Universiteit
Leiden
The Netherlands

Comparative effectiveness of surgery for traumatic acute subdural hematoma

Essen, T.A. van

Citation
Essen, T. A. van. (2023, June 8). *Comparative effectiveness of surgery for traumatic acute subdural hematoma*. Retrieved from https://hdl.handle.net/1887/3619981

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3619981

Note: To cite this publication please use the final published version (if applicable).

Chapter 14

Discussion

DISCUSSION

The overall aim of this thesis was to evaluate the effectiveness of surgical treatment approaches for patients with traumatic acute subdural hematoma (ASDH). The main research questions were:

I. What is the current evidence on the effectiveness of surgical treatment of ASDH?
II. What is the current practice in treatment of patients with ASDH in Europe?
III. Which study designs and analyses are suited to determine the effectiveness of surgical treatment of ASDH?
IV. What is the effectiveness of different treatment approaches (surgery versus initial conservative treatment and decompressive craniectomy versus craniotomy) for ASDH?

In this chapter I first discuss the key findings of the thesis. Thereafter, I reflect on the key findings, critically appraise surgical intervention research using observational data, and discuss limitations. Finally, I propose recommendations for clinical practice and future research.

Key findings

I first studied the **current evidence** (Part I) of surgery for ASDH in a systematic review. Available literature was limited to observational cohorts or case series, with selected populations, mostly performed more than two decades ago. There were no comparative studies. In a meta-analysis comparing contemporary surgical prognosis with historical conservative prognosis I concluded a strong effect of surgery in comatose patients.

Surveys of **current practice** (Part II) showed large practice variation with regard to whether to operate or not in ASDH. In addition, the willingness among neurosurgeons to randomize to surgery or conservative treatment was low.

In Part III we studied the methodological **challenges of studying acute neurosurgical interventions** using observational data. I evaluated which analysis strategy would be best suited to minimize the confounding in studies on acute neurosurgical interventions. Instrumental variable (IV) analysis might provide a more valid estimate of the treatment effect compared to conventional analytical methods. The methodological justification of this design and the data collection and the analysis were documented in a protocol for the actual effectiveness study.

In Part IV I determined **the effectiveness of neurosurgical interventions in acute subdural hematoma**. In a retrospective comparative effectiveness study comparing treatment preferences of two centers, I concluded that a treatment strategy favoring emergency hematoma evacuation might be associated with better outcome. However, residual confounding might explain the results and, therefore, larger,

prospective, comparative studies with more hospitals were needed. I then performed a prospective comparative effectiveness study that showed that similar patients with ASDH were indeed treated differently due to varying treatment preferences. A treatment strategy preferring an aggressive approach of acute surgical evacuation over initial conservative treatment was not associated with better outcome. Finally, I showed that a primary decompressive craniectomy (DC) does not seem to lead to a better outcome than ASDH evacuation by craniotomy although the results are uncertain due to a small sample.

Reflection on key findings

The thesis is the result of a stepwise, tinkering process in trying to find ways to determine the effectiveness of a longstanding surgical intervention, acute surgery for ASDH, that is considered well-established among the expert community. I have critically appraised the proof of concept (IDEAL framework stage 1) by developing and exploring methods (IDEAL stage 2) to asses and falsify the supposed benefit of acute surgery (IDEAL stage 3).

Especially as neurosurgeons, we know that acute surgery in ASDH can be highly effective in specific cases. My first study came about in my wish to integrate this neurosurgical common sense with evidence-based medicine. I presented the first published systematic review of ASDH surgery. Despite the lack of comparative studies, I could reliably demonstrate that surgical evacuation of ASDH greatly reduces mortality risk in comatose patients (**Chapter 3**). The methodology stems from a paper using meta-analysis with 'dramatic effect' methodology to deduce the effectiveness of a parachute.[11] This methodology resonated with us because the comparison of research in ASDH surgery to a parachute in skydiving had often been made.[2]

This study also highlights the improved care for patients with severe TBI and ASDH over the past century in modern well-resourced hospitals.[3] Targeted and efficient neurosurgery, advances in neurocritical care and the widespread introduction of CT-scanners and ICP monitors have decreased mortality in severe TBI dramatically.[4] Most cases of ASDH, however, do not present in comatose state and most cases are not treated surgically (Table 1).

In **Chapter 3** I conclude that the current evidence base has not improved since the publication of the widely known Brain Trauma Foundation (BTF) guideline in 2006. 'To operate or not' therefore remains a decision surrounded with uncertainty. In **Chapter 2** I, therefore, have tried to move the needle among the scientific community and policy makers from the focus on descriptive small studies, reporting on prognosis of either treatment arm, to well-designed comparative effectiveness studies that aim to answer the actual relevant intervention question.

Subgroup	Number who had surgery/total number
Age	
<65 years	202/725
≥65 years	90/435
TBI severity	
Mild TBI	62/542
Moderate TBI	58/193
Severe TBI	172/425
Haematoma size	
Large	221/309
Small	71/851
Total	**292/1160**

Table 1. Proportion patients surgically treated of the CENTER-TBI cohort of patients with acute subdural hematoma, stratified for specific subgroups

The surveys (**Chapter 4 and 5**) showed large variation for treatment strategies in TBI, specifically for everyday ASDH cases (Figure 1). Although the regional variation in surgical treatment of ASDH has not been shown in the literature, other studies have shown that variability in treatment of TBI exists.[5,6] Moreover, in only 17% of a random sample of (brain) trauma patients care delivered according to the BTF guidelines, suggesting a variable approach.[7] In addition, there exists a difference in point of view among neurosurgeons with respect to combining the evacuation of an ASDH with a DC.[8]

The existing variation might be a sign of unwarranted care differences, but also a reflection of real-life neurosurgical practice in TBI. Decisions are made under large uncertainty, constrained not only by limited evidence, but also by time, in an emotionally charged environment, at moments when regular consultation with colleagues is difficult. Furthermore, clinicians need to make accurate prognostic estimates to inform surgical decision-making but this process is at best unequivocal.[9] What makes matters worse is that there seems to be a seemingly reasonable quality of life after severe TBI, which might be attributed to the disability paradox.[10]

Subsequently, I explored whether an RCT, the superior methodological approach for causal inference in clinical epidemiology,[11] would be possible. Unfortunately, the survey showed not enough willingness among neurosurgeons to randomize patients with an ASDH. Thus, I concluded that there is at present not sufficient clinical equipoise while, at the same time, opposite treatment strategies for similar patients are employed. An explanation for this contradiction might be that neurosurgeons extent the clear role of surgery in specific cases of ASDH to other - more uncertain

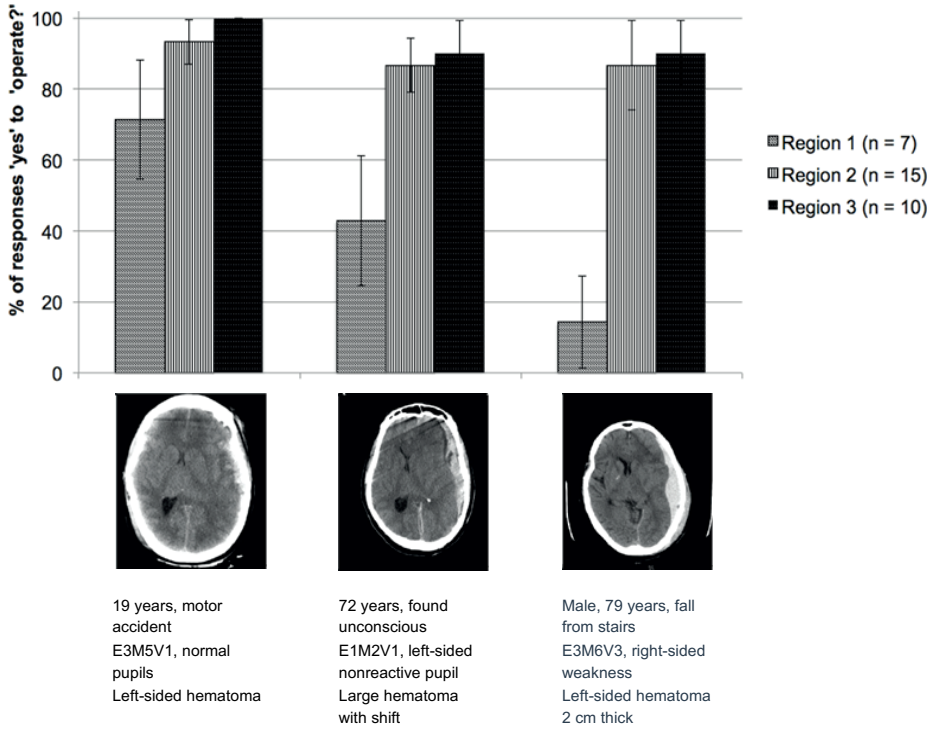

Figure 1. Regional variation with respect to operate or not in ASDH
Bar graph with percentages 'yes'-answers to the question whether surgery is indicated for these hypothetical clinical cases, stratified in three regions. Adapted from published paper in Chapter 4 and presented at the '11th Symposium of the International Neurotrauma Society' in Budapest, Hungary, march 2014.

