Reliability and validity of the Therapy Intensity Level (TIL) scale: analysis of clinimetric properties of a novel approach to assess management of intracranial pressure in traumatic brain injury

Patrick Zuercher[1,2], Justus L Groen[1,3], Marcel JH Aries[1,4], Ewout W Steyerberg[5], Andrew IR Maas[6], Ari Ercole[1]*, David K. Menon[1]*

1. Division of Anaesthesia, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
2. Department of Intensive Care Medicine, University Hospital Inselspital, Bern, Switzerland
3. Department of Neurosurgery, UMC Leiden, Leiden, the Netherlands
4. Department of Critical Care, University of Groningen, UMC Groningen, Groningen, the Netherlands
5. Center for Medical Decision Sciences, Department of PH, Erasmus MC, Rotterdam, the Netherlands
6. Department of Neurosurgery, University Hospital Antwerp - University of Antwerp, Edegem, Belgium

*Both these authors contributed equally to supervision of this study

Patrick Zuercher (corresponding author)

Department of Intensive Care Medicine, University Hospital Bern, Inselspital, Freiburgstrasse, 3010 Bern, Switzerland, Tel: +41 316325300, Fax: +41 316321771, email: pz257@cam.ac.uk / patrick.zuercher@insel.ch

Justus L Groen

Department of Neurosurgery, Leiden University Medical Center, Leiden, the Netherlands, Tel: +31 715262109, Fax: 31 715266987, Email: J.L.Groen@lumc.nl

Marcel JH Aries

Department of Critical Care, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands, Tel: +31 503612315, Fax: +31 503616161, Email: m.j.h.aries@umcg.nl

Ewout W Steyerberg

Center for Medical Decision Sciences, Department of Public Health, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands, Tel: +31 107038470, Fax: +31 107038465, Email: e.steyerberg@erasmusmc.nl

Andrew IR Maas

Department of Neurosurgery, University Hospital Antwerp - University of Antwerp, Wilrijkstraat 10, 2650 Edegem, Belgium, Tel: +32 38213328, Fax: +32 38214185, Email: Andrew.Maas@uza.be

Ari Ercole

Division of Anaesthesia, University of Cambridge, Box 93, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK, Tel: +44 1223217889, Fax: +44 1223217887, Email: ae105@cam.ac.uk

David K Menon (corresponding author)

Division of Anaesthesia, University of Cambridge, Box 93, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK, Tel: +44 1223217889, Fax: +44 1223217887, Email: dkm13@wbic.cam.ac.uk

Abstract

We aimed to assess the reliability and validity of the Therapy Intensity Level scale (TIL) for intracranial pressure (ICP) management. We reviewed the medical records of 31 patients with traumatic brain injury (TBI) in two European intensive care units (ICU). The ICP TIL was derived over a 4-day period for 4-hour (TIL4) and 24-hour epochs (TIL24). TIL scores were compared with historical schemes for TIL measurement, with each other, and with clinical variables. TIL24 scores in ICU TBI patients were compared to two control groups: patients with extracranial trauma requiring intensive care (Trauma_ICU; n=20) and patients with TBI not requiring ICU care (TBI_WARD; n=19), to further determine the discriminative validity of the TIL for ICP-related ICU interventions. Inter-rater and intra-observer agreement were excellent for TIL4 and TIL24 (Cohen's κ: 0.98 -0.99; intraclass correlation coefficient: 0.99 – 1; $p < 0.0005$). The mean + SD TIL24 in the ICU TBI cohort was significantly higher than the Trauma_ICU patients and the TBI_WARD patients (8.2±3.2 vs. 2.2±0.9 and 0.1±0.1, respectively; $p<0.005$ for both comparisons). Correlations between the TIL scale scores and historical TIL scores, between TIL24 and the Glasgow Coma Scale (GCS), and between a range of TIL metrics and summary measures of ICP over the 4-day period, were all highly significant ($p < 0.01$). The results were consistent with the expected direction. A linear mixed effect analysis, accounting for within-subjects repeated measures, showed strong correlation between TIL4 and 4-hourly ICP ($p < 0.0000005$). The TIL scale is a reliable measurement instrument with a high degree of validity for assessing the therapeutic intensity level of ICP management in patients with TBI.

Key words: TIL scale, TBI, ICP, Clinimetrics

Introduction

Intracranial pressure (ICP) is a strong driver of outcome in patients after traumatic brain injury (TBI).[1] Many index pathophysiological processes, such as vasogenic oedema, cytotoxic oedema, and increased cerebral blood volume cause ICP elevation. Consequently, the ICP has been suggested as a biomarker of the activity of these pathophysiological processes. Indeed, the severity and/or duration of intracranial hypertension have been listed as secondary outcome measures in TBI trials.[2] However, this approach is confounded by the fact that modern intensive care mitigates intracranial hypertension through escalating interventions that seek to normalise ICP,

thus reducing its sensitivity as an early endpoint. Given this context, there has been a growing desire to use the intensity of ICP-directed therapy as an alternative biomarker in this context.[3–5] Many different therapies may be used for the control of ICP, often simultaneously. This poses major difficulties in clinical TBI research, because the individual effect of a study therapy can be obscured by adjustments in any or all other therapies.[4] Integration of all known and relevant ICP directed treatments into a single summary score could therefore be useful in conducting research studies, allowing better comparison between management approaches and outcome variables between centres and countries.

In 1987, Maset et al. proposed a Therapy Intensity Level scale (TIL) to assess the intensity of ICP lowering management with a 15-point scale. This scale has been used in several trials since as a secondary outcome despite important limitations[5–10] such as showing a ceiling effect (i.e. scoring at maximal levels whenever barbiturates are given), not including a complete range of interventions, and being labour-intensive due to the need for hourly assessments. A novel approach to assessing TIL, which sought to address some of these issues, was developed as part of the Interagency Common Data Elements scheme.[11] The summary score was designed to be consistent with the paediatric intensity level of therapy (PILOT) scale, proposed by Shore et al.[4] The novel adult TIL has been broadly accepted by the neurotrauma research community, but has not as yet, been subjected to validation. Such analysis is important to confirm that it effectively documents therapeutic intensity of ICP directed measures, rather than a diagnosis of TBI, injury severity, non-ICP specific ICU procedures, or clinical outcome. Further, such validation would need to address its consistency across repeated measurements.

In the current study, we aimed to assess the reliability and validity of the therapeutic intensity level scale (TIL) for intracranial pressure (ICP) management.

METHODS

TIL Scale

Individual ICP-targeting therapies were assigned a score based on published estimates of their relative efficacy and risks of morbidity.[11] The TIL includes eight ICP-treatment modalities, termed items (Table 1). We calculated the following TIL scale scores:

- TIL_4, the numerical summary TIL scale score for every 4-hr epoch.
- TIL_{24}, a daily TIL score based on the highest score in each item per day, to provide a metric of the maximal therapeutic intensity for ICP management for the day.
- TIL_{max}, the highest TIL_{24} score in the assessed 4-day period.
- TIL_{mean}, the mean between the four TIL_{24} in the assessed 4-day period.

For certain calculations, e.g. correlation analysis between the different TIL scale scores (i.e. TIL_4 vs TIL_{24}, TIL_{mean}, TIL_{max}), we used a processed value and calculated a mean TIL_4 per day or per 4-day period.

Patients and data acquisition

Data from TBI patients admitted to the neurocritical care unit of the University Hospital Groningen (*n*=16) and Addenbrooke's Hospital Cambridge (*n*=32) from May 2012 until December 2013 were collected and screened. Patients in the latter cohort were part of an existing approved research study (29 REC 97/291), while use of data from patients in the former cohort was permitted following local medical ethical committees review, which and waived the need for informed consent. Only patients older than 16 years of age, with at least four consecutive days of continuous high resolution ICP monitoring starting within 48h of the incident were initially included for the ICU stratum (TBI_ICU). 31 patients fulfilled these criteria. All ICU patients in the TBI group were sedated, intubated and mechanically ventilated. Both centres used (different) protocol-based ICP management strategies. We collected patients' demographics and baseline clinical data for the prognostic assessment using the extended IMPACT scores.[12] Clinical outcome was measured in most TBI patients at 6 months using the Glasgow Outcome Scale (GOS).[13–16] In nine patients (none of whom had severe TBI), a favourable outcome was charted at an earlier time point (at the time of clinical or research follow up), and patients did not return for

further follow up. These patients were assumed to have carried their favourable outcome out to the six month follow up point.

For the calculation of TIL, medical and surgical interventions for managing intracranial hypertension were extracted from clinical records, including amount of CSF drained, treatment of fever, and whether or not an active cooling protocol was in place. Points for surgical interventions were assigned to the period when the intervention took place and included in every successive period onward. We additionally extracted dosages of all sedative and vasoactive drugs, neuromuscular blockers, hyperosmolar agents and barbiturate administration. $PaCO_2$ values were directly obtained from measurements made during each four-hour epoch in the majority of instances. However, in less than 10% of epochs (73 of > 800 epochs), no contemporaneous $PaCO_2$ value was available within the four hour period. In such instances, the $PaCO_2$ was derived from the end-tidal CO_2 ($etCO_2$) measurement, using the nearest $PaCO_2$ value to correct for the $PaCO_2$-$etCO_2$ gradient, and making the assumption that this had not changed.

