Brier score

The Brier score is a proper score function that measures the accuracy of probabilistic predictions. It is applicable to tasks in which predictions must assign probabilities to a set of mutually exclusive discrete outcomes. The set of possible outcomes can be either binary or categorical in nature, and the probabilities assigned to this set of outcomes must sum to one (where each individual probability is in the range of 0 to 1). It was proposed by Glenn W. Brier in 1950.

The Brier score can be thought of as either a measure of the "calibration" of a set of probabilistic predictions, or as a "cost function". More precisely, across all items {\displaystyle i\in {1...N}} i\in{1...N} in a set N predictions, the Brier score measures the mean squared difference between:

The predicted probability assigned to the possible outcomes for item i The actual outcome {\displaystyle o_{i}} o_i Therefore, the lower the Brier score is for a set of predictions, the better the predictions are calibrated. Note that the Brier score, in its most common formulation, takes on a value between zero and one, since this is the largest possible difference between a predicted probability (which must be between zero and one) and the actual outcome (which can take on values of only 0 or 1). In the original (1950) formulation of the Brier score, the range is double, from zero to two.

The Brier score is appropriate for binary and categorical outcomes that can be structured as true or false, but is inappropriate for ordinal variables which can take on three or more values.

Cervical spine injuries can occur in military scenarios from events such as underbody blast events. Such scenarios impart inferior-to-superior loads to the spine. The objective of this study is to develop human injury risk curves (IRCs) under this loading mode using Post Mortem Human Surrogates (PMHS). Twenty-five PMHS head-neck complexes were obtained, screened for pre-existing trauma, bone densities were determined, pre-tests radiological images were taken, fixed in polymethylmethacrylate at the T2-T3 level, a load cell was attached to the distal end of the preparation, positioned end on custom vertical accelerator device based on the military-seating posture, donned with a combat helmet, and impacted at the base. Posttest images were obtained, and gross dissection was done to confirm injuries to all specimens. Axial and resultant forces at the cervico-thoracic joint was used to develop the IRCs using survival analysis. Data were censored into left, interval, and uncensored observations. The Brier score metric was used to rank the variables. The optimal metric describing the underlying response to injury was associated with the axial force, ranking slightly greater than the resultant force, both with BMD covariates. The results from the survival analysis indicated all IRCs are in the "fair" to "good" category, at all risk levels. The BMD was found to be a significant covariate that best describes the response of the helmeted head-neck specimens to injury. The present experimental protocol and IRCs can be used to conduct additional tests, matched-pair tests with the WIAMan and/or other devices to obtain injury assessment risk curves (IARCs) and injury assessment risk values (IARVs) to predict injury in crash environments, and these data can also be used for validating component-based head-neck and human body computational models¹⁾.

Outcomes from craniosynostosis surgeries performed between 2012 and 2016 at our academic pediatric hospital were evaluated using the NSQIP-P risk calculator. Descriptive statistics were performed comparing predicted 30-day postoperative events and clinically observed outcomes. The

performance of the calculator was evaluated using the Brier score and receiver operating characteristic curve (ROC).

RESULTS: A total of 202 craniosynostosis surgeries were included. Median age was 0.74 years (range 0.15-6.32); 66% were males. Blood transfusion occurred in 162/202 patients (80%). The following clinical characteristics were statistically correlated with surgical complications: American Society of Anesthesiologists physical status classification >1 (P<0.001), central nervous system abnormality (P<0.001), syndromic craniosynostosis (P=0.001), and redo operations (P=0.002). Postoperative events occurred in <3%, including hardware breakage, tracheal-cartilaginous sleeve associated with critical airway, and surgical site infection. The calculator performed well in predicting any complication (Brier=0.067, ROC=73.9%), and for pneumonia (Brier=0.0049, ROC 99%). The calculator predicted a low rate of cardiac complications, venous thromboembolism, renal failure, reintubation, and death; the observed rate of these complications was 0.

CONCLUSIONS: The risk calculator demonstrated reasonable ability to predict the low number of perioperative complications in patients undergoing craniosynostosis surgery with a composite complications outcome. Efforts to improve the calculator may include further stratification based on procedure-specific risk factors²⁾.

