Evolution of Evidence and Guideline Recommendations for the Medical Management of Severe Traumatic Brain Injury

Victor Volovici,1,2 Ewout W. Steyerberg,2,3 Maryse C. Cnossen,2 Iain K. Haitsma,1 Clemens M. F. Dirven,1 Andrew I. R. Maas,4 and Hester F. Lingsma2

Abstract

Brain Trauma Foundation (BTF) Guidelines for medical management of severe traumatic brain injury (TBI) have become a global standard for the treatment of TBI patients. We aim to explore the evolution of the guidelines for the management of severe TBI. We reviewed the four editions of the BTF guidelines published over the past 20 years. The 1996 and 2000 editions were merged because of minimal differences, and are referred to as the 1996 edition. We described changes in topics and recommendations over time, and analyzed predictors of survival of recommendations with logistical regression. The guidelines contained 27 recommendations on 18 topics in 2016, 35 recommendations on 15 topics in 2007, and 22 recommendations on 10 topics in 1996. Substantial delays were found between the search for evidence and the guideline publication, ranging from 18 to 34 months. The overall body of evidence comprised 189 studies on 18 topics in 2016, compared with 156 studies on 15 topics in 2007 and 180 studies on 10 topics in 1996. Over time, a total of 175 studies were discarded from the evidence base following more rigorous grading of evidence. A total of 15/23 (65%) of the 1996/2000 recommendations were discarded over time. Out of 12 new recommendations introduced in the 2007 edition, 8 (66%) were discarded in 2016. Survival of recommendations varied between 33% and 100% for level I recommendations and between 11% and 31% for level II and III recommendations. No predictors of survival of recommendations were found. Substantial delays exist between literature search and publication, and survival rate of TBI guideline recommendations is poor. These factors may adversely affect currency and adherence to guidelines. The TBI community should take responsibility for improving the quality of the evidence base and ensuring that the translation of the evidence into guidelines supports clinicians in daily clinical practice.

Keywords: BTF Guidelines; evidence base; guidelines for severe TBI; survival of recommendations

Introduction

Purported aims of clinical practice guidelines include the promotion of an evidence-based delivery of healthcare and reduction of inappropriate variations in practice.1 To date, >6800 clinical guidelines have been developed, available via the Guidelines International Network.2 However, in the early 2000s, attention was drawn to the fact that a previous decade of published guidelines reflected poor quality data, poor reporting, and dubious overall methodology.3 In the next decade, the quality of guidelines and their underlying evidence came under fierce scrutiny.4 In response to this, in 2011, the Institute of Medicine (IOM) developed standards to be followed by future guideline development in order to ensure their proper foundation and form.5 However, a recent critical overview shows that from 18 standards, more than half of the guidelines follow ≤8 standards.6 The authors concluded that there have been “two decades of little, if any, progress.”

In the field of traumatic brain injury (TBI), the Brain Trauma Foundation (BTF) Guidelines for medical management of severe TBI have become a global standard. The first edition of these guidelines was published in 1996 – > 20 years ago – and was highly welcomed. In a recent survey of mostly academic, preponderantly level I trauma centers, 92% of centers that used any guidelines in the management of their patients either used the BTF Guidelines or a modified version of them.7 The initial guidelines were first updated in 2000. The evidence base included animal studies, case reports, and letters to the editor. Mechanistic proof-of-principle studies were included if it was felt that they provided sufficient
information on specific topics to be useful for clinical decision making.9

The methodology drastically changed in 2007 and 2016. In the 2016 version, the evidence base was critically reviewed, and low-quality studies were discarded. As a result, the latest edition of the guidelines on the management of severe TBI has been praised for its methodological rigor and criticized for its lack of clinical appeal.10,11

In this study, we aim to explore the evolution of the BTF guidelines for the management of severe TBI and its supporting body of evidence over the past 20 years.

Methods

Methodology of the guidelines and study extraction

The basis for this study were the 1996, 2000, 2007, and 2016 versions of the BTF Guidelines. The time between running the last literature search and publication of the version was calculated. Also, we extracted the number of recommendations per topic and the number of articles in the evidence base per recommendation. The number of studies that make up the evidence base as well as their design was extracted in order to map out the evolution of the evidence underpinning the recommendations for every edition of the guidelines, differentiated by classes of evidence.