A range of summary metrics of ICP were calculated for correlations:

- ICP_4, mean ICP within 4-hr period
- ICP_{24}, mean ICP within 24-hr period
- ICP_{max}, highest ICP_{24} in the assessed 4-day period
- ICP_{mean}, mean between the four ICP_{24} in the assessed 4-day period

Two control groups were defined, selected from patients admitted to Addenbrooke's Hospital in Cambridge between November 2012 and April 2015. Patients with extracranial trauma admitted to the NCCU were randomly selected, screened and included if TBI could be reasonably excluded based on history, examination and neuroimaging findings (Trauma_ICU, *n*=20). In most of these patients we had a reliable post injury GCS of 15 and normal neuroimaging. In a minority (see Table 2) severe extracranial injury meant that we were either unable to obtain a post resuscitation GCS, as the patient was intubated for cardiorespiratory instability, or received substantial doses of opioids or ketamine for analgesia before a reliable GCS could be recorded. None

of these patients had any TBI-directed therapy, and all had neuroimaging and a subsequent clinical course that excluded any significant TBI. A second group consisted of mild/moderate TBI patients directly admitted to a ward for observation and treatment of extracranial trauma (*n*=19), and who required no ICP- specific therapies.

Demographic data from all three groups were compared using the chi-square test for categorical variables and analysis of variance with Bonferroni's post hoc test for continuous variables. Whenever the criteria for a chi-square test were not met, Fisher's exact test was used instead.

Reliability Assessment

We assessed inter-rater and intra-rater reliability in a random subset of 10 TBI patients. TIL_4 and TIL_{24} were calculated independently by two blinded investigators (PZ, JLG; inter-rater). After a washout interval of 3 months and blinded to the initial TIL scale scores, one investigator (JLG, intra-rater) repeated the measurement. We calculated Cohen's kappa and intraclass correlation coefficients (ICC) to better compare our data with the literature.[4]

Validity Assessment

In accordance with the recommendations from the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) taxonomy[17] and clinimetric literature,[18] we evaluated *content validity*, *criterion validity* and *construct validity* (assessed by c*onvergent, discriminant and discriminative validity).*

Content validity is 'the degree to which the content of an instrument is an adequate reflection of the construct to be measured'.[18] It is a subjective measure of how appropriate the items seem to a set of experts. This is not quantified with statistics, and given the derivation and acceptance of our TIL scheme by experts, this was assumed to exist. *Criterion validity* is 'the degree to which the scores of a measurement instrument are an adequate reflection of a gold-standard'.[18] We compared the TIL scale with the grading system suggested by[19]

(TIL_Maset), mindful of its limitations. We calculated TIL_4 and TIL_{24} in a subset of TBI_ICU patients and tested our hypothesis of a positive correlation of moderate to strong magnitude using Spearman's rho.

Construct validity was quantified to evaluate our expectations regarding how the measurement instrument related to known parameters.[18] We evaluated construct validity by assessing convergent (assessing positive or negative correlations with similar constructs), discriminant (assessing for correlations with measurement instruments measuring different constructs) and discriminative (assessing for ability to differentiate between known groups) validity as follows:

Convergent validity was evaluated by testing our expectations of a negative correlation between TIL_{mean} and TIL_{max} with GCS of moderate to strong magnitude, a positive correlation between $TIL_4/TIL_{24}/TIL_{max}/TIL_{mean}$ with $ICP_4/ICP_{24}/ICP_{max}/ICP_{mean}$ of moderate magnitude. Additionally we expected positive correlations of strong magnitude between TIL subtypes, TIL_{24} vs TIL_4 (daily mean), TIL_{mean} vs TIL_4 (averaged over 4 days), TIL_{mean} vs TIL_{max} and TIL_{max} vs TIL_4 (averaged over 4 days). The TIL scale quantifies therapeutic intensity and is not intended to predict clinical outcome. Any clear correlations between outcome and the TIL scale are therefore not expected. To assess discriminant validity we therefore hypothesized that there is no correlation between the TIL_{mean} and TIL_{max} with outcome (GOS, IMPACT) in the TBI_ICU group alone and at best a weak negative correlation in the combined TBI group (TBI_ICU and TBI_WARD) in agreement with the PILOT study.[4] For most of the variables used to assess convergent validity, simple non-parametric statistical tests were used (Spearman's rho). However, this approach was not appropriate for examining the relationship between TIL_4 and ICP_4, since the multiple estimates of ICP in each individual were not independent. We therefore additionally used linear mixed effects (LME) regression techniques to examine this relationship, using the lme4 package (v. 1.1-8) in R (v. 3.2.1; R Foundation for Statistical Computing, Vienna, Austria).

Discriminative validity was assessed by testing the hypothesis that the TIL scale can accurately discriminate cases (the TBI_ICU cohort) from controls (the TBI_WARD and Trauma_ICU cohorts) by comparing TIL_{mean} and TIL_{max}

between groups using the Kruskal-Wallis test with follow-up testing for pairwise comparison with adjusted *p*-values.

All statistical analyses, other than the LME analysis, were undertaken using IBM SPSS Statistics, v. 22. Where multiple analyses were undertaken using the same pairs of data, or their derivatives, we applied a Bonferroni correction to our *p*-values.

RESULTS

Patient demographics, along with GCS, IMPACT, GOS (where appropriate), and the different TIL scale scores are shown in Table 2. Data were collected from patients in each of the three groups (total *n*=70). The groups did not differ in terms of age or sex.

The instrument showed high intra-rater reliability for both TIL_{24} and TIL_4 measurements. An assessment of inter-rater reliability resulted in a Cohens κ of 0.981 with and ICC of 0.999 ($p < 0.0005$) for the TIL_4, and perfect agreement for the TIL_{24} (Table 3).

Validity metrics for our TIL score are shown in Table 4, Figure 1&2 and Supplementary Figure 1. The TIL_{24} and TIL_4 showed moderate correlations with the corresponding historical scores (TIL_Maset, showing criterion validity), and with the GCS and ICP (showing convergent validity). The direction and strength of these correlations were all in keeping with our *a priori* predictions.

A random intercepts linear mixed effects modelling, grouped by patient, undertaken to correct for the non-independence of multiple measurements of TIL_4 and ICP_4 within patients, showed a significant positive association ($p < 0.0000005$). This parameter was still significant ($p < 0.0005$) with both random intercepts and slopes, providing strong evidence for an underlying population-level relationship between these parameters.

Patients in the TBI_ICU stratum showed significantly higher TIL_{mean} and TIL_{max} values than the two control cohorts (TIL_{mean} ± SD: 8.2±3.2 vs. 2.2±0.9 and 0.1±0.1; TIL_{max} ± SD: 9.9±3.7 vs. 3.4±1.4 and 0.2±0.4, for groups TBI_ICU, Trauma_ICU and TBI_WARD, respectively). Kruskal-Wallis test to compare multiple independent samples for TIL_{mean} showed $H(2) = 60.55$, $p < 0.0005$ and for TIL_{max} $H(2) = 59.39$, $p<0.0005$. Pairwise comparisons with adjusted p-values showed that there are significant differences between all groups: TBI_ICU vs Trauma_ICU ($p < 0.005$), TBI_ICU vs TBI_Ward ($p < 0.0005$) and Trauma_ICU vs TBI_Ward ($p = 0.007$) for TIL_{mean} and TIL_{max}, but the correlations between TIL and predicted or observed outcome did not survive Bonferroni correction for multiple comparisons.

The TIL_4 and TIL_{24} showed high correlation within our scheme, showing that the daily measure of TIL (TIL_{24}) was an acceptable summary metric of ICP therapy intensity. The TIL_{max} showed strong correlation with TIL_{mean} and less strong, but still highly significant correlations with the TIL_4 (mean over 4 days), suggesting that abstraction of the higher intensity interventions performed in each 24 hour period still provided an acceptable (albeit less faithful) measure of the TIL (Supplementary Figure 2&3 and Supplementary Table 1).

DISCUSSION

We show that the TIL scale score can be obtained retrospectively in TBI patients and that it has excellent inter- and intra-rater reliability with minimal measurement error, both for 4-hr and 24-hr assessments. These results are in agreement with the PILOT study[4] which showed an ICC of 0.91 for inter- and 0.94 for intra-rater reliability.

We judge the amount of content validity of the TIL scale as very high, since it was based on the consensus paper on standardizing data collection in TBI[11] developed by an international expert panel.

As a fundamental prerequisite, the test-of-interest is judged against a 'gold-standard' for assessing the same variable, or the same concept. This measure of criterion validity is less likely to find a suitable "gold standard" with more abstract concepts. In this case, we compared our TIL scale against the "best available alternative" comparator - the TIL grading system described by Maset et al.,[19] despite its recognised limitations and lack of formal validation. We found a high correlation between these two variables, as expected, adding further criterion validity to the iterative process of validation. Although we had no patients in our cohort who received metabolic suppression with barbiturates or other anaesthetic agents, our scoring system would (self-evidently) not suffer from the ceiling effects that have been seen to be a problem with the TIL_Maset, even if this intervention was deployed.[10,6,9]

Convergent validity was demonstrated by showing a correlation of expected magnitude and direction between TIL vs. GCS and TIL vs. ICP. These correlations were broadly in agreement with corresponding figures in the PILOT study. Our expectation of only a moderate to strong correlation between GCS and TIL was based on the fact that some patients (e.g. those with diffuse axonal injury) who present with low GCS may experience no problems with intracranial hypertension, and hence achieve low TIL scores, even with prolonged ICU stays. Critically, the relationship between TIL and ICP was retained with the application of a linear mixed effects modelling approach, which accounted for the repeated measures of ICP within individuals, thus ensuring the reliability of our results.