In automotive events, head injuries (skull fractures and/or brain injuries) are associated with head contact loading. While the widely-used head injury criterion is based on frontal bone fracture and linear accelerations, injury risk curves were not developed from original datasets.

OBJECTIVES: Develop skull fracture-based risk curves for using previously published data and apply resampling techniques to assess their qualities.

METHODS: Force, deflection, energy, and stiffness data from thirteen human cadaver head impact tests were used to develop risk curves using parametric survival analysis. Injuries occurred to all specimens. Data points were treated as uncensored. Variables were ranked, and the variable best explaining the underlying fracture response was determined using the Brier Score Metric (BSM). The qualities of the risk curves were determined using normalized confidence interval sizes. Statistical resampling methods were used to assess the quality of the risk curves and the impact of the sample size by conducting 2000 simulations. Sample sizes ranged from 13 to 26.

FINDINGS: The Weibull distribution was optimal for all the response variables, except deflection (loglogistic). The quality of the risk curves was the highest for deflection. This variable best explained the underlying head injury response, based on BSM. Improvements in the quality of the risk curves were achieved with additional samples of force and deflection (<13), while energy and stiffness variables required more size. Individual risk curves are given.

INTERPRETATION: These probability curves from head contact loading add to the understanding skull fractures and can be used to improve safety in injury producing environments ³⁾.

The aim of this study was to develop an effective surgical site infection (SSI) prediction model in patients receiving free-flap reconstruction after surgery for head and neck cancer using artificial neural network (ANN), and to compare its predictive power with that of conventional logistic regression (LR).

MATERIALS AND METHODS: There were 1,836 patients with 1,854 free-flap reconstructions and 438 postoperative SSIs in the dataset for analysis. They were randomly assigned tin ratio of 7:3 into a training set and a test set. Based on comprehensive characteristics of patients and diseases in the absence or presence of operative data, prediction of SSI was performed at two time points (pre-operatively and post-operatively) with a feed-forward ANN and the LR models. In addition to the calculated accuracy, sensitivity, and specificity, the predictive performance of ANN and LR were assessed based on area under the curve (AUC) measures of receiver operator characteristic curves and Brier score.

RESULTS: ANN had a significantly higher AUC (0.892) of post-operative prediction and AUC (0.808) of pre-operative prediction than LR (both P<0.0001). In addition, there was significant higher AUC of post-operative prediction than pre-operative prediction by ANN (p<0.0001). With the highest AUC and the lowest Brier score (0.090), the post-operative prediction by ANN had the highest overall predictive performance.

CONCLUSION: The post-operative prediction by ANN had the highest overall performance in predicting SSI after free-flap reconstruction in patients receiving surgery for head and neck cancer ⁴⁾.

The All Patient Refined Diagnosis Related Group (APR-DRG) is an inpatient visit classification system that assigns a diagnostic related group, a Risk of Mortality (ROM) subclass and a Severity of Illness (SOI) subclass. While extensively used for cost adjustment, no study has compared the APR-DRG subclass modifiers to the popular Charlson Comorbidity Index as a measure of comorbidity severity in models for perioperative in-Hospital mortality. In this study we attempt to validate the use of these subclasses to predict mortality in a cohort of surgical patients. We analyzed all adult (age over 18 years) inpatient non-cardiac surgery at our institution between December 2005 and July 2013. After exclusions, we split the cohort into training and validation sets. We created prediction models of inpatient mortality using the Charlson Comorbidity Index, ROM only, SOI only, and ROM with SOI. Models were compared by receiver-operator characteristic (ROC) curve, area under the ROC curve (AUC), and Brier score. After exclusions, we analyzed 63,681 patient-visits. Overall in-Hospital mortality was 1.3%. The median number of ICD-9-CM diagnosis codes was 6 (Q1-Q3 4-10). The median Charlson Comorbidity Index was 0 (Q1-Q3 0-2). When the model was applied to the validation set, the c-statistic for Charlson was 0.865, c-statistic for ROM was 0.975, and for ROM and SOI combined the c-statistic was 0.977. The scaled Brier score for Charlson was 0.044, Brier for ROM only was 0.230, and Brier for ROM and SOI was 0.257. The APR-DRG ROM or SOI subclasses are better predictors than the Charlson Comorbidity Index of in-Hospital mortality among surgical patients⁵⁾.