- cases. Another explanation might be that this absence of clinical equipoise among the neurosurgical community implies a practice sometimes justified by the unique nature of surgery, an idea referred to as (neuro)surgical exceptionalism, which renders commonly used research methods from other medical disciplines unsuitable.[12] Anyway, starting an RCT would clearly not have been worthwhile, at least at the time in 2016.

Instead, I explored study designs that could reach the same level of validity. The idea was to setup a study that was specifically designed to exploit the existing variation. The resulting natural experiment mimics the counterfactual ideal of an RCT.[13,14] The methodological framework is called an IV analysis. This method originates from econometrics. An IV is a variable that is correlated to the treatment, but not to outcome. My hypothesis was that regional treatment preference is a valid IV.

In **Chapter 6**, we estimated, that, given the non-randomized data, an IV analysis is the superior approach because of its ability to account for unmeasured confounding but needs relatively large sample sizes to obtain reliable estimates. In addition, I

concluded that alternative methods should be used simultaneously to strengthen the credibility of effect estimation. Both conclusions are in line with an abundance of literature in other fields.[13,15-17]

In **Chapter 7** I suggested this approach in a letter to the authors of an observational study investigating the effect of surgical evacuation for another – but related – condition, the spontaneous intracerebellar hematoma.[18] Intracerebellar hematoma is pathophysiologically and anatomically different but the decision whether to operate or not represents a similar decision leading to similar considerations regarding potential study designs. I felt that the findings, based on propensity score matching, might be spurious due to (unmeasured) confounding. Instead I proposed a center-preference based IV because the cohort stemmed from 64 centers, with likely differing practice culture among institutions. In their response the authors indicated that they performed this analysis and that it led to similar results as with their original method.[19] They showed a considerable increase of the point estimate, albeit with a large confidence interval (including 1; no effect). Therefore, I contest their conclusion that the new IV analysis supports the robustness of the analysis. Instead, I suggest that it might point to an overall beneficial effect for early surgery.

After this intermezzo on intracerebellar hematoma I shift the focus back to traumatic ASDH. I collected data in the centers with the most divergent treatment preference from our national survey (region 1 and 3 in Figure 1) to compare patient characteristics, treatments and outcomes between them (**Chapter 9**). I show that indeed the strategy of regional data collection and analyses resulted in a comparable baseline prognosis. Despite this comparability, patients were more likely to be operated in the center a-priori preferring acute surgery. It appeared that the region preferring surgery actually had a better outcome. However, the small sample, residual confounding on center level with differences in referral patterns, and the inability to sufficiently take immortal time bias into account were major limitations.[20] Further, generalization to clinic should take into account the relatively high age of 68 (median). So, we would need larger cohorts with more detailed data collection that allowed for more rigorous adjustment and extrapolation.

The preceding studies shaped the methodology of my studies performed in the international Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study, and the Dutch Neurotraumatology Quality Registry (Net-QuRe) initiative.[21,22] In the protocol I described the methodological choices (**Chapter 8**). Specifically, immortal time bias has been addressed through the design, in which I defined treatment groups after the first CT scan (showing the acute subdural hematoma), thereby aligning the start of the follow-up with treatment assignment. As a primary analysis an IV analysis was chosen to account for the confounding by indication.

The thesis culminates in the comparative effectiveness study (**Chapter 10**). The widely differing surgical practices coming from centers that on average treat similar patients enabled exploration of effectiveness of the immediate treatment strategies in IV analysis. I showed that for most patients with an ASDH conservative treatment is preferable. This conclusion has gone through ample evaluation to verify the assumptions but challenged the neurosurgical status quo.[40,41] In **Chapter 11**, an 'author's reply' to a letter to the editor, I respond to some of these opposing arguments, discuss concerns that might have arisen among neurosurgeons globally and emphasize the results of our effectiveness study of acute surgery in ASDH. I will further discuss the validity and generalizability in the next section of this discussion. Uncertainty also exists on the benefits of a DC, a procedure that aims to mitigate effects of raised intracranial pressure by removing part of the skull. DC is frequently performed in combination with the evacuation of focal lesions, but is associated with severe complications and its additional benefits are only mechanistically clear. DC can be performed as primary procedure (combined with evacuation of a hematoma) or as secondary procedure (to decompress the brain). In Europe and Australia, inferred from the CENTER-TBI and the harmonized Australian OzENTER cohort studies, DC was performed in 320 patients, and in most conducted as primary procedure (258/320; 81%).[23] In other parts of the world, DC is performed even more frequently: In the China CENTER-TBI registry, DC was performed in 48% of patients with severe TBI (1354/2804), mostly as a primary procedure.[24] Previous trials have investigated the effectiveness of DC as secondary procedure, but until recently no evidence was available to support the use of DC as primary procedure. The advantage of a DC is that it more effectively controls ICP elevation, which could lead to secondary brain injury and poor clinical outcome. However, DC necessitates an additional operation (cranioplasty) to reconstruct the cranium and the associated risk of complications such as infection, cerebral edema and bone-flap reabsorption.[25] Literature on the effect of primary DC versus craniotomy to treat ASDH is inconsistent, with mostly worse outcomes for primary DC.[1,3,6,7,21] In a similar analysis as our preceding study on acute surgery vs conservative treatment in ASDH, I showed that there is no difference in outcome between a primary DC vs craniotomy. Hence, a craniotomy, a less invasive procedure, is preferred (**Chapter 12**). An important limitation is the small sample size of the study (although still the largest to date), resulting in a wide confidence interval possibly obscuring a clinically relevant effect.

Challenges around analysis and interpretation of observational studies on acute neurosurgery

In the process of interpreting, reporting and publishing the results of our main effectiveness study (Chapter 10), extensive discussions took place regarding

the (causal) interpretation. I have adhered to several guidelines for the conduct and reporting of IV analysis (Table 2),[13,15,26,27] but these discussions highlight the challenges in disseminating results from non-randomized interventions studies

Confounding

An IV analysis is preferable in a setting were strong confounding is expected because of its ability to control for unmeasured confounding. I performed preparatory methodological studies and showed the same holds true for our case study of acute neurosurgical decisions in TBI.[28,29] However, IV also relies on strong assumptions itself, which are not always met. Our IV, surgery treatment rate as a proxy for surgeon's preference, fulfils these assumptions; the centers' treatment preference predicts the likelihood to get acute surgery and the preference does not have an effect on outcome other than through surgery (Chapter 7 and 10).[30,31] Furthermore, I found similar effects in propensity score matching and multivariable regression adjustment. The consensus is when multiple methods give similar results, like in our study, it tends to strengthen the credibility of the findings. It may be argued that confounding is fully controlled.[16,17]

An instrumental variable is a variable in nonexperimental data that can be thought to mimic the coin toss in a randomized trial.
If an appropriate and valid instrument is found, then the effects of measured and unmeasured confounding can be mitigated.
An IV analysis always has an experimental analog, however absurd the experiment sounds. The IV analysis is therefore based on "a natural experiment."
Assumption (1): The IV must predict treatment but that prediction does not have to be perfect. An IV that does a poor job of prediction is said to be weak.
Assumption (2): A valid IV will not be directly related to outcome, except through the effect of the treatment.
Assumption (3): A valid IV will also not be related to outcome through either measured or unmeasured paths.
In a randomized trial, the assumptions are met by design in the act of randomization. In an IV analysis, these assumptions must be empirically checked to the extent possible or assumed based on context and subject matter knowledge.
In cases of treatment-effect homogeneity, IV studies estimate the effect on the marginal subject, the average treatment effect for patients whose treatment was determined by the instrument

Abbreviation: IV, instrumental variable.