Recommendations

We extracted the following variables in a database: the number of randomized controlled trials (RCTs) used to underpin each recommendation when it was issued for the first time; the number of prospective observational (PO) and retrospective observational (RO) studies; the grading of the recommendation (according to levels of evidence) in the 1996, 2000, 2007, and 2016 editions; whether the recommendation was formally restated or downgraded; the number of patients for each study; and the total number of patients per study design (e.g., RCT, PO, RO). Finally, we recorded for each study that underpinned a recommendation how it would be graded according to the 2016 methodology: class 1 and 2 (high-quality studies) or class 3 and discarded from the evidence base (low-quality studies). If an RCT was used to underpin a specific recommendation, but the data used to draw the conclusion stated in the recommendation were not the primary aim of the RCT, we considered it a PO study. For analysis, we combined the 1996 and 2000 versions because of their similarity.

If a recommendation was no longer restated or if it was downgraded to a lower level in a subsequent version of the guidelines, this was recorded as the outcome of interest; that is, non-survival. If a recommendation was merged with another one, the level of the newly merged recommendation was used to determine the outcome. Likewise, if a recommendation was split into two different ones, we judged the outcome for each of the split parts (seen as independent recommendations). Three readers (V.V., M.C.C., and I.K.H.) extracted data independently and then assessed whether the recommendation was positive “+” (intervention recommended), negative “−” (intervention should be avoided), or “±” neutral (no clear recommendation, the risks considered equal to the benefit).

Predictors of survival of recommendations

We used logistical regression analysis to test for predictors of survival of a recommendation; namely, the number of high- or low-quality studies that underpinned a particular recommendation, judging the studies using the strictest methodology from 2016.

Results

Evidence base

The delay between the last search in MEDLINE® and the publication date of the guideline edition was between 18 and 34 months (Fig. 1).

The overall body of evidence comprised 180 studies for 10 topics in 1996, compared with 156 studies for 15 topics in 2007, and 189 studies for 18 topics in 2016. The 180 studies in 1996 included, among others, case reports, mechanistic and animal studies, and letters to the editor. Most RCTs were considered class 1 evidence, and there were 30 class 1 studies in total. All 30 of these class 1 studies were reappraised: some were reclassified as class 3 and others were removed from the evidence base in subsequent editions. The 5 class 1 studies in 2016 are all new evidence. Of the 180 studies in the 1996/2000 evidence base, only 53 remained in the 2016 evidence base (29%), and were mostly classified as class 3 evidence. From the 156 studies in the evidence base in 2007, 98 (63%) remained in the evidence base of 2016. Of the 189 studies in the evidence base of 2016 (Fig. 2), 5 were classified as class 1.

Methodological assessment of evidence base

Methodological changes occurred in 2007 and again in 2016 with the incorporation of a team of methodologists into the guideline committee. The patient population targeted by the guidelines became more clearly defined, animal studies and case series with <25 patients were no longer regarded as evidence, and hence the guidelines became more restricted in scope.12 Studies were reclassified according to the new methodology: an RCT would be classified as lower evidence (class 2) if it violated one or more criteria for a good quality RCT.12 This trend toward methodological rigor continued and was augmented in the most recent update, in 2016.13 To make positive treatment recommendations, the studies in the evidence base needed to show the effectiveness of interventions in terms of mortality or functional outcome of TBI patients. Secondary outcomes were no longer considered proof of effectiveness.

Survival of recommendations and predictors of survival of recommendations

In 1996, the guidelines contained 22 recommendations on 10 topics, of which 3 were graded as level I, 9 were graded as level II, and 10 were graded as level III. In 2000, for the same 10 topics, there were 23 recommendations, of which 3 were level I, 9 were level II, and 11 were level III.

Sixteen new recommendations were presented in 2007 (4 for the original topics and 12 for 5 new topics), yielding a total of 35 recommendations (for 15 topics), of which 1 was level I, 15 were level II, and 19 were level III (Fig. 1).

From the original 10 topics comprising a total of 23 recommendations (1996/2000 edition), 15 recommendations (65%) were discarded (6 [26%] in 2007 and another 9 [39%] in 2016).

Regarding the five new topics added in 2007, 8 of the 12 recommendations (66%) were discarded in 2016.

