiving no treatment and BP > 160 receiving treatment

[3] Age ≤ 72 receiving no treatment and age > 72 receiving treatment

[4] Age ≤ 74 receiving no treatment and age > 74 receiving treatment

Table 4. RCT and RD analyses in the PROSPER trial (n=5581)

Analysis	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for three month cholesterol level	Standard error (SE) of treatment effect estimate	Design and patients	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for three month cholesterol level	Standard error (SE) of treatment effect estimate
		RCT						**RCT**			
Linear adjustment	2790	-	75	-1.31 (-1.35, -1.27)	0.021	*Linear adjustment*	2790	-	35	-1.28 (-1.35, -1.22)	0.033
		RD – assignment on cholesterol level[1]						**RD – assignment on age[3]**			
Linear adjustment	2787	2066 / 721	34	-0.99 (-1.06, -0.92)	0.034	*Linear adjustment*	2792	586 / 2206	26	-1.24 (-1.34, -1.13)	0.051
RCS adjustment	2787	2066 / 721	35	-1.14 (-1.25, -1.03)	0.057	*RCS adjustment*	2792	586 / 2206	26	-1.27 (-1.47, -1.08)	0.099
Local linear regression	2787	2066 / 721	NA	-1.04 (-1.16, -0.93)	0.058	*Local linear regression*	2792	586 / 2206	NA	-1.16 (-1.39, -0.93)	0.117
		RD – assignment on cholesterol level[2]						**RD – assignment on age[4]**			
Linear adjustment	2824	1532 / 1292	32	-1.14 (-1.21, -1.07)	0.034	*Linear adjustment*	2818	1175 / 1643	33	-1.22 (-1.33, -1.11)	0.057
RCS adjustment	2824	1532 / 1292	34	-1.28 (-1.40, -1.17)	0.065	*RCS adjustment*	2818	1175 / 1643	33	-1.27 (-1.49, -1.05)	0.112
Local linear regression	2824	1532 / 1292	NA	-1.29 (-1.40, -1.18)	0.055	*Local linear regression*	2818	1175 / 1643	NA	-1.26 (-1.45, -1.07)	0.097

Patient selections:

[1] Cholesterol ≤ 5.0 receiving no treatment and cholesterol > 5.0 receiving treatment

[2] Cholesterol ≤ 5.5 receiving no treatment and cholesterol > 5.5 receiving treatment

[3] Age ≤ 72 receiving no treatment and age > 72 receiving treatment

[4] Age ≤ 74 receiving no treatment and age > 74 receiving treatment

Table 5a. Relative efficiency of global treatment effect estimates in RD design in terms of required sample size (compared to global treatment effect estimate in RCT design) for different validation studies in preDIVA and PROSPER*.

	preDIVA	PROSPER
RCT (linear adjustment) vs RD (RCS adjustment)	6.25 [1]	9.0 [1]
RCT (linear adjustment) vs RD (RCS adjustment)	8.45 [2]	11.52 [2]

**Formula: $(SE_{RD} / SE_{RCT})^2$*

[1] Patient selection age ≤ 72 Tx- and BP > 72 Tx+

[2] Patient selection age ≤ 74 Tx- and BP > 74 Tx+

Table 5b. Relative efficiency of local treatment effect estimates in RD design in terms of required sample size (compared to local treatment effect estimate in RCT design) for different validation studies in preDIVA and PROSPER*.

	preDIVA	PROSPER
RCT (RCS adjustment) vs RD (local linear regression)	3.56 [1]	1.04 [3]
RCT (RCS adjustment) vs RD (local linear regression)	3.78 [2]	0.72 [4]

**Formula: $(SE_{RD} / SE_{RCT})^2$*

[1] Patient selection BP ≤ 140 Tx- and BP > 140 Tx+

[2] Patient selection BP ≤ 160 Tx- and BP > 160 Tx+

[3] Patient selection cholesterol ≤ 5.0 Tx- and cholesterol > 5.0 Tx+

[4] Patient selection cholesterol ≤ 5.5 Tx- and cholesterol > 5.5 Tx+

Analysis with restricted cubic spline adjustment in regression discontinuity in PROSPER showed only slightly different treatment effect estimates compared to the results from the regression discontinuity setting analyzed with local linear regression (Table 4).

In the hypothetical regression discontinuity design in PROSPER with assignment on the age variable, both analyses with local linear regression and normal linear regression with adjustment for age as a restricted cubic spline variable showed similar treatment effect estimates compared to the estimates from the RCT (Table 4). This is consistent with Figure 1d, which shows a constant effect of treatment over the whole range of age.

In terms of statistical precision, the regression discontinuity with restricted cubic spline adjustment was 1 to 4 times less efficient than the adjusted RCT for the local effects estimated, and 6 to 12 times less efficient for the global effects estimated in regression discontinuity compared to the RCT. (Table 5)

DISCUSSION

In this study we performed simulation studies and analyzed data from two large cardiovascular RCTs with the aim to assess the validity of the regression discontinuity design as a prospective quasi-experimental design compared to a traditional RCT. In the

simulations studies we found unbiased but substantially less precise treatment effect estimates from an regression discontinuity design compared to an RCT design.

In the two validation studies we found somewhat different results. In the case of the treatment assignment on baseline BP and baseline cholesterol level we assumed treatment effect heterogeneity over the baseline assignment variable. Therefore we assumed the treatment effect estimates in these regression discontinuity designs to be local. The estimates from the analyses with local linear regression are not comparable with the RCT estimates in terms of bias since these effects are local, in contrast to the global RCT estimates. In terms of statistical precision, the regression discontinuity with restriction cubic spline adjustment was 1 to 4 times less efficient than the adjusted RCT for the local effects estimated.

When treatment was assigned on the age variable, we assumed no interaction between age and treatment. Therefore, the treatment effect estimate for the regression discontinuity analysis was assumed to be a global treatment effect. Estimates from the analyses with flexible functions (restricted cubic splines) to obtain an optimal fit of the adjustment model were comparable to the RCT estimate. In PROSPER the treatment effect estimates were consistent over the different cut-offs, which confirms the assumption of no interaction. In preDIVA, the estimates were quite different over the different cut-offs. This gives the impression that there is interaction and that the treatment effects are local over the baseline age range. Further, we found that sample size needs to be at least 6 times larger to make regression discontinuity as precise as an RCT to estimate global treatment effect estimates.

Position of regression discontinuity design in epidemiology

The main goal of our study was to assess the regression discontinuity design as a prospective design, and compare it with an RCT design. The regression discontinuity design could be implemented as an observational study design, but we focused on the situation of prospective enrollment of patients, with a predefined cut-off. While estimation of treatment effects with retrospectively collected data is could be hampered by selection bias and confounding by indication, which is difficult to fully control for since unmeasured confounding cannot be accounted for ('residual confounding'), in the prospective application of the regression discontinuity design the confounding variable is measured and controlled for by design.

Local or global treatment effect and model specification

A very important question in designing a regression discontinuity study is whether a global or a local treatment effect should be assumed. The assignment variable in an RCT is random allocation and therefor would not interact with treatment in an RCT. This results in a global or average treatment effect. In contrast, the assignment variable in

regression discontinuity could interact with treatment. Therefor the estimated effect is not always a global effect, but a local treatment effect estimate. Only when the assumption of no interaction between the baseline assignment variable and treatment can be made, the estimated effect of regression discontinuity can be interpreted as a global treatment effect. Our analyses with normal linear regression models with restricted cubic spline adjustment assume such a global treatment effect. In our RCT data we could test the assumption of no interaction as we had treatment and control patients across the complete range of the assignment variable. In practice, however, this assumption cannot be tested in regression discontinuity designs, as treatment and control patients have no overlap in the assignment variable.

When the aim is to estimate the treatment effect on a certain threshold value of the assignment variable, local linear regression should be used. This estimate will be unbiased for that certain cut-off, but it is only valid for a small subset of patient around the cut-off and is not generalizable for the whole population.

Thus, if one aims to estimate an overall treatment effect from RD, using a normal regression models is preferred over local linear regression, as supported by our finding that the treatment effect estimates from regression discontinuity with restricted cubic splines more were more similar to the RCT estimate than the local linear regression estimates. However, this could be expected because local linear regression uses only data around the cut-off while normal regression uses all data, as happens in the RCT. In fact the local linear regression and restricted cubic splines estimates cannot be judged as more or less biased. They estimate a different effect (global or local) and their validity is dependent on the assumptions made.

In our case of blood pressure the assumption of no interaction was not met: graphical inspection showed qualitative treatment effect heterogeneity (Figure 1a). The intervention is a multi-component intervention tailored according to patients' risk profile, and the intensity of the intervention (both medical and lifestyle) is thus higher for participants with a higher risk profile.(25) This difference in intensity might explain the interaction of treatment with baseline risk for outcome, i.e. a stronger effect in patients with higher baseline risk. This led to a different treatment effect when we misspecified the adjustment model.

Optimal modeling of the effect of the assignment variable is extremely important in regression discontinuity. We used restricted cubic spline functions and local linear regression. Restricted cubic spline functions are attractive for their flexibility with low degrees of freedom. They and not driven by extreme values in both ends of the fit, which is an advantage over ordinary cubic spline functions.(29) We suggest that flexible functions should be used for optimal adjustment since this function accounts for differential effects of the baseline variable that is used as the assignment variable, on outcome.

Choice of regression discontinuity versus randomized clinical trials

Requirements for informed consent for clinical trials are often more stringent than for treatment outside of the setting of an RCT.(31) Patient enrollment may hence be easier due to avoidance of the randomization process. More lenient inclusion criteria and easier enrollment when using the RD design most likely will result in more representative cohorts for analysis.

A regression discontinuity design is attractive when random assignment of treatment is problematic or not possible. This may occur when a medical intervention is already standard practice for a part of the patients in clinic but the effectiveness has not yet been assessed. For instance, the effect of blood pressure and cholesterol lowering on incident dementia has only been studied in randomized controlled trials using dementia as a secondary outcome and with inconclusive results.(32) At present performing such RCTs is no longer considered ethical, as there is a clinical imperative to treat those with high blood pressure and cholesterol. However, there is circumstantial evidence that there is a beneficial effect of BP reduction on dementia risk, which is not translated to clinical guidelines in the absence of evidence to substantiate this claim. In this situation RD could be a solution to assess a treatment effect using less affected patients for whom an intervention is not deemed indicated as control patients. In fact, because this strategy already closely resembles clinical practice, it may well be feasible to include a number of participants in such a trial that is 6-12 fold higher than in a classical RCT. Adherence to assignment of treatment according to the threshold value is crucial, and both participating clinicians and patients should therefore be well aware that they are participating in a comparative study. A possible threat when using a regression discontinuity design is selection bias near the threshold value. When physicians selectively treat patients just below the threshold value and vice versa, selection bias occurs due to confounding by indication. It is thus very important to avoid such protocol violations in a prospective regression discontinuity study.

Strengths and limitations

A limitation of this study is that in the simulation study we assumed the ideal condition of a linear treatment effect and no residual confounding, which may not reflect real life practice. However, we showed that even in such an ideal setting regression discontinuity is less efficient than RCT. Further, we only studied continuous outcomes and therefore cannot draw conclusions on the performance of the regression discontinuity design for dichotomous outcome parameters. The relative inefficiency may be different for such settings. Furthermore, we assessed the regression discontinuity design in RCT data, in which we artificially set the threshold. This results in perfect adherence to the defined threshold, which is unlikely to occur in real life.

On the other hand, the possibility to change the cut-off was a major strength in our study. We were able to study interaction by varying the cut-off, which would be impossible in a 'real' regression discontinuity design. Moreover, with the RCT data we were able to study the effect of different hypothetical variables to assign treatment, which is a unique feature of this study.

There might be measurement error in the assignment variable. Many claim that regression to the mean caused by such measurement error is a possible threat to the validity of regression discontinuity.(14). A high baseline measurement will on expectation regress down to a lower value and a low baseline measurement will on expectation regress up to a higher value. However, as this will occur equally on both sides of the cut-off in the assignment variable, the measurement error will in the end be irrelevant for the correct estimation of the treatment effect.(21)

Conclusion

We conclude that the regression discontinuity design has perfect theoretical validity and may have reasonable validity in real life situations compared to RCT. Regression discontinuity may provide similar estimates of treatment effects to RCT estimates, but requires the assumption of a global treatment effect over the range of the assignment variable. Controlling for the assignment variable is essential and may be achieved by an optimal fit of the adjustment model, with for example restricted cubic splines when the assumption of a global treatment effect over the range of the assignment variable can be made. Regression discontinuity is, however inefficient, requiring sample sizes which are over 6 times higher than for conventional RCTs to obtain the same statistical precision for a global treatment effect estimate. When considering a study with regression discontinuity versus RCT design, in addition to a risk of bias due to wrong assumptions, researchers need to weigh better recruitment against the substantial loss in precision.

REFERENCES

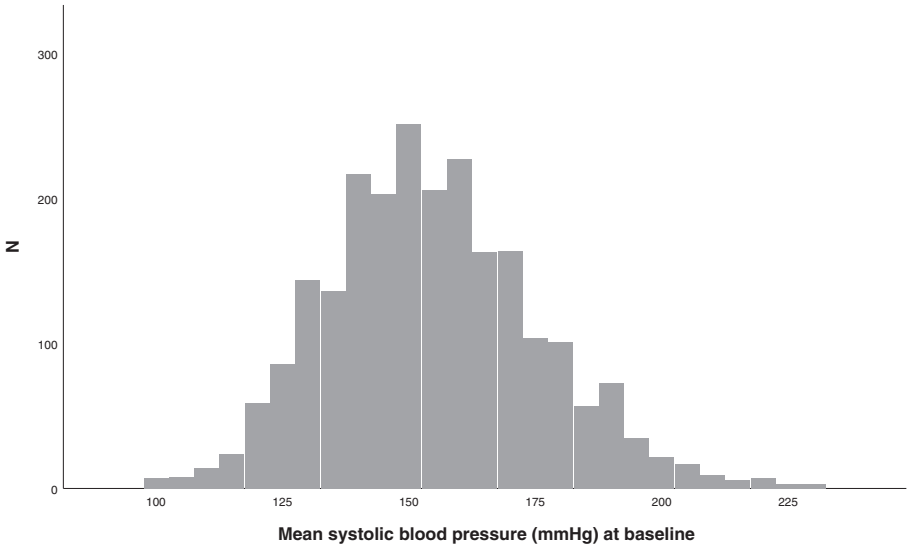
1. Wheeler JL, Segal AD, Ledoux WR, et al. Patient willingness to enroll in a randomized surgical trial. Presented at the 25th Annual Summer Meeting of the American Orthopaedic Foot and Ankle Society. July 15-18, 2009. Vancouver, British Columbia.
2. Jenkins V, Fallowfield L. Reasons for accepting or declining to participate in randomized clinical trials for cancer therapy. Br J Cancer. 2000 Jun;82(11):1783-8.
3. Gotay CC. Accrual to cancer clinical trials: directions from the research literature. Soc Sci Med. 1991;33(5):569-77.
4. Harrison JD, Solomon MJ, Young JM, Meagher A, Hruby G, Salkeld G, et al. Surgical and oncology trials for rectal cancer: who will participate? Surgery. 2007 Jul;142(1):94-101.
5. Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. Surgery. 2006 Apr;139(4):469-83.
6. Creel AH, Losina E, Mandl LA, Marx RJ, Mahomed NN, Martin SD, et al. An assessment of willingness to participate in a randomized trial of arthroscopic knee surgery in patients with osteoarthritis. Contemp Clin Trials. 2005 Apr;26(2):169-78.
7. Chang BH, Hendricks AM, Slawsky MT, Locastro JS. Patient recruitment to a randomized clinical trial of behavioral therapy for chronic heart failure. BMC Med Res Methodol. 2004 Apr 17;4:8.
8. Martinson BC, Crain AL, Sherwood NE, Hayes MG, Pronk NP, O'Connor PJ. Population reach and recruitment bias in a maintenance RCT in physically active older adults. J Phys Act Health. 2010 Jan;7(1):127-35.
9. Vozdolska R, Sano M, Aisen P, Edland SD. The net effect of alternative allocation ratios on recruitment time and trial cost. Clin Trials. 2009 Apr;6(2):126-32.
10. Halpern SD. Prospective preference assessment: a method to enhance the ethics and efficiency of randomized controlled trials. Control Clin Trials. 2002 Jun;23(3):274-88.
11. Institute of Medicine Forum on Drug Discovery D, Translation. 2010 Opportunities: Workshop Summary.
12. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: who are these patients and how did they get there? Ann Intern Med. 2002 Jul 2;137(1):10-6.
13. Beckie TM, Mendonca MA, Fletcher GF, Schocken DD, Evans ME, Banks SM. Examining the challenges of recruiting women into a cardiac rehabilitation clinical trial. J Cardiopulm Rehabil Prev. 2009 Jan-Feb;29(1):13-21; quiz 2-3.
14. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton-Mifflin; 2002. p 229.
15. Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. J Policy Anal Manag. 2008 Fal;27(4):724-50.
16. Wong VC, Cook TD, Barnett WS, Jung K. An effectiveness-based evaluation of five state prekindergarten programs. J Policy Anal Manag. 2008 Win;27(1):122-54.
17. Aiken LS, West SG, Schwalm DE, Carroll JL, Hsiung S. Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation - Efficacy of a university-level remedial writing program. Evaluation Rev. 1998 Apr;22(2):207-44.
18. Goldberger A.S. Selection bias in evaluating treatment effects: Some formal illustrations. Working Paper, Economics Department, University of Wisconsin. 1972.
19. Buddelmeyer H, Skoufias E. An evaluation of the Performance of Regression Discontinuity Design on PROGRESA. World Bank Policy Research. Aug 2004.

20. Black D, Galdo J, Smith J. Estimating the selection bias of the regression discontinuity design using experimental data. Working paper, University of Chicago. 2007.
21. Vandenbroucke JP, le Cessie S. Commentary: regression discontinuity design: let's give it a try to evaluate medical and public health interventions. Epidemiology. 2014 Sep;25(5):738-41.
22. Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014 Sep;25(5):729-37.
23. O'Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. BMJ. 2014;349:g5293.
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 Feb;68(2):122-33.
25. Richard E, Van den Heuvel E, Moll van Charante EP, Achthoven L, Vermeulen M, Bindels PJ, et al. Prevention of dementia by intensive vascular care (PreDIVA): a cluster-randomized trial in progress. Alzheimer Dis Assoc Disord. 2009 Jul-Sep;23(3):198-204.
26. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet. 2002 Nov 23;360(9346):1623-30.
27. White HD, Simes RJ, Anderson NE, Hankey GJ, Watson JD, Hunt D, et al. Pravastatin therapy and the risk of stroke. N Engl J Med. 2000 Aug 3;343(5):317-26.
28. Port S, Demer L, Jennrich R, Walter D, Garfinkel A. Systolic blood pressure and mortality. Lancet. 2000 Jan 15;355(9199):175-80.
29. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
30. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst. 1988 Oct 5;80(15):1198-202.
31. Ellis PM, Butow PN, Tattersall MH. Informing breast cancer patients about clinical trials: a randomized clinical trial of an educational booklet. Ann Oncol. 2002 Sep;13(9):1414-23.
32. Peters R, Beckett N, Forette F, Tuomilehto J, Clarke R, Ritchie C, et al. Incident dementia and blood pressure lowering in the Hypertension in the Very Elderly Trial cognitive function assessment (HYVET-COG): a double-blind, placebo controlled trial. Lancet Neurol. 2008 Aug;7(8):683-9.