Table 2. Key points about instrumental variable analysis
Adapted from [13]

However, the analysis through center-specific surgery rates might have induced new confounding on center level. The question is whether the effect of specific neurosurgical treatment choices/policies on outcome can be based on between-center variation without being residually confounded by other associated center characteristics. Unfortunately, I cannot exclude that other (unmeasured) regional differences might have distorted the findings,[13] but I tried to make reasonable that this confounding is minimal. I have accounted for these center differences through various ways. The primary analysis, with the IV as treatment variable and GOSE as dependent outcome variable in regression modelling, also included a random effects term for center, thereby adjusting for center clustering of other care processes or hospital structures. These random-effects accounts for other between-center differences than the factors included in the model. I calculated the post-hoc intraclass correlation coefficient (ICC) to be 0.018, so around 2% of the variance in outcome is explained by center. This confirms the small contribution of other center differences than acute surgery to outcome. Furthermore, another CENTER-TBI study showed that specific treatment choices did not correlate with other treatment choices within the centers.[32] This is in line with what would be expected, because treatment decisions are made by different medical specialties.

Monotonicity

Center preference relates to the policy in a specific center, e.g. a more aggressive or a more conservative approach to surgical indications. The decision in individual patients is primarily based on patient symptomatology, but is of course influenced by the centers' treatment preference. As mentioned, an IV analysis relies on strong assumptions. One of these is whether the center preference is actually employed in the cohort in which we estimate the treatment effect. There can be a discrepancy between the center preference and what treatment is actually initiated. In RCT terms, what is the compliance to the center policy?

This is a relevant methodological assumption in IV analysis, commonly referred to as the monotonicity assumption.[31] Monotonicity in our case refers to consequent decision-making of the neurosurgeon: a patient treated by a neurosurgeon that prefers acute surgery, 'should' predominantly treat with acute surgery. When patients do not 'comply' to the surgeon's preference or the preference is not consequently employed, the monotonicity assumption does not hold. In other words, true monotonicity requires an absence of defiers. To determine the possible discrepancy of the center preference versus the actual individual patient decision we need to know whether a patient has been treated according to the preference of the surgeon (or the center). However, surgeon's preference is not measured per patient, and, even more elusive, preference in itself might not be measurable. In an RCT it is clear when a study

patient did not receive the treatment that was determined by the randomization. Thus, the monotonicity assumption in an IV analysis cannot be tested and is normally not empirically verified.[26]

However, I tried to examine this assumption in our study. We compared predefined preferences towards immediate ASDH treatment (measured with a questionnaire) to actual surgery rates and found these to correspond well. Thus, the stated preferences directly mapped to patient treatment, i.e. we are comparing consistent preferences. We also have some other indirect measures from which we can deduce consistency. For example, the majority of neurosurgeons (at least in our study) confirmed the statement that structural practice variation for our research question exists (Chapter 4 and 5).

Generalizability and treatment effect heterogeneity

Generalizability boils down to the question to whom the effect estimate applies. In epidemiological terms: what is the "marginal patient" in this IV analysis (Table 2)? Similar to other study designs (including randomized clinical trials), in an IV analysis with a proxy for physician's preference as IV, the effect estimate also arises from the full study population, with individuals contributing with unknown weights.[30,31] In IV this means that the treatment effect stems from those patients in whom practice variation is present. We found that practice variation is present even in patients with an extreme prognosis. So, the effect estimate may well apply to the full population.

An important nuance is that the effect estimate pertains less to the patients with an extreme prognosis - because they might either never be treated or always be treated – but the precise cutoffs as to what entails an extreme prognosis cannot be well defined. I repeated the main analysis in a group of patients excluding the extreme patients at either end of the treatment spectrum (patients with one or more unreactive pupil, or who are GCS 15), and found a similar effect estimate. We, therefore, conclude that the treatment effect does apply to the total study population. And importantly, the treatment effect is robust for a middle group, which, although ill-defined, probably contributes the most to the effect estimate.

Another related aspect is that the treatment effect may differ across patients. Whether we can apply the estimate to the full study population, also depends on whether we can expect homogeneity of treatment effects, in which case, like in an RCT, the estimate effect is valid among the full study population, i.e. the estimand is the average treatment effect (ATE). We do have circumstantial evidence of heterogeneity of acute surgery. Comatose patients with a large ASDH greatly benefit because their risk of dying is greatly reduced (Chapter 3). Our retrospective study (Chapter 9) gives a suggestion of benefit in patients older than 65 years. And subgroup analyses of the main study (Chapter 10) suggest some treatment effect heterogeneity: older patients

and patients with large hematoma might benefit (Chapter 10). Nevertheless, without a convincing suggestion of treatment effect heterogeneity, the overall treatment effect is interpreted as being applicable to all patients in the study, as in other study designs.

On the other hand, when we do assume treatment effect heterogeneity, the generalizability becomes more abstract. We have to rely more on whether the application of the surgical preference is consistent across all (types of) patients; i.e. on monotonicity. In that case the estimand becomes the local average treatment effect (LATE); the treatment effect is applicable to those patients of which the treatment has been determined by the IV.[33] Monotonicity complicates matters because it is unclear whether a patient is treated under the preference that is similar to that of the instrumental variable. This inability to define who is a complier or a defier makes inferring to whom the effect estimate applies indeterminate. Thus, the inherent difficulty to define the patient population in IV effect estimation precludes recognizing the exact average treatment effect.

Importantly, patient characteristics always play a role when deciding on surgical treatment and knowledge of clinical practice suggests that for some patients even a (valid) IV plays no role in determining treatment assignment; they always or never will be treated. And just as clinicians use caution in generalizing findings from randomized trials, clinicians interpreting our study should use caution in determining whether the IV estimate applies to their patient. Treatments can have different effects in different patients. Positive and negative effects might even cancel each other.[34] IV analysis is even more complicated to interpret. A valid request might be a more extensive evaluation of the treatment heterogeneity than provided until now. These studies have not been performed yet and will be challenging since they require very large sample sizes.

We conclude that the effect estimate applies to those patients for which the neurosurgeon may hypothetically be in clinical equipoise, or in other words for which there seems to be more than one valid treatment option.

Limitations

Terminology

The presented studies are observational and inherently prone for bias. Reviewers have often argued that it would be more appropriate to use an associative denotation to describe our results ('association' instead 'does not lead to' for example). On most occasions we accepted and resigned ourselves.

We faced a considerable challenge in formulating the conclusions, finding the appropriate balance between describing the findings and causally inferring our association estimates. However, though the observational nature of our studies

certainly imposes difficult methodological challenges to infer causality, that does has not influence the initial aim of our research: to study causal effects. The use of the word 'association' in intervention research has no benefits and only confuses.[35] Or as causal inference researcher Miguel Hernan put it: "If your thinking clearly separates association from causation, make sure your writing does too."[36]

Power

For the power calculation of the IV analysis we assumed an odds ratio of 0.6, which is a fairly large treatment effect. We based this treatment effect on the available evidence and on a pilot study (Chapter 8).[28-32] We actually employed a smaller effect size than our pilot study showed (Chapter 9) to account for the uncertainty due to its small sample and the possibility residual confounding. The assumptions for the power calculation may, nevertheless, raise the question whether the absence of an effect of acute surgery or DC may be explained by an insufficient sample size to exclude small but clinically meaningful effects.

I consider that unlikely for acute surgery versus conservative treatment. All the analyses show robust odds ratios close to 1. We reported the uncertainty in these estimates through confidence intervals; not by claiming non-significance in the p-values. For DC versus craniotomy it might be the case that a smaller effect might have gone unnoticed since the confidence interval is wide (i.e. 0.5-1.5 for the original primary analysis excluding centers < 15 patients). Nevertheless, I analyzed the data since it might benefit future meta-analysis. Whereas sample size calculations are worthwhile in deciding whether a prospective study (e.g. RCT) can realistically be achieved in terms of ethics and logistics, not analyzing already collected data is not a rational strategy.[37]

Obviously, larger sample sizes are preferable to reduce statistical uncertainty. However, the current results are highly relevant for clinical practice. Practically speaking, CENTER-TBI is one of the largest studies in TBI performed so far and it is unlikely that in the near future larger studies will address these research questions.