Head contact-induced loads can result in skull fractures and/or brain injuries. While skull fractures have been produced from post-mortem human cadaver surrogates (PMHS), injury probability curves describing their structural responses have not been developed. The objectives of this study were to develop skull fracture-based injury risk curves and describe human tolerances using survival analysis. Published PMHS data in this journal were used. Mean age, stature, and weight of 12 PMHS were: 66.6 ± 2.3 years, 1.71 ± 2.9 m, and 76.4 ± 4.6 kg. A testing device applied contact loading to the head. Failure force, deflection, energy, and linear and secant stiffness variables were used to develop probability curves. Parametrical survival analysis included identifying most optimal distribution, ensuring that the chosen distribution is not significantly different from the nonparametrical model, determining $\pm 95\%$ confidence interval bounds and Normalized Confidence Interval Sizes (NICS), obtaining quality indices for each risk curve, and determining their hierarchical sequence using the

Brier score metric (BSM). Lognormal distribution was the most optimal distribution for all variables, except failure force, for which Weibull distribution was optimal. Tightness-of-fit of risk curves for failure force, energy, and deflection were better than linear and secant stiffness variables. Force best represented skull fracture response based on BSM and NCIS, followed by deflection and energy, while two stiffness variables were least preferred metrics. These structural response-based set of risk curves, hitherto not reported, form a fundamental dataset for validating/assessing accuracy of outputs from computational models and serve as hierarchical skull fracture injury criteria under head contact loads ⁶⁾.

The risk calculator of the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) has been shown to be useful in predicting postoperative complications. In this study, we aimed to evaluate the predictive value of the ACS-NSQIP calculator in geriatric patients undergoing lumbar surgery. A total of 242 geriatric patients who underwent lumbar surgery between January 2014 and December 2016 were included. Preoperative clinical information was retrospectively reviewed and entered into the ACS-NSQIP calculator. The predictive value of the ACS-NSQIP model was assessed using the Hosmer-Lemeshow test, Brier score (B), and receiver operating characteristics (ROC, also referred C-statistic) curve analysis. Additional risk factors were calculated as surgeon-adjusted risk including previous cardiac event and cerebrovascular disease. Preoperative risk factors including age (P = .004), functional independence (P = 0), American Society of Anesthesiologists class (ASA class, P=0), dyspnea (P=0), dialysis (P=.049), previous cardiac event (P=.001), and history of cerebrovascular disease (P=0) were significantly associated with a greater incidence of postoperative complications. Observed and predicted incidence of postoperative complications was 43.8% and 13.7% (\pm 5.9%) (P<.01), respectively. The Hosmer-Lemeshow test demonstrated adequate predictive accuracy of the ACS-NSQIP model for all complications. However, Brier score showed that the ACS-NSQIP model could not accurately predict risk of all (B=0.321) or serious (B = 0.241) complications, although it accurately predicted the risk of death (B = 0.0072); this was supported by ROC curve analysis. The ROC curve also showed that the model had high sensitivity and specificity for predicting renal failure and readmission. The ACS-NSQIP surgical risk calculator is not an accurate tool for the prediction of postoperative complications in geriatric Chinese patients undergoing lumbar surgery $^{7)}$.

The WHO classification of brain tumours describes 15 subtypes of meningioma. Nine of these subtypes are allotted to WHO grade I, and three each to grade II and grade III. Grading is based solely on histology, with an absence of molecular markers. Although the existing classification and grading approach is of prognostic value, it harbours shortcomings such as ill-defined parameters for subtypes and grading criteria prone to arbitrary judgment. In this study, we aimed for a comprehensive characterisation of the entire molecular genetic landscape of meningioma to identify biologically and clinically relevant subgroups.