We show that our TIL scale can accurately discriminate between different treatment groups in intensive care (TBI_ICU vs. Trauma_ICU), and between TBI patients treated within and outside the ICU environment (TBI_ICU vs. TBI_WARD), suggesting that the items that it includes are specific for the ICU management of TBI (which is prominently targeted at intracranial hypertension). One drawback of our TRAUMA_ICU population was the fact that a reliable GCS could not be obtained pre-sedation in a minority, although the overall clinical course in these patients excluded a significant TBI. In any case, if these patients did have an undetected mild TBI, this confound would work against the discriminative validity of the TIL. Additional evidence supporting its discriminant validity come from the demonstration that the negative correlation between TIL and outcome (using a dichotomized GOS) is present but weak in the combined group of TBI patients (n=40 with mild to severe TBI), a finding that is

concordant with the PILOT study results, but no correlation between TIL and outcome within the TBI_ICU cohort. The former relationship detects ICP lowering therapies in the TBI_ICU group that are absent in the TBI_WARD patients, while the absence of this relationship in the TBI-ICU group on its own suggests that TIL itself does not correlate with outcome, since successful treatment of refractory intracranial hypertension (with very high TIL scores) can be associated with good outcomes.

The choice of 4-hr or 24-hr epochs was arbitrary. Our results show no difference in reliability or validity between these two approaches, with strong and statistically significant correlations both between the two TIL scores. In addition, the correlations that we show for demonstrating validity apply equally to both scores. Concerns about practicability, burden of data capture, and time investment may predicate the use of the TIL_{24} in many settings. However, where the research question requires this (e.g. evaluation of a new pharmaceutical approach to lowering ICP), a more frequent assessment of TIL may be justified. On the other hand, we also showed less strong, but still reasonable correlation between the TIL_{max} and the TIL_4. Given that the TIL_{max} is relatively easy to abstract from clinical notes, this may provide a reasonable alternative in resource-limited research settings.

The positive correlation of TIL_4 with four-hourly ICP values is interesting. While ICP control generally remained acceptable throughout the range of TIL, this correlation implies that the actual ICP value rose slightly as therapy was progressively intensified. This perhaps suggests that clinicians balance risks and benefits in a pragmatic manner, and when applying more aggressive therapies, accept slightly higher ICP values. These data provide insights into clinician behaviour that merit further investigation.

While the TIL score that we describe here has many pragmatic benefits, it also has limitations. The TIL scale per se is arbitrary. Point assignment to each TIL item reflects a presumed weighting of inter-item differences in therapy intensity that is impossible to determine objectively with the limited evidence available. This is inevitable, however, and reflects the very nature of any data reduction exercise.

Another limitation of this present study is the fact that it is a two-centre, retrospective study, analysing a specific - and small - population in a specific context that may not be truly representative.

In an initial pilot assessment of the scheme (data not shown) we obtained less satisfactory κ value of 0.455 in a 4-hr inter-rater reliability assessment. A review of our internal rules for assigning points showed the possibility of ambiguity. A revision of these rules was therefore undertaken prior to implementing the study (see footnote to Table 1). This discussion shows that it is critical to ensure the availability of a precise and unambiguous protocol for implementation of a TIL scale. In addition, they merit a cautionary note, that the results from the inter-rater reliability assessment may be theoretically contaminated by possible practice effects.

CONCLUSION

We have presented evidence that the TIL scale is a reliable measurement with a high degree of validity for assessing Therapy Intensity Level of ICP management in patients with TBI on an ICU. Further studies are warranted, ideally prospective and across heterogeneous centres in a larger population with different preferential therapeutic approaches, to evaluate the generalizability of this measurement instrument.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Balestreri, M., Czosnyka, M., Hutchinson, P., Steiner, L.A., Hiler, M., Smielewski, P., and Pickard, J.D. (2006). Impact of intracranial pressure and cerebral perfusion pressure on severe disability and mortality after head injury. Neurocrit. Care 4, 8–13.

2. Maas, A.I., Murray, G., Henney 3rd, H., Kassem, N., Legrand, V., Mangelus, M., Muizelaar, J.P., Stocchetti, N., Knoller, N., and Pharmos, T.B.I. investigators. (2006). Efficacy and safety of dexanabinol in severe traumatic brain injury: results of a phase III randomised, placebo-controlled, clinical trial. Lancet Neurol 5, 38–45.

3. Marmarou, A., Nichols, J., Burgess, J., Newell, D., Troha, J., Burnham, D., and Pitts, L. (1999). Effects of the bradykinin antagonist Bradycor (deltibant, CP-1027) in severe traumatic brain injury: results of a multi-center, randomized, placebo-controlled trial. American Brain Injury Consortium Study Group. J. Neurotrauma 16, 431–44.

4. Shore, P.M., Hand, L.L., Roy, L., Trivedi, P., Kochanek, P.M., and Adelson, P.D. (2006). Reliability and validity of the Pediatric Intensity Level of Therapy (PILOT) scale: a measure of the use of intracranial pressure-directed therapies. Crit. Care Med. 34, 1981–1987.

5. Biestro, A., Alberti, R., Galli, R., Cancela, M., Soca, A., Panzardo, H., and Borovich, B. (1997). Osmotherapy for increased intracranial pressure: Comparison between mannitol and glycerol. Acta Neurochir. (Wien). 139, 725–733.

6. Marion, D.W., Obrist, W.D., Carlier, P.M., Penrod, L.E., and Darby, J.M. (1993). The use of moderate therapeutic hypothermia for patients with severe head injuries: a preliminary report. 354-362 p.

7. Clifton, G.L., Allen, S., Barrodale, P., Plenger, P., Berry, J., Koch, S., Fletcher, J., Hayes, R.L., and Choi, S.C. (1993). A phase II study of moderate hypothermia in severe brain injury. 263-271; discussion 273 p.

8. Lemeshow, S., and Hosmer, D.W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. Am. J. Epidemiol. 115, 92–106.

9. Münch, E., Horn, P., Schürer, L., Piepgras, A., Paul, T., and Schmiedek, P. (2000). Management of severe traumatic brain injury by decompressive craniectomy. Neurosurgery 47, 315–322; discussion 322–323.

10. Clifton, G.L., Miller, E.R., Choi, S.C., Levin, H.S., McCauley, S., Smith, K.R., Muizelaar, J.P., Wagner, F.C., Marion, D.W., Luerssen, T.G., Chesnut, R.M., and Schwartz, M. (2001). Lack of effect of induction of hypothermia after acute brain injury. 556-563 p.

11. Maas, A.I., Harrison-Felix, C.L., Menon, D., Adelson, P.D., Balkin, T., Bullock, R., Engel, D.C., Gordon, W., Langlois-Orman, J., Lew, H.L., Robertson, C., Temkin, N., Valadka, A., Verfaellie, M., Wainwright, M., Wright, D.W., and Schwab, K. (2011). Standardizing data collection in traumatic brain injury. J Neurotrauma 28, 177–187.

12. Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., Murray, G.D., Marmarou, A., Roberts, I., Habbema, J.D.F., and Maas, A.I.R. (2008). Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 5, e165; discussion e165.

13. Teasdale, G.M., Pettigrew, L.E., Wilson, J.T., Murray, G., and Jennett, B. (1998). Analyzing outcome of treatment of severe head injury: a review and update on advancing the use of the Glasgow Outcome Scale. J. Neurotrauma 15, 587–597.

14. Wilson, J.T., Pettigrew, L.E., and Teasdale, G.M. (1998). Structured interviews for the Glasgow Outcome Scale and the extended Glasgow Outcome Scale: guidelines for their use. J. Neurotrauma 15, 573–585.

15. Wilson, J.T.L., Slieker, F.J.A., Legrand, V., Murray, G.D., Stocchetti, N., and Maas, A.I.R. (2007). Observer Variation in the Assessment of Outcome in Traumatic Brain Injury: Experience from a Multicenter, International Randomized Clinical Trial. Neurosurgery 61, 123–129.

16. JENNETT, B. (1975). ASSESSMENT OF OUTCOME AFTER SEVERE BRAIN DAMAGE A Practical Scale. Lancet 305, 480–484.

17. Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., and de Vet, H.C.W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63, 737–745.

18. De Vet, H.C.W., Mokkink, L.B., Terwee, C.B., and Knol, D.L. (2011). Measurement in Medicine. Cambridge University Press, 338 p.

19. Maset, A.L., Marmarou, A., Ward, J.D., Choi, S., Lutz, H.A., Brooks, D., Moulton, R.J., DeSalles, A., Muizelaar, J.P., and Turner, H. (1987). Pressure-volume index in head injury. J. Neurosurg. 67, 832–840.