In total, between the 1996 and 2016 editions, 35 recommendations (70%) were either discarded or downgraded, and only 15 (30%) were kept at the same level or upgraded (Fig. 1). The survival of recommendations for level II and III recommendations varied between 11% and 33% (Table 1).

Of all the recommendations in the various editions of the guidelines, 27 (54%) were not underpinned by an RCT when they
FIG. 1. Evolution of the guidelines from one edition to the next. Descriptive data concerning the number of recommendations, as well as the number of recommendations per topic and the numbers of articles/recommendations for each edition of the guidelines. The 1996/2000 guidelines are pooled together because of minimal differences. The shaded rectangles refer to the changes in recommendations.

FIG. 2. Evolution of the evidence base. The number of articles included in the evidence base, according to the original criteria (A) and the more stringent 2016 methodological criteria (B). According to the 2016 criteria, class II and class III evidence increased substantially between 1996/2000 and 2007, and between 2007 and 2016. Color image is available online.
were issued, whereas 18 (36%) were underpinned by at least one class 1 or 2 study, not necessarily an RCT (Tables S1, S2, and S3).

We assessed whether the number of high- and low-quality studies – as defined by the 2016 methodology – underpinning a recommendation at the moment it was issued, were predictors of survival of recommendations. The number of low-quality studies did not predict survival (odds ratio [OR] = 1.01, 95% confidence interval [CI] = [0.85–1.20], \( p = 0.88 \)), but there seemed to be a positive association between the number of high quality studies and survival (OR = 1.24, 95% CI = [0.81–1.90], \( p = 0.31 \)) (Table 2).

**Table 1. Survival of Recommendations**

<table>
<thead>
<tr>
<th>Edition</th>
<th>Level I (%)</th>
<th>Level II (%)</th>
<th>Level III (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996/2000, survival in 2007</td>
<td>1/3 (33%)</td>
<td>1/9 (11%)</td>
<td>3/11 (27%)</td>
</tr>
<tr>
<td>2007, survival in 2016</td>
<td>1/1 (100%)</td>
<td>3/15 (20%)</td>
<td>6/19 (31%)</td>
</tr>
</tbody>
</table>

Number of recommendations that were neither discarded nor downgraded from one edition to the next.

Discussion

**Summary of findings**

The overall body of evidence comprised 189 studies on 18 topics in 2016, compared with 156 studies on 15 topics in 2007, and 180 studies on 10 topics in 1996. Over time, a total of 175 studies were discarded from the evidence base, following more rigorous grading of evidence. At the same time, the guidelines contained 27 recommendations on 18 topics in 2016, 35 recommendations on 15 topics in 2007, and 22 recommendations on 10 topics in 1996. A total of 15/23 (65%) of the 1996/2000 recommendations were discarded over time. For level II and level III recommendations, the likelihood that they would be carried forward and not downgraded in a new edition was between 11% and 33%. When searching for predictors of survival of recommendations, none were found.

**FIG. 3.** Evolution of the recommendations. The number of recommendations, their level according to the original (A) and 2016 criteria (B); that is, what happened to the original recommendations in 2016, after introducing the stricter methodological criteria. The number of topics increases, but the number of recommendations does not. Color image is available online.
The Cochrane Collaboration\textsuperscript{17} recommends that the time between publication and date of the last search should be \(\leq 6\) months or ideally \(\leq 3\) months. In the 2016 version of the guidelines this was 3 years, meaning that at the time of publication, the Guidelines were at least 3 years outdated.

### Survival of recommendations

 Compared with other guidelines, severe TBI recommendations have very low survival. According to existing data in the literature, a recommendation underpinned by one or more RCTs would have an 81\% chance of surviving in a subsequent version.\textsuperscript{18} There were no predictors of survival of recommendations in the analysis we performed when looking either at the number of low- or high-quality studies, although there seemed to be a positive association between the number of high-quality studies and survival. The majority of recommendations (77\%) were discarded or downgraded because of the change in methodology and reappraisal of existing evidence. The rest were downgraded or discarded because of new evidence. The change in methodology is, therefore, the most likely cause of low survival of recommendations.