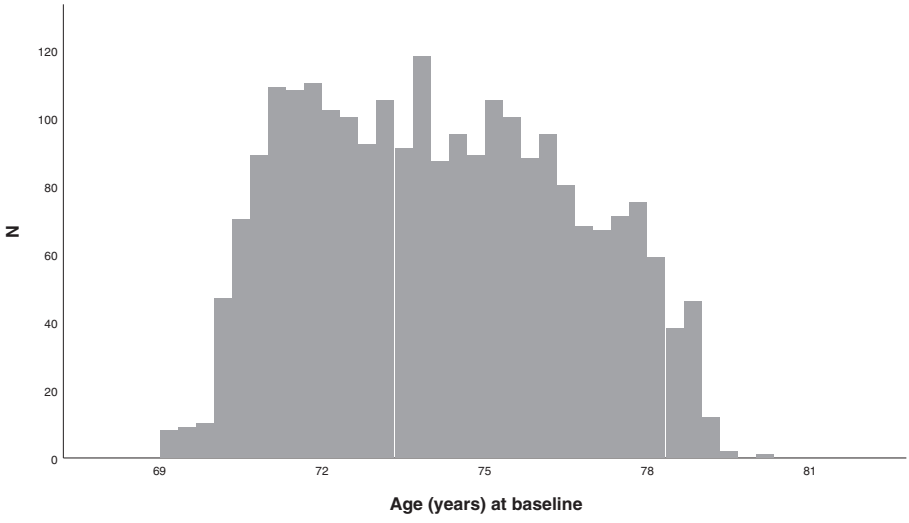
APPENDIX 1: R CODE.

```
# Activation of required libraries
library(rms)
library(mvtnorm)
library(arm)
library(rdd)

# Hypothetical dataset
n_patients          <- 100
R2          <- 0.0
corr        <- sqrt(R2)
treatment_effect  <- -10
n_sim     <- 10000
Est_Effect          <- matrix(nrow = n_sim, ncol=6)
Std_error           <- matrix(nrow = n_sim, ncol=6)
DF          <- matrix(nrow = n_sim, ncol=6)

# Variance of prognosis (sigma1), outcome (sigma2)
sigma1   <- 4
sigma2   <- 100
#Covariantie prognosis and outcome
sigma12 <- corr * sqrt(sigma1) * sqrt(sigma2)

# Mean of prognosis and outcome
mu          <- c(10, 90)
# Covariance Matrix
sigma     <- matrix(c(sigma1, sigma12, sigma12, sigma2), nrow = 2, byrow = TRUE)

for(i in 1:n_sim){

dataset   <- rmvnorm(n_patients, mean = mu, sigma = sigma)

dataset   <- as.data.frame(dataset)
names(dataset)    <- c("prognosis", "outcome")

## Randomized Controlled Trial, all patients randomized

# Treatment "Randomize all patients"
```

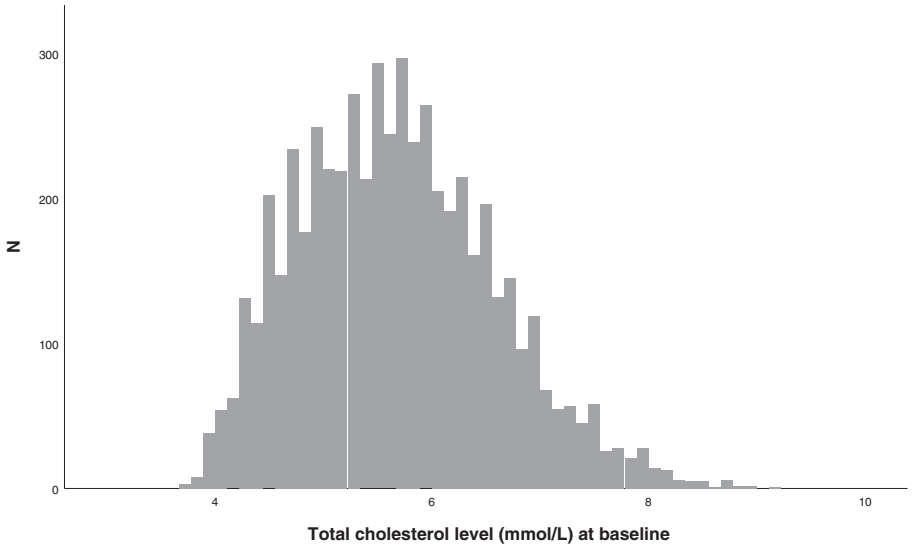
```
dataset$T_RCT    <- as.numeric(runif(n_patients) <= 0.5)

# Outcome "Randomize all patients"
dataset$O_RCT    <- dataset$outcome
dataset$O_RCT[dataset$T_RCT == 1] <- dataset$outcome[dataset$T_RCT == 1] + treat-
ment_effect

## Regression discontinuity, good prognosis control, poor prognosis treatment

#Treatment "Regression discontinuity design"
dataset$T_RDC    <- as.numeric(dataset$prognosis>10)

#Outcome "Regression discontinuity design"
dataset$O_RDC    <- dataset$outcome
dataset$O_RDC[dataset$T_RDC == 1] <- dataset$outcome[dataset$T_RDC == 1] +
treatment_effect

fit_RCT   <- lm(O_RCT ~ prognosis + T_RCT, data = dataset)
fit_RDC   <- lm(O_RDC ~ prognosis + T_RDC, data = dataset)

fit_RCT_rcs          <- ols(O_RCT ~ rcs(prognosis) + T_RCT, data = dataset, x=T, y=T)
fit_RDC_rcs          <- ols(O_RDC ~ rcs(prognosis) + T_RDC, data = dataset, x=T, y=T)

fit_RCT_llr          <- RDestimate(O_RCT ~ prognosis, cutpoint = 10, data = dataset)
fit_RDC_llr          <- RDestimate(O_RDC ~ prognosis, cutpoint = 10, data = dataset)

Est_Effect[i, ]      <-    c(fit_RCT$coefficients[3],    fit_RDC$coefficients[3],    fit_RCT_
rcs$coefficients[6], fit_RDC_rcs$coefficients[6], fit_RCT_llr$est[1], fit_RDC_llr$est[1])

}

#Mean Effect estimate of treatment
colMeans(Est_Effect)
#Mean squared error of effect estimate treatment
colMeans((Est_Effect - treatment_effect)^2)
```



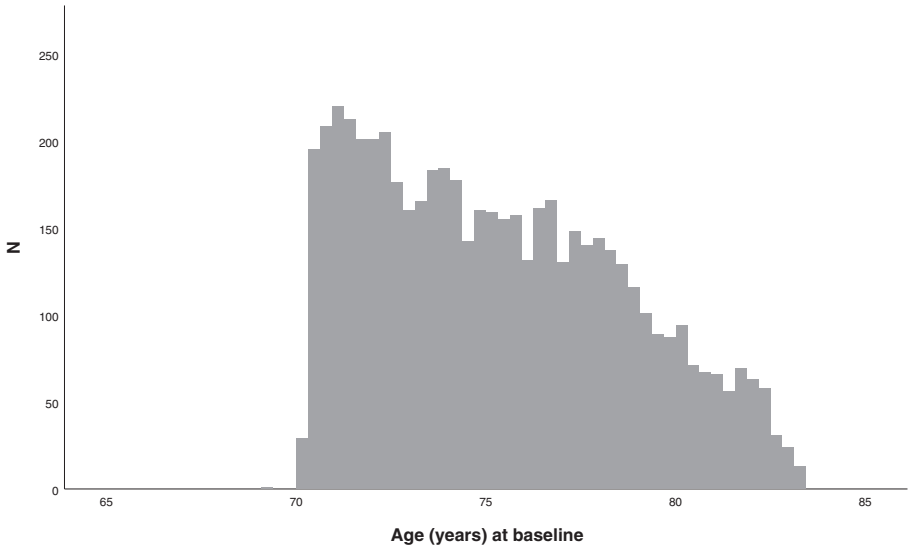
a) Histogram of mean systolic blood pressure (mmHg) at baseline in PreDIVA.



b) Histogram of age (years) at baseline in PreDIVA.



c) Histogram of total cholesterol level (mmol/L) at baseline in PROSPER.



d) Histogram of age (years) at baseline in PROSPER.

Appendix 2. Histogram of the assignment variable in both preDIVA and PROSPER in RCT data.

Chapter **6**

Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials

Nikki van Leeuwen
Hester F Lingsma
Simon P Mooijaart
Daan Nieboer
Stella Trompet
Ewout W Steyerberg

Journal of Clinical Epidemiology 2018

ABSTRACT

Introduction

Regression discontinuity (RD) is a quasi-experimental design that may provide valid estimates of treatment effects in case of continuous outcomes. We aimed to evaluate validity and precision in the RD design for dichotomous outcomes.

Methods

We performed validation studies in three large RCTs (CRASH, GUSTO and PROSPER). To mimic the RD design, we selected patients above and below a cut-off (e.g. age 75 years) randomized to treatment and control respectively. Adjusted logistic regression models using restricted cubic splines (RCS) and polynomials, and local logistic regression models estimated the odds ratio (OR) for treatment, with 95% confidence intervals to indicate precision.

Results

In CRASH, treatment increased mortality with OR 1.22 [95% CI 1.06; 1.40] in the RCT. The RD estimates were 1.42 [0.94; 2.16] and 1.13 [0.90; 1.40] with RCS adjustment and local regression respectively. In GUSTO, treatment reduced mortality (OR 0.83 [0.72; 0.95]), with more extreme estimates in the RD analysis (OR 0.57 [0.35; 0.92] and 0.67 [0.51; 0.86] respectively). In PROSPER, similar RCT and RD estimates were found, again with less precision in RD designs.

Conclusion

We conclude that the RD design provides similar but substantially less precise treatment effect estimates compared to an RCT, with local regression being the preferred method of analysis.

INTRODUCTION

Randomized clinical trials (RCTs) provide the most reliable evidence of effectiveness of medical interventions.(1) Nevertheless, recruitment of sufficient numbers of patients is a challenge in RCTs; it is estimated that less than 50% of the RCTs meet their recruitment targets.(2, 3) Patients' treatment preferences and clinicians equipoise are often cited as barriers to recruitment in RCTs.(2, 4-7) Patients participating in trials may poorly represent the population of interest.(8, 9) Especially, under-representation of older participants and women is well known in RCTs.(8, 10)

The quasi-experimental "regression discontinuity" (RD) design is an alternative epidemiological design to assess effectiveness of treatment. It has been suggested that RD is the observational design that most resembles an RCT.(11, 12) In the RD design, treatment is not assigned randomly, but is allocated to a subset of patients, based on a baseline assignment variable, often related to the outcome. The control group consists of a complementary subset of patients, not receiving treatment. E.g. all patients with an age over 75 years receive treatment and patients with an age below 75 years do not receive treatment and are considered as the control group. Such treatment assignment method may closely resemble clinical practice and may thus facilitate patient inclusion. In the analysis of the treatment effect, a regression model is used to compare treatment to the control group, while adjusting for the treatment assignment variable, in this example age.

The RD design is attractive because some of the challenges of the randomization process are avoided. However, the estimates from this quasi-experimental design may be substantially less efficient compared to an RCT.(13) The validity of RD estimates on continuous outcomes are well studied(13-15), but the validity of the RD design with binary outcomes is less known. Only a few examples have been described before(16, 17), while many health outcomes are dichotomous. Moreover, the efficiency of modeling approaches is unclear, i.e. the precision of estimated treatment effects. The aim of this study was to assess validity and precision of the RD design in studies with dichotomous outcome compared to an RCT. We hereto analyzed data from three large RCTs.

METHODS

Patients

Three trials were used to validate the RD design in empirical data. To assess the internal validity of the RD design we compared RD estimates with the estimates resulting from the RCT data. For the RD design we used a continuous baseline variable as assignment variable and the dichotomous endpoints of the RCTs.

The "Corticosteroid Randomisation After Significant Head injury" (CRASH) trial studied the effect of corticosteroids on death and disability after head injury.(18) CRASH enrolled 10,008 patients between 1999 and 2005. The primary outcome in CRASH was 14-day mortality. We included 9,554 patients with complete outcome data of whom 2,323 died before 14 days (24%). The median age was 33 years (IQR: 23 – 47 years).

Second, we analyzed 30,510 patients from "The Global Utilization of Streptokinase and Tissue plasminogen activator for Occluded coronary arteries" trial (GUSTO). Patients were entered between 1990 and 1993. 10,348 patients were assigned to treatment (accelerated tissue plasminogen activator, t-PA) and 20,162 patients were used as control patients receiving streptokinase.(19) The primary outcome was 30-day mortality. The median age was 61 (IQR: 52 – 69) and mortality occurred in 2,128 (7%). For both CRASH and GUSTO, age was used as the treatment allocation variable.

Third, we analyzed data from "PROspective Study of Pravastatin in elderly individuals at risk of vascular disease" (PROSPER).(20) This study enrolled 5,804 patients between December 1997 and May 1999, who were assigned to pravastatin (n = 2,891) or placebo (n = 2,913) to reduce the risk of coronary disease in elderly individuals. The outcome was a composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke at 3.2 years on average after randomization. 881 (15%) of the patients experienced the composite endpoint. The median total cholesterol level was 5.6 mmol/L (IQR: 5.0 – 6.3 mmol/L) at baseline (Table 1). For PROSPER, we considered baseline total cholesterol as the treatment allocation variable.

Table 1. Patient characteristics of CRASH (n = 9554), GUSTO (n = 30,510) and PROSPER (n = 5804).

Characteristic	CRASH	GUSTO	PROSPER
N in treatment arm (%)	4800 (50)	10348 (34)	2891 (50)
N in control arm (%)	4454 (50)	20162 (66)	2913 (50)
Median (IQR) of baseline variable for treatment assignment*	33 (23 - 47) *years*	61 (52 – 69) *years*	5.6 (5.0 – 6.3) *mmol/L*
N outcome (%)**	2323 (24)	2128 (7)	881 (15)

** Baseline measurement is age in years in CRASH and GUSTO and total cholesterol in mmol/L in PROSPER.*
*** Outcome is 14-day all-cause mortality in CRASH, 30-day all-cause mortality in GUSTO and a composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke at 3.2 years on average in PROSPER.*

Statistical analysis

To analyze the data as an RD design, we selected those patients with a baseline value above the median of the assignment variable, who were assigned to treatment in the original RCT as the intervention group, and those with a baseline value below the median and not assigned to treatment in the RCT as control group. Histograms of the baseline assignment variables for each study were plotted, as well as binned scatterplots for

outcome means for treated and controls at each baseline assignment value. The analysis was based on the intention-to-treat principle. This led to inclusion of approximately half of the RCT patients. The treatment effect was expressed as odds ratios (OR) with 95% confidence intervals (95% CI), with adjustment for the baseline variable in a logistic regression model. To compare the RD estimates to the RCT estimates in comparable sample sizes, random samples of 50% from the complete RCT data were drawn (5000 times). To compare the designs in terms of efficiency we calculated the ratio of variances between both designs based on estimated standard errors (SEs) of the estimated treatment effects: $(SE_{design\ 2} / SE_{design\ 1})^2$.

Previous work has shown that the validity of the RD design is highly dependent on the quality of the adjustment in the analysis phase, and on assumptions of a local or global effect of the treatment.(13) All analyses (RCT and RD) were adjusted for the baseline variable that was used to attribute treatment; age in both CRASH and GUSTO and baseline cholesterol in the PROSPER trial. We assessed non-linearity of the effect of the baseline variable with non-linear restricted cubic splines (RCS) functions. An RCS function is a smooth function that consists of pieced-together cubic splines that are restricted to be linear in the tails. We used three knots for adequate flexibility.(21) Consequently we used the RCS of the baseline variable in the adjustment model for optimal adjustment. To consider a different approach to estimate RD estimates, we also used polynomials of the baseline variables in the adjustment model. R^2 statistics were calculated to indicate the explained variance of the adjustment model.

The approach described above assumes a global treatment effect. It has been argued that this assumption is hard to make and can never be proven.(11) We therefore also analyzed the RD design with local logistic regression models. In local logistic regression, only patients around the cut-off were used in the analysis to estimate the treatment effect. For the local estimations, the *gam* package in R was used, in which a default span of 0.5 is set. Gaussian kernel was used for the local logistic regression analysis. Using this kernel, the observations outside the span have lower influence on the estimation, but all the data are used in the analysis. To assess differential treatment effects, we studied interaction between the baseline variable and the treatment in the RCT data. For all three trials we assessed treatment effect heterogeneity in the complete RCT data, using interaction terms between treatment and the assignment variable. Moreover, to study the stability of the estimates for all three validation studies, we added RD analyses on an additional cut-off.

All statistical analyses were performed in R statistical software version 2.15.3 (R Foundation for Statistical Computation, Vienna, Austria) using the *rms* and *gam* package.

RESULTS

In CRASH the treatment was harmful. The adjusted OR was 1.22 [95% CI: 1.06; 1.40] for the effect of treatment on mortality in the 50% subset of the RCT. For the hypothetical RD design, the estimated OR was 1.42 [0.94; 2.16], with RCS adjustment for age. When analyzed with polynomial adjustment the OR for treatment was 1.09 [0.81; 1.46]. The alternative method to analyze this hypothetical RD design, local logistic regression, resulted in an estimated OR of 1.13 [0.90; 1.40] (Table 2).

Table 2. RCT and RD analyses in the CRASH trial (n = 9554).