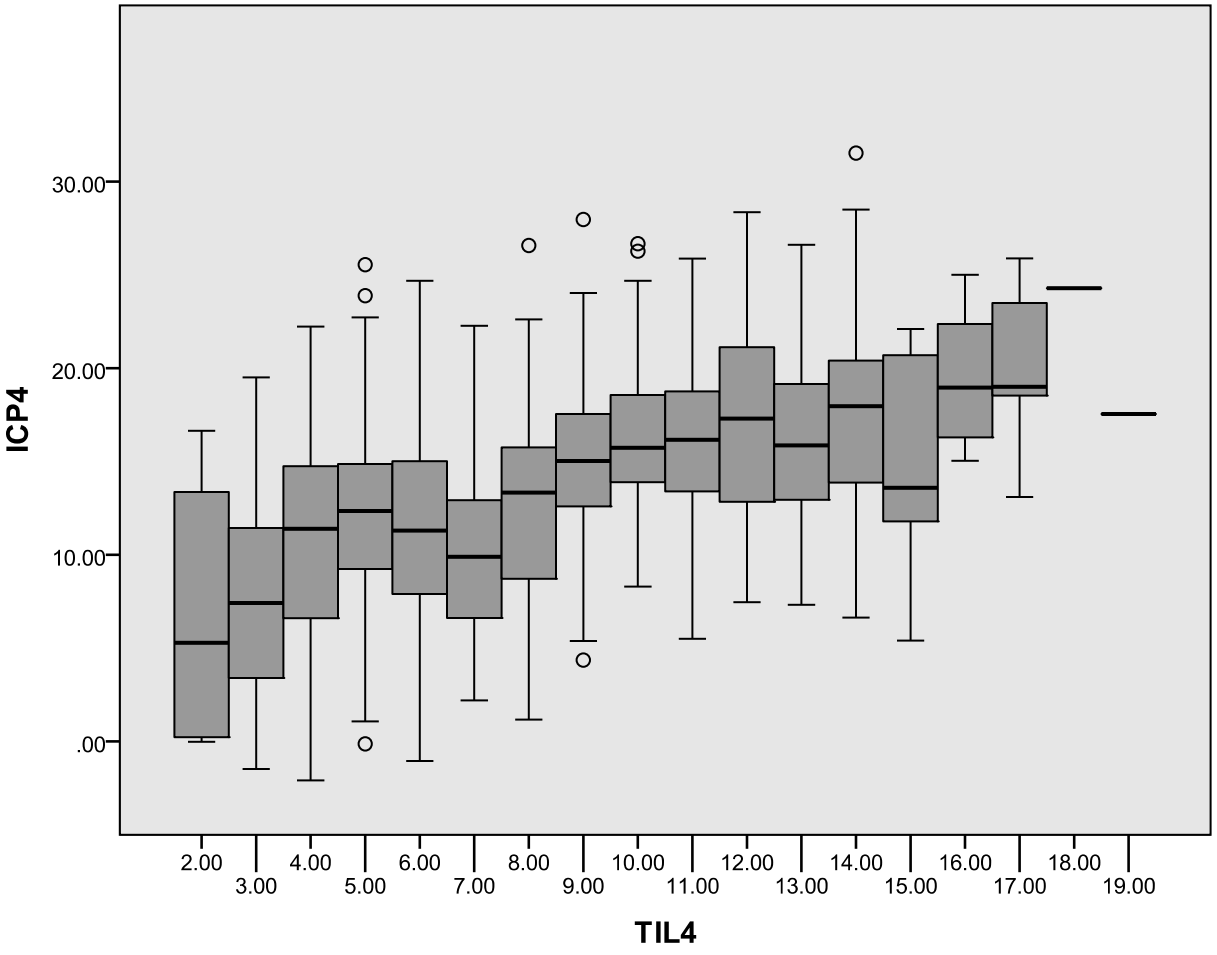
ICP4
30.00
20.00
10.00
.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
16.00
17.00
18.00
19.00
TIL4

Fig. 1a.

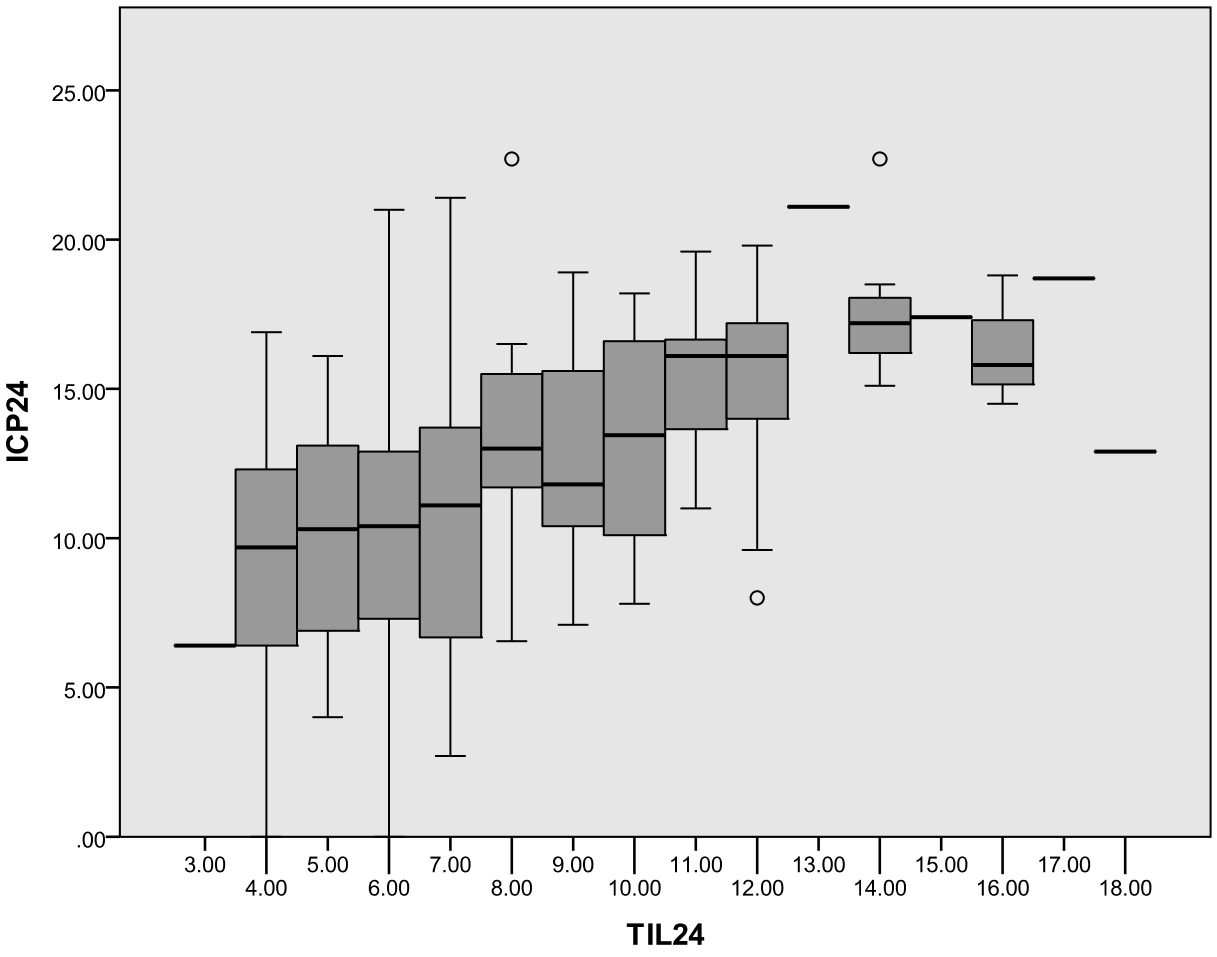
ICP24
25.00
20.00
15.00
10.00
5.00
.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
16.00
17.00
18.00
TIL24

Fig. 1b.

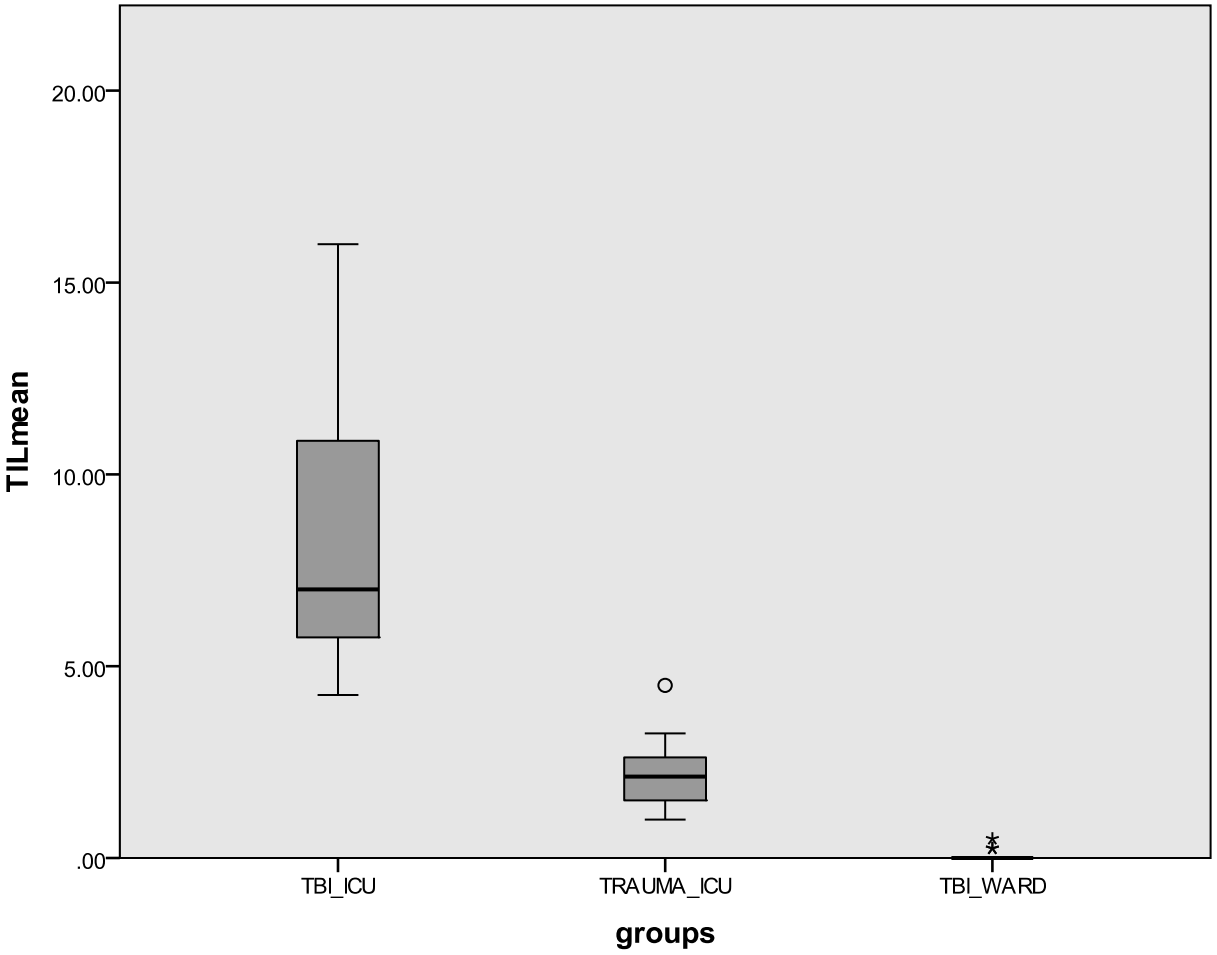
20.00
15.00
10.00
5.00
.00
TILmean
TBI_ICU
TRAUMA_ICU
TBI_WARD
groups

Fig. 2a.

