The Utility of Scoring Systems in Critically Ill Patients

Manuel Boller, Dr. med. vet., MTR, DACVECC
Faculty of Veterinary Science, University of Melbourne, Melbourne, Victoria, Australia

Introduction

Severity of illness scores for human ICU patients were developed in the 1980s in order to objectively characterize populations according to the level of overall disease severity. The most widely used scoring systems in human critical care are the Acute Physiology and Chronic Health Evaluation (APACHE), the Mortality Prediction Model (MPM) and the Simplified Acute Physiology Score (SAPS) for point assessment, and the Multiorgan Dysfunction Score (MODS) and Sequential Organ Failure Assessment (SOFA) for longitudinal severity of illness determination. Veterinary scoring systems relevant to emergency and critical care include the Animal Trauma Triage (ATT), the Survival Prediction Index (SPI), the Modified Glasgow Coma Scale (MGCS) and the Acute Patient Physiologic and Laboratory Evaluation (APPLE). The principle design idea is the same for all scoring systems: they attempt to relate a quantitative assessment of a patient’s physiologic response to disease and comorbidities with patient outcome, most commonly mortality or duration of hospitalization. The intention of scoring systems is in line with the long-standing desire and possibly duty of every clinician to provide an estimate of what likely course a patient’s ailment will take. On the first view, severity of illness scores thus appear to represent an effort to systematically and objectively substantiate the subjective prediction of outcome of an individual, medical or veterinary professional whose judgment could be influenced by all kinds of biases. The development, validation and therefore interpretation of the objective models that describe this relationship between patient (independent) and outcome (dependent) variables can however be quite complex. This presentation will provide insight and deeper understanding of these complexities. It will further explain the expected benefits of scoring systems in the domains of research (i.e., control of confounding; indication of effective randomization; external validity), clinical medicine (i.e., individual patient risk prediction) and administration (i.e., resource allocation; system performance surveillance).

Development of Illness Severity Scores

The underlying principle of score development consists of modelling within the population of interest the relationship between patient variables and patient outcomes and thus all three elements (population, patient and outcome variables), need to be carefully chosen. If the population is too narrowly defined, external applicability (external validity) may be limited. Variables selected for model development have to be accompanied by clear operational definition to reduce errors of score application. Moreover, variables should be objective, easy and affordable to measure, clinically relevant, known to relate to outcome, and independent of institution or clinician-specific processes. Clearly, a model, however accurate, that includes variables that are clinically not feasible to collect is useless. In a next step, suitable variable candidates will be assessed by multivariate logistic regression analysis in order to select the variables that most significantly inform a specific outcome. Typical general outcomes used in veterinary medicine are survival to discharge or duration of intensive care, both of which are confounded by economic considerations. A subsequent validation step is important to determine how well the score system performs in predicting the outcome it was modelled after. The validation dataset used for that purpose should be different from the model construction data set and ideally, datasets from other institutions should be used to assess for external validity. Validation focuses on two characteristics. Score system calibration describes how the model performs over the available range of outcomes. Good calibration indicates that a model is accurate in predicting good outcomes as well as poor outcomes. Model discrimination indicates how the model correctly distinguishes, or discriminates between survivors and non-survivors.

Usefulness in Clinical Medicine

As severity of illness scores are typically derived from clinically relevant physiological variables they can be used for patient evaluation at a particular time point (e.g., at admission to hospital or ICU) or serially over the course of hospitalization for trending. As such these scores allow an overall objective assessment of a patient’s severity of illness that is based on a large patient population surpassing the clinician’s own personal experience. However, for the same reason it is problematic to utilize these scores in isolation of conventional clinical assessment for therapeutic decision making regarding the care of individual patients. The scores are modelled towards predicting a certain outcome such as mortality on the basis of patient variables obtained from a large population. While, for example, a severity of illness score may allow one to determine that a particular patient is part of a population in which the mortality is 70%, it does not permit to predict whether this specific patient will be in the survivor or non-survivor cohort. As the outcome is binary (death or alive), each patient can only be one or the other. Score users also must be familiar with the conditions under which the score model was developed in order to determine to which extend it is valid under the clinical conditions of concern. For example, was the context under which the physiological variables were collected in the score development data set? Were the variables collected at the time of presentation to the emergency room, or was the most
abnormal value during the first 24-hours of ICU admission selected? Overall, great caution needs to be applied when considering scores to guide clinical decision making regarding a single patient.