Analysis	N total	R2 (%)	OR (95% CI) for 14-day mortality	Standard error (SE) of treatment effect estimate
RCT				
Linear adjustment*	4777	7	1.22 (1.06; 1.40)	0.071
RD – assignment: age ≤ 33 Tx-, age > 33 Tx+				
RCS adjustment*	4844	10	1.42 (0.94; 2.16)	0.212
Polynomial adjustment*	4844	10	1.09 (0.81; 1.46)	0.151
Local logistic regression	4844	NA	1.13 (0.90; 1.40)	0.112
RD – assignment: age ≤ 40 Tx-, age > 40 Tx+				
RCS adjustment*	4806	10	1.04 (0.68; 1.60)	0.218
Polynomial adjustment*	4806	10	0.94 (0.72; 1.23)	0.138
Local logistic regression	4806	NA	1.02 (0.80; 1.32)	0.129

**Linear, RCS or polynomial adjustment means that baseline age was used as a linear, RCS or polynomial term in the regression analysis to control for age.*

In GUSTO the estimated OR for mortality was 0.83 [0.72; 0.95] in a subset of 50% of the patients. The estimated OR, in the RD scenario was 0.57 [0.35; 0.92] adjusted with RCS for age. The OR for treatment from RD estimated with polynomial adjustment for age was 0.82 [0.63; 1.07]. The analysis with local logistic regression resulted in an estimated OR of 0.67 [0.51; 0.86] (Table 3).

In the PROSPER trial, the adjusted OR for the composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke was 0.85 [95% CI; 0.69; 1.04] when assessed in the subset of 50% of the RCT. The estimated OR was 0.80 [0.46; 1.38] in the hypothetical RD design adjusted for baseline cholesterol with RCS. The OR for treatment from RD estimated with polynomial adjustment for age was 0.81 [0.56; 1.16]. The RD design analyzed with local logistic regression showed an OR for treatment of 0.79 [0.56; 1.13] (Table 4).

In none of the RCTs we found statistically significant interaction between treatment and the assignment variable. However, this interaction test has limited statistical power. In all three trials there appeared to be a differential treatment effect over the range of

the assignment variable, (Figure 1d, 2d and 3d). This is confirmed in the additional RD analysis with treatment assignment based on a different cut-off (Table 2, 3 and 4). In these validation studies we see slightly different RD estimates between the two different assignment approaches in all three studies.

In terms of efficiency, the RD with adjustment was 7.2 to 12.1 times less efficient than the adjusted RCT, compared to 3.1 to 4.5 less efficient estimates from RD with polynomial adjustment. The RD design analyzed with local logistic regression was 2.5 to 3.5 times less efficient than the adjusted RCT (Table 5).

Table 3. RCT and RD analyses in the GUSTO trial (n = 30,510).

Analysis	N total	R2 (%)	OR (95% CI) for 30-day mortality	Standard error (SE) of treatment effect estimate
RCT				
Linear adjustment*	15255	12	0.83 (0.72; 0.95)	0.071
RD – assignment: age ≤ 62 Tx-, age > 62 Tx+				
RCS adjustment*	15423	11	0.57 (0.35; 0.92)	0.246
Polynomial adjustment*	15423	11	0.82 (0.63; 1.07)	0.133
Local logistic regression	15423	NA	0.67 (0.51; 0.86)	0.132
RD – assignment: age ≤ 70 Tx-, age > 70 Tx+				
RCS adjustment*	17846	10	0.94 (0.72; 1.22)	0.133
Polynomial adjustment*	17846	10	0.95 (0.75; 1.21)	0.121
Local logistic regression	17846	NA	0.90 (0.74; 1.10)	0.102

**Linear, RCS or polynomial adjustment means that baseline age was used as a linear, RCS or polynomial term in the regression analysis to control for age.*

Table 4. RCT and RD analyses in the PROSPER trial (n = 5804).

Analysis	N total	R2 (%)	OR (95% CI) for composite endpoint	Standard error (SE) of treatment effect estimate
RCT				
Linear adjustment	2902	0.4	0.85 (0.69; 1.04)	0.104
RD – assignment: cholesterol ≤ 5.6 Tx-, cholesterol > 5.6 Tx+				
RCS adjustment	2919	0.7	0.80 (0.46; 1.38)	0.279
Polynomial adjustment	2919	0.7	0.81 (0.56; 1.16)	0.185
Local logistic regression	2919	NA	0.79 (0.56; 1.13)	0.181
RD – assignment: cholesterol ≤ 6.2 Tx-, cholesterol > 6.2 Tx+				
RCS adjustment	2969	0.7	1.30 (0.71; 2.40)	0.311
Polynomial adjustment	2969	0.6	1.03 (0.69; 1.53)	0.205
Local logistic regression	2969	NA	1.07 (0.75; 1.56)	0.187

**Linear, RCS or polynomial adjustment means that baseline cholesterol level was used as a linear, RCS or polynomial term in the regression analysis to control for cholesterol level.*

Table 5. Relative efficiency in terms of required sample size for different designs in CRASH, GUSTO and PROSPER*.

	CRASH	GUSTO	PROSPER
RCT adjusted vs RD RCS adjustment	8.9	12.1	7.2
RCT adjusted vs RD polynomial adjustment	4.5	3.5	3.1
RCT adjusted vs RD local logistic regression	2.5	3.5	3.0

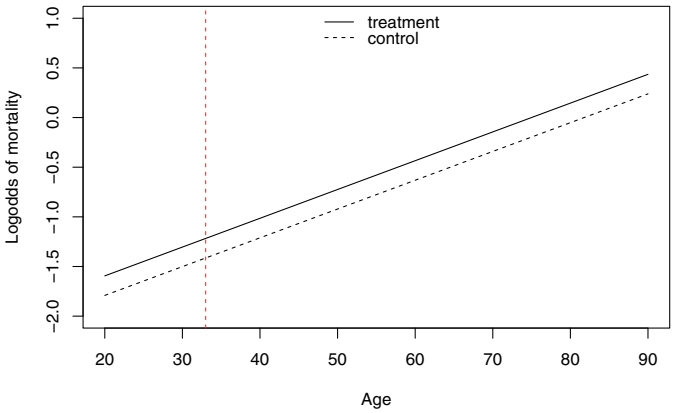
**Formula: $(SE_{design\ 2} / SE_{design\ 1})^2$*

DISCUSSION

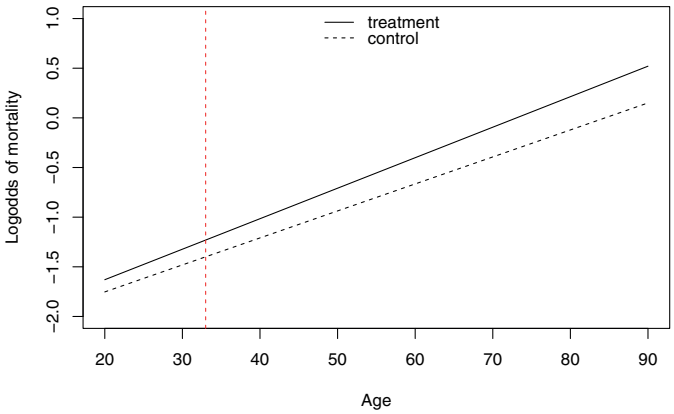
This validation study, with data from three large RCTs, showed that the treatment effect estimates from the hypothetical RD were similar to the treatment effect estimates from the RCTs, either with RCS and polynomial adjustment or local logistic regression. In all three studies the confidence interval of all RD estimates overlapped with the point estimate of the RCT. However, RD estimates were substantially less precise.

Causality in regression discontinuity design

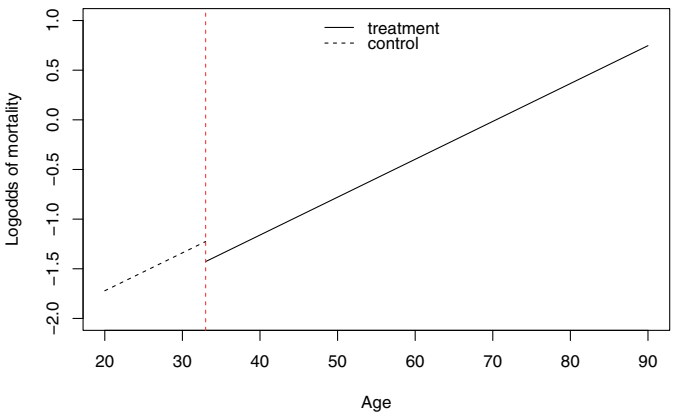
The advantage of a quasi-experimental, prospective, RD design over an observational study is the controlled assignment of treatment. This property is shared with an RCT. As Labrecque et al. stated, in both an RCT as in an RD design, we have good knowledge of the assignment mechanism.(11) In RCTs, treatment is randomly allocated and in RD treatment is assigned to patients using a baseline assignment variable. The treated and untreated patients in an RCT are unconditionally exchangeable. Therefore, RCTs are accepted to make causal inference. In an RD design the treated and the control patients are not exchangeable across the whole baseline range since they have a systematically different baseline characteristic. In RD the treated and untreated are only exchangeable close to the cut-off of the baseline assignment variable.(11, 12) Therefore, causal inference can only be made around the cut-off in an RD design, where patients can be considered to be exchangeable. The causal treatment effect estimated in RD is a local treatment effect estimate. This means that comparing estimates from RCT and RD may not be completely straightforward, even with comparable RCT and RD data.(11) Therefor it may not be entirely fair to interpret the concordance between local RD estimates and global RCT estimates as a measure of validity of RD estimates. The overall RCT estimate is the average treatment effect in the whole RCT population, although we can condition on the assignment variable for more efficient analysis.(22-25) An RD estimate is a local treatment effect among patients at the cut-off and may vary dependent on the cut-off chosen.(13) At the end of the day, it is the RCT estimate that is the average of local estimates across the distribution of the assignment variable.



a Linear function of the baseline variable over the outcome variable in RCT data. The space between both lines indicates the main treatment effect in the RCT.

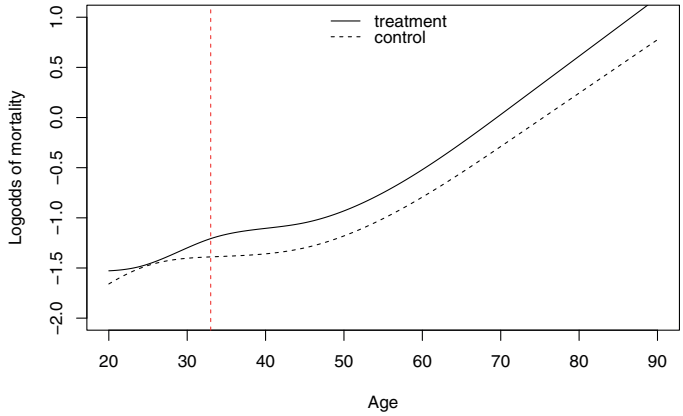


b Linear interaction function of the treatment effect over the baseline variable in RCT data.

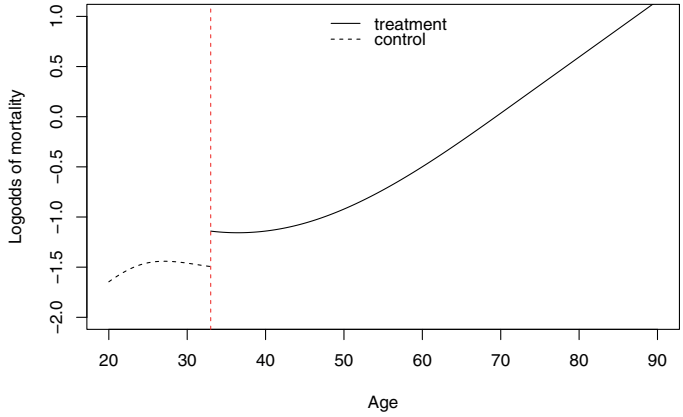


c Linear function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

Figure 1. CRASH



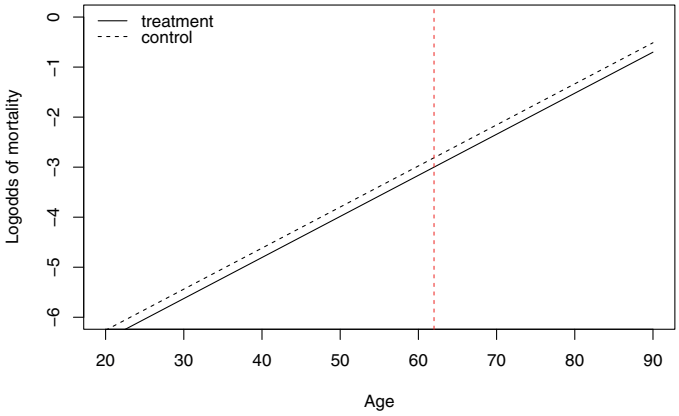
d RCS interaction function of the treatment effect over the baseline variable in RCT data.



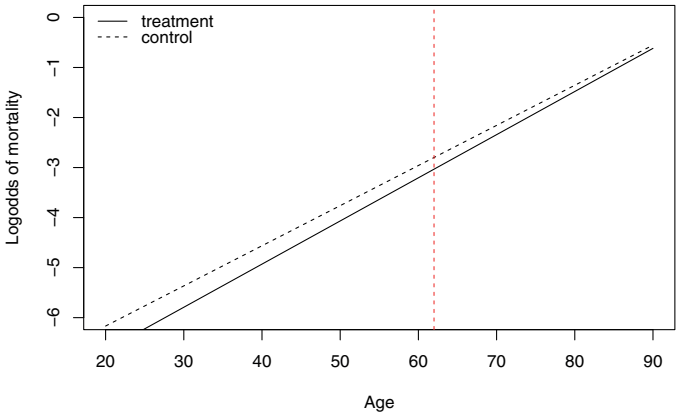
e RCS function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.
Figure 1. CRASH

Global vs. local treatment effects

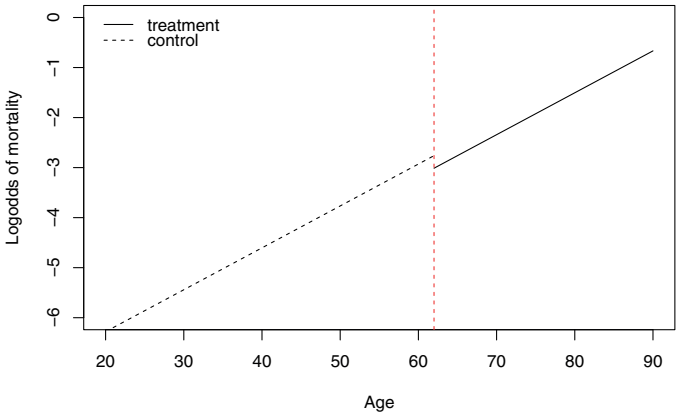
Only when treatment does not interact with the baseline assignment variable the estimate from an RD design can be interpreted as a global treatment effect estimate.[11] In order to estimate a global treatment effect estimate in RD, one would have to feel confident modeling the relationship between the assignment variable and the outcome even where it is not observed in the data.(11, 26, 27) In other words, the model for the assignment variable-outcome relationship in both the treated and untreated groups would have to be extrapolated to the side of the cutoff where they were not observed. (11)



a Linear function of the baseline variable over the outcome variable in RCT data. The space between both lines indicates the main treatment effect in the RCT.

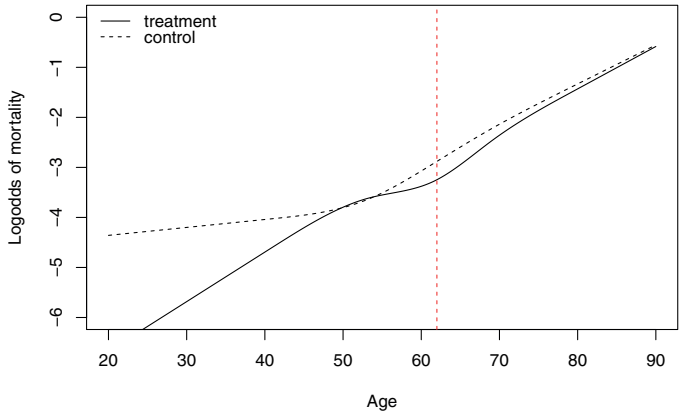


b Linear interaction function of the treatment effect over the baseline variable in RCT data.

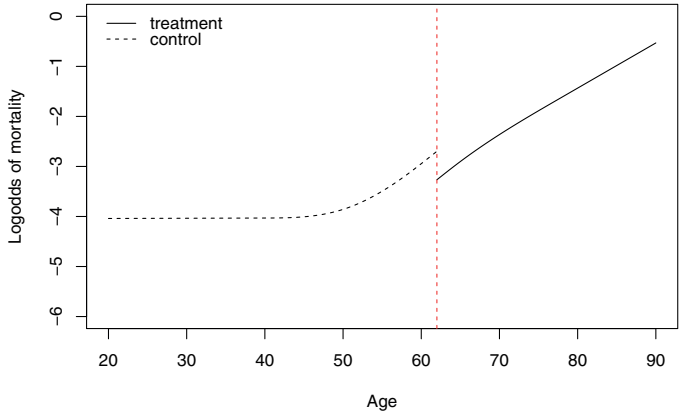


c Linear function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

Figure 2. GUSTO

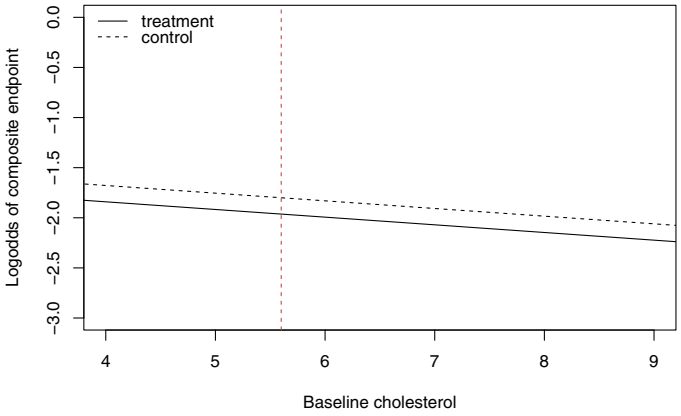


d RCS interaction function of the treatment effect over the baseline variable in RCT data.