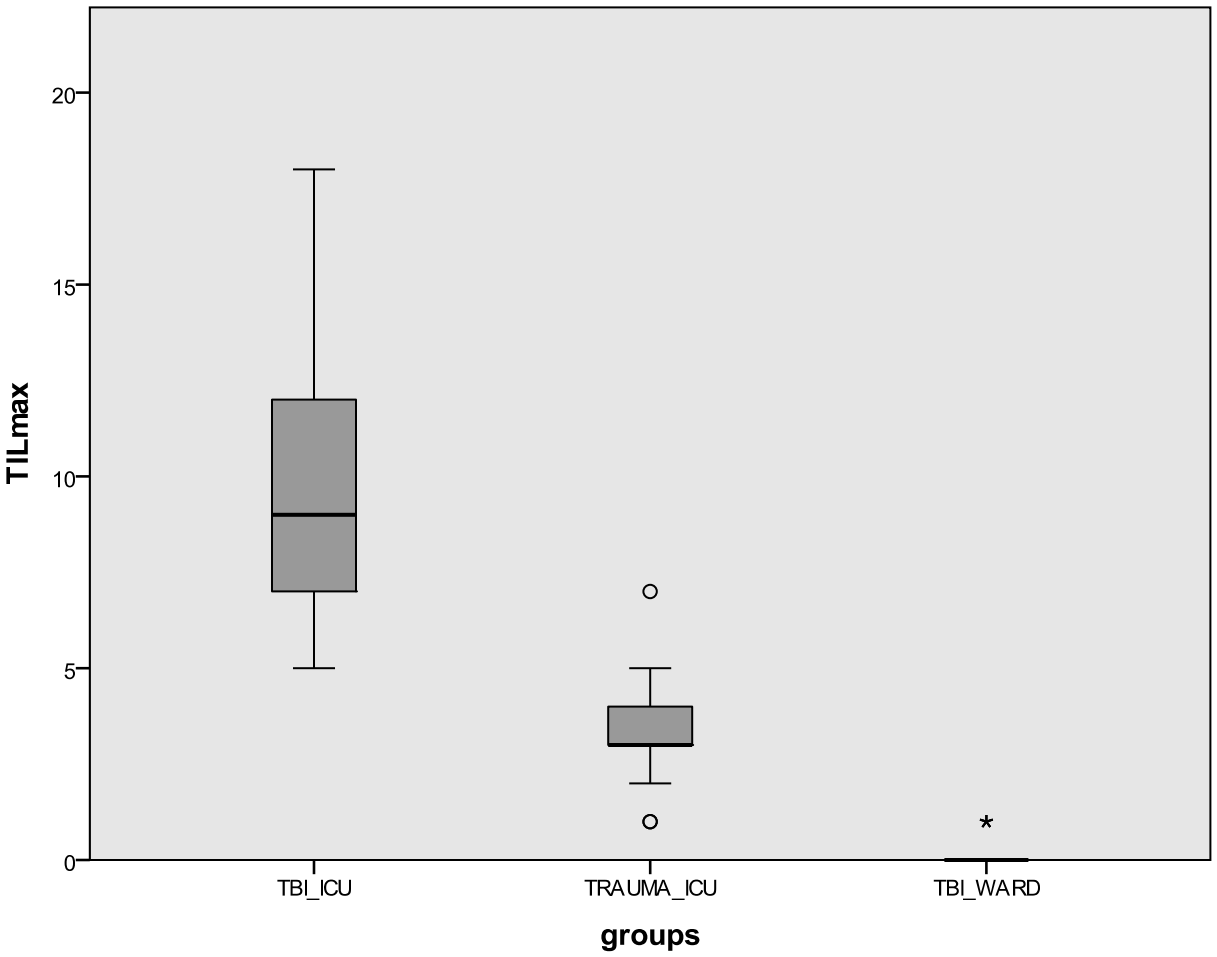

Fig. 2b.

Table 1. Therapy Intensity Level (TIL) scale				
ITEM	DETAILS	SPECIFICS	SCORE	MAX
Positioning	head elevation for ICP control		1	1
	nursed flat (180°) for CPP management		1	
Sedation and neuromuscular blockade	low dose sedation (as required for mechanical ventilation)		1	8
	higher dose sedation for ICP control (but not aiming for burst suppression)		2	
	high dose propofol or barbiturates for ICP control (metabolic suppression)		5	
	neuromuscular blockade (paralysis)		3	
CSF drainage	CSF drainage – low volume	< 120ml/day (< 5ml/h)	2	3
	CSF drainage – high volume	≥ 120ml/day (≥ 5ml/h)	3	
CPP management	fluid loading for maintenance of cerebral perfusion		1	2
	vasopressor therapy required for management of cerebral perfusion		1	
Ventilatory management(*)	mild hypocapnia for ICP control, based on arterial CO2 in mmHg	≥ 35, < 40	1	4
	moderate hypocapnia for ICP control	≥ 30, < 35	2	
	intensive hypocapnia for ICP control	< 30	4	
Hyperosmolar therapy(#)	mannitol	≤ 2g/kg/24h	2	6
	mannitol	> 2g/kg/24h	3	
	hypertonic saline	≤ 0.3g/kg/24h	2	
	hypertonic saline	> 0.3g/kg/24h	3	

Temperature control	treatment of fever (T> 38°C or spontaneous T < 34.5°C)		1	5
	cooling for ICP control, ≥ 35°C		2	
	hypothermia < 35°C		5	
Surgery for intracranial hypertension	intracranial operation for progressive mass lesion, NOT scheduled on admission		4	9
	decompressive craniectomy		5	
Maximum total possible score				38

Table 2. Demographics of cases and controls						
		TBI_ICU	Trauma_ICU	TBI_WARD	*p-value*	Bonferroni corrected *p*-value
n		31	20	19	ns	
Age (years)		38.5±17.3	45.1±21.1	51.4±18.5	ns	
Sex –male (n)		23 (74%)	15 (75%)	12 (63%)	ns	
Marshall Score	6	2 (6.5%)	-	-		
	5	2 (6.5%)	-	-		
	4	2 (6.5%)	-	-		
	3	1 (3.2%)	-	-		
	2	22 (71%)	-	-		
	1	2 (6.5%)	-	-		
GCS	3-8	27 (87%)	0 (0%)	0 (0%)	<.0005 (a)	<.005 (a)
	9-12	3 (9.7%)	0 (0%)	1 (5.3%)		
	13-15	1 (3.2%)	17 (85%)	18(94.7%)		
	Unknown*		3 (15%)			
IMPACT	Mort	23.8±12.3	-	-		
	UO	44.6±18.4	-	-		
GOS	1-3	17 (55%)	-	3 (15.8%)	.008 (a)	<.005 (a)
	4-5	14 (45%)		16 (84.2%)[y]		
TIL_{24} day 1		7.7±3.2	2.1±1.1	0.1±0.3	<.0005 (b)	<.005 (b)
TIL_{24} day 2		9.2±3.6	2.6±1.7	0.1±0.3	<.0005 (b)	<.005 (b)
TIL_{24} day 3		8.1±3.5	2.1±1	0	<.0005 (b)	<.005 (b)
TIL_{24} day 4		8±3.4	1.9±1.3	0	<.0005 (b)	<.005 (b)
TIL_{max}		9.9±3.7	3.4±1.4	0.2±0.4	<.0005 (b)	<.005 (b)
TIL_{mean}		8.2±3.2	2.2±0.9	0.1±0.1	<.0005 (b)	<.005 (b)

Table 3. Results: Reliability assessment					
Reliability		Cohen's κ	*p*-value	$ICC_{agreement}$	*p*-value
Inter-rater	TIL_4	0.981	<0.0005	0.999	<0.0005
	TIL_{24}h	1	<0.0005	1.000	<0.0005
Intra-rater	TIL_4	0.988	<0.0005	1.000	<0.0005
	TIL_{24}	0.971	<0.0005	0.999	<0.0005

Table 4. Results: Validity assessment					
type of validity	correlations	Spearman's rho (rs)	*p*-value	Bonferroni corrected *p-value*	Concordance with prediction*
criterion	TIL_Maset(4-hr) vs TIL_4	0.677	<.0005	<0.01	yes
criterion	TIL_Maset(24-hr) vs TIL_{24}	0.563	<.0005	<0.01	yes
convergent	GCS vs TIL_{mean}	-0.733	<.0005	<0.01	yes
convergent	GCS vs TIL_{max}	-0.713	<.0005	<0.01	yes
convergent	ICP_4 vs TIL_{24}	0.405	<.0005	<0.01	yes
convergent	ICP_{24} vs TIL_{24}	0.547	<.0005	<0.01	yes
convergent	ICP_{mean} vs TIL_{mean}	0.633	<.0005	<0.01	yes
convergent	ICP_{max} vs TIL_{max}	0.606	<.0005	<0.01	yes
discriminant	GOS(TBI) vs TIL_{mean}	-0.376	0.007	<0.5	yes
discriminant	GOS(TBI_ICU) vs TIL_{mean}	-0.080	0.669	ns	yes
discriminant	GOS(TBI) vs TIL_{max}	-0.385	0.006	<0.5	yes
discriminant	GOS(TBI_ICU) vs TIL_{max}	-0.098	0.598	ns	yes
discriminant	IMPACT(UO) TIL_{mean}	-0.107	0.565	ns	yes
discriminant	IMPACT(mort) vs TIL_{mean}	0.051	0.784	ns	yes
discriminant	IMPACT(UO) vs TIL_{max}	-0.079	0.673	ns	yes
discriminant	IMPACT(mort) vs TIL_{max}	0.079	0.671	ns	yes