METHODS: In this multicentre, retrospective analysis, we investigated genome-wide DNA methylation patterns of meningiomas from ten European academic neuro-oncology centres to identify distinct methylation classes of meningiomas. The methylation classes were further characterised by DNA copy number analysis, mutational profiling, and RNA sequencing. Methylation classes were analysed for progression-free survival outcomes by the Kaplan-Meier method. The DNA methylation-based and WHO classification schema were compared using the Brier prediction score, analysed in an independent cohort with WHO grading, progression-free survival, and disease-specific survival data available, collected at the Medical University Vienna (Vienna, Austria), assessing methylation patterns with an alternative methylation chip.

FINDINGS: We retrospectively collected 497 meningiomas along with 309 samples of other extra-axial skull tumours that might histologically mimic meningioma variants. Unsupervised clustering of DNA methylation data clearly segregated all meningiomas from other skull tumours. We generated genome-wide DNA methylation profiles from all 497 meningioma samples. DNA methylation profiling distinguished six distinct clinically relevant methylation classes associated with typical mutational, cytogenetic, and gene expression patterns. Compared with WHO grading, classification by individual and combined methylation classes more accurately identifies patients at high risk of disease progression in tumours with WHO grade I histology, and patients at lower risk of recurrence among WHO grade II tumours (p=0.0096) from the Brier prediction test). We validated this finding in our independent cohort of 140 patients with meningioma.

INTERPRETATION: DNA methylation-based meningioma classification captures clinically more homogenous groups and has a higher power for predicting tumour recurrence and prognosis than the WHO classification. The approach presented here is potentially very useful for stratifying meningioma patients to observation-only or adjuvant treatment groups. We consider methylation-based tumour classification highly relevant for the future diagnosis and treatment of meningioma⁸⁾.

Breast cancer (BC) is the second most common cause of brain metastases (BM). Optimal management of BM from BC is still debated. In an attempt to provide appropriate treatment and to assist with optimal patient selection, several specific prognostic classifications for BM from BC have been established. We evaluated the prognostic value and validity of the 6 proposed scoring systems in an independent population of BC patients with BM.

METHODS: We retrospectively reviewed all consecutive BC patients referred to our institution for newly diagnosed BM between October 1995 and July 2011 (n = 149). Each of the 6 scores proposed for BM from BC (Sperduto, Niwinska, Park, Nieder, Le Scodan, and Claude) was applied to this population. The discriminative ability of each score was assessed using the Brier score and the Cindex. Individual prognostic values of clinical and histological factors were analyzed using uni- and multivariate analyses.

RESULTS: Median overall survival was 15.1 months (95% CI,11.5-18.7). Sperduto-GPA (P < .001), Nieder (P < .001), Park (P < .001), Claude (P < .001), Niwinska (P < .001), and Le Scodan (P = .034) scores all showed significant prognostic value. The Nieder score showed the best discriminative ability (C-index, 0.672; Brier score error reduction, 16.1%).

CONCLUSION: The majority of prognostic scores were relevant for patients with BM from BC in our independent population, and the Nieder score seems to present the best predictive value but showed a relatively low positive predictive value. Thus, these results remain insufficient and challenge the routine use of these scoring systems ⁹.

study aimed to identify models that predicted the short-term outcome after traumatic brain injury (TBI) from the literature and to evaluate their clinical significance.

METHODS: Literatures from PubMED were reviewed. Regression coefficients and intercepts were extracted. A group of 229 cases was used for validation and the unfavourable rate was calculated to assess the validity of these models by the area under receiver operating. Characteristic curve (AUC),

C-statistic and Brier score.

MAIN RESULTS: In total, 13 studies of 18 different models were included. Data from the validation group were in accordance with the indicators of the studies reviewed. All models got an AUC value ranging from 0.644-0.890 except two (AUC value <0.6) and their Brier scores were near zero. However, the calibration of most studies was insufficient (p < 0.05).

CONCLUSIONS: Most of the models included in this study have a good discriminatory power while lacking sufficient calibration. However, they all predict with relative accuracy at the level of individuals. Therefore, current models can be used to predict the survival rate of individual patients and may be useful to inform patients and relatives about the likelihood of a beneficial outcome ¹⁰.

1)

Yoganandan N, Chirvi S, Pintar FA, Banerjee A, Voo L. Injury Risk Curves for the Human Cervical Spine from Inferior-to-Superior Loading. Stapp Car Crash J. 2018 Nov;62:271-292. PubMed PMID: 30608997.