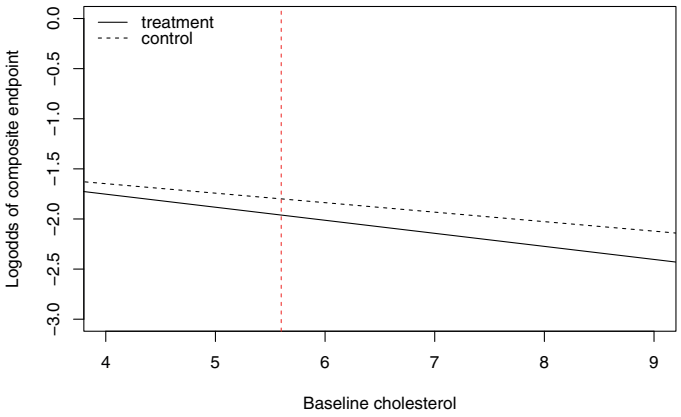


e RCS function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.
Figure 2. GUSTO

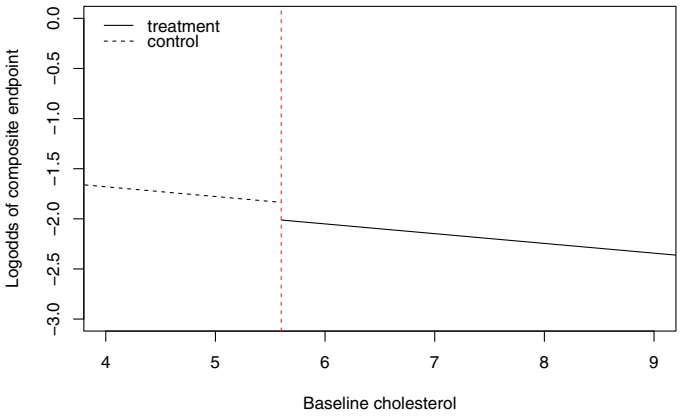
When using RCS or polynomial adjustment, the treatment effect in CRASH was slightly different compared to the RCT. Graphical inspection showed qualitative interaction between treatment and the adjustment variable age (Figure 1d). At the cut-off (age 33 years) the treatment effect – the difference between the plotted line for the control patients and the plotted line for the treated patients – was larger than the global RCT effect which is shown in Figure 1a. This explains the difference between the RD estimate and the RCT. The presence of a heterogeneous treatment effect over the range of age was confirmed in the RD analysis with treatment based on a different cut-off, resulting in less similar treatment estimates compared to the RCT estimates. Qualitative interaction was also observed in GUSTO (Figure 2d), and could have led to more extreme RD estimates (0.57 and 0.67) compared to the OR estimated in the RCT (0.83). At the cut-off of 62 years



a Linear function of the baseline variable over the outcome variable in RCT data. The space between both lines indicates the main treatment effect in the RCT.

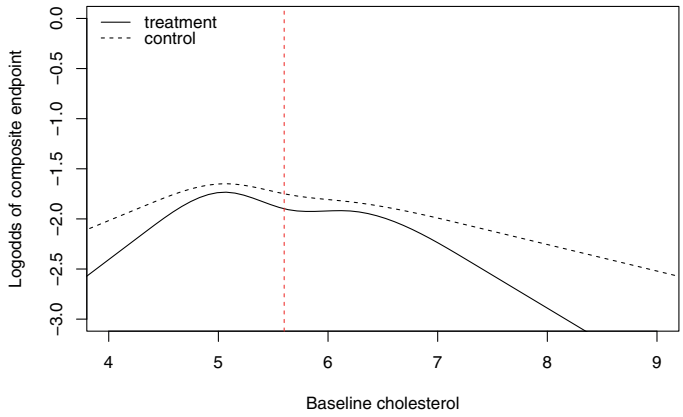


b Linear interaction function of the treatment effect over the baseline variable in RCT data.

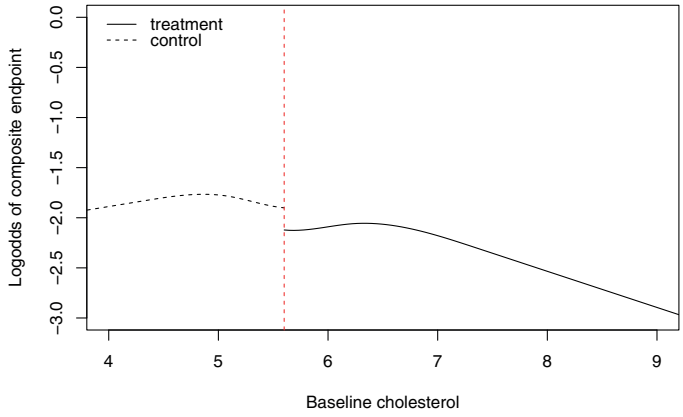


c Linear function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

Figure 3. PROSPER



d RCS interaction function of the treatment effect over the baseline variable in RCT data.



e RCS function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.
Figure 3. PROSPER

in Figure 2d a larger treatment effect is shown compared to the global treatment effect in Figure 2a. However, in RD with polynomial adjustment for age, the treatment effect is similar (0.82) to the RCT estimate. A smaller treatment effect was estimated when the cut-off for treatment assignment was set at 70 years. This is also confirmed in Figure 2d; after the age of 62 the treatment effect decreases.

In PROSPER, also qualitative interaction was found and shown in Figure 3d. However, RD with treatment assignment set at cholesterol 5.6 mmol/L, RD estimates (0.80, 0.81 and 0.79) and the RCT estimate (0.85) were quite similar. When the treatment assignment rule was set at cholesterol 6.2 mmol/L for RD, the RD estimates were slightly different from the RCT estimate. These results confirm that the RD estimate is not equal to the

global RCT treatment effect estimate when the treatment effect is heterogeneous across the baseline assignment variable.

In a prospective RD design, it is hard to know whether the baseline assignment variable interacts with treatment. It can be formally tested but since the treatment groups each have data on only one side of the cut-off, the result only represents possible interaction at a small range of the assignment variable, around the cut-off. Thus, the assumptions required to estimate the global treatment effect are not verifiable within the RD design. This is why it has been suggested that global treatment effect estimates from RD designs should only be presented secondary to local average treatment effect estimates and never as the primary parameter of interest.(11, 12)

In this study we also assessed and compared RCS and polynomials for adjustment in RD. The advantage of an RCS function over polynomial adjustment is the restriction of the function to be linear in the tails. This is important when using this for optimal adjustment in for example RCTs, to estimate global effects over the whole range of the population studied. However, in RD we are primarily interested in local estimates and thus optimal adjustment around the cut-off for treatment assignment. So the advantage of RCS spline functions over polynomial adjustment in for example RCTs, may be less applicable to optimal adjustment in RD.

Our results suggest 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 and polynomial adjusted analyses and local logistic regression are more similar to each other than when there is interaction. If there was some interaction between the assignment variable and treatment, the results from local logistic regression and the RCS and polynomial adjusted analyses were less similar. So, the comparison of both RD estimates could be a way to have more information on the assignment variable – treatment relationship.

Efficiency of RD design

The RD estimates with adjustment appeared to be substantially less efficient than the RCT estimates. An RD design analyzed with adjusted logistic regression using RCS adjustment implies that 7.2 to 12.1 times more patients need to be included in the study compared to an RCT design. RD with polynomial adjustment would need 3.1 to 4.5 more patients compared to an RCT. If one would analyze the RD design with local logistic regression, this study would need about 2.4 to 3.6 times more patients than an RCT. So, the local regression approach was more efficient compared to the adjusted logistic regression. Also in terms of efficiency, local logistic regression would be preferred to analyze an RD design.

In absolute numbers an RD design needs more patients to obtain similar efficiency, compared to an RCT to estimate global treatment effect estimates. Although RD is de-

scribed as less efficient than RCT in identifying the global average causal effect, it may be nearly as good in identifying the local causal effect, which may be of interest depending on the context. From a power perspective, it would be a fair comparison if the RCTs were powered to estimate treatment effects in the assignment variable subgroups around the discontinuity and compare these with the local RD treatment effect estimates. However, in our study we focus on the comparison between global RCT estimates and estimates from an RD design, and the efficiency of an RD design to estimate global treatment effect estimates.

Also, an RD design could facilitate patient recruitment, especially when the cut-off for treatment assignment closely resembles clinical practice. In these specific cases an RD design may be cheaper and less-time intensive than an RCT. Besides, RD designs could be conducted in different settings than RCTs; one can assume that RD design have less stringent inclusion criteria. This would be especially the case in a retrospective RD design when data from (clinical) registries are used. Therefore, some argue that data used in RD designs could lead to more external validity.(28, 29)

Linear versus logistic models in RD

In this study we specifically assess the performance of RD vs RCT in the context of dichotomous outcomes and logistic regression, which is not the standard in RD designs, but is common in health research. RD is underused with logistic regression models; only a few examples are described before.(16, 17) RD can be easily extended to generalized linear models like logistic regression.(30) When using dichotomous endpoints in RD it is straightforward to obtain more interpretable parameters like risk differences and risk ratios even in the logistic regression context, because the predicted probabilities at the threshold can be obtained directly from the model. The only barrier using logistic models in RD would be the absence of a data driven optimal bandwidth selector for the logistic model, like Imbens-Kalyanamaran(31) optimal bandwidth calculation is available for local linear regression models. For the local estimations in this study the *gam* package in R was used, in which a default span a 0.5 proportion of the observations over the assignment range is included. This can be adjusted specifying "span" in the gam function, for example span=0.2. When one is interested in a local treatment effect estimate, extending the span will in theory decrease validity but also increase reliability.

Strengths and limitations

We used RCT data to evaluate a hypothetical RD design, in which we artificially set the cut-off to "assign" treatment. This resulted in perfect adherence to the defined cut-off. This is unlikely to be the case in real life where which patients are prospectively assigned to treatment. A strength of this study is the use of data from three large RCTs to be able to compare the RD results with the RCT estimates. Moreover, because of the RCT

data we were able to carefully assess interaction between the assignment variable and treatment.

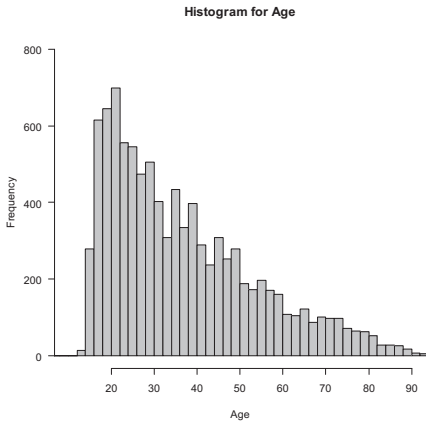
Conclusion and recommendations

Our findings for dichotomous outcomes are in line with previous work on RD for continuous outcomes.(13) The RD design may provide similar treatment effect estimates compared to RCT estimates for dichotomous outcome measures, but has some strong disadvantages that should be carefully considered when choosing an RD design to assess the effectiveness of a medical intervention. First, to be able to estimate the same global treatment effect in an RD design as in an RCT, the assumption of a global treatment effect over the range of the assignment variable is required. In prospectively collected RD data this assumption of a global treatment effect cannot be proven. Global treatment effect estimates from RD designs should therefore only be reported secondary to local treatment effect estimates. Second, the RD design is substantially less efficient than an RCT, requiring sample sizes at least three times higher than for the conventional RCT to obtain the same precision for the treatment effect estimate. In this study we found local logistic regression would be most efficient to analyze an RD design. Future research on the RD design should focus on more efficient application of the RD design, considering different approaches to estimate treatment effects from an RD design and examining their properties.

REFERENCES

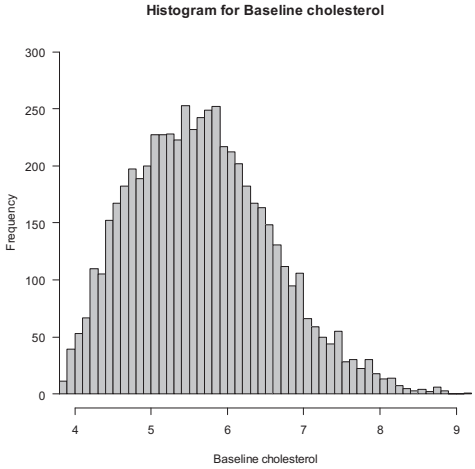
1. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006;7:9.
2. Mills N, Blazeby JM, Hamdy FC, Neal DE, Campbell B, Wilson C, et al. Training recruiters to randomized trials to facilitate recruitment and informed consent by exploring patients' treatment preferences. Trials. 2014;15:323.
3. Treweek S, Mitchell E, Pitkethly M, Cook J, Kjeldstrom M, Taskila T, et al. Strategies to improve recruitment to randomised controlled trials. Cochrane Database Syst Rev. 2010(1):MR000013.
4. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. J Clin Epidemiol. 1999 Dec;52(12):1143-56.
5. Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. Lancet Oncol. 2006 Feb;7(2):141-8.
6. King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. JAMA. 2005 Mar 2;293(9):1089-99.
7. Fayter D, McDaid C, Eastwood A. A systematic review highlights threats to validity in studies of barriers to cancer trial participation. J Clin Epidemiol. 2007 Oct;60(10):990-1001.
8. Martinson BC, Crain AL, Sherwood NE, Hayes MG, Pronk NP, O'Connor PJ. Population reach and recruitment bias in a maintenance RCT in physically active older adults. J Phys Act Health. 2010 Jan;7(1):127-35.
9. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: who are these patients and how did they get there? Ann Intern Med. 2002 Jul 2;137(1):10-6.
10. Beckie TM, Mendonca MA, Fletcher GF, Schocken DD, Evans ME, Banks SM. Examining the challenges of recruiting women into a cardiac rehabilitation clinical trial. J Cardiopulm Rehabil Prev. 2009 Jan-Feb;29(1):13-21; quiz 2-3.
11. Labrecque JA, Kaufman JS. Commentary: Can a quasi-experimental design be a better idea than an experimental one? Epidemiology. 2016 Jul;27(4):500-2.
12. Lee DS, Lemieux T. Regression discontinuity designs in economics. J Econ Lit. 2010;48:281–355.
13. Van Leeuwen N, Lingsma HF, De Craen AJ, Nieboer D, Mooijaart SP, Richard E, Steyerberg EW. Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes. Epidemiology. 2016 Jul;27(4):503-11.
14. Maas IL, Nolte S, Walter OB, Berger T, Hautzinger M, Hohagen F, Lutz W, Meyer B, Schröder J, Späth C, Klein JP, Moritz S, Rose M. The regression discontinuity design showed to be a valid alternative to a randomized controlled trial for estimating treatment effects. J Clin Epidemiol. 2017;82:94-102.
15. Hansen H, Klejnstrup NR, Andersen OW. A Comparison of Model-Based and Design-Based Impact Evaluations of Interventions in Developing Countries. Am J Eval. 2013;34(3):320-338.
16. Berk RA, De Leeuw J. An Evaluation of California's Inmate Classification System Using a Generalized Regression Discontinuity Design. JASA. 1999;94(448):1045-1052.
17. Imbens GW, Lemieux T. Regression discontinuity designs: A guide to practice. J Econom. 2008;142(2):615-635.

18. Edwards P, Farrell B, Lomas G, Mashru R, Ritchie N, Roberts I, et al. The MRC CRASH Trial: study design, baseline data, and outcome in 1000 randomised patients in the pilot phase. Emerg Med J. 2002 Nov;19(6):510-4.
19. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO investigators. N Engl J Med. 1993 Sep 2;329(10):673-82.
20. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet. 2002 Nov 23;360(9346):1623-30.
21. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst. 1988 Oct 5;80(15).
22. Hernandez AV, Steyerberg EW, Butcher I., Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema JD, Maas AI. 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.
23. 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.
24. Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? Am Heart J. 2000;139(5):745-751.
25. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. Int Stat Rev.1991;58(2):227-240.
26. Bor J, Moscoe E, Bärnighausen T. Three approaches to causal inference in regression discontinuity designs. Epidemiology. 2015;26:e28–30; discussion e30.
27. Rubin DB. Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational and Behavioral statistics. 1977;6:377–401.
28. Moscoe E, Bor J, Bärnighausen 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.
29. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. BMJ. 2016;352:i1216.
30. Bor J, Moscoe E, Mutevedzi P, Newell ML, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25(5):729-37.
31. Imbens G, Kalyanaraman K, Optimal Bandwidth Choice for the Regression Discontinuity Estimator. https://scholar.harvard.edu/files/imbens/files/rd_09feb3.pdf.



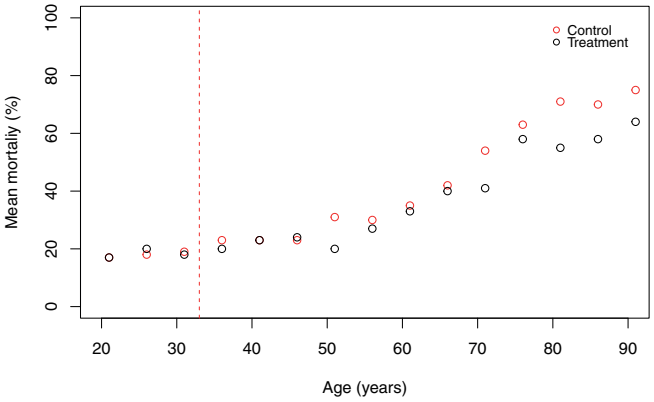
a. Distribution of baseline age in years in CRASH.

b. Distribution of baseline age in years in GUSTO.

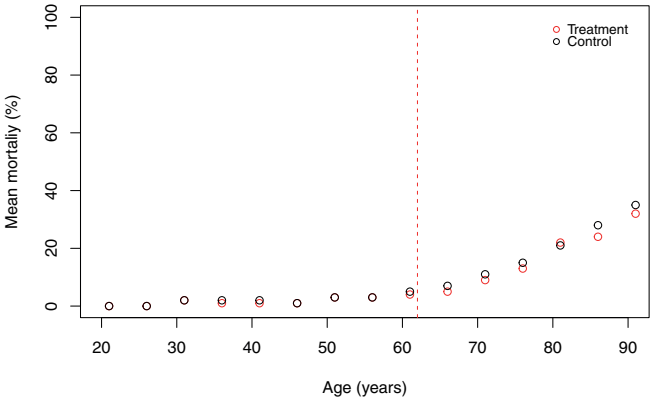


c. Distribution baseline cholesterol level in mmol/L in PROSPER.

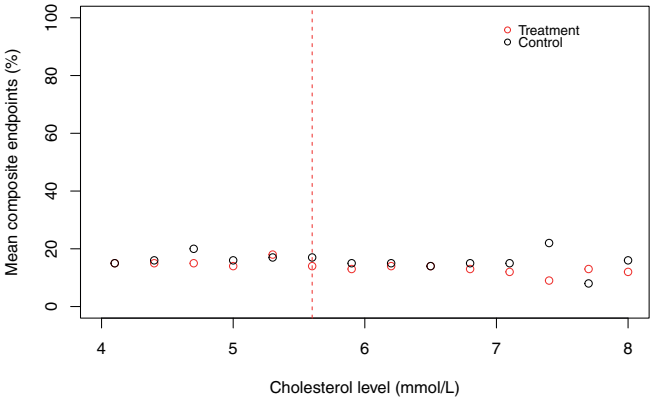
Appendix 1.



a. Binned scatterplot for mortality average, over the baseline age range in CRASH.



b. Binned scatterplot for mortality average, over the baseline age range in the GUSTO.



c. Binned scatterplot for composite endpoint average, over the baseline cholesterol level in mmol/L range in PROSPER.

Appendix 2.