Examining treatment preference

Center as a surrogate marker for treatment preference is an assumption of our IV analysis. Comparing the international survey to the case-mix adjusted surgery rates did allow falsifying the consistency of the treatment preferences. However, the treatment preference could have been explored more in detail. A solution could have been to present the hypothetical cases of the Dutch/Belgian survey also to neurosurgeons in Europe. Another alternative for defining the instrument would have been to determine how centers and neurosurgeons within each center treated patients with ASDH the year before patient inclusion. Such external instruments are

different from the center's case-mix adjusted surgery rate, which is an instrument defined internally, i.e. through the same data in which we estimate the effect.
Nevertheless, we estimated that both validation strategies of treatment preference, hypothetical cases to every neurosurgeon of all 65 participating centers and collection of prior treatment decisions, would not work for pragmatic reasons. Specifically, the number of questions already required a lot from the respondents. Increasing the survey would have risked noncompliance leading to missing data. The quantification of prior treatment strategies was also considered to be too time intensive.

Exploration of monotonicity

An IV study fulfilling all assumptions with homogeneous treatment effects results in an ATE estimand. However, when there is treatment effect heterogeneity, the estimand due to the monotonicity assumption becomes the LATE. External instruments as the primary IV have an advantage over defining the instrument internally; it is easier to determine the deviation from monotonicity by allowing to quantify the proportion of compliers and defiers and the impact on the effect estimate.[33,38]
Of note, the degree to which results apply to a specific patient depends not only on the proportion of the study population that consisted of these specific patients, but also on the strength of the instrument in this group. The stronger the instrument, the higher the relative contribution to the estimate. When interpreting heterogenous treatment effects in the setting of IV analysis we assume so-called stochastic monotonicity with the estimand being the strength-of-IV weighted treatment effect.[38]

Data quality and information bias

The granularity of the datasets allowed for our specific study design and analysis but also resulted in an overwhelming amount of data. Harmonization and data curation took a considerable amount of time and resources.[39] In this process arguably arbitrary decisions were made on many specific variables and data points. To check the robustness of the results for these decisions many sensitivity analyses had to be performed. Finally, the radiological data were also analyzed centrally by a dedicated radiological team, based on the National Institute of Neurological Disorders and Stroke TBI Common Data Elements,[40] which led to an enriched dataset but also to discrepancies between local and central data. Fortunately, there was good agreement for both diagnoses 'ASDH' and 'contusion/ICH' between the local (treating physicians) and central readings respectively, making misclassification bias unlikely.[41] In general, all aforementioned points in this paragraph should have been accommodated (more) in the protocol, although some could not have been anticipated.

Implications

The responses of the neurosurgical community to the comparative effectiveness study of surgery versus conservative treatment have varied (Chapter 11).[40,41] Some contest the absence of effect; after all, ever since the beginning of the previous century acute surgical evacuation for ASDH was considered common sense. Others said that it merely confirms what we already know from daily practice. The following points can be learned from our studies.

First, it will directly impact clinical practice. Our findings in part support the continued use of interventions already routinely applied in daily practice; surgical evacuation of acute subdural hematoma remains the cornerstone of treatment in life-threatening neurological deterioration. The clear effect from the meta-analysis with dramatic effect might also imply that we should operate comatose patients with a poor prognosis more often. However, this decision requires objectifying salvageability, and thus prognostication which is not easy.[42-44] On the other hand, the comparative effectiveness study in Chapter 10 shows that in certain surgical cases conservative treatment could have been employed. The 'aggressive' centers could safely treat their patients more often conservatively

Second, the inferences of the effects rely on strong evidence, derived from the largest cohorts of ASDH. This thesis is an important addition to the current low-quality literature in ASDH. The absence of a positive or negative association with outcome overall is important knowledge for all clinicians treating patients with ASDH. The findings regarding type of surgery (DC versus craniotomy) for ASDH are compatible with the current best evidence.[45] It is the largest study to date, and provides robust guidance for neurosurgeons.

Third, the wide treatment variation, from all across Europe, and an absence of benefit of either treatment preference is important knowledge for neurosurgeons treating patients with ASDH. Neurosurgeons are often 'traditional' physicians who, mostly due to the lack of appropriate evidence, understandably use clinical experience and mechanistic reasoning.[46] This mindset of 'eminence-based medicine' trained over the years and consequently passed on to the next generation, leads to a culture of having no doubts about the 'best' treatment for a specific patient, despite the lack of evidence. This intuitive decision-making under uncertainty and in an emergency setting, leads to biases that might not be recognized by the decision maker.[47,48] We hope to contribute to the neurosurgeon to become aware of these cognitive fallacies.

Fourth, this thesis presents an illustration of medical specialists taking the initiative to judiciously reflect on their practice. We live in a time of 'do-not-do lists'.[49] The lack of evidence might be interpreted unjustly as a lack of effectiveness, not in the least by stakeholders such as health care insurance companies and policy makers. With

studies like those presented in this thesis we take matters in our own hands and contribute to evidence-based decision-making on policy level.
Thus, this thesis reflects a systematic approach to evaluate common neurosurgical interventions, using high quality data and analytical approaches, and delivered results that can directly impact clinical practice.

Recommendations for clinical practice

Our studies assess effectiveness of common interventions in everyday practice in relatively unselected populations and under flexible conditions, in order to inform decisions about practice in a "real-world" setting. The inherent consequence of the IV analysis is that it is not immediately clear to what patient population the results pertain. I acknowledge this difficult interpretation and, therefore, propose specific recommendations for clinical practice. I will describe how neurosurgical practice could be affected with several cases of ASDH and propose an update of the Brain Trauma Foundation (BTF) guideline.

Applicable patient populations

The practice variation mostly affects the patients who are at neither of the extremes in terms of clinical presentation. Because an identical patient may be operated upon in one center but not in another, it naturally follows that there is more than one valid treatment option. The varying center policies thus lead to a group of patients for whom we infer clinical equipoise. The middle group is our study population (Figure 2). Because center policies are not measured per patient, it is not immediately clear what patient population the equipoise actually pertains to. However, practically, the physician in charge of the decision should make an estimation whether he/she thinks the other treatment could be a valid alternative. If not, it could be argued that the effect estimate does not apply.
There is something inherently contradictory in this line of reasoning; the lack of critical appraisal required to employ our study' conclusion is exactly what led to led to the practice variation in the first place. I expect, however, that the confrontation with this strong and consistent practice variation offers physicians the confidence to apply this kind of scrutinous, abstract thinking. I adhere hereby to the definition of clinical equipoise as the 'genuine uncertainty within the expert medical community-not necessarily on the part of the individual investigator – about the preferred treatment'.[50]

Evidence-based guideline

The BTF guideline is the most widely applied guideline in TBI.[51,52] The last version, published in 2016 was focused on severe TBI. In contrary to the previous version