Nonetheless, the change in methodology was necessary because the 1996/2000 guideline recommendations were improperly classified. As such, despite being issued on poor quality evidence, the recommendations were graded higher than they would have been when assessed according to the strictest methodology. This approach might lead to an improperly high degree of confidence in the findings of poor-quality studies. The downside of the reduction in the number of recommendations is some loss of clinical appeal.

#### Translation of evidence into guidelines

The challenge of translating the evidence base into clinically appealing guidelines is clearly illustrated in the case of the “hyperosmolar therapy” topic. Of the 20 original studies included in the evidence base in the 2000 edition, 1 was still considered evidence in 2016. A total of six studies in the evidence base in 2016 could be translated to 0 recommendations. The recommendations in the 2007 edition were restated in 2016, but at the same time, a warning was included that they were no longer supported by evidence according to the new methodology. The authors of the BTF Guidelines wanted to retain awareness of the role of hyperosmolar therapies in treating high intracranial pressure (ICP). In the absence of any explicit statement on what the use of hyperosmolar therapies should be, the clinician is left without a formal recommendation. The accompanying text does contain recommendations no longer supported by evidence and a statement that hyperosmolar therapies are indeed important, but no actual guidelines, which leads to confusion.

### Table 2. Predictors of Survival

<table>
<thead>
<tr>
<th>Covariate</th>
<th>OR and 95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of high quality studies (classes 1 and 2)</td>
<td>1.24 (0.81–1.90)</td>
<td>0.31</td>
</tr>
<tr>
<td>Number of low quality studies (class 3 and discarded)</td>
<td>1.01 (0.85–1.20)</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Results of the logistical regression analysis for survival of a recommendation based on the number of high (classes 1 and 2) and low (class 3 and discarded from the evidence base) quality studies it was issued on, graded according to the 2016 criteria.

OR, odds ratio; CI, confidence interval.

### Evidence base and search updates

Looking at the evidence base of the severe TBI guidelines through the lens of more rigorous methodology (from the 2016 edition), there is consistent growth over the years, especially for class 2 and class 3 studies. Moreover, between the last two editions, employing the strictest methodology, the evidence base has increased with 98 studies (52\% of the entire current evidence base). Almost half of these, 48 (49\%), concerned the seven new topics, indicating more research interest in new topics rather than in strengthening the evidence for older recommendations, which was not particularly strong to begin with.

TBI is one of the fields in neurosurgical research in which quite a large number of RCTs have been conducted.\textsuperscript{14,15} However, of the 207 RCTs, only 26 across 18 interventions were robust, according to a previously conducted overview of research in TBI.\textsuperscript{14} In other areas of neurosurgery than TBI, ~1 in 10 trials is considered “robust.”\textsuperscript{16} (multi-center, low risk of bias, \(n > 100\) patients). Although the evidence base shows significant growth in absolute numbers overall, class 1 evidence remains scarce. Surprisingly, the 5 class 1 and 48 class 2 studies could only be translated into eight level I and level II A recommendations (Table 3), of which only three were positive, which denotes treatments with considered proven efficiency that should be used for severe TBI patients.

A substantial and increasing delay was observed between the date of the last literature search and publication of the guidelines.

### Table 3. Levels I and IIA Recommendations in the 2016 Edition

<table>
<thead>
<tr>
<th>Positive (treat with)</th>
<th>Negative (do not use)</th>
<th>Neutral (risks and benefits are similar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutrition</td>
<td>Steroids (Level I)</td>
<td>Seizure prophylaxis to treat early PTS</td>
</tr>
<tr>
<td>Early tracheostomy</td>
<td>Seizure prophylaxis to treat late post-traumatic seizures (PTS)</td>
<td></td>
</tr>
<tr>
<td>Performing a large rather than a small decompressive craniectomy</td>
<td>Povidone-iodine use</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Performing a bifrontal decompressive craniectomy instead of conservative treatment in diffuse injury</td>
<td></td>
</tr>
</tbody>
</table>

The highest level recommendations that were available to the clinician in 2016, based on high-quality studies from the evidence base.
A similar situation is present in the case of ventilation therapies. For that topic, the evidence base regressed from 28 studies in the 2000 edition to 1 study in 2016, an RCT from 1991. One recommendation remained, and the other three recommendations were restated in the text (with the warning that they are no longer “formal” recommendations and that they are no longer supported by evidence meeting current standards).