Chapter **7**

Efficient treatment assignment in a regression discontinuity design? Simulations and validation in a large randomized controlled trial

Nikki van Leeuwen
Daan Nieboer
Hester F Lingsma
David van Klaveren
Ewout W Steyerberg

Submitted

ABSTRACT

Introduction

The quasi-experimental regression discontinuity (RD) design may provide valid treatment effect estimates but is inefficient compared to a randomized controlled trial (RCT). We aimed to compare different assignment approaches to increase the statistical efficiency of the RD design.

Methods

In Monte Carlo simulations, a random (R^2=0), low (R^2=7%) and highly (R^2=31%) correlating variable with outcome was used for treatment assignment. Patients were sampled from the CRASH trial, with a dichotomous outcome simulated. The treatment effect was analyzed with both local logistic regression and logistic regression with spline adjustment. To assess the relative statistical efficiency, standard errors (SE) of the different treatment assignment strategies were compared with an RCT of the same total sample size. This procedure was repeated in CRASH (n=9,554) as a case study.

Results

In the simulations, treatment effect estimates were unbiased. To obtain the same efficiency as an unadjusted RCT, RD required 2.8 times as many patients when using an assignment variable not correlating with outcome, and approximately 3.3 times as many patients when using an assignment variable highly correlating with outcome, using local regression. Compared to an adjusted RCT, the relative efficiency was not dependent on the correlation between the assignment variable and outcome since the adjustment affects the efficiency of an RCT as well. In the case study similar results were found.

Conclusion

The relative efficiency of the RD design is not dependent on the correlation between the assignment variable and outcome. We recommend researchers to use assigment variables that are feasible in clinical practice.

INTRODUCTION

The regression discontinuity (RD) design is a quasi-experimental design to study effectiveness of treatment. In the RD design, treatment is assigned to a subset of patients based on a baseline variable; e.g. older patients receive treatment. The RD design has been described as the next best design after a randomized controlled trial (RCT)(1): it enables causal inference of the treatment effect without randomizing patients to a treatment- or control group. The crucial feature of the RD design compared to observational designs is the exchangeability of patients around the cut-off of the assignment variable, making causal inference between treatment and outcome possible.(2-4) In some cases, the RD design might be attractive because randomization is avoided and the RD strategy closely resembles clinical practice.

A substantial drawback is that it requires far larger numbers of patients compared to an RCT.(5-7) Goldberger proved that this reduced precision stems from the correlation between the assignment variable and (binary) treatment indicator. This is because the treatment indicator is itself a function of the assignment variable and both must be included to model the outcome in RD.(7, 8) Different assignment variables can be used for treatment allocation in an RD design. Bor et al. suggested that "the assignment variable could be any continuous pretreatment measure including the outcome measure at baseline, or another measure of risk; a baseline covariate that is loosely correlated with the outcome; or even a random number, in which case regression discontinuity is identical to an RCT".(9) Since it is known that an RCT is more efficient than RD, we hypothesize that RD based on a poorly correlating assignment variable with outcome - thus more similar to treatment allocation in an RCT – results in more efficient treatment effect estimates compared to treatment effect estimates from an RD with treatment assignment based on a variable highly correlating with outcome.

In this study, we aim to compare different assignment approaches to increase the statistical efficiency of the RD design. We hereto performed a simulation study and analyzed data from a large RCT.

METHODS

Simulation study set up

According to the key steps and decisions in simulation studies described by Morris et al.(10), we performed Monte Carlo simulations to compare the efficiency of different assignment strategies in RD. For 5,000 patients, baseline prognostic characteristics were drawn from the "Corticosteroid Randomisation After Significant Head injury" (CRASH) trial.(11) A dichotomous outcome measure was simulated for each patient, with odds

ratios for treatment of 0.8 and 1.0, corresponding to a small true effect and no effect respectively. We evaluated three approaches to assign treatment in RD. A random (R^2 = 0%), low (R^2 = 7%) - and a higher (R^2 = 31%) correlating variable with outcome was used to assign treatment. For all three strategies, treatment was assigned to the 50% of patients above the median value of the assignment variable. RD was analyzed with local logistic regression analysis to estimate a local treatment effect for the area around the cut-off for treatment assignment. Also, logistic regression models were used to estimate the treatment effect, adjusted for the assignment variable in a restricted cubic spline (RCS) term. Simulations were repeated 10,000 times. The standard errors (SEs) of the effect estimates from the RCT and the different RD approaches were compared as a measure of efficiency. The ratio of variances between an RCT and different RD assignment approaches were calculated with the following formula: $(SE_{RD} / SE_{RCT})^2$. The simulation code is provided in the Appendix.

Case study

The CRASH trial(11) was also used to illustrate the potential effect of different assignment approaches on the efficiency of the RD design in empirical data. The CRASH trial assessed the effect of corticosteroids on death and disability after head injury. CRASH enrolled patients between 1999 and 2005, of which 9,554 patients had complete outcome data. Of 10,008 patients included, 5,007 patients were allocated to treatment and 5,001 patients were control patients. To resemble the RD design, we used patients' baseline measures as assignment variable and the primary dichotomous endpoint (14-day all-cause mortality) of CRASH as outcome measure.

Efficient RD assignment approach

Nagelkerke R^2 statistics for all baseline characteristics and the full prediction model with outcome were calculated. The R^2 statistic between treatment allocation - which was completely at random in CRASH - and outcome, in the absence of a treatment effect, would be 0. Next, we assessed three hypothetical treatment assignment variables. First, RD based on a hypothetical completely random assignment variable was performed. In the second assignment strategy, a poorly correlating variable with outcome, age, was used to assign treatment. Finally, in the last setting, the linear predictor of a full prediction model highly correlating with outcome was used to assign treatment. This hypothetical assignment variable was constructed with a logistic regression model containing the most important known predictors for 14-day mortality, namely age, pupillary reactivity and motor score.(12-14) The medians of the assignment variables were used as the cut-off for treatment assignment. To imitate an RD design within the RCT data, we selected treated patients with a value of the assignment variable above the cut-off, and control patients with a value of the assignment variable below the cut-off.

Table 1. Monte Carlo simulations (n=5000, simulations: 10,000 times) comparing RCT with different treatment assignment strategies in regression discontinuity design.

Design	Analysis	Adjustment	Covariate for adjustment	Simulated treatment effect OR = 1.0		Simulated treatment effect OR = 0.8	
				Treatment effect estimate (OR)	Standard error	Treatment effect estimate (OR)	Standard error
Assignment based on (random) variable with no correlation with outcome							
RCT	Logistic regression	-		1.00	0.0660	0.80	0.0681
RCT	Logistic regression	Linear	Random, R^2=0	1.00	0.0660	0.80	0.0682
RD	Logistic regression	-		1.00	0.0660	0.80	0.0681
RD	Logistic regression	RCS	Random, R^2=0	1.00	0.1668	0.80	0.1722
RD	Local logistic regression	-		1.00	0.1095	0.80	0.1129
Assignment based on low correlating (R^2 = 7%) assignment variable with outcome							
RCT	Logistic regression	-		1.00	0.0660	0.81	0.0681
RCT	Logistic regression	Linear	Low correlating assignment variable, R^2=7%	1.00	0.0678	0.80	0.0698
RD	Logistic regression	RCS	Low correlating assignment variable, R^2=7%	1.01	0.2088	0.81	0.2129
RD	Local logistic regression	-		1.00	0.1097	0.80	0.1137
Assignment based on high correlating (R^2 = 31%) assignment model with outcome							
RCT	Logistic regression	-		1.00	0.0597	0.83	0.0607
RCT	Logistic regression	Linear	High correlating assignment variable, R^2=31%	1.00	0.0654	0.80	0.0666
RD	Logistic regression	RCS	High correlating assignment variable, R^2=31%	1.00	0.1898	0.80	0.1924
RD	Local logistic regression	-		1.00	0.1088	0.80	0.1098

So, for example, in the first RD assignment strategy, treated patients with an age > 33 years and control patients with an age ≤ 33 years, were selected to analyze the data as if it was an RD design.

In each scenario, the treatment effect was estimated with both local logistic regression(15) and a logistic regression model with treatment and adjusted for the assignment variable in an RCS term. The treatment effect was expressed as ORs with 95% confidence intervals (95% CI). Analyses were repeated 5,000 times. Random samples of 50% from the complete RCT data were drawn (5,000 times), to calculate the treatment effect from the RCT as a reference estimate. In this way we were able to compare the RD and RCT estimates in the same sample sizes.

To assess the heterogeneity of the treatment effect over the baseline assignment variable, we fitted a model with an interaction term between treatment and the different assignment models to the complete RCT data.

All statistical analyses were performed in R statistical software version 2.15.3 (R Foundation for Statistical Computation, Vienna, Austria) using the *rms* and *gam* packages.

RESULTS

Monte Carlo simulations

Simulations showed that treatment assignment based on a random or poorly correlating variable with outcome, resulted in higher relative efficiency compared to an unadjusted RCT, than RD with treatment assignment based on a higher correlating variable with outcome (Table 2). To obtain the same efficiency as an unadjusted RCT, RD required 2.8 times as many patients when using an assignment variable not correlating with outcome, and approxiamately 3.3 times as many patients when using an assignment variable highly (R^2 31%) correlating with outcome, when RD was analyzed with local regression. The relative efficiency was up to approximately 10 times as low with the strongly correlating assignment variable compared to an unadjusted RCT, when using logistic regression using adjustment. However, compared to an adjusted RCT, the relative efficiency was not dependent on the correlation between the treatment assignment variable and outcome. With local logistic regression, RD required at most 2.8 times as many patients compared to an adjusted RCT in all three assignment strategies (Table 2). In all three treatment assignment approaches, the estimated treatment effects were similar to the simulated treatment effect.

Case study

The median age in CRASH was 33 years (inter quartile range (IQR) 23-47), 2323 (24%) patients died within 14-days after injury. The correlation between the assignment variable

Table 2. Relative efficiency in terms of required sample size in an RD design for different baseline risk assessments compared to an RCT*.

	Assignment based on random variable		Assignment based on low (R^2=7%) correlating variable with outcome		Assignment based on high (R^2 =31%) correlating variable with outcome	
Simulated Odds Ratio	0.8	1.0	0.8	1.0	0.8	1.0
Compared to an unadjusted RCT						
RD no adjustment	1.00	1.00	-	-	-	-
RD RCS adjustment	6.39	6.39	9.78	9.99	10.04	10.12
RD local logistic regression	2.75	2.76	2.79	2.76	3.27	3.32
Compared to an adjusted RCT						
RD no adjustment	1.00	1.00	-	-	-	-
RD RCS adjustment	6.38	6.39	9.31	9.48	8.34	8.42
RD local logistic regression	2.74	2.75	2.65	2.62	2.72	2.77

**Formula used to calculate the relative efficiency: $(SE_{RD} / SE_{RCT})^2$*

age and 14-day mortality was low (R^2 7%). The correlation between the full model for treatment assignment including age, pupillary reactivity and motor score and mortality was stronger (R^2 31%) (Table 3).

In CRASH treatment had a negative effect on outcome overall. The mean unadjusted OR for treatment on 14-day mortality over in 5,000 subsets of 50% of the RCT was 1.20 [95% CI; 1.05-1.37] and 1.27 [1.08; 1.48] adjusted for the linear predictor defined by age, motor score and pupil reactivity. In the RCT data we found no statistically significant interaction between treatment and both of the assignment variables. However, non-linear RCS functions of the treatment effect over the assignment variable were plotted and showed in Figure 1 and suggests some interaction over the range of the assignment variable. The (local) estimates of the treatment effect in RD varied according to the assignment variable and corresponding cut-off. RD based on a random treatment assignment variable resulted in similar point estimates as the RCT, with and without RCS adjustment and with local logistic regression. The RD estimates were more similar to the global RCT estimates in the approach with assignment based on a poorly correlating variable with outcome compared to RD assignment based on a highly correlated variable with outcome; with assignment based on only age the adjusted ORs for treatment were 1.44 [0.95; 2.19] and 1.13 [0.91; 1.41] estimated with RCS adjusted logistic regression and local logistic regression respectively. In the RD design with assignment based on the higher correlating assignment variable, the estimates were less similar to the RCT effect estimate for treatment; the adjusted OR for treatment estimated with logistic regression was 1.69 [1.01; 2.82]. The estimated OR with local logistic regression was 1.41 [1.12; 1.76]) (Table 4).

Table 3. Patient characteristics and explained variance with 14-day mortality in the CRASH trial (n = 9,554)

Characteristic	N (%)	$R^{2\#}$
Random treatment allocation	4800 (50)	0^
Age, median (IQR)	33 (23 - 47)	7
Motor score*		22
1	785 (8)	
2	515 (5)	
3	659 (7)	
4	1181 (12)	
5/6	6414 (67)	
Pupillary reactivity		19
Both responsive	8100 (85)	
One responsive	597 (6)	
Both unresponsive	857 (9)	
Predicted probability from full prediction model**, median (IQR)	0.15 (0.08 – 0.31)	31
14-day mortality	2323 (24)	NA

 # Explained variance with 14-day mortality for CRASH

^ independent of the treatment effect (in the absence of treatment effect)

** 1 Makes no movements, 2 Extension to painful stimuli, 3 Abnormal flexion to painful stimuli, 4 Flexion/withdrawal to painful stimuli, 5/6 Localizes painful stimuli / Obeys commands*

***Age, motor score and pupillary reactivity*

DISCUSSION

We investigated the impact on efficiency of different associations of the treatment assignment variable with the outcome under study in the RD design. 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.

Table 4. RCT and RD analyses in the CRASH (n=9 554), repeated 5000 times.

Analysis	Adjustment	Covariate for adjustment	N total	OR (95% CI) for 14-day mortality	Standard error
Randomized controlled trial (50% subset)					
Logistic regression	-		4777	1.20 (1.05-1.37)	0.07
Logistic regression	Linear	Age	4777	1.22 (1.06-1.40)	0.07
Logistic regression	RCS	Age	4777	1.22 (1.07-1.40)	0.07
Logistic regression	Linear	Linear predictor full model	4777	1.27 (1.09-1.48)	0.08
Logistic regression	RCS	Linear predictor full model	4777	1.27 (1.09-1.48)	0.08
Regression discontinuity: assignment based on random variable with no correlation with outcome					
Logistic regression	-		4777	1.20 (1.05-1.37)	0.07
Logistic regression	RCS	Random	4777	1.21 (0.86-1.69)	0.17
Local logistic regression	-		4777	1.20 (0.97-1.50)	0.11
Regression discontinuity: assignment based on low (R^2 = 0.07) correlating assignment model with outcome*					
Logistic regression	RCS	Age	4777	1.44 (0.95-2.19)	0.21
Local logistic regression	-		4777	1.13 (0.91-1.41)	0.11
Regression discontinuity: assignment based on high (R^2 = 0.31) correlating assignment model with outcome**					
Logistic regression	RCS	Linear predictor full model	4777	1.69 (1.01-2.82)	0.26
Local logistic regression	-		4777	1.41 (1.12-1.76)	0.11

** The median age was used as a cut-off for treatment assignment. Age ≤ 33 receiving no treatment and age > 33 receiving treatment.*

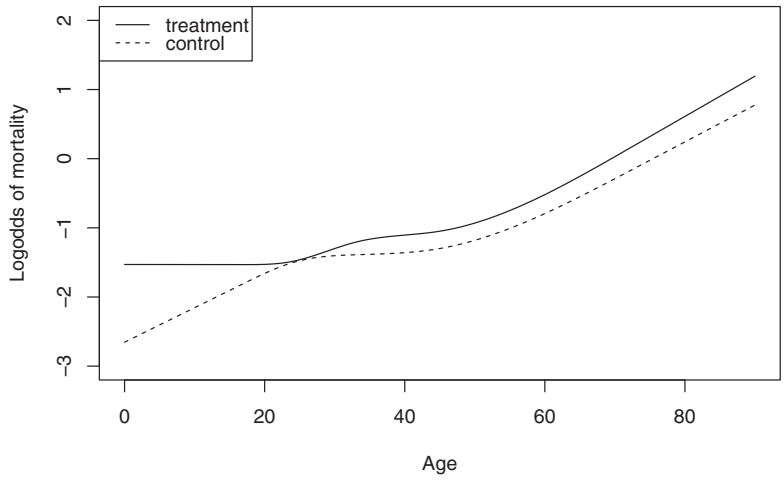
***Assignment based on the linear predictor of age, pupillary reactivity and motor score as predictors for outcome. The median linear predictor was used as a cut-off for treatment assignment.*

Efficiency of different RD assignment strategies

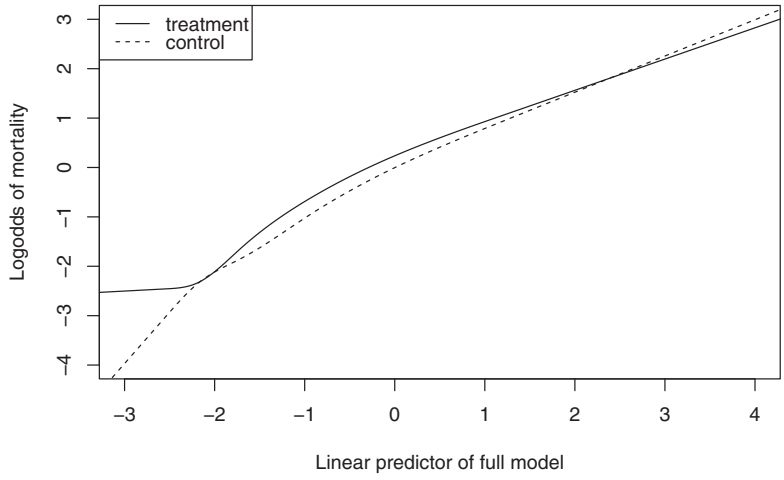
First, the simulation study shows higher relative efficiency of RD based on a random or poorly correlating variable than RD based on a higher correlating assignment variable, when compared to an unadjusted RCT. We found that RD based on a random variable needs 2.75 times as many patients to have the same statistical power as in an unadjusted RCT, which is similar to what has been described in other studies on the efficiency of RD compared to RCTs.(7, 8, 16, 17)

We note that treatment effect estimates in RD are conditional on the assignment variable. Thus, comparing RD estimates with conditional estimates from an RCT with adjustment for the assignment variable would be the more appropriate comparison. A feature of covariate adjustment with nonlinear models, such as logistic regression, is an increase of the standard error of the conditional treatment effect estimate from an RCT.(18-20) The increase in relative efficiency of RD based on a higher correlating assignment variable is eliminated when the standard errors of estimates are compared to the increased standard errors of the treatment effect estimate from the adjusted RCT,

since the increase of standard error of the treatment effect estimated in an adjusted RCT is higher in case of a high correlating assignment variable. In other words, when the RCT estimates are conditioned on the same covariates as used in the RD design, the relative efficiency is independent of the correlation between the assignment variable and the outcome.