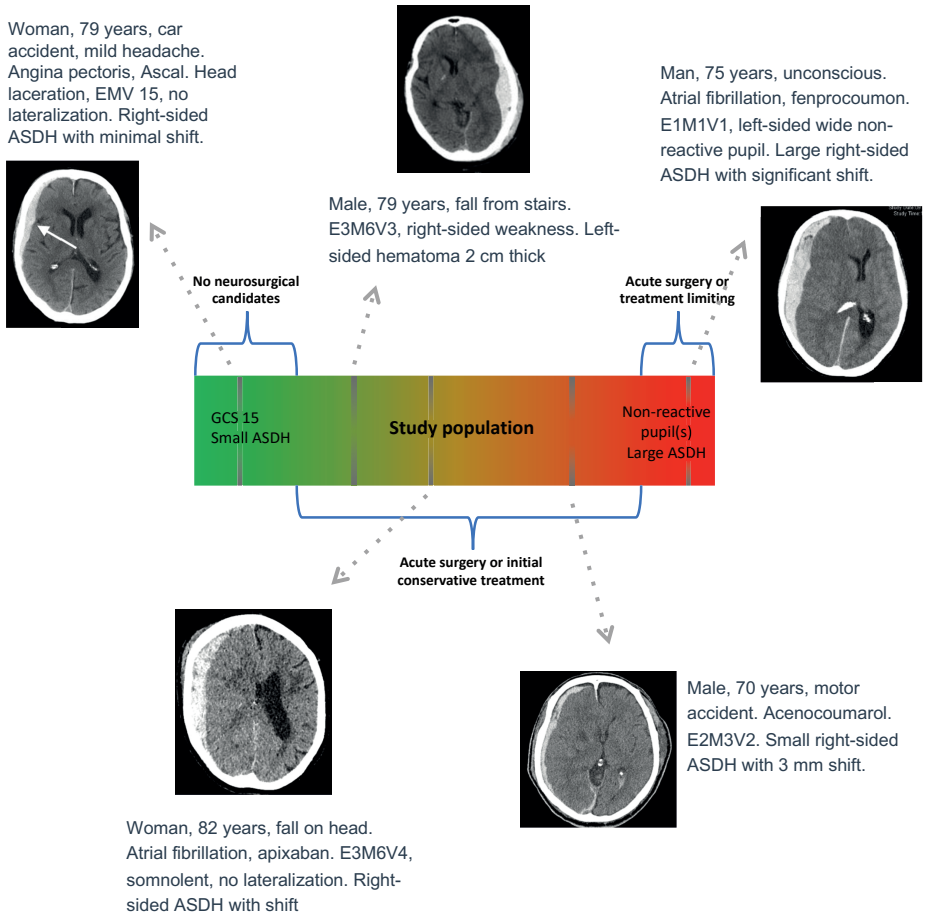

Figure 2. Acute subdural hematoma cases depicted on a scale representing the clinical spectrum with varying preference for acute surgery or (initial) conservative treatment

The right - red - end, represents patients with poor clinical characteristics for whom clinicians may perform acute (rescue) surgery to avoid death or choose for comfort measures in a treatment-limiting setting. At the left – green – end, all patients will not be operated upon due to an anticipated good prognosis without need for acute evacuation. The window between these extremes represents our study population. The exact cut-offs cannot be objectively characterized because the window is applicable to those patients for whom the neurosurgeon may hypothetically be in clinical equipoise.

of 2006, this latest version did not cover the surgical management for intracranial hematomas, probably due to the fact that no higher quality evidence has been generated since 2006.[53] We propose an extension of the latest guideline (i.e. of 2006) based on the studies of this thesis.

The underlying evidence base needs to be responsibly translated into clinically applicable, accurately graded recommendations in order to help clinicians properly

treat patients with TBI. We adhere to the methodology of the BTF on how to formulate treatment recommendations (Panel).[54] The BTF has formulated an hierarchy in evidence quality, summarized in three classes. These classes translate to recommendations while considering the methodological principles consistency, directness and precision (Panel; legend Table). The BTF guideline for the indication for surgery in ASDH can both be extended with Level IIA recommendations based on level 2 evidence. Some of the 2006 recommendations have been retained and now labeled Level III recommendations.

The BTF guideline working group in 2006 already recognized the special conditions that apply to the efficacy of acute surgical evacuation in extra-axial lesions. Specifically, a neurological deteriorating patient with a large epidural hematoma should always be surgically treated immediately based on common clinical sense while there is very little evidence to support this recommendation. In an effort to integrate obvious effects with evidence based medicine, Glasziou et al. coined the term 'dramatic effect' and explored methods to infer causality of several interventions.[55] Inspired by this approach, I evaluated the effect of surgery in ASDH for comatose patients in a meta-analysis using a historical control group (Chapter 3). The Oxford Centre for Evidence Based Medicine has proposed to upgrade observational studies with a dramatic effect to a higher evidence level.[56]

Our comparative studies (Chapter 3, 9, 10 and 12) are higher quality studies whose findings supersede those of the lesser quality investigations on which the previous recommendations for ASDH were based.

With regard to Chapter 9 and 10: though both studied acute surgery for the treatment of ASDH, a key difference is the median age difference of 68 in our Dutch study and 54 years in our European study. Indeed, both studies enrolled TBI patients consecutively but apparently in The Netherlands the average age is considerably higher. We adhere to these difference patient domains in constructing the recommendations.

Thus, the BTF guideline for the indication for surgery can be extended with level 2 recommendations. The Level IIA recommendation is based on our meta-analysis (Chapter 3) and the observational comparative effectiveness study (Chapter 10). We feel that the comparative observational study on the effectiveness of primary DC (Chapter 12) does not allow for Level IIA conclusions given the limited sample size. Therefore, we distill a Level II B recommendation.

Furthermore, our research does not support evidence-based recommendations for other surgical techniques or perioperative interventions. As the studies were observational, we recommend neurosurgeons to perform the surgical procedure according to their regular practice.

The Level III recommendations were retained from the previous guideline in 2006 and supplemented with the observational comparative effectiveness study in Chapter

9.[57] These recommendations are based on consistent effects shown in subgroups and class 3 cohort studies.

Panel: Guidelines for the Surgical management of Traumatic Brain Injury: Proposal for an Update of the Acute Subdural Hematoma Recommendations

A guideline is typically reached through consensus among a group of clinicians that are experts on the subject matter. The following paragraph should be read as a first draft for such a process. I like to emphasize therefore that this is a proposal.

RECOMMENDATIONS

Level I

• There was insufficient evidence to support a Level I recommendation for this topic.

Level II A

• Acute surgical evacuation is recommended to greatly reduce mortality among salvageable comatose patients (severe TBI).

• Acute surgical evacuation is not recommended to improve outcomes as measured by the Glasgow Outcome Scale–Extended (GOS-E) score at 6 months post-injury when considering patients with all TBI severities for which the neurosurgeon sees no clear superiority of either treatment

Level II B

• Primary DC over craniotomy in acute surgery is not recommended to improve GOS-E at 6-months.

Level III

• Acute surgical evacuation is recommended for large ASDHs, regardless of the patient's Glasgow Coma Scale score

• Acute surgical evacuation is recommended for salvageable patients older than 65

EVALUATION OF THE EVIDENCE

Quality of the Body of Evidence

With regard to whether to operate or not, three studies (Chapter 3, 9 and 10) are qualified as Class 2 evidence. The meta-analysis (Chapter 3) used a dramatic effect design. The effect was dramatic enough that confounding and immortal time bias could be sufficiently ruled out. Therefore, it was considered a Class 2 study with moderate quality of the evidence. The observational study in Chapter 9 addressing acute surgery versus conservative treatment was a retrospective comparative study comparing two centers, one with a preference for surgery and the other with a preference for conservative treatment. Residual confounding could not be sufficiently ruled out. It is considered a low-quality study with low precision and replication (in the older patient population) is advised.

For the study of Chapter 10, the quality of the evidence was moderate, because it was a multicenter, non-randomized prospective study with a very large sample size (n=1407) with multiple analyses to address confounding.

The observational study that compared primary DC to craniotomy (Chapter 12) was rated Class 2 but this study was low quality; it was a single study, with low precision, and replication is needed for higher confidence (Table).

Table. Quality of the body of evidence (acute surgery for ASDH)

Components of Overall Quality – Class 2								
Topic	**Number of Studies**	**Meta-Analysis**	**Number of Subjects**	**Class of Studies***	**Consistency (High, Moderate, Low) ****	**Directness (Direct or indirect) *****	**Precision (High, Moderate, Low) *******	**Quality of Evidence (High, Moderate, Low, or Insufficient)**
Acute surgery vs. initial conservative treatment	3 observational	1	- 12,287 (meta-analysis) - 195 - 1,407	2	Moderate	Direct	High	Moderate
Primary DC vs. craniotomy	1 observational	NA	336	2	NA	Direct	Low	Low

** Class 1 Evidence is derived from randomized controlled trials. However, some may be poorly designed, lack sufficient patient numbers, or suffer from other methodological inadequacies that render them Class 2 or 3.*
Class 2 Evidence is derived from cohort studies including prospective, retrospective, and case- control. Comparison of two or more groups must be clearly distinguished. Class 2 evidence may also be derived from flawed RCTs.
Class 3 Evidence is derived from case series, databases or registries, case reports, and expert opinion. Class 3 evidence may also be derived from flawed RCTs, cohort, or case-control studies.
*** Consistency: Consistency is the extent to which the results and conclusions are similar across studies. It is rated High (all are similar), Moderate (most are similar), Low (no one conclusion is more frequent). It is NA (not applicable) when the body of evidence consists of a single study.*
**** Directness: Directness can have different definitions. We define it as whether the study population is the same as the population of interest and whether the study includes clinical rather than intermediate outcomes. Indirect is noted if the population differs; for example if the study includes both moderate and severe TBI or patients with stroke or TBI and does not separate the results by these population characteristics, or if the outcomes are not mortality or neurological function. As outlined in Methods, indirect evidence was only included if no direct evidence was found.*
***** Precision: Precision is the degree of certainty surrounding the effect estimate for a given outcome. Precision is rated as High, Moderate, and Low. How this is determined depends on the type of analysis used in a specific study but may include consideration of the range of confidence intervals or the significance level of p-values.*