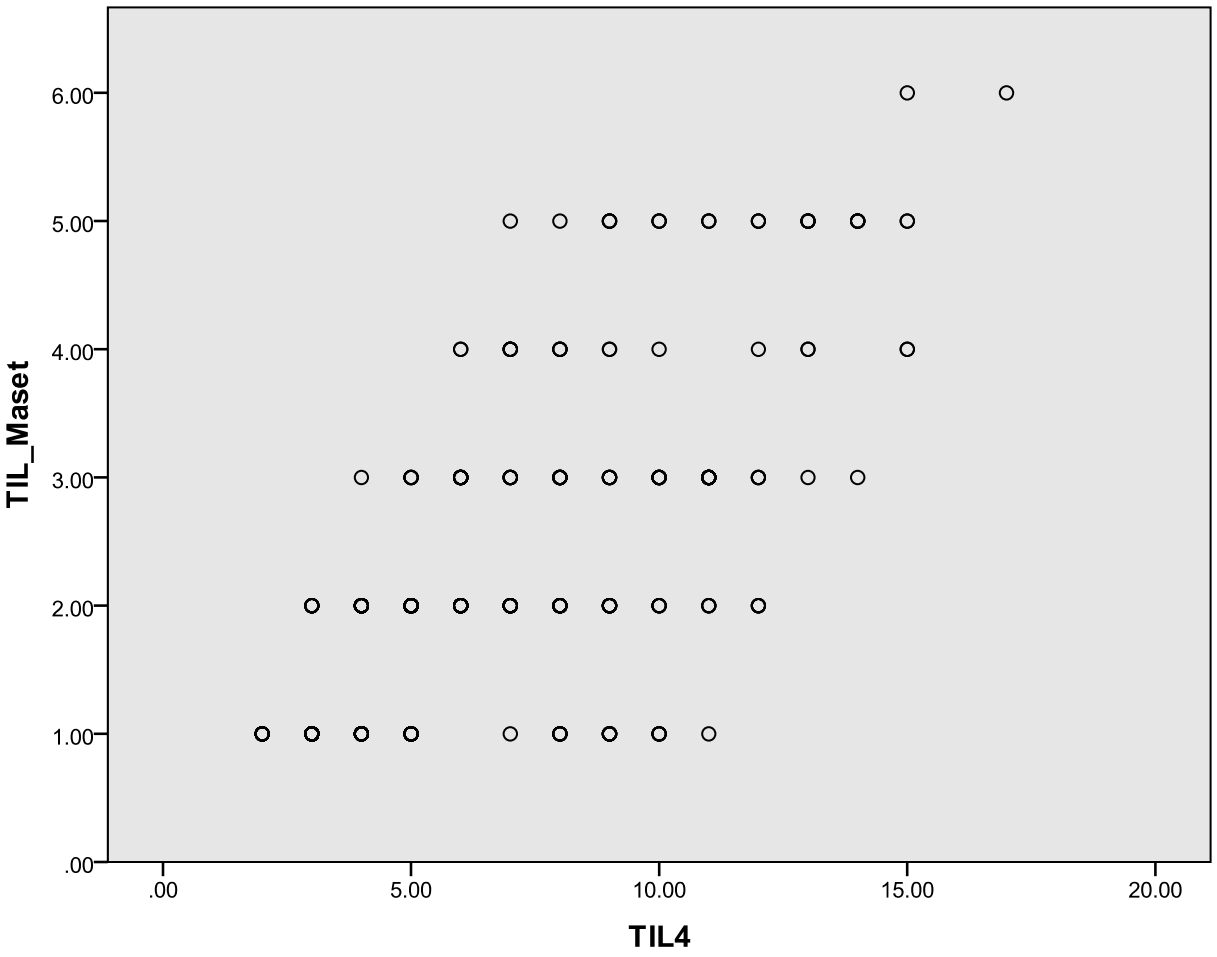

FIG. S1A.

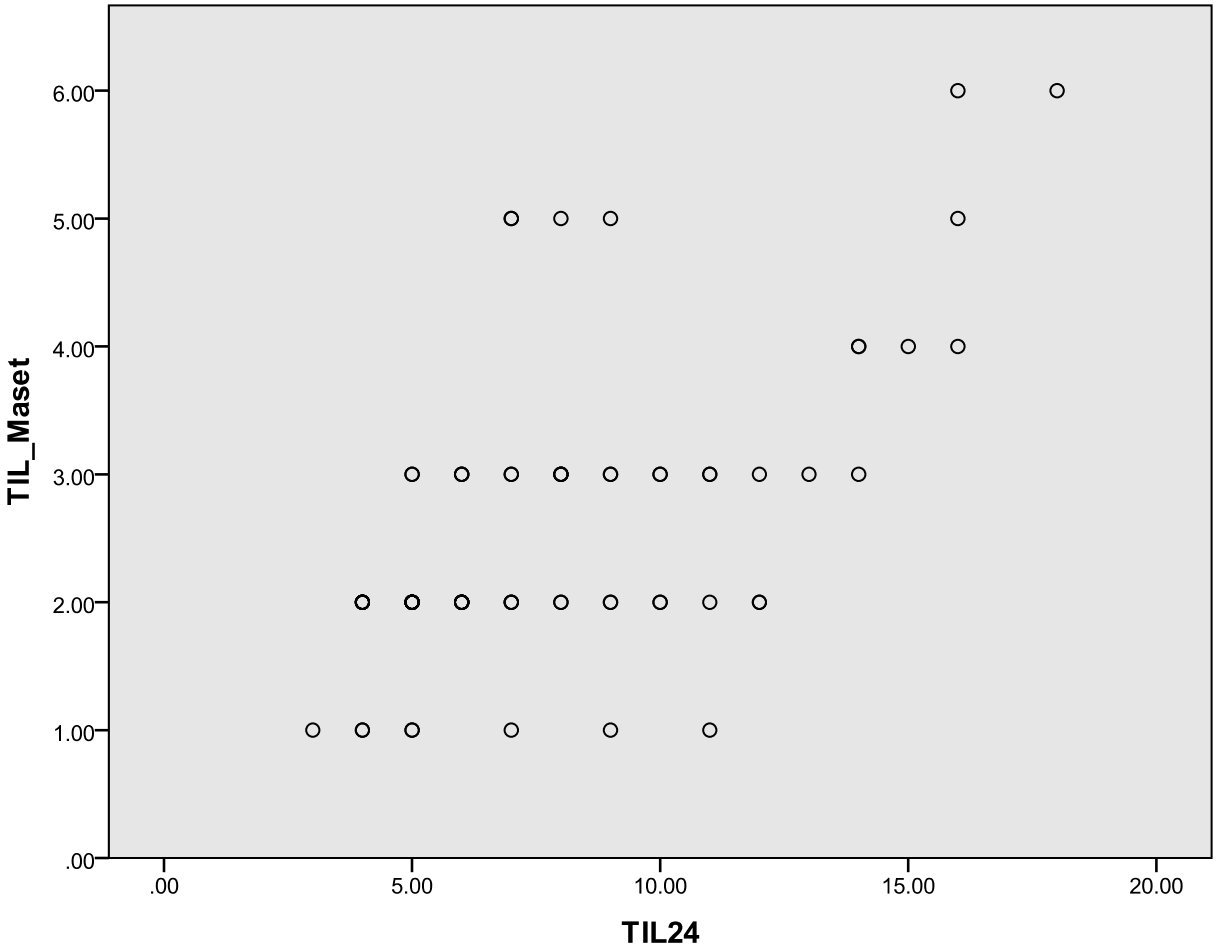
TIL_Maset
6.00
5.00
4.00
3.00
2.00
1.00
.00
.00
5.00
10.00
15.00
20.00
TIL24

FIG. S1B.