*A) Nonlinear rcs function (the function fitted is 14 day mortality ~ treatment * rcs [age]) of the interaction of the treatment effect over the range of age in the CRASH trial. Interaction test of age * treatment was not significant (p = 0.17).*



*B) Nonlinear rcs function (the function fitted is 14 day mortality ~ treatment * rcs [linear predictor of the full model]) of the interaction of the treatment effect over the range of the linear predictor in the CRASH trial. Interaction test of the linear predictor * treatment was not significant (p = 0.99).*

Figure 1. Density plot of the assignment variables and nonlinear restricted cubic spline functions of the interaction of the treatment effects over the range of the assignment variables in the CRASH trial (n=9,554).

In the simulations and in the case study two different methods for the analysis of the treatment effect estimate were used: local logistic regression and logistic regression with adjustment for the assignment variable using restricted cubic splines (RCS). The two methods both provided valid treatment effect estimates. However, estimation of the treatment effect using local logistic regression would be the preferred method to use in an RD design since both in the simulations and the case study this method resulted in lower standard errors; local logistic regression provides more efficient estimations.

Validity of different RD assignment strategies

The treatment effects estimated in RCTs can be interpreted as global treatment effects. Interpreting RD estimates as global treatment effect estimates requires the assumption of an identical treatment effect over the full range of the assignment variable. This implies that treatment does not interact with the baseline assignment variable.(2) However, the treatment effect could vary over de range of the assignment variable, as is shown in Figure 1. We were able to plot this effect since we had the RCT data available. In contrast, in a prospective RD design, the assumption of no interaction between treatment and the assignment variable cannot be tested, since the treatment groups each have data on only one side of the cut-off. The RD estimates should thus primarily be interpreted as local treatment effects at the assignment cut-off.(9) This is also illustrated in our case study in CRASH. As expected, RD based on a random treatment assignment variable resulted in the same treatment effect estimates as in the RCTs. The estimates from RD based on a poorly correlating variable with outcome were more similar to the global RCT estimates, compared to the RD estimates with assignment based on a variable highly correlating with outcome. Indeed, the treatment effect varied over the range of the higher correlating variable (Figure 1b). This might reflect a more general explanation that treatment effect heterogeneity over the range of the baseline assignment variable is less likely when an assignment variable has a no correlation with outcome. When treatment assignment is based on a random or poorly correlating variable with outcome in RD, it may be more acceptable to assume a global treatment effect over the range of the assignment variable and the estimates from RD can be interpreted as an average treatment effect.

Implications

We can debate the applicability of a prospective RD design and choosing a variable for treatment assignment. There might be more clinical support to assign treatment to high-risk patients, because these patients have the highest absolute benefit of treatment, when the relative benefit is similar over the whole range of the assignment variable.(21) In an RD design this approach would not increase efficiency. Thus, in RD we do not necessarily have to aim for an assignment variable that strongly correlates

with outcome, such as a prognostic model that combines multiple predictors. It could be more practical to apply RD on a single baseline measurement, such as blood pressure, cholesterol level or age. The simplicity of this approach is an advantage. Besides, compared to an unadjusted RCT this approach is more efficient. In clinical practice, it is not uncommon that treatment is assigned based on a single baseline measurement; this treatment assignment strategy highly resembles treatment assignment in an RD design. For example, intensified medical treatment given to very low-birth-weight-babies (weighing less than 1,500 g).(4, 22) Also a CD4 count threshold is used in HIV patients to determine treatment assignment for immediate vs. deferred antiretroviral therapy.(4, 9) In traumatic brain injury, it is recommended to treat patients with intracranial pressure monitoring above 22 mmHg.(23) These are examples of treatment assignment in daily clinical practice that resemble a 'natural' application of the RD design. In theory, observational data of these examples could be used to assess the (local) effectiveness of treatment. Thus, RD based on one single measure as an assignment variable may be a good trade-off between efficiency and feasibility in clinical practice.

Conclusion and recommendations

In conclusion, 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. 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.

REFERENCES

1. 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.
2. Labrecque JA, Kaufman JS. Commentary: Can a Quasi-experimental Design Be a Better Idea than an Experimental One? Epidemiology. 2016;27(4):500-2.
3. Bor J, Moscoe E, Barnighausen T. Three approaches to causal inference in regression discontinuity designs. Epidemiology. 2015;26(2):e28-30; discussion e.
4. 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.
5. 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.
6. van Leeuwen N, Lingsma HF, Mooijaart SP, Nieboer D, Trompet S, Steyerberg EW. Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials. J Clin Epidemiol. 2018;98:70-9.
7. Goldberger AS. Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations: University of Wisconsin--Madison; 1972.
8. Tang Y, Cook TD, Kisbu-Sakarya Y. Statistical Power for the Comparative Regression Discontinuity Design With a Nonequivalent Comparison Group. Psychol Methods. 2018;23(1):150-68.
9. 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.
10. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019.
11. Edwards P, Farrell B, Lomas G, Mashru R, Ritchie N, Roberts I, et al. The MRC CRASH Trial: study design, baseline data, and outcome in 1000 randomised patients in the pilot phase. Emerg Med J. 2002;19(6):510-4.
12. Reith FCM, Lingsma HF, Gabbe BJ, Lecky FE, Roberts I, Maas AIR. Differential effects of the Glasgow Coma Scale Score and its Components: An analysis of 54,069 patients with traumatic brain injury. Injury. 2017;48(9):1932-43.
13. Steyerberg EW, Mushkudiani N, Perel P, Butcher I, Lu J, McHugh GS, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008;5(8):e165; discussion e.
14. Collaborators MCT, Perel P, Arango M, Clayton T, Edwards P, Komolafe E, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. BMJ. 2008;336(7641):425-9.
15. Cleveland WS, Devlin SJ, Grosse E. Regression by local fitting: Methods, properties, and computational algorithms. Journal of Econometrics. 1988;37(1):87-114.
16. Schochet PZ. Statistical Power for Regression Discontinuity Designs in Education Evaluations. J Educ Behav Stat. 2009;34(2):238-66.
17. Senn S. Statistical Issues in Drug Development (Second edition). Senn S, editor: John Wiley; 2007.
18. 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.

19. 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.
20. 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.
21. Califf RM, Woodlief LH, Harrell FE, Jr., Lee KL, White HD, Guerci A, et al. Selection of thrombolytic therapy for individual patients: development of a clinical model. GUSTO-I Investigators. Am Heart J. 1997;133(6):630-9.
22. 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.
23. 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.

APPENDIX 1

```
## Activation of required libraries
library(rms)
library(foreign)
library(gam)

set.seed(100)

n_patients          <- 5000
treatment_effect <- 0
n_sim     <- 10000
Est_Effect          <- matrix(nrow = n_sim, ncol=24)
colnames(Est_Effect)          <- c("Tx-effect RCT1", "se RCT1",
          "Tx-effect RCT2", "se RCT2",
          "Tx-effect RCT3", "se RCT3",
          "Tx-effect RCT4", "se RCT4",
          "Tx-effect RCT5", "se RCT5",
          "Tx-effect RDrandom1", "se RDrandom1",
          "Tx-effect RDrandom2", "se RDrandom2",
          "Tx-effect RDrandom3", "se RDrandom3",
          "Tx-effect RDlow1", "se RDlow1",
          "Tx-effect RDlow2", "se RDlow2",
          "Tx-effect RDhigh1", "se RDhigh1",
          "Tx-effect RDhigh2", "se RDhigh2")

for(i in 1:n_sim){
index     <- sample(1:nrow(data), replace = TRUE, size = n_patients)
data$motor          <- as.factor(data$motor)
data$pupils_i       <- as.factor(data$pupils_i)
data$random         <- rnorm(10008, 50, 12.5)
CRASH_sim <- data[index, c("age", "motor","pupils_i", "random")]
CRASH_sim$pupils_i1       <- CRASH_sim$pupils_i==1
CRASH_sim$pupils_i2       <- CRASH_sim$pupils_i==2
CRASH_sim$motor1          <- CRASH_sim$motor==2
CRASH_sim$motor2          <- CRASH_sim$motor==3
CRASH_sim$motor3          <- CRASH_sim$motor==4
CRASH_sim$motor4          <- CRASH_sim$motor==5
CRASH_sim$lp1    <- with(CRASH_sim, 0)
```

```
CRASH_sim$lp2   <- with(CRASH_sim, 0.0289 * age)
CRASH_sim$lp3   <- with(CRASH_sim, 0.0336 * age + 0.7878 * motor1 + 0.2312 * motor2
+ -0.2916 * motor3 + -1.3765 * motor4 + 0.9090 * pupils_i1 + 1.7841 * pupils_i2)

## Randomized controlled trial
# Treatment "Randomize all patients"
CRASH_sim$T_RCT          <- as.numeric(runif(n_patients) <= 0.5)
# Outcome "Randomize all patients"
CRASH_sim$O_RCTlp1       <- with(CRASH_sim, plogis(-1.1355 + lp1 + treatment_effect
* T_RCT)> runif(nrow(CRASH_sim)))
CRASH_sim$O_RCTlp2       <- with(CRASH_sim, plogis(-2.2671 + lp2 + treatment_effect
* T_RCT)> runif(nrow(CRASH_sim)))
CRASH_sim$O_RCTlp3       <- with(CRASH_sim, plogis(-1.9675 + lp3 + treatment_effect
* T_RCT)> runif(nrow(CRASH_sim)))
## Regression discontinuity, assignment with random variable
#Treatment "Regression discontinuity design"
CRASH_sim$T_RDD_R        <-          as.numeric(median(CRASH_sim$random)<CRASH_
sim$random)

#Outcome "Regression discontinuity design"
CRASH_sim$O_RDD_R        <- with(CRASH_sim, plogis(-1.1355 + lp1 + treatment_effect
* T_RDD_R)> runif(nrow(CRASH_sim)))

## Regression discontinuity, assignment with low correlating variable
#Treatment "Regression discontinuity design"
CRASH_sim$T_RDD_L        <- as.numeric(median(CRASH_sim$lp2)<CRASH_sim$lp2)

#Outcome "Regression discontinuity design"
CRASH_sim$O_RDD_L        <- with(CRASH_sim, plogis(-2.2671 + lp2 + treatment_effect
* T_RDD_L)> runif(nrow(CRASH_sim)))

## Regression discontinuity, assignment with high correlating variable
#Treatment "Regression discontinuity design"
CRASH_sim$T_RDD_H        <- as.numeric(median(CRASH_sim$lp3)<CRASH_sim$lp3)

#Outcome "Regression discontinuity design"
CRASH_sim$O_RDD_H        <- with(CRASH_sim, plogis(-1.9675 + lp3 + treatment_effect
* T_RDD_H)> runif(nrow(CRASH_sim)))
```

```
#fit RCT
fit_RCT1 <- lrm(O_RCTlp1 ~ T_RCT, data = CRASH_sim, x=T, y=T)
fit_RCT2 <- lrm(O_RCTlp2 ~ T_RCT, data = CRASH_sim, x=T, y=T)
fit_RCT3 <- lrm(O_RCTlp2 ~ T_RCT + lp2, data = CRASH_sim, x=T, y=T)
fit_RCT4 <- lrm(O_RCTlp3 ~ T_RCT, data = CRASH_sim, x=T, y=T)
fit_RCT5 <- lrm(O_RCTlp3 ~ T_RCT + lp3, data = CRASH_sim, x=T, y=T)
#fit RD with assignment based on random variable
fit_RDD_random1          <- lrm(O_RDD_R ~ T_RDD_R, data = CRASH_sim, x=T, y=T)
fit_RDD_random2          <- lrm(O_RDD_R ~ rcs(random) + T_RDD_R, data = CRASH_
sim, x=T, y=T)
fit_RDD_random3          <- gam(O_RDD_R ~ lo(random) + T_RDD_R, family = bino-
mial, data = CRASH_sim)
#fit RD with assignment based on low correlating variable
fit_RDD_low1     <- lrm(O_RDD_L ~ rcs(lp2) + T_RDD_L, data = CRASH_sim, x=T, y=T)
fit_RDD_low2     <- gam(O_RDD_L ~ lo(lp2) + T_RDD_L, family = binomial, data =
CRASH_sim)
#fit RD with assignment based on high correlating variable
fit_RDD_high1    <- lrm(O_RDD_H ~ rcs(lp3) + T_RDD_H, data = CRASH_sim, x=T, y=T)
fit_RDD_high2    <- gam(O_RDD_H ~ lo(lp3) + T_RDD_H, family = binomial, data =
CRASH_sim)

Est_Effect[i, ]      <- c(fit_RCT1$coefficients["T_RCT"],
          sqrt(fit_RCT1$var["T_RCT","T_RCT"]),

          fit_RCT2$coefficients["T_RCT"],
          sqrt(fit_RCT2$var["T_RCT","T_RCT"]),

          fit_RCT3$coefficients["T_RCT"],
          sqrt(fit_RCT3$var["T_RCT","T_RCT"]),

          fit_RCT4$coefficients["T_RCT"],
          sqrt(fit_RCT4$var["T_RCT","T_RCT"]),

          fit_RCT5$coefficients["T_RCT"],
          sqrt(fit_RCT5$var["T_RCT","T_RCT"]),



          fit_RDD_random1$coefficients["T_RDD_R"],
```

```
            sqrt(fit_RDD_random1$var["T_RDD_R", "T_RDD_R"]),

            fit_RDD_random2$coefficients["T_RDD_R"],
            sqrt(fit_RDD_random2$var["T_RDD_R", "T_RDD_R"]),

            fit_RDD_random3$coefficients["T_RDD_R"],
            (sqrt(diag(vcov(fit_RDD_random3)))["T_RDD_R"]),



            fit_RDD_low1$coefficients["T_RDD_L"],
            sqrt(fit_RDD_low1$var["T_RDD_L", "T_RDD_L"]),

            fit_RDD_low2$coefficients["T_RDD_L"],
            (sqrt(diag(vcov(fit_RDD_low2)))["T_RDD_L"]),



            fit_RDD_high1$coefficients["T_RDD_H"],
            sqrt(fit_RDD_high1$var["T_RDD_H", "T_RDD_H"]),

            fit_RDD_high2$coefficients["T_RDD_H"],
            (sqrt(diag(vcov(fit_RDD_high2)))["T_RDD_H"]))

}

#Mean Effect estimate of treatment and standard error
colMeans(Est_Effect,na.rm=T)
```

Chapter 8

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.

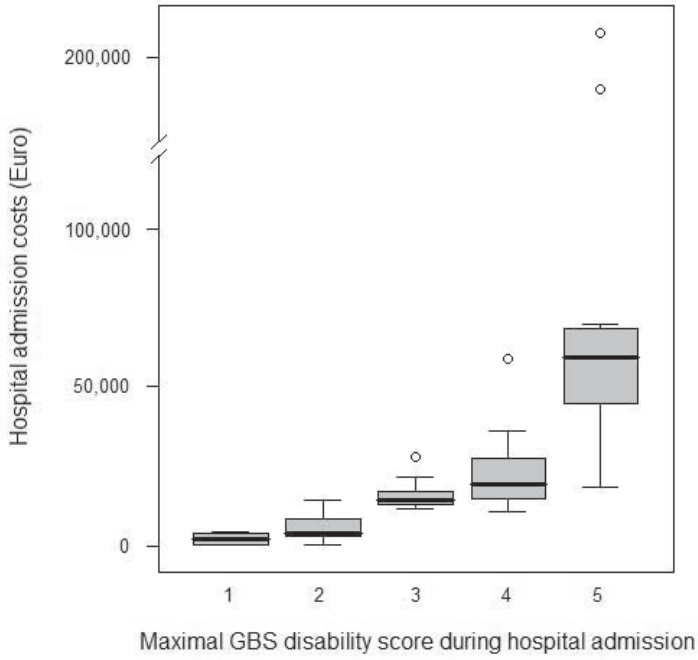


Figure 1. Interquartile ranges (grey boxes), 95% confidence intervals (whiskers) and median (dark lines in middle of the boxes) of costs of hospital admission for different maximal GBS disability scores.
Excluded was one patient who died. Circles are (extreme) outliers. Maximal GBS disability score during 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

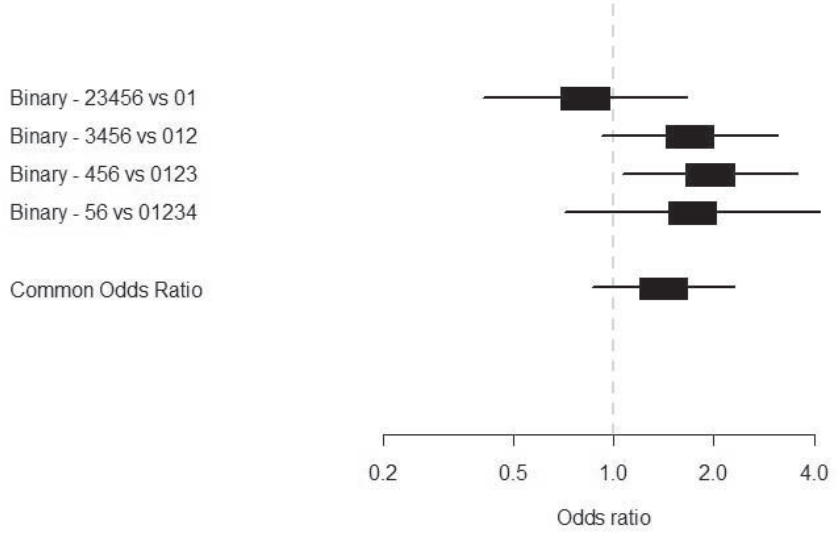


Figure 2. Treatment effect analysis: forest plots of the adjusted binary and proportional odds logistic regression in the IVIg + placebo vs IVIg + Methylprednisolon (IVIg vs MP) trial

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.