Applicability
The large comparative observational study comparing acute surgery to conservative treatment took place in 65 European centers over a 3-year period, and included 1407 patients across all TBI severities. The study comparing primary DC to initial conservative management was conducted in 65 countries over a 3-year period, and included 336 patients. Both studies are considered applicable to current practice.
The second, smaller, study comparing acute surgery to conservative treatment was performed in three centers in The Netherlands. Although including all consecutive patients, the average age was 68 years which is older than expected in current practice. The conclusion of benefit of surgery in this elderly population, is in line with the subgroup analysis of the other, larger, observational study. Furthermore, in the previous version of the guideline, surgery was recommended in ASDH with thickness > 1 cm or midline shift > 0.5 cm regardless of GCS, based on Class 3 evidence. This is also in line with the suggestion of benefit in these subgroups in the larger observational study. Treatment effect heterogeneity could not formally be determined through subgroup analyses, but large hematomas showed to be associated with a strong, albeit nonsignificant, effect estimate (OR 2.7 95% CI 0.9 - 8.3). Therefore, this recommendation is retained as a Level III recommendation.

Future research

Bryan Jennett (1926-2008) knew that it was a privilege when he was asked by his chief if he would like to pursue a research project next to his residency neurosurgery. Talented residents were offered scientific research, at the time mostly fundamental research in the neurophysiology laboratory. His chief asked whether he would want to investigate how the physiological change of the reflexes in patients with a slow acting thyroid comes about. Although appreciating the honor, he answered decisively that he could not care less about the reflexes in thyroid disease. The chief reacted surprised: "Don't you want to invest in research?". "Yes" answered Jennett "but merely into clinical dilemma's I run into daily." His chief had no clue what Jennett was suggesting. Research on clinical problems? Is that scientific research at all? Jennett explained that he noticed major variability among his supervisors on which patient after a head injury to admit to the hospital. That kind of questions he would like to research.[58]

We have come a long way since then. All neurosurgical departments heavily invest in clinical research. However, still, neurosurgery has a weak evidence base.[52,59,60] Clinical decisions in neurosurgery have a history of strongly relying on personal or departmental dogma, often deduced from fundamental and basic science, and continue to do so.[61] The first RCTs addressing common surgical procedures in TBI have been conducted only the last decennium.[52,62,63] Since the famous 1996 editorial in the *British Medical Journal* on evidence-based medicine by Sackett et al.,[64] another editorial in 2008 addressed this issue for neurosurgery: "Twenty-two years later, the impact of evidence-based medicine in British neurosurgical practice has been variable and the management of some common neurosurgical conditions has not been subjected to the same scrutiny as some less common conditions. The National Health Service has tended to look more critically at expensive procedures and technology, while ignoring conditions that are dealt with in everyday practice." [65] Naturally this does not only pertain to the United Kingdom.

High-quality research begins with recognizing relevant clinical questions. Research within neurosurgery should be done by neurosurgeons because they have the unique position to appreciate the management decisions and recognize clinical relevance. There is otherwise nobody else who will investigate the clinical practice important for patients with neurosurgical diseases. Obvious constraints that explain the lagging behind is the limit in dedicated time of neurosurgeons due to the drain of clinical work and the operating theatre. Furthermore, neurosurgeons by training - and possible also by character - are highly trained decision makers that are not uncertain about their treatment preference. They might be less amenable for critical appraisal on their own actions, which is of course pivotal to clinical research. For example, the basic premise for RCTs, clinical equipoise is still often misinterpreted as doubt

or uncertainty, which are terms neurosurgeons tend to avoid in their decisions for TBI patients.[50,66] Acknowledging clinical equipoise, especially in acute settings, might be unconsciously be mistaken for uncertain or even insecure decision making. However, clinical equipoise is not at odds with treatment preferences of the individual investigators but rather reflects the collective will of a body of expert clinicians.[50] Naturally, this perceived absence of equipoise might be an important barrier for conducting RCTs. In fact, the STITCH-trauma trial that formalized the required equipoise in a inclusion criterium might attest to this problem because it was prematurely halted due to insufficient patient recruitment.[67]

There are other challenges in surgical TBI research. Case series and single treatment group cohorts are generally not suited to infer causality and there are plenty enough for common neurosurgical interventions.[61] Comparative studies especially RCTs, however, are difficult to conduct. The challenges mainly are the heterogeneity of the populations, the traditionally eminence-based neurosurgical culture, inadequate research budgets, and difficulties related to obtaining patient informed consent in an emergency situation.[12,68,69]

I propose several characteristics of future research. RCTs are preferred and should be as pragmatic as possible, directed by proper guidelines.[70] Most neurosurgical inventions have clear efficacy (they can work) but are yet to be clearly examined in terms of effectiveness (do they actually work in clinical reality?). When RCTs are difficult to conduct, observational studies with an IV analysis, when possible, should be considered to circumvent the substantial confounding by indication of acute neurosurgical interventions. Part of the solution for better neurosurgical research may be the increase of methodological knowledge among neurosurgeons. While RCTs are relatively straightforward methodologically with the major challenge being logistics and infrastructure, causal inference in observational (big) data is much harder.[71,72] Observational studies should not be considered inferior to RCTs per se.[73,74] Neurosurgeons might benefit from a strong, durable cooperation with epidemiologists and statisticians that are dedicated to the specificities of neurosurgical research.

Moreover, TBI researchers should be aware of the available informed consent alternatives to optimize patient recruitment.[68] Furthermore, I support the early assessment of new neurosurgical interventions, so that their effectiveness can be established before wide adoption. Preferably, the IDEAL framework should guide evaluation of new interventions.[75] Finally, embedding TBI research into standard clinical practice could reduce expenses and lower the threshold for study participation. Concerted efforts should be undertaken by stakeholders to develop research agendas that prioritize clinical research and allocates the scarce grants accordingly. The limited public attention through policy and targeted grants sharply contrasts the major impact TBI has on society.[76,77]

Specific research questions

When it comes to acute neurosurgical evacuation in traumatic intracranial lesions, there are several research questions I consider sufficiently pressing to make a claim on further resources. First, it is clear these large-scale studies show that the studied interventions are not uniformly beneficial. An explanation could be the inherent heterogeneity of TBI with different responses to surgery and, therefore, a neutral effect due to the mixing of systematic positive and negative effects.[34] The subgroup analyses presented in this thesis were mostly exploratory. Therefore, targeted subgroup analysis should be performed to more closely determine who benefits from surgical treatment strategies. To reduce chance findings there should be a predefined consensus on logical subgroups. I advise individual participant data meta-analysis of CENTER-TBI/Net-QuRe data and other large multicenter cohorts such as the Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) project.[78] Advanced risk-stratification tools, that stratifies treatment outcome on baseline risk instead of the "one-variable-at-a-time" subgroup analysis, should be used to estimate the heterogeneous treatment effects, which allows personalize management.[79] Naturally, pragmatic trials in specific, more homogenous subgroups serves the same goal of personalize neurosurgical care.

Second, primary DC warrants further investigation in a larger study. Our own analysis (Chapter 11) was imprecise. More definite evidence will come from the RESCUE-ASDH trial, a multi-center randomized trial comparing craniotomy against DC for patients undergoing evacuation of a traumatic ASDH. Last, the effects are being considered on a limited time-frame and outcome set. I would like to have performed a costing analysis alongside our comparative studies. An hypothesis is that - costly - aggressive acute or early surgical treatment strategy might induce earlier recovery and less burden on ICU care and post-acute care rehabilitation. Although we did collect indirect measures (such as the hospital length of stay) and we described costs in our own centers,[10] we did not collect data to conduct a formal economical evaluation. Thus, the effectiveness studies on ASDH should be followed by cost-effectiveness, safety and long-term outcome analyses.