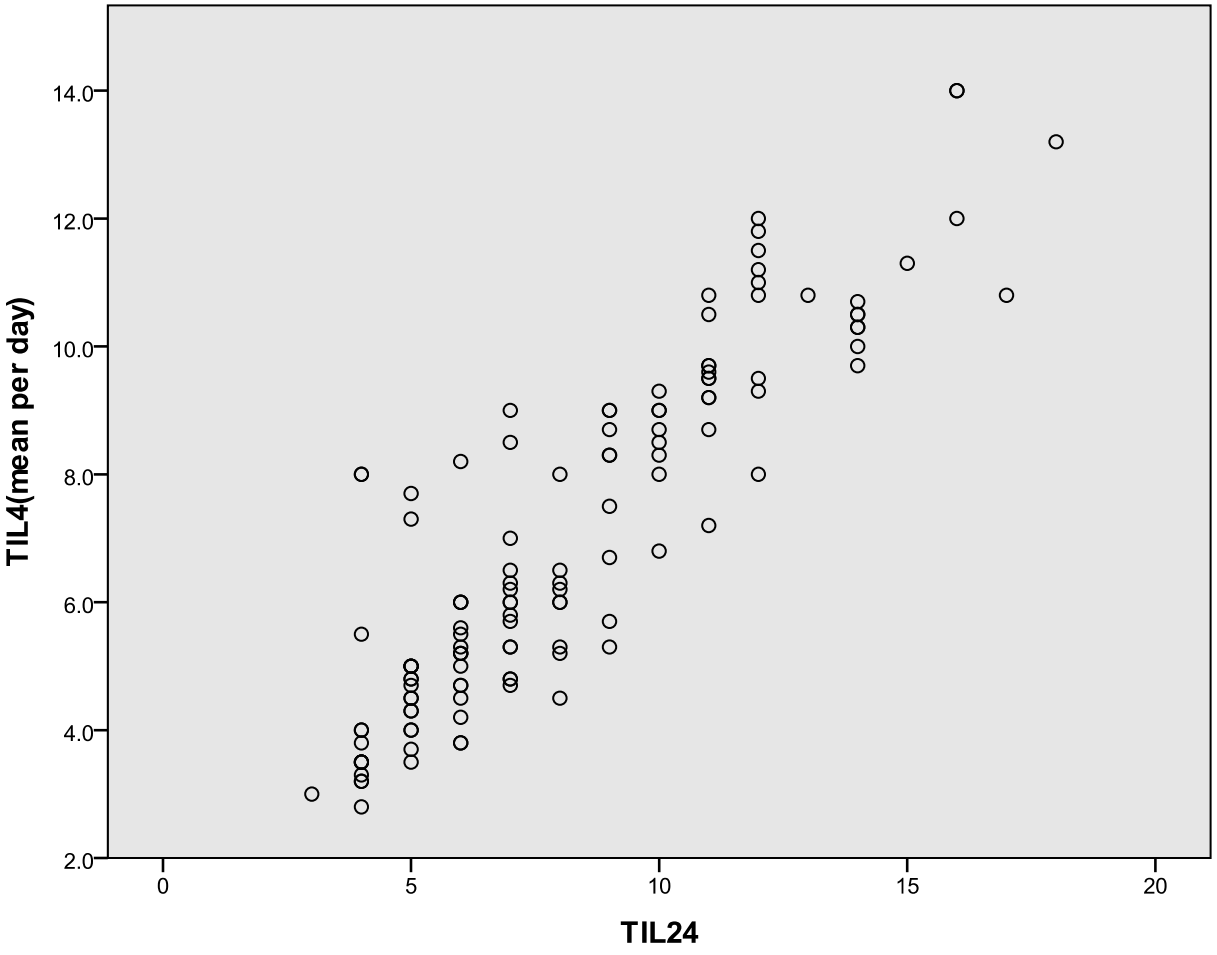

FIG. S2A.

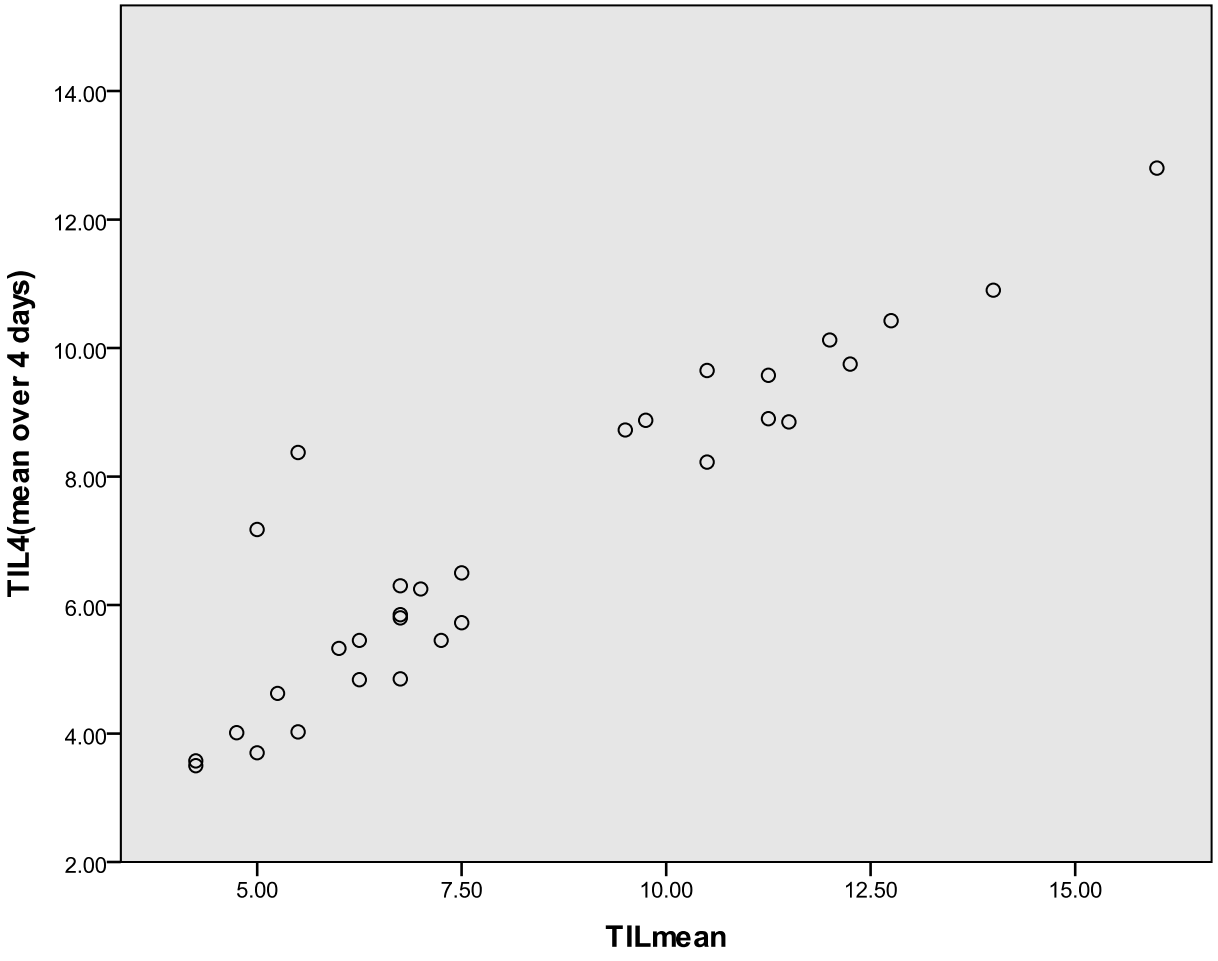
TIL4(mean over 4 days)
14.00
12.00
10.00
8.00
6.00
4.00
2.00
5.00
7.50
10.00
12.50
15.00
TILmean

FIG. S2B.

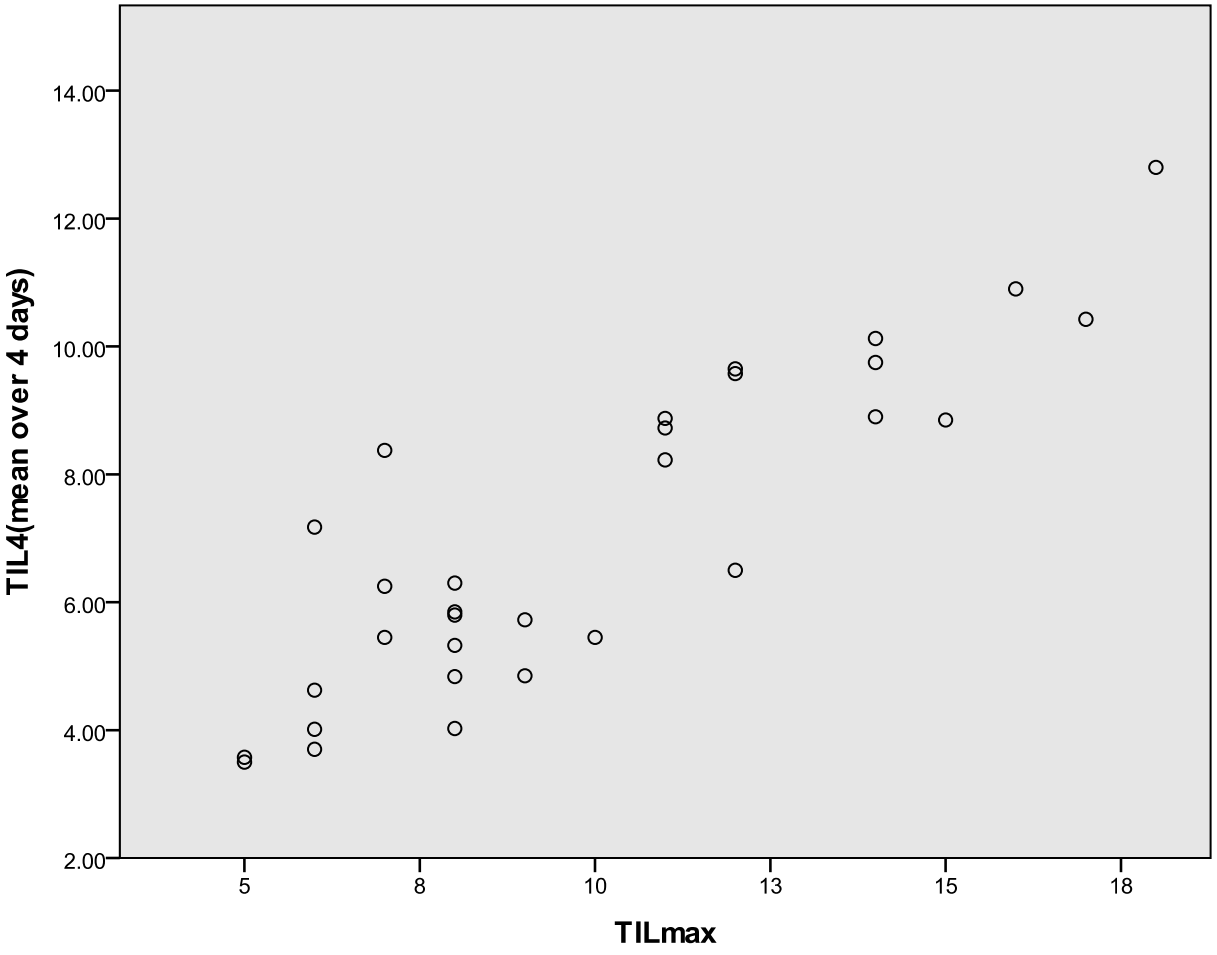
TIL4(mean over 4 days)
14.00
12.00
10.00
8.00
6.00
4.00
2.00
5
8
10
13
15
18
TILmax

FIG. S3A.