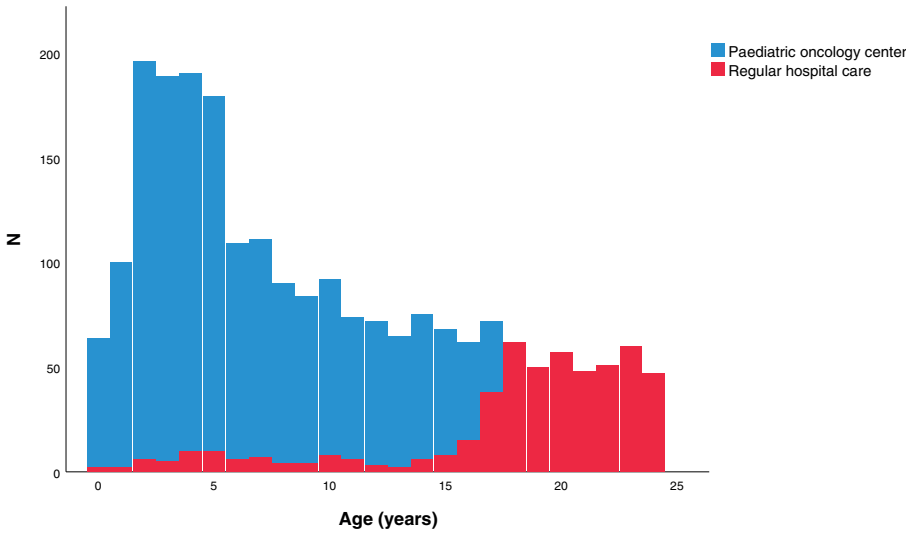


Figure 3. Histogram of the distribution of baseline age of patients treated in either pediatric oncology center or with regular hospital care.

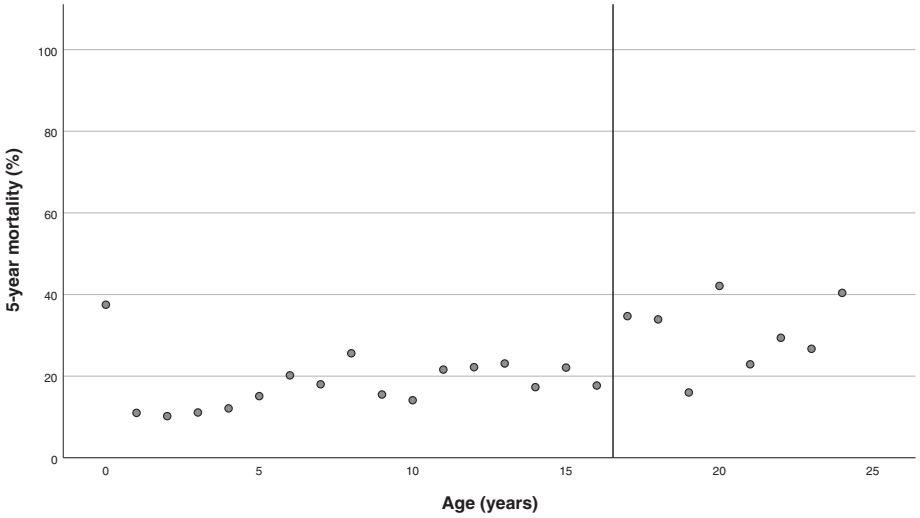


Figure 4. Scatterplot of the probability on 5-year mortality (ignoring censoring) per age.

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.

Summary

Chapter 1, the general introduction, gives an overview of the background and aims addressed in this thesis. Randomized clinical trials (RCTs) provide the most reliable evidence of effectiveness of medical interventions. Specific challenges with regard to efficiency arise when conducting RCTs in rare diseases in heterogeneous populations are challenging. Despite the random allocation between treatment- and control group, differences in baseline risk on outcome can arise between the treatment arms, simply due to chance. Also, in diseases with large heterogeneity in natural disease course, severity and outcome, small differences in baseline risk on outcome between the treatment arms may have influence on the treatment effect estimated.

When performing an RCT is impossible, the quasi-experimental "regression discontinuity" (RD) design is an alternative epidemiological design to study effectiveness of a medical intervention. In the RD design, treatment is not assigned randomly like in an RCT, but is allocated to a subset of patients, based on a cut-off of a baseline assignment variable. A subset of patients below the cut-off, not receiving a medical intervention, is considered as the control group. Due to the controlled treatment assignment, an RD design achieves balance on unobserved factors between the treatment- and control group, just like in an RCT. RD may thus provide an opportunity to obtain unbiased causal treatment effect estimates, when an RCT is not feasible.

The aim of this thesis was to assess the benefits of more advanced statistical analyses to estimate treatment effects from RTCs in heterogeneous diseases (part I; chapter 2, 3 and 4). Besides, the validity and reliability of the RD design compared to an RCT to estimate causal treatment effects was studied (part II; chapter 5, 6 and 7).

In **chapter 2** we found that hospital admissions for Guillain Barré syndrome (GBS) patients were heterogeneous, especially with regard to number of transfers and costs. GBS is a complex disorder because of the various stages in the clinical course and diversity in clinical course between patients. 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 admission. Moreover, in 25% 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. The large heterogeneity should be taken into account when designing an RCT in GBS.

In **chapter 3**, we studied the prognostic value of major extracranial injury (MEI) on mortality in traumatic brain injury (TBI) patients. Our results 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. Our meta-analysis 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.

In **chapter 4** the benefits of both covariate adjustment and proportional odds analysis in RCTs in GBS were assessed. On expectation, covariate adjustment leads to more extreme (further away from $\beta = 0$ *or odds ratio* = 1) treatment effect estimates and larger standard errors. Indeed, 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 vs Intravenous Immunoglobulin (PE vs IVIg) trial. In the IVIg and placebo versus IVIg and Methyl-Prednisolone (MP) (IVIg vs MP) trial, 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. The potential gain of proportional odds analysis was also assessed. 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. Because the ordinal analysis uses the full ordinal outcome scale instead of one dichotomy, the variability will be smaller compared to binary analysis. This was confirmed in our study, where the proportional odds resulted in lower standard errors compared to the binary approaches.

Chapter 5 describes simulations and a validation study to assess the validity and efficiency of the RD design with continuous outcomes, compared to an RCT. In both the simulations and the validation study the treatment effect estimates from an RCT were used as the reference for a prospectively performed RD. We estimated the treatment effect using linear regression adjusting for the assignment variable both as linear terms and restricted cubic spline (RCS) and using local linear regression models. In the first validation study, the estimated treatment effect β from a cardiovascular RCT was −4.0 mmHg blood pressure (95% confidence interval (CI): −5.4, −2.6) at 2 years after inclusion. the estimated effect in RD was −5.9 mmHg (95% CI: −10.8, −1.0) with RCS adjustment.

RD showed different, local effects when analyzed with local linear regression. In the second RCT, RD treatment effect estimates on total cholesterol level at 3 months after inclusion were similar to RCT estimates, but at least six times less precise. We concluded that RD may provide similar estimates of treatment effects to RCT estimates but requires the assumption of a global treatment effect over the range of the assignment variable. In addition to a risk of bias due to wrong assumptions, researchers need to weigh better recruitment against the substantial loss in precision when considering a study with RD versus RCT design.

In **Chapter 6**, we aimed to evaluate validity and efficiency in the RD design for dichotomous outcomes compared to an RCT. We hereto performed validation studies in three large RCTs. To mimic the RD design, we selected patients above and below a cutoff (e.g., age 75 years) randomized to treatment and control, respectively. Adjusted logistic regression models using RCS and polynomials and local logistic regression models estimated the odds ratios (ORs) for treatment, with 95% CIs to indicate precision. In the first RCT, treatment increased mortality with OR 1.22 [95% CI 1.06e1.40] in the RCT. The RD estimates were 1.42 (0.94 - 2.16) and 1.13 (0.90 - 1.40) with RCS adjustment and local regression, respectively. In the second RCT, treatment reduced mortality (OR 0.83 [0.72 - 0.95]), with more extreme estimates in the RD analysis (OR 0.57 [0.35 - 0.92] and 0.67 [0.51 - 0.86]). In the third RCT, similar RCT and RD estimates were found, again with less precision in RD designs. We concluded that the RD design provides similar but substantially less precise treatment effect estimates compared with an RCT.

Although we know that the RD design may provide valid treatment effect estimates, the design is inefficient. In **chapter 7** we aimed to compare different assignment approaches to increase the statistical efficiency in RD. In Monte Carlo simulations, a random (R^2=0), low (R^2=7%) and highly (R^2=31%) correlating variable with outcome was used for treatment assignment. Patients were sampled from the CRASH trial, with a dichotomous outcome simulated. The treatment effect was analyzed with both local logistic regression and logistic regression with spline adjustment. To assess the relative statistical efficiency, standard errors (SE) of the different treatment assignment strategies were compared with an RCT of the same total sample size. This procedure was repeated in CRASH (n=9,554) as a case study. In the simulations, treatment effect estimates were unbiased. To obtain the same efficiency as an unadjusted RCT, RD required 2.8 times as many patients when using an assignment variable not correlating with outcome, and approximately 3.3 times as many patients when using an assignment variable highly correlating with outcome, using local regression. Compared to an adjusted RCT, the relative efficiency was not dependent on the correlation between the assignment variable and outcome since the adjustment affects the efficiency of an RCT as well. In the case study similar results were found.

Chapter 8, focusses on implications and recommendations when designing an RCT or RD to study the effectiveness of a medical intervention. When designing a future RCT in heterogeneous diseases we recommend de following:

- 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. These methods 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 variables for 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 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.

However, when an RCT is impossible, an RD design can be considered and is preferred over an observational design to assess effectiveness of a medical intervention. Summary implications and recommendations to use RD in epidemiologic and clinical research can be made:

- In an RD design we have full understanding of the allocation of treatment, in contrast to observational studies. The treated- and control patients are replaceable around the cut-off of the assignment variable. This enables local causal inference.
- The RD design may result in similar treatment effect estimates compared to an RCT but are substantially less efficient than the RCT estimates. A prospective RD design needs much higher patient inclusion than RCTs. Otherwise, large observational registry data should be available to apply a retrospective RD.
- RD estimates should primarily be interpreted as local treatment effects and global treatment effect estimates should only be presented secondary to local treatment effect estimates.
- 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.

Samenvatting

In **hoofdstuk 1**, de algemene inleiding, wordt de achtergrond van het onderzoek uiteengezet en worden het doel en de onderzoeksvragen die in dit proefschrift worden beantwoord beschreven. Gerandomiseerde klinische studies (RCT's) bieden het meest betrouwbare bewijs van de effectiviteit van medische interventies. Specifieke uitdagingen met betrekking tot efficiëntie doen zich voor bij het uitvoeren van RCT's bij zeldzame ziekten in heterogene populaties. Ondanks de willekeurige toewijzing tussen de behandel- en controlegroep in een RCT, kunnen door toeval verschillen in baseline risico op de uitkomst optreden tussen de behandelarmen. Ook kunnen bij ziekten met grote heterogeniteit in natuurlijk ziekteverloop, ernst en uitkomst, kleine verschillen in baseline risico op de uitkomst tussen de behandelarmen van invloed zijn op het geschatte behandeleffect.

Wanneer het uitvoeren van een RCT onmogelijk is, is het quasi-experimentele "regression discontinuity" (RD) design een alternatief epidemiologisch design om de effectiviteit van een medische interventie te onderzoeken. In het RD design wordt de behandeling niet willekeurig toegewezen, zoals in een RCT, maar wordt deze toegewezen aan een subgroep van patiënten, op basis van een afkapwaarde van een baseline variabele. Een subgroep van patiënten onder de cut-off, die geen medische interventie krijgt, wordt beschouwd als de controlegroep. Vanwege de gecontroleerde toewijzing van behandeling in een RD design zijn de niet gemeten factoren tussen de behandel- en controlegroep in evenwicht, net als in een RCT. Een RD design biedt daarom de mogelijkheid om valide schattingen van het behandeleffect te krijgen, wanneer een RCT niet haalbaar is.

Het doel van dit proefschrift was om te onderzoeken wat de voordelen zijn van geavanceerde statistische analyses, om behandeleffecten met RTC's in heterogene ziektes te schatten (deel I, hoofdstuk 2, 3 en 4). Daarnaast werd de validiteit en betrouwbaarheid van het RD design om causale behandeleffecten te schatten, in vergelijking met een RCT, onderzocht (deel II, hoofdstuk 5, 6 en 7).

In **hoofdstuk 2** wordt beschreven dat ziekenhuisopnames voor patiënten met de Guillain Barré-syndroom (GBS) heterogeen zijn, met name wat betreft het aantal transfers en kosten. GBS is een complexe aandoening vanwege de verschillende stadia in het klinische beloop en de diversiteit in klinisch beloop tussen patiënten. De complexiteit wordt weerspiegeld in de hoge frequentie van overplaatsingen tussen afdelingen en ziekenhuizen, vooral kort na de initiële opname. Overplaatsingen binnen en tussen ziekenhuizen kwamen frequent voor: 40% van de patiënten werd ten minste één keer overgeplaatst en de helft daarvan werd binnen twee dagen na opname overgeplaatst. Bovendien was de ziekenhuisopname in 25% mogelijk suboptimaal vanuit een perspectief van kosteneffectiviteit, vanwege opname op een andere afdeling dan (pediatrische) neurologie of IC, opname van mild aangedane patiënten op IC's en overplaatsingen kort na de initiële opname. De gerelateerde kosten waren zeer variabel tussen patiënten en

vooral geassocieerd met de ernst van de ziekte. Bij het ontwerpen van een RCT in GBS moet rekening gehouden worden met de heterogeniteit van de populatie.

In **hoofdstuk 3** bestudeerden we de prognostische waarde van groot extracranieel letsel (MEI) op mortaliteit bij patiënten met traumatisch hersenletsel (TBI). Onze resultaten tonen aan dat MEI een belangrijke prognostische factor is voor mortaliteit bij TBI patiënten. Het prognostische effect is echter afhankelijk van de TBI populatie die bekeken wordt. Ten eerste is de sterkte van het effect afhankelijk van de ernst van het hersenletsel. Het prognostische effect van MEI is groter bij patiënten met een mild TBI. Bovendien vonden we dat het effect afhankelijk is van de tijd van inclusie in een onderzoek. In de registratie data die we in onze studie hebben gebruikt, is MEI sterk geassocieerd met mortaliteit na correctie voor leeftijd, Glasgow Coma Scale motorscore en pupilreactiviteit. In breed geselecteerde observationele studies en een RCT, waaronder TBI patiënten die het vroege stadium na hun letsel overleefden, was de incrementele prognostische waarde van MEI in vergelijking met bekende voorspellers van mortaliteit beperkt. Deze resultaten zijn bijvoorbeeld belangrijk om prognostische variabelen voor 'covariate adjustment' te selecteren, bij het ontwerpen van toekomstige TBI trials. Onze meta-analyse impliceert dat MEI een belangrijke prognostische factor is om voor te corrigeren als het effect van pre-ziekenhuisinterventies wordt onderzocht, omdat in deze studies patiënten geïncludeerd worden direct vanaf het moment van het optreden van het letsel. Daarentegen zou het minder noodzakelijk zijn om MEI te overwegen mee te nemen in onderzoeken waarin interventies in ziekenhuizen worden beoordeeld, waarin voornamelijk patiënten geïncludeerd zijn met ernstig hersenletsel die de vroege fase na letsel overleefden.

In **hoofdstuk 4** werden de voordelen van zowel 'covariate adjustment' als 'proportional odds analyse' in RCT's in GBS onderzocht. In theorie leidt het toepassen van 'covariate adjustment' tot extremere (verder weg van $\beta = 0$ of odds ratio = 1) schattingen van behandeleffecten en grotere standaard errors. Inderdaad vonden we verhoogde standard errors in alle analyses met 'covariate adjustment' vergeleken met ongecorrigeerde analyses. De betere prognose in de behandelgroep verkleinde de schatting van het behandeleffect β na 'covariate adjustment' in de Plasma Exchange vs Intravenous Immunoglobulin (PE vs IVIg) trial. In de IVIg and placebo versus IVIg and Methyl-Prednisolone (MP) (IVIg vs MP) studie had de behandelgroep een lagere kans op een gunstige uitkomst. Daarom leidde 'covariate adjustment' in de IVIg vs MP studie tot een grotere schatting van het behandeleffect β en een kleinere p-waarde. De potentiële winst van 'proportional odds analyse' werd ook onderzocht. 'Proportional odds analyse' schat het behandeleffect op elke afkappunt van de ordinale uitkomstschaal, in tegenstelling tot bij lineaire regressie waar het behandeleffect wordt geschat op het verschil tussen de gemiddelde scores in de beide behandelingsarmen. Omdat de ordinale analyse de volledige ordinale uitkomstschaal gebruikt in plaats van één dichotomie, zal de variabiliteit

kleiner zijn in vergelijking met binaire analyse. Dit werd bevestigd in onze studie, waar de 'proportional odds analyse' resulteerde in lagere standaard errors in vergelijking met de binaire aanpak.

Hoofdstuk 5 beschrijft simulaties en een validatiestudie om de validiteit en efficiëntie van het RD design te onderzoeken met continue uitkomsten, vergeleken met een RCT. In zowel de simulaties als de validatiestudie werden de schattingen van het behandeleffect van een RCT gebruikt als referentie voor een prospectief uitgevoerd RD design. We hebben het behandeleffect geschat met lineaire regressie gecorrigeerd voor de baseline variabele waarop behandeling is toegewezen. Dit deden we zowel met lineaire termen als met 'restric cubic splines' (RCS). Ook gebruikten we lokale lineaire regressiemodellen. In de eerste validatiestudie was het geschatte behandeleffect β van een cardiovasculaire RCT -4.0 mmHg (95% betrouwbaarheidsinterval (CI): -5.4, -2.6) op bloeddruk na 2 jaar. Het geschatte effect in RD was -5.9 mmHg (95% CI: -10.8, -1.0) met RCS adjustment. RD liet verschillende, lokale effecten zien wanneer lokale lineaire regressie werd gebruikt in de analyse. In de tweede RCT waren de RD schattingen van het behandeleffect op het totale cholesterolniveau na 3 maanden vergelijkbaar met de RCT schattingen, maar waren minstens zes keer minder nauwkeurig. We concludeerden dat RD vergelijkbare schattingen van het behandeleffect kan geven in vergelijking met een RCT, maar dit vereist de aanname van een globaal behandeleffect over het de gehele range van de variabele waarop behandeling wordt toebedeeld in RD. Naast een risico van bias in schattingen van het behandeleffect als gevolg van verkeerde aannames, moeten onderzoekers een makkelijkere inclusie van patiënten in een RD design afwegen tegen het aanzienlijke verlies aan precisie bij het overwegen van een onderzoek met RD versus RCT design.

In **hoofdstuk 6** evalueerden we de validiteit en efficiëntie van het RD design voor dichotome uitkomsten in vergelijking met een RCT. We voerden validatie studies uit in drie grote RCT's. Om het RD design na te bootsen, selecteerden we patiënten boven en onder een afkapwaarde (bijv. leeftijd 75 jaar) die in de RCT gerandomiseerd waren naar respectievelijk de behandel- en controlegroep. Met zowel logistische regressiemodellen gecorrigeerd met RCS en 'polynomial' termen, als lokale logistische regressiemodellen schatten we de odds ratio's (ORs) en bijbehorende CI voor de behandeling. In de eerste RCT verhoogde de behandeling de mortaliteit met OR 1.22 (95% CI 1.06 -1.40) in de RCT. De schattingen in de RD designs waren 1.42 (0.94 – 2.16) en 1.13 (0.90 – 1.40) met respectievelijk RCS adjustment en lokale regressie. In de tweede RCT verminderde de behandeling mortaliteit (OR 0.83 (0.72 – 0.95)), met extremere schattingen in de RD analyse (OR 0.57 (0.35 – 0.92) en 0.67 (0.51-0.86)). In de derde RCT werden vergelijkbare RCT- en RD schattingen gevonden, opnieuw met minder precisie in RD designs. We concludeerden dat het RD design vergelijkbare maar aanzienlijk minder nauwkeurige schattingen van het behandeleffect oplevert in vergelijking met een RCT.