Thus, pragmatic trials targeting specific subgroups should be an important focus for future surgical TBI research. We have, therefore, grasped the nettle and have proposed an RCT, namely the Randomized Evaluation of Surgery in Elderly with Traumatic Acute SubDural Hematoma (RESET-ASDH trial), a study protocol for a pragmatic RCT. The Randomized Evaluation of Surgery in Elderly with Traumatic Acute SubDural Hematoma (RESET-ASDH) trial is an international multicenter RCT on the (cost-)effectiveness of acute neurosurgical hematoma evacuation versus initial conservative treatment in elderly with an ASDH. The study is in line with

current epidemiological developments of the rising incidence of elderly with TBI.[80,81] The studies presented in this thesis have paved the way for this trial.

Conclusion

The uncertainty about the benefits of surgery inspired the research presented in this thesis, in which I aimed to assess the effectiveness of treatment strategies for ASDH. First, I showed that good quality evidence of acute surgery versus conservative treatment to underpin guidelines is lacking, although among comatose patients, acute surgery has a clear benefit, because it leads to a large mortality reduction. Second, I demonstrated strong and consistent practice variation with regard to the fundamental question to operate or not in ASDH. Third, I argue that the value of observational studies on neurosurgical intervention for acute hematomas for practice is limited as causal inference is precluded by strong confounding by indication. One of the few analytical approaches that might allow causal interpretation is IV analysis, which requires assumptions that are not easily met. Fourth, I showed that patients treated in centers that prefer acute surgery (over conservative treatment) have an equal outcome to patients treated in centers that prefer (initial) conservative treatment. Therefore, I conclude that acute surgical evacuation of an ASDH in patients for whom equipoise exists on surgical indication may not lead to a better functional outcome compared to a strategy favoring (initial) conservative treatment. Finally, with regard to surgical technique, primary DC might not lead to better outcomes as compared to craniotomy in acute surgery in ASDH.

In light of the absence of uniform benefit, the restrictions certain neurosurgeons place on surgery for ASDH and primary DC appear justified. In light of my findings, I propose guideline recommendations for patients with ASDH. Finally, targeted subgroup analysis, risk-based prediction of heterogeneous effects and pragmatic trials should further disentangle heterogeneous treatment effects. Such future studies should be performed by neurosurgeons in collaboration with clinical epidemiologists and should have well-defined and realistic objectives to directly inform clinical practice.

REFERENCES

1. Smith GCS, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003;327(7429):1459–61.
2. Compagnone C, Murray GD, Teasdale GM, et al. The Management of Patients with Intradural Post-Traumatic Mass Lesions: A Multicenter Survey of Current Approaches to Surgical Management in 729 Patients Coordinated by the European Brain Injury Consortium. *Neurosurgery* 2005; 1183–92.
3. Fountain DM, Kolias AG, Lecky FE, et al. Survival Trends After Surgery for Acute Subdural Hematoma in Adults Over a 20-year Period. *Annals of Surgery* 2017;265(3):590–6.
4. Stein SC, Georgoff P, Meghan S, et al. 150 Years of Treating Severe Traumatic Brain Injury: A Systematic Review of Progress in Mortality. *Journal of Neurotrauma* 2010;27(7):1343–53.
5. Cnossen MC, Polinder S, Lingsma HF, et al. Variation in Structure and Process of Care in Traumatic Brain Injury: Provider Profiles of European Neurotrauma Centers Participating in the CENTER-TBI Study. *PLoS ONE* 2016;11(8):e0161367.
6. Huijben JA, van der Jagt M, Cnossen MC, et al. Variation in blood transfusion and coagulation management in Traumatic Brain Injury at the Intensive Care Unit: A survey in 66 neurotrauma centers participating in the Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study. *Journal of Neurotrauma* 2017; neu.2017.5194.
7. Rayan N, Barnes S, Fleming N, et al. Barriers to compliance with evidence-based care in trauma. *J Trauma Acute Care Surg* 2012; 72(3):585–92–discussion592–3.
8. Kolias AG, Belli A, Li LM, et al. Primary decompressive craniectomy for acute subdural haematomas: results of an international survey. *Acta Neurochir* 2012;154(9):1563–5.
9. Bonds B, Dhanda A, Wade C, et al. Prognostication of Mortality and Long-Term Functional Outcomes Following Traumatic Brain Injury: Can We Do Better? *Journal of Neurotrauma* 2021;38(8):1168–76.
10. Van Dijck J, van Essen TA, Dijkman MD, et al. Functional and patient-reported outcome versus in-hospital costs after traumatic acute subdural hematoma (t-ASDH): a neurosurgical paradox? *Acta Neurochir* 2019;161(5):875–84.
11. Sackett DL, Haynes RB, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. *Lippincott Williams and Wilkins*; 1985.
12. George P. The struggle for better research in surgery. *The Lancet* 2016;387(10032):1970.
13. Rassen JA, Brookhart MA, Glynn RJ, et al. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Journal of Clinical Epidemiology* 2009; 62(12):1226–32.
14. Hernán M, Robins J. Causal Inference: What If. *Boca Raton: Chapman & Hall/CRC;* 2020.
15. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidem Drug Safe* 2010;19(6):537–54.
16. Laborde-Castérot H, Agrinier N, Thilly N. Performing both propensity score and instrumental variable analyses in observational studies often leads to discrepant results: a systematic review. *Journal of Clinical Epidemiology* 2015; 68(10):1232–40.
17. Wagenmakers E-J, Sarafoglou A, Aczel B. One statistical analysis must not rule them all. *Nature* 2022; 605(7910):423–5.
18. Kuramatsu JB, Biffi A, Gerner ST, et al. Association of Surgical Hematoma Evacuation vs Conservative Treatment With Functional Outcome in Patients With Cerebellar Intracerebral Hemorrhage. *JAMA* 2019; 322(14):1392–12.
19. Kuramatsu JB, Sheth KN, Huttner HB. Unmeasured Confounding in Observational Studies of Management of Cerebellar Intracranial Hemorrhage-Reply. *JAMA* 2020; 323(7):666.
20. Boef AGC, Dekkers OM, Vandenbroucke JP, le Cessie S. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *Journal of Clinical Epidemiology* 2014; 67(11):1258–64.
21. Maas AIR, Menon DK, Steyerberg EW, et al. Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI): a prospective longitudinal observational study. *Neurosurgery* 2015; 76(1):67–80.
22. Nederlands Trial Register [Internet]. Amsterdam: Academic Medical Center (The Netherlands). 2004 Oct 26. Identifier Trial NL5761; 2016 July 28. Available from: https://www.trialregister.nl/trial/5761.