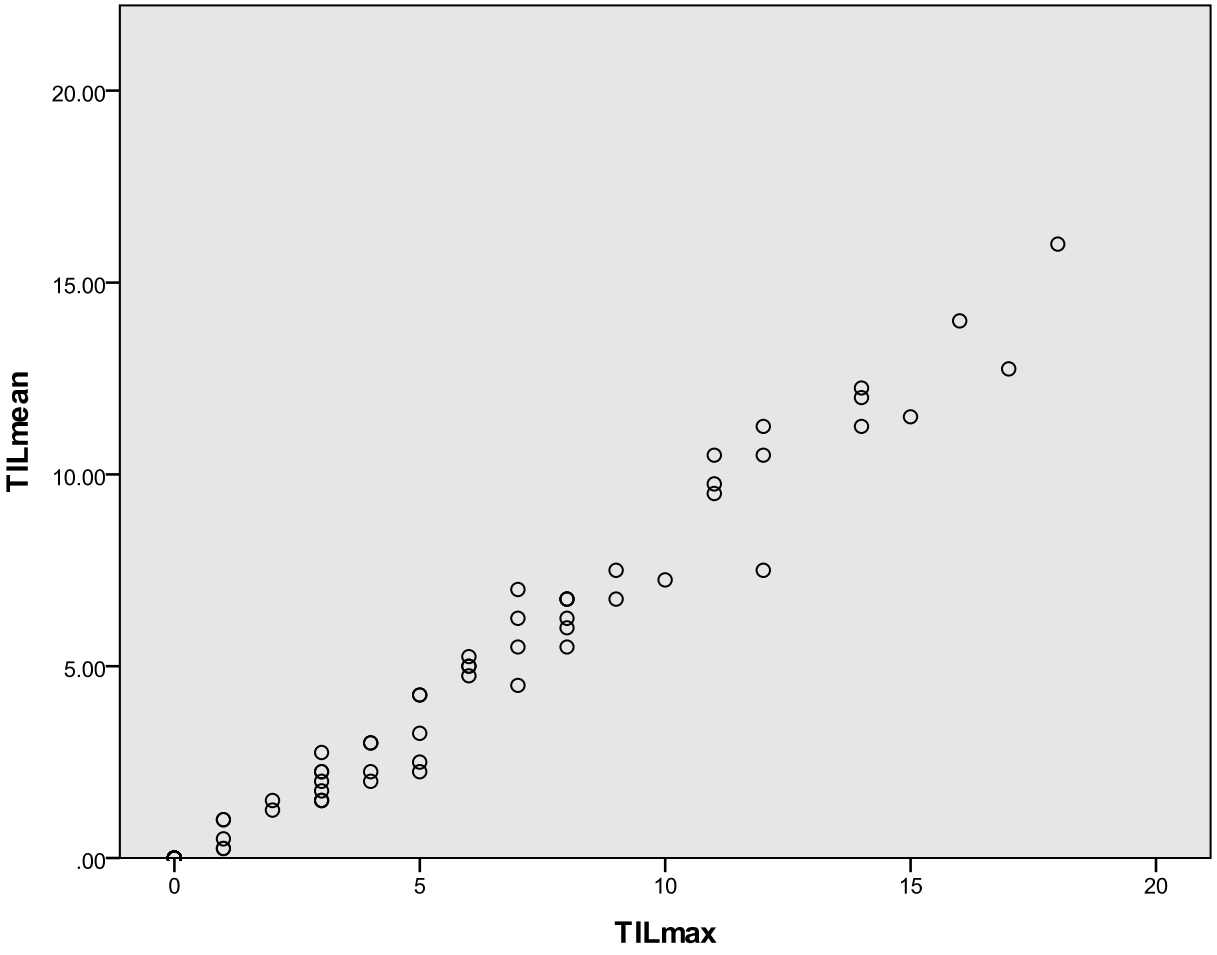

FIG. S3B.

Table S1. Correlations between TIL subtypes: TIL_4, TIL_{24}, TIL_{mean} and TIL_{max}				
	TIL_4	TIL_{24}	TIL_{mean}	TIL_{max}
TIL_4		r_s=0.882 (a)	r_s=0.892 (a)	r_s=0.847 (a)
TIL_{24}	r_s=0.882 (a)		(*)r_s=0.872-0.938	(*)r_s=0.803-0.951
TIL_{mean}	r_s=0.892 (a)	(*)r_s=0.872-0.938		r_s=0.992 (b)
TIL_{max}	r_s=0.847 (a)	(*)r_s=0.803-0.951	r_s=0.992 (b)	

Legends to figures and tables – Main Manuscript

Figure 1A&1B.

Correlation between the TIL scale score and ICP (1A, 4-hr assessments, *n*=872; 1B, 24-hr assessments, *n*=124).

Figure 2A&2B.

Construct, discriminative validity of the TIL scale. Kruskal-Wallis test, independent samples. Boxplot shows TIL_{mean} (2A) and TIL_{max} (2B). The TIL scale can significantly discriminate between groups, for TIL_{mean}, $H(2)=60.55$, for TIL_{max} $H(2)=59.39$, $p < 0.0005$ for both. Pairwise comparisons with adjusted p-values showed that there are significant differences between all groups: TBI_ICU vs Trauma_ICU ($p < 0.005$), TBI_ICU vs TBI_Ward ($p < 0.0005$) and Trauma_ICU vs TBI_Ward ($p = 0.007$) for both.

Table 1.

The scheme for Therapy Intensity Level assessment, based on Maas et al.,[11] minimally adapted. Initial problems in the pilot phase which were subsequently addressed were:

(*) Conversions between kPa and mmHg for $PaCO_2$ were ambiguous because of "rounding up" errors . We consequently decided to base our calculations on mmHg, which resulted in less ambiguous cut-offs.

(#) For the 4-hr assessments we used 0.33 g/kg/4h for mannitol and 0.05 g/kg/4h for hypertonic saline to assign a score value. However, in the pilot phase of the study, these cutoffs were calculated from the 24 hour thresholds (inconsistently) by individual raters. In addition, because of lack of clarity in scoring instructions, some cases scored maximally in this category for the 4-hr assessment, but were wrongly not scored as maximal for the 24-hr assessment, as total dose of hyperosmolar agent did not exceed thresholds when averaged over 24 hours. In a revised version, we explicitly stated that if dose of hyperosmolar agent exceeded a given threshold in any 4-hour epoch, the same score should apply to the 24-hour period in which that 4-hour epoch was contained.

Table 2.

TBI_ICU: Patients with TBI and elevated intracranial pressure in need of intensive care;

Trauma_ICU: patients with extracranial trauma in need of intensive care;

TBI_WARD: patients with mild/moderate TBI not requiring ICP directed therapy or ICU admission;

GOS: Glasgow Outcome Scale;

IMPACT: Outcome prediction from IMPACT scheme (Mort: mortality; UO: unfavourable outcome; both calculated from extended IMPACT calculation (core model +CT +Lab);

GCS: Glasgow Coma Scale; data and age shown are mean ± SD or frequency and percentage; (-) not assessed;

* in three patients with extracranial trauma, no reliable pre-sedation GCS was available;

(y) for 10 patients 6-month GOS data was not available at the time of data analysis but since favourable outcome was achieved at an earlier follow up point, this was assumed to have been maintained subsequently for the purposes of this analysis.

(a) Fisher's Exact Test

(b) *p*-value & Bonferroni corrected *p*-value for TBI_ICU vs. all other groups

Table 3

Reliability assessment with Cohen's κ and Intraclass Correlation Coefficients (aimed for agreement) after pilot phase testing with subsequent rulebook optimization.

Table 4.

Correlation coefficients regarding criterion and construct validity. The results were all in agreement to the predictions (*) made prior to data analysis.

Legends to figures and tables – Supplemental Material

Figure S1A&S1B.

TIL vs TIL_Maset (S1A, 4-hr, *n*=426; S1B, 24-hr, *n*=76), showing positive correlation, no ceiling effect.

Figure S2A&2B.

(S2A) Correlation between TIL_{24} and TIL_4 (mean per day) in the TBI_ICU group (patient days *n*=124). Spearman's rho r_s=0.882, p<0.0005. (S2B) Correlation between TIL_{mean} and TIL_4 (mean over 4 days) in the TBI_ICU group (*n*=31). Spearman's rho r_s=0.892, p<0.0005.

Figure S3A&S3B.

Correlation analysis for TIL_{max} vs TIL_4 (mean over 4 days) (S3A, *n*=31, Spearman's rho r_s = 0.847, p < 0.0005) and vs TIL_{mean} (S3B, *n*=70, Spearman's rho r_s = 0.992, p < 0.0005).

Table S1.

Data shown: Spearman's rho r_s, p<0.0005 for every correlation.

(*) plotted against TIL_{24} for day 1 to day 4, lowest and highest correlations showed

(a) plotted in TBI_ICU group only, n=31

(b) plotted in all groups, n=70