Hoewel het RD design valide schattingen van het behandeleffect kan opleveren, is het design inefficiënt. In **hoofdstuk 7** vergeleken we verschillende manieren om behandeling toe te wijzen om de statistische efficiëntie in RD te vergroten. In Monte Carlo simulaties, een random (R^2=0), laag (R^2=7%) en hoog (R^2=31%) gecorreleerde variabele met uitkomst werd gebruikt om behandeling toe te wijzen. Dichotome uitkomsten werden gesimuleerd voor patienten gesampeled uit de CRASH trial. Het behandeleffect werd geanalyseerd met zowel locale logistische regressie en logistische regressie met splie correctie. Om de relatieve statisische efficientie te bepalen, werden standard errors (SE) van de verschillende behandeleffect schattingen vergeleken met een schatting uit een RCT met vergelijkbare sample size. Deze procedure werd herhaald in CRASH (n = 9,554), als case study. In de simulaties waren de behandeleffecten unbiased. Om dezelfde efficiëntie als een niet-gecorrigeerde RCT te verkrijgen, vereiste RD 2,8 keer zoveel patiënten bij gebruik van een toewijzingsvariabele die niet correleerde met de uitkomst, en ongeveer 3,3 keer zo veel patiënten bij gebruik van een toewijzingsvariabele die sterk correleerde met de uitkomst, gebruik makende van lokale regressie. Vergeleken met een gecorrigeerde RCT was de relatieve efficiëntie niet afhankelijk van de correlatie tussen de toewijzingsvariabele en de uitkomst, aangezien 'covariate adjustment' ook de efficiëntie van een RCT beïnvloedt. In de case study werden vergelijkbare resultaten gevonden.

Hoofdstuk 8, richt zich op de implicaties en aanbevelingen bij het ontwerpen van een RCT of RD om de effectiviteit van een medische interventie te onderzoeken. Bij het ontwerpen van een toekomstige RCT in heterogene ziektes bevelen we het volgende aan:

- Met behulp van 'covariate adjustment' en 'proportional odds analyse' worden de beschikbare onderzoeksgegevens het meest efficiënt gebruikt en wordt voor evenwicht tussen de behandel- en controlegroep gezorgd om betrouwbare en valide schattingen van het behandeleffect te krijgen. Beide methoden dienen worden toegepast in toekomstige trials met zeldzame en heterogene neurologische aandoeningen zoals GBS.
- Om 'covariate adjustment' toe te passen in toekomstige studies is een goede kennis van de prognostische waarde van baseline karakteristieken van cruciaal belang, zodat het mogelijk is de variabelen voor 'covariate adjustment' van tevoren te specificeren. Deze variabelen kunnen worden geïdentificeerd op basis van klinische ervaring en de wetenschappelijke literatuur over de prognostische waarde van baseline karakteristieken.
- De common OR van een 'proportional odds analyse' is een juiste weergave van het effect van de behandeling op een ordinale uitkomst. Bovendien is deze methode efficiënter dan logistische regressie op een dichotomie van de ordinale schaal. Daarom

raden we aan om de volledige ordinale uitkomstschaal te gebruiken in toekomstige onderzoeken met zeldzame en heterogene neurologische aandoeningen.

Wanneer een RCT niet mogelijk is, kan een RD design worden overwogen. Als de effectiviteit van een medische interventie wordt onderzocht heeft een RD design de voorkeur boven een observationele studie. Samenvattende implicaties en aanbevelingen om RD te gebruiken in epidemiologisch en klinisch onderzoek kunnen worden gemaakt:

- In een RD design hebben we volledig inzicht in het mechanisme van toewijzing van de behandeling, in tegenstelling tot in observationele studies. De behandelde en controlepatiënten zijn uitwisselbaar rond de afkapwaarde van de variabele die gebruikt wordt om behandeling toe te wijzen. Dit maakt lokale causale gevolgtrekking mogelijk.
- Het RD design kan vergelijkbare schattingen van het behandeleffect opleveren in vergelijking met een RCT, maar zijn aanzienlijk minder efficiënt dan de RCT schattingen. Een prospectief RD design vereist veel hogere patiëntaantallen dan RCT's. Een alternatief kan zijn om grote observationele registratiegegevens te gebruiken om een retrospectief RD toe te passen.
- RD schattingen moeten in de eerste plaats worden geïnterpreteerd als lokale behandeleffecten. Schattingen van het globale behandeleffect dienen alleen secundair te worden gepresenteerd aan lokale schattingen van het behandeleffect.
- De relatieve efficiëntie in vergelijking met het behandelingseffect in een RCT met 'covariate adjustment' was niet afhankelijk van de correlatie tussen de behandelingsvariabele en de uitkomst, omdat 'covariate adjustment' ook de efficiëntie van een RCT beïnvloedt.
- Bij het opzetten van een prospectieve RD studie bevelen we onderzoekers aan om toewijzingsvariabelen te gebruiken die haalbaar zijn in de klinische praktijk, maar die niet noodzakelijkerwijs een hoge correlatie met de uitkomst hebben, om de patientinclusie te vergemakkelijken en de efficiëntie van het RD design te optimaliseren.

Dankwoord

Na vele jaren is het eindelijk zo ver: mijn proefschrif is af! Een woord van dank ben ik verschuldigd aan vele mensen, waarvan een aantal in het bijzonder.

Allereerst mijn beide promotoren prof.dr. Ewout Steyerberg en prof.dr. Bart Jacobs.

Beste Ewout, ik ken weinig mensen die zo slim zijn als jij. Je denksnelheid en scherpte zijn bewonderenswaardig. Je bent een uitzonderlijk goede onderzoeker en ik voel mij vereerd dat ik bij je mocht promoveren. Ik heb ontzettend veel van je geleerd, daar ben ik je erg dankbaar voor. Ook heb ik je interesse in mijn sportieve activiteiten altijd zeer gewaardeerd.

Beste Bart, ik heb onze samenwerking van de afgelopen jaren als zeer prettig ervaren. Ik heb het enorm fijn gevonden dat je altijd geduldig de tijd nam om mij, als niet-dokter, uitgebreid uit te leggen hoe het nu precies zat met het Guillain-Barré syndroom. Daarnaast dank ik je voor je opbouwende commentaar op mijn manuscripten en de aanmoedigende en opbeurende woorden, onder andere als papers werden afgewezen.

Dan mijn co-promotor: dr. Hester Lingsma. Beste Hester, ik voel mij een bevoorrecht mens dat jij de afgelopen jaren mijn co-promotor bent geweest. Naast dat onze overleggen altijd erg gezellig zijn, heb ik vooral ook heel veel van je geleerd. Als ik het even niet meer zag zitten kon ik altijd rekenen op je scherpe inzichten en een portie positieve energie. Veel dank ook voor alle kansen die je me tot nu toe gegeven hebt.

De leden van de commissie wil ik danken voor het plaatsnemen in de oppositie en het verdiepen in mijn proefschrift. Speciale dank aan de leden van de leescommissie prof.dr. Pieter van Doorn, prof.dr. Olaf Dekkers en prof.dr. Eric Boersma.

Daan, zonder jou was dit proefschrift er nu nòg niet geweest en was ik nog steeds bezig geweest alle errors in R op te lossen. Enorm veel dank voor al je hulp en geduld hierbij en ook voor je uitleg over allerlei ingewikkelde wiskundige zaken.

David, bedankt voor je hulp bij de laatste loodjes van hoofdstuk 7, dat heeft mij geholpen om dit proefschrift af te ronden. En ik vind het erg leuk dat we weer collega's zijn!

Alle CMB-ers bedankt voor de samenwerking en gezelligheid! Veel dank ben ik verschuldigd aan al het ondersteunende personeel op MGZ. Heren van de ICT helpdesk, bedankt voor jullie hulp als ik weer eens een apparaat had laten vastlopen. En speciale dank aan Sanne, Farsia en Judith, jullie zijn top!

Jitske, Astrid, Elise, Kirsten, Linda en Nanda, volgens mij ben ik nu de laatste van ons groepje die promoveert, of hebben jullie nog plannen Jits en Elis? ;-) Bedankt voor alle koffie, gezelligheid, en het delen van alle life-events!

Esther, vanaf dag één dat we elkaar leerden kennen in het Erasmus MC waren we maatjes. Regelmatig kwam het voor dat we in dezelfde outfit op het werk verschenen. Ondanks dat we geen collega's meer zijn ben ik blij dat we nu bevriend zijn.

De eerste jaren dat ik aan dit proefschrift werkte, was ik veel op de atletiekbaan te vinden. Alle atletiekvrienden, van binnen en buiten Rotterdam, die ik aan deze periode over heb gehouden, bedankt voor alle zware trainingen, trainingsstages, wedstrijden, toernooien, feestjes. Het was, mede dankzij jullie, een mooie periode in mijn leven!

De Prethoek mag natuurlijk niet ontbreken in mijn dankwoord. Michiel de Boer lag niet alleen ten grondslag aan mijn interesse in onderzoek en statistiek, maar ook aan onze naam. Lieve Malou, An, Kris en Marjo, bedankt voor jullie aanmoedigingen en interesse in mijn proefschrift, en voor het aanhoren van mijn eindeloze gezeur… En voor het bedenken van de naam van mijn eenmanszaak: N = 1. Onze etentjes zijn altijd een groot feest en ik verheug me nu al op onze volgende reis.

Eva en Marije, bedankt voor jullie onvoorwaardelijke vriendschap. Jullie zijn schatten!

Lieve Jas, waar moet ik beginnen, al heel lang zijn we vriendinnen. Interrailen door Italië, op safari in Tanzania, vele andere mooie tripjes, samen op de bank op de Statenweg, afzien op de atletiekbaan, samen NK medailles winnen, en ga zo maar door. Aan een half woord hebben wij genoeg. Inmiddels ben je samen met Daan je eigen gezin aan het bouwen, en daar geniet ik met volle teugen van mee. Bedankt dat je altijd voor me klaar staat. Ik kijk er naar uit nog veel mooie momenten in onze levens samen te delen! En misschien schrijven we toch ooit nog wel eens een paper samen?

Lieve Es, wat hebben we al veel meegemaakt samen: lief en leed gedeeld op de Statenweg, trainingsstages, wedstrijden, vele vakanties (Ensjoi!), ontelbaar veel legendarische (huis)feestjes, en nog heel veel andere avonturen... Ik bewonder en geniet van je open blik op het leven, niks is voor jou te gek. Bedankt dat je mijn paranimf wil zijn!

Lieve pap en mam, van jullie heb ik geleerd om altijd door te zetten: "als je ergens aan begint, moet je het ook afmaken". Mijn proefschrift is zeker een resultaat van deze instelling die ik van jullie heb meegekregen. Ik kan altijd op jullie onvoorwaardelijke steun en liefde rekenen, niets is jullie te veel. Bedankt voor alles. Jullie zijn geweldig!

Sam, mijn lieve broer, en paranimf. Van jongs af aan zijn wij twee handen op een buik. Ik bewonder je talent om te genieten van het leven, naar je gevoel te luisteren en buiten de gebaande paden te gaan. Ik ben super trots op je!

Lieve Bob, ik heb heel veel zin in de toekomst samen met jou!

Curriculum Vitae

Nikki van Leeuwen was born in Rotterdam, the Netherlands, on the 4th of August in 1989. After finishing secondary school at the Emmauscollege in Rotterdam, she started a Bachelor of Science in Health Sciences at the Vrije Universiteit in Amsterdam in 2007. She obtained her BSc degree with a thesis on the prognostic value of major extracranial injury in traumatic brain injury, under supervision of dr. Hester Lingsma and prof.dr. Ewout Steyerberg. In 2010, she both started a Master of Science in Clinical Epidemiology at the Netherlands Institute for Health Sciences (NIHES) as well as started working at the department of Public Health of the Erasmus University Medical Center. During this period Nikki also spent much time at the athletics track as a track and field athlete. From 2003 until 2013 Nikki competed in several (inter)national championships as a 100 meter sprinter. Her highest achievement was being part of the 4x100 relay team at the World Championships in Daegu, South Korea in 2011. In 2013 she obtained her MSc degree. During her time at the department of Public Health she was involved in several research projects. As a junior researcher she worked on the development of a quality registration for acute stroke patients, a cost effectiveness study for a second course of immunoglobulins in Guillain-Barré syndrome patients and an alternative methodological design to estimate treatment effect estimates: the regression discontinuity design resulting in this thesis, under supervision of prof.dr. Ewout Steyerberg, prof.dr. Bart Jacobs and dr. Hester Lingsma. Nikki worked as an epidemiologist at the Quality department at the Erasmus University Medical Center from October 2015 until July 2017. In August 2017 she started working as a postdoctoral researcher in the Center for Medical Decision Making at the Department of Public Health in the Erasmus University Medical Center.

List of publications

Van Leeuwen N, Lingsma HF, Perel P, Lecky F, Roozenbeek B, Lu J, Shakur H, Weir J, Steyerberg EW, Maas AI; International Mission on Prognosis and Clinical Trial Design in TBI Study Group; Corticosteroid Randomization After Significant Head Injury Trial Collaborators; Trauma Audit and Research Network. Prognostic value of major extracranial injury in traumatic brain injury: an individual patient data meta-analysis in 39,274 patients. Neurosurgery. 2012;70(4):811-8.

Lingsma H, Nauta S, van Leeuwen N, Borsboom G, Bruining N, Steyerberg E. Tools & Techniques: Analysis of clustered data in interventional cardiology: current practice and methodological advice. EuroIntervention. 2013;9(1):162-4.

Kandimalla R, Masius R, Beukers W, Bangma CH, Orntoft TF, Dyrskjot L, van Leeuwen N, Lingsma H, van Tilborg AA, Zwarthoff EC. A 3-plex methylation assay combined with the FGFR3 mutation assay sensitively detects recurrent bladder cancer in voided urine. Clin Cancer Res. 2013;19(17):4760-9.

Darweesh SK, Polinder S, Mulder MJ, Baena CP, van Leeuwen N, Franco OH, Jacobs BC, Van Doorn PA. Health-related quality of life in Guillain-Barré syndrome patients: a systematic review. J Peripher Nerv Syst. 2014;19(1):24-35.

Van Leeuwen N, Lingsma HF, De Craen AJ, Nieboer D, Mooijaart SP, Richard E, Steyerberg EW. Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes. Epidemiology. 2016;27(4):503-11.

Van Leeuwen N, Lingsma HF, Vanrolleghem AM, Sturkenboom MCJM, Van Doorn PA, Steyerberg EW, Jacobs BC. Hospital admissions, transfers and costs of Guillain-Barré syndrome. PLoS One. 2016;11(2):e0143837.

Kahlmann V, Roodbol J, van Leeuwen N, Ramakers CRB, Van Pelt D, Neuteboom RF, Catsman-Berrevoets CE, De Wit MCY, Jacobs BC. Validated age-specific reference values for CSF total protein levels in children. Eur J Paediatr Neurol. 2017;21(4):654-660.

Van Leeuwen N, Lingsma HF, Mooijaart SP, Nieboer D, Trompet S, Steyerberg EW. Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials. J Clin Epidemiol. 2018;98:70-79.

Lagendijk M, van Egdom LSE, Richel C, van Leeuwen N, Verhoef C, Lingsma HF, Koppert LB. Patient reported outcome measures in breast cancer patients. Eur J Surg Oncol. 2018;44(7):963-968.

Lagendijk M, van Egdom LSE, van Veen FEE, Vos EL, Mureau MAM, van Leeuwen N, Hazelzet JA, Lingsma HF, Koppert LB. Patient-Reported Outcome Measures May Add Value in Breast Cancer Surgery. Ann Surg Oncol. 2018;25(12):3563-3571.

Van Leeuwen N, Walgaard C, Lingsma HF, Van Doorn PA, Jacobs BC, Steyerberg EW. Efficient design and analysis of randomized controlled trials in rare neurological diseases: an example in Guillain-Barré syndrome. PLoS One. 2019;14(2):e0211404.

Van Leeuwen N, Nieboer, Lingsma HF, Van Klaveren D, Steyerberg EW. Efficient treatment assignment in a regression discontinuity design? Simulations and validation in a large randomized controlled trial. Submitted

PhD training and teaching

PhD Student	Nikki van Leeuwen
Erasmus MC	Department of Public Health, Medical Decision Making
Promotor	Prof.dr. E.W. Steyerberg, Prof.dr. B.C. Jacobs
Co-promotor	Dr. H.F. Lingsma

	Year	Workload (ECTs)
1. PhD Training		
General academic training		
Scientific writing course	2011	1
Methodology of patient orientated research and preparation of application for grants (CPO)	2013	1
Timemanagement course	2013	1
Research Integrity	2014	1
Seminars, workshops and courses		
Research seminars Public Health, Erasmus MC	2010-2018	5
'Absolute risk prediction' Mitchell Gail & Ruth Pfeiffer, National Cancer Institute	2012	1
'Causal inference', University for Health Sciences, Medical Informatics and Technology, Austria	2018	2
Female Talent Class, Erasmus MC	2018	2
(Poster) Presentations and International Conferences		
European Association of Neurosurgical Association	2010	1
Inflammatory Neuropathy Consortium - Peripheral Nerve Society	2012	2
International Society for Clinical Biostatistics	2013	2
Society for Clinical Trials	2015	2
Presentations within Erasmus MC	2010-2017	3
National conferences / oral presentations	2010-2016	3
Society for Medical Decision Making	2018	2
Reviewing papers		
Reviewing activities for Journal of Clinical Epidemiology	2016	1
2. Teaching activities		
Lecturing		
'Measuring quality of Care', Clinical Technology BSc course	2017	1
'Quality of Care', Erasmus Summer Programme (NIHES)	2017	1
Preparing and supervising practicals		
'Prognosis research', NIHES course	2011/2012	1
'Clinical epidemiology', NIHES course	2017	1
Supervising		
Supervising Community Projects, 3rd year Medical Students, Erasmus University Medical Center	2010-2015/ 2018/2019	7
Examination Bachelor Essays, 3rd year Medical Students, Erasmus University Medical Center	2014	1