In contrast to the former examples, highly precise recommendations are made about ICP thresholds. The threshold was changed from 20 to 22 mm Hg, and this was not well received by the TBI research community, as they argued that this change of 2 mm Hg was not clinically relevant and suggested a “cookbook” approach to care of severe TBI patients.

The most striking change among all four editions is the fact that the first editions were focused on lowering ICP as the mainstay treatment for severe TBI, whereas the 2016 version contains no recommendations on lowering ICP except for a large decompressive craniectomy instead of a smaller one, and not using a prophylactic bifrontal decompressive craniectomy instead of medical management in diffuse injury.

The best available evidence, graded according to current stringent approaches, leaves the clinician with three positive treatment recommendations: providing adequate nutrition to decrease mortality, performing an early tracheostomy to reduce mechanical ventilation days (but not to reduce mortality), and performing a larger rather than a smaller decompressive craniectomy in order to improve outcomes. The rest of the high-quality recommendations advise against the use of steroids (the only level I recommendation available), against the use of povidone-iodine, and against the use of seizure prophylaxis for prevention of late post-traumatic seizures. They also advise against the prophylactic use of a bifrontal decompressive craniectomy in diffuse injury.

The few treatment recommendations are the reason for critique of the current guidelines. However, they do justice to the absence of strong evidence. Making treatment recommendations on little evidence carries the risk of recommending treatments that might be redundant or even harmful. Additionally, it might suggest that research into these topics is not needed anymore. On the other hand, it has been argued that it is desirable to standardize care even when a knowledge hiatus exists, to create the opportunity to run randomized trials with a uniform control group that represents the current standard of care.

Potential solutions for the future

The challenge is to timely summarize and translate the available evidence into guideline recommendations in the most comprehensive way possible. Being “lost in translation” between clinical studies and clinical practice is not specific to TBI alone.

The first priority is to generate high-quality evidence. The TBI research community strongly supports comparative effectiveness research in addition to clinical trials as a way of evidence generation. Concerning the outdated searches, solutions have been proposed by the BTF Guidelines authors, among others, in the form of living guidelines, which would be updated online periodically. Living Guidelines, however, do not solve the underlying problem of the poor evidence base, and, therefore, if no high-quality studies are available, no guideline recommendations can be derived.

One potential solution for a complete, up-to-date evidence base might be to employ living systematic reviews for individual topics. This solves the problem of outdated guidelines by providing clinicians with a “current” knowledge base on every topic, properly graded according to strict methodology. It would also indicate where a knowledge hiatus exists in order to stimulate the TBI community to perform research in those particular areas. Moreover, the TBI community itself could prioritize these knowledge hiatuses, leading to a more effective collaboration among various groups on TBI research and more rapid generation of high-quality evidence. The Guideline Committee would then use the living systematic reviews, which should apply a strict methodological appraisal of the quality of the evidence, to make clinically relevant “living recommendations,” branded with a clear level of confidence. This will avoid the situation that occurred in 1996/2000, when recommendations were graded too high while relying on poor quality evidence.

In this way, sound methodology and making useful treatment recommendations can be reconciled.

Conclusion

Despite considerable interest in TBI research, evidence for the management of severe TBI remains limited, with few robust studies and even fewer studies showing the benefit of a particular intervention. However, there are more high-quality studies in the 2016 version of the Guidelines than in the 1996/2000 versions. Therefore, the evidence base is improving slowly, but the TBI research community should take responsibility for generating more high-quality evidence. The underlying evidence base needs to be responsibly translated into clinically applicable, accurately graded recommendations in order to help clinicians properly treat severe TBI patients. These two efforts should be complementary and stem from a unified vision of evidence, guidelines, and implementation for the benefit of the patient.

Acknowledgments

This work was supported by the European Union Framework 7 program (grant 602150) – CENTER TBI. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Table S1
Supplementary Table S2
Supplementary Table S3

References


Address correspondence to:
Victor Volovici, MD
Erasmus MC University Medical Center
Erasmus MC Stroke Center
Department of Neurosurgery and Medical Decision Making
Doctor Molewaterplein 40
3015 GD, Rotterdam
The Netherlands

E-mail: v.volovici@erasmusmc.nl