23. Gantner D, Wiegers E, Bragge P, et al. Decompressive Craniectomy Practice following Traumatic Brain Injury in Comparison with Randomized Trials: Harmonized, Multi-Center Cohort Studies in Europe, the United Kingdom, and Australia. *Journal of Neurotrauma* 2022 Jun; 39(11-12):860-869.
24. Gao G, Wu X, Feng J, et al. Clinical characteristics and outcomes in patients with traumatic brain injury in China: a prospective, multicentre, longitudinal, observational study. *The Lancet Neurology* 2020;19(8):670–7.
25. Stiver SI. Complications of decompressive craniectomy for traumatic brain injury. *Neurosurgical FOCUS* 2009; 1–16.
26. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology* 2013;24(3):363–9.
27. Boef AGC, Dekkers OM, Le Cessie S, Vandenbroucke JP. Reporting instrumental variable analyses. *Epidemiology* 2013;24(6):937–8.
28. Cnossen MC, van Essen TA, Ceyisakar IE, et al. Adjusting for confounding by indication in observational studies: a case study in traumatic brain injury. *Clinical Epidemiology* 2018; 10:841–52.
29. Van Essen TA, Volovici V, Cnossen MC, et al. Comparative effectiveness of surgery in traumatic acute subdural and intracerebral haematoma: study protocol for a prospective observational study within CENTER-TBI and Net-QuRe. *BMJ Open* 2019;9(10):e033513–8.
30. Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. *Health Serv Res* 1998; 33(5 Pt 1):1337–60.
31. Swanson SA, Hernán MA. The challenging interpretation of instrumental variable estimates under monotonicity. *International Journal of Epidemiology* 2017; 47(4):1289–97.
32. Ceyisakar IE, Huijben JA, Maas AIR, et al. Can We Cluster ICU Treatment Strategies for Traumatic Brain Injury by Hospital Treatment Preferences? *Neurocrit Care* 2022; 36(3):846-856.
33. Swanson SA, Miller M, Robins JM, Hernán MA. Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology* 2015; 26(3):414–20.
34. Gregson BA, Mitchell P, Mendelow AD. Surgical Decision Making in Brain Hemorrhage. *Stroke* 2019; 50(5):1108–15.
35. Haber NA, Wieten SE, Rohrer JM, et al. Causal and Associational Language in Observational Health Research: A Systematic Evaluation. *American Journal of Epidemiology* 2022; 19;191(12):2084-2097.
36. Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data. *Am J Public Health* 2018;108(5):616–9.
37. Hernán MA. Causal analyses of existing databases: no power calculations required. *Journal of Clinical Epidemiology* 2022; 144:203–5.
38. Boef AGC, le Cessie S, Dekkers OM, et al. Physician's Prescribing Preference as an Instrumental Variable: Exploring Assumptions Using Survey Data. *Epidemiology* 2016; 27(2):276–83.
39. Maas AIR, Ercole A, De Keyser V, et al. Opportunities and Challenges in High-Quality Contemporary Data Collection in Traumatic Brain Injury: The CENTER-TBI Experience. *Neurocritical Care* 2022; 37(Suppl 2):192-201
40. Haacke EM, Duhaime A-C, Gean AD, et al. Common data elements in radiologic imaging of traumatic brain injury. *J Magn Reson Imaging* 2010;3 2(3):516–43.
41. Vande Vyvere T, Wilms G, Claes L, et al. Central versus Local Radiological Reading of Acute Computed Tomography Characteristics in Multi-Center Traumatic Brain Injury Research. *Journal of Neurotrauma* 2019; 36(7):1080–92.
42. Honeybul S, Janzen C, Kruger K, Ho KM. Decompressive craniectomy for severe traumatic brain injury: is life worth living? *Journal of Neurosurgery* 2013; 119(6):1566–75.
43. Van Dijck JT, Reith FC, van Erp IA, et al. Decision making in very severe traumatic brain injury (Glasgow Coma Scale 3-5): a literature review of acute neurosurgical management. *Journal of Neurosurgical Sciences* 2018; 62(2):153–77.
44. Van Dijck J, van Essen TA, Dijkman MD, et al. Functional and patient-reported outcome versus in-hospital costs after traumatic acute subdural hematoma (t-ASDH): a neurosurgical paradox? *Acta Neurochir* 2019;161(5):875–84.
45. Hartings JA, Vidgeon S, Strong AJ, et al. Surgical management of traumatic brain injury: a comparative-effectiveness study of 2 centers. *Journal of Neurosurgery* 2014;120(2):434–46.
46. Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *The Lancet* 2009; 374(9695):1097–104.

47. Kelly ML, Sulmasy DP, Weil RJ. Spontaneous intracerebral hemorrhage and the challenge of surgical decision making: a review. *Neurosurgical FOCUS* 2013;34(5):E1.
48. Kahneman D. Thinking, Fast and Slow. *Macmillan* 2011.
49. Wammes JJG, van den Akker-van Marle ME, Verkerk EW, et al. Identifying and prioritizing lower value services from Dutch specialist guidelines and a comparison with the UK do-not-do list. *BMC Med* 2016; 14(1):196.
50. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987; 317(3):141–5.
51. Volovici V, Ercole A, Citerio G, et al. Variation in Guideline Implementation and Adherence Regarding Severe Traumatic Brain Injury Treatment: A CENTER-TBI Survey Study in Europe. *World Neurosurg* 2019; 125:e515–20.
52. Volovici V, Steyerberg EW, Cnossen MC, et al. Evolution of Evidence and Guideline Recommendations for the Medical Management of Severe Traumatic Brain Injury. Journal of Neurotrauma 2019; 36(22):3183–9.
53. Bullock MR, Chesnut R, Ghajar J, et al. Surgical Management of Traumatic Brain Injury Guidelines. *Neurosurgery* 2006; 58(Supplement):S2–1–S2–3.
54. Carney N, Totten AM, O'Reilly C, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. *Neurosurgery* 2017; 80(1):6–15.
55. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589):349–51.
56. Howick J, Library ICJL, Glasziou P, et al. "The Oxford Levels of Evidence 2." Oxford Centre for Evidence-Based Medicine. Available from: http://www.cebm.net/index.aspx?o=5653
57. Bullock MR, Chesnut R, Ghajar J, et al. Surgical management of acute subdural hematomas. *Neurosurgery* 2006; 58(3 Suppl):S16–24–discussionSi–iv.
58. Stam J, Renckens C, Bouma J, Hijdra A, Van Gijn J, Vermeulen R. De Kritieken. Amsterdam: *Off Page* 2011.
59. Mansouri A, Cooper B, Shin SM, Kondziolka D. Randomized controlled trials and neurosurgery: the ideal fit or should alternative methodologies be considered? Journal of *Neurosurgery* 2016; 124(2):558–68.
60. Martin E, Muskens IS, Senders JT, et al. Randomized controlled trials comparing surgery to non-operative management in neurosurgery: a systematic review. *Acta Neurochir (Wien)* 2019; 161(4):627-634.
61. Yarascavitch BA, Chuback JE, Almenawer SA, et al. Levels of evidence in the neurosurgical literature: more tribulations than trials. *Neurosurgery* 2012; 71(6):1131–7–discussion1137–8.
62. Vranos G, Tatsioni A, Polyzoidis K, Ioannidis JPA. Randomized trials of neurosurgical interventions: a systematic appraisal. *Neurosurgery* 2004; 55(1):18–25–discussion25–6.
63. Azad TD, Veeravagu A, Mittal V, et al. Neurosurgical Randomized Controlled Trials-Distance Travelled. *Neurosurgery* 2018; 82(5):604–12.
64. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312(7023):71–2.
65. Farah JO, Varma TRK. Evidence-based neurosurgery—is it possible? *Br J Neurosurg* 2008; 22(4):461–2.
66. Helmy A, Timofeev I, Santarius T, Hutchinson P. What constitutes Clinical Equipoise? *Br J Neurosurg* 2009; 23(5):564–5.
67. Mendelow AD, Gregson BA, Rowan EN, et al. Early surgery versus initial conservative treatment in patients with traumatic intracerebral haemorrhage [STITCH(Trauma)]: the first randomised trial. *Journal of Neurotrauma* 2015; 32(17):1312-1323.
68. Kompanje EJO, van Dijck JTJM, Chalos V, et al. Informed consent procedures for emergency interventional research in patients with traumatic brain injury and ischaemic stroke. *The Lancet Neurology* 2020; 19(12):1033–42.
69. Roozenbeek B, Lingsma HF, Perel P, et al. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Critical Care* 2011; 15(3):R127.
70. Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015; 350.
71. Bergstrom CT, West JD. Calling Bullshit. *New York: Penguin Random House 2021.*
72. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions – a primer for the clinician. *Clinical Epidemiology* 2017; Volume 9:185–93.
73. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med* 2016; 21(4):125–7.
74. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy

of research designs. *N Engl J Med* 2000; 342(25):1887–92.
75. McCulloch P, Altman DG, Flum DR, et al. No surgical innovation without evaluation: the IDEAL recommendations. *The Lancet* 2009; 374(9695):1105–12.
76. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. *The Lancet Neurology* 2017; 16(12):987–1048.
77. Bowman K, Matney C, Berwick DM. Improving Traumatic Brain Injury Care and Research: A Report From the National Academies of Sciences, Engineering, and Medicine. *JAMA* 2022; 327(5):419–20.
78. Yue JK, Vassar MJ, Lingsma HF, et al. Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. *Journal of Neurotrauma* 2013; 30(22):1831–44.
79. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Ann Intern Med* 2020;172(1):35–45.
80. Roozenbeek B, Maas AI, Menon DK. Changing patterns in the epidemiology of traumatic brain injury. *Nat Rev Neurol* 2013 ;9(4):231-6.
81. Steyerberg EW, EW, Sewalt C, et al. Case-mix, care pathways, and outcomes in patients with traumatic brain injury in CENTER-TBI: a European prospective, multicentre, longitudinal, cohort study. *The Lancet Neurology* 2019; 18(10):923–34.