Gadgil N, Pan IW, Babalola S, Lam S. Evaluating the National Surgical Quality Improvement Program-Pediatric Surgical Risk Calculator for Pediatric Craniosynostosis Surgery. J Craniofac Surg. 2018 Sep;29(6):1546-1550. doi: 10.1097/SCS.000000000004654. PubMed PMID: 29877982.

DeVogel N, Banerjee A, Yoganandan N. Application of resampling techniques to improve the quality of survival analysis risk curves for human frontal bone fracture. Clin Biomech (Bristol, Avon). 2018 Apr 21. pii: S0268-0033(18)30346-2. doi: 10.1016/j.clinbiomech.2018.04.013. [Epub ahead of print] PubMed PMID: 29753560.

4)

Kuo PJ, Wu SC, Chien PC, Chang SS, Rau CS, Tai HL, Peng SH, Lin YC, Chen YC, Hsieh HY, Hsieh CH. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. Oncotarget. 2018 Feb 9;9(17):13768-13782. doi: 10.18632/oncotarget.24468. eCollection 2018 Mar 2. PubMed PMID: 29568393; PubMed Central PMCID: PMC5862614.

McCormick PJ, Lin HM, Deiner SG, Levin MA. Validation of the All Patient Refined Diagnosis Related Group (APR-DRG) Risk of Mortality and Severity of Illness Modifiers as a Measure of Perioperative Risk. J Med Syst. 2018 Mar 22;42(5):81. doi: 10.1007/s10916-018-0936-3. PubMed PMID: 29564554.

Yoganandan N, Banerjee A. Survival Analysis-Based Human Head Injury Risk Curves: Focus on Skull Fracture. J Neurotrauma. 2018 Jun 1;35(11):1272-1279. doi: 10.1089/neu.2017.5356. Epub 2018 Mar 29. PubMed PMID: 29409390.

Wang X, Hu Y, Zhao B, Su Y. Predictive validity of the ACS-NSQIP surgical risk calculator in geriatric patients undergoing lumbar surgery. Medicine (Baltimore). 2017 Oct;96(43):e8416. doi: 10.1097/MD.000000000008416. PubMed PMID: 29069040; PubMed Central PMCID: PMC5671873.

Sahm F, Schrimpf D, Stichel D, Jones DTW, Hielscher T, Schefzyk S, Okonechnikov K, Koelsche C, Reuss DE, Capper D, Sturm D, Wirsching HG, Berghoff AS, Baumgarten P, Kratz A, Huang K, Wefers AK, Hovestadt V, Sill M, Ellis HP, Kurian KM, Okuducu AF, Jungk C, Drueschler K, Schick M, Bewerunge-Hudler M, Mawrin C, Seiz-Rosenhagen M, Ketter R, Simon M, Westphal M, Lamszus K, Becker A, Koch A, Schittenhelm J, Rushing EJ, Collins VP, Brehmer S, Chavez L, Platten M, Hänggi D, Unterberg A, Paulus W, Wick W, Pfister SM, Mittelbronn M, Preusser M, Herold-Mende C, Weller M, von Deimling A. DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. Lancet Oncol. 2017 May;18(5):682-694. doi: 10.1016/S1470-2045(17)30155-9. Epub 2017 Mar 15. PubMed PMID: 28314689.

9)

Tabouret E, Metellus P, Gonçalves A, Esterni B, Charaffe-Jauffret E, Viens P, Tallet A. Assessment of prognostic scores in brain metastases from breast cancer. Neuro Oncol. 2014 Mar;16(3):421-8. doi: 10.1093/neuonc/not200. Epub 2013 Dec 4. PubMed PMID: 24311640; PubMed Central PMCID: PMC3922513.

7/7

10)

Xu XY, Liu WG, Yang XF, Li LQ. Evaluation of models that predict short-term outcome after traumatic brain injury. Brain Inj. 2007 Jun;21(6):575-82. Review. PubMed PMID: 17577708.

From: https://neurosurgerywiki.com/wiki/ - **Neurosurgery Wiki**

Permanent link: https://neurosurgerywiki.com/wiki/doku.php?id=brier_score



Last update: 2024/06/07 02:49