**Usefulness in research**

Severity-of-illness scores play an important role in research. For one, severity of illness acts as an important extraneous variable for commonly used primary outcome measures in critical care, such as time until hospital discharge or survival to hospital discharge. Statistical control for severity of illness will help isolating the effect of the study intervention or variable of interest. For the same reason severity of illness scores are commonly found as data elements in clinical registries in both veterinary and medical medicine. Severity of illness scores are commonly used to determine whether randomization was effective in creating comparable populations in each study arm, but are also utilized for treatment/exposure group stratification in observational studies. Finally, score-driven population characterization will help determining how appropriate it is to transfer research results to another population, and aid in putting clinical research data originating from different studies into context with each other.

**General application in health care administration**

Adjusting death rate and other outcomes for severity of illness is of great interest in human health care administration as it allows comparing hospital or ICU performance within and among human medical centres, and over time. It may also guide resource allocation within an institution. Risk prediction scores obtained at initial presentation may help with directing patient allocation to the appropriate level of care, functioning similar to a triage system.

**Illness severity scores relevant to veterinary emergency and critical care**

The most relevant illness severity score currently used in veterinary emergency and critical are the SPI2, APPLE, ATT and MGCS.

The SPI2 was modelled based on physiologic and biochemical variables, disease chronicity and source problem category (surgical vs. medical) (i.e., MAP, respiratory rate, creatinine, PCV, albumin, age, medical vs. surgical) obtained from dogs during the first 24 hours of ICU admission, against the outcome of 30 day survival. The severity of the population used can be estimated by the fact that 243 of the 624 dogs in the dataset did not survive to hospital discharge. Data was prospectively collected from four different sites, which may improve external validity of the SPI2 score. All sites were referral hospitals. 499 dogs were used to construct the model and data from 125 animals were included into the validation data set. Score calibration is acceptable but discrimination is poor. In addition, the score is not easy to calculate at the bedside as the model output is calculated from a logit equation: Logit P = 0.3273 + (0.0108 • MAP) – (0.0102 • respiratory rate) – (0.2183 • creatinine) + (0.0164 • PCV) + (0.3553 • albumin) – (0.1184 • age) – (0.8069 • medical vs surgical status).

Nevertheless, to this author’s knowledge the SPI2 was reported in at least 10 research veterinary studies.

The APPLE score includes biochemical, physical exam and imaging variables collected during the first 24 hours of ICU admission of dogs or cats, and can be used in a short (APPLEshort) or long (APPLEfull) version. The data set used for the model development originates from a single referral hospital and includes prospective data obtained from 810 dogs admitted to an ICU. Patient variables include 10 variables in the full version, and 5 variables in the short version. These variables were modelled against the outcome of survival to discharge. Survival was achieved by 81.6% of dogs, indicating that the ICU population at this hospital was overall likely less severely ill than the population used to develop the SPI2 index. Of those animals that did not survive, 96% were euthanized. This is relevant, as euthanasia may occur for reasons other than severity of illness. Overall, the APPLE score is the most meticulously developed score in veterinary ECC. A construction data set included 598 dogs, and a separate smaller validation data set of 212 dogs was used for determination of calibration and discrimination, both of which were good. In addition to better validation results compared to the SPI2, rapid cage side calculation is simple as integer scores are obtained for all variables. The short version of the APPLE score (APPLEshort) still performs well and only requires five variables (glucose, albumin, lactate, platelet count and a mentation score) to be obtained. All these variables are obtained during the first 24 hours of admission to ICU. If several measurements of the same variable were available within that time frame, the most abnormal value was used to construct the model. The mentation score should be used at first presentation and under consideration of the confounding effects of sedatives and analgesics. A total score of 50 can be obtained for APPLEshort, with 0 (best) and 50 (worst). At least six research studies have been published that report an APPLE score.

The ATT scoring system integrates physiological variables from various body systems obtained during triage from dogs and cats presenting with trauma. It can be determined by applying a score ranging from 0-3 to each of six body system categories (i.e., perfusion, cardiac, respiratory, eye/muscle/integument, skeletal, neurological), and adding them up to reach a final ATT ranging from 0 (best) to 18 (worst). The score was developed with a retrospective and prospective data set containing 138 dogs and 51 cats admitted to an emergency service of an academic referral veterinary hospital. The overall 7-day mortality rate was 22.8% and 14.8% in the retrospective and prospective data set, respectively. Logistic regression was used to determine the relationship between ATT score and outcome, but no separate validation set was used and calibration and
discrimination were not determined. The score is an essential data variable in the Veterinary Committee on Trauma (VetCOT) registry that collected around 10,000 canine and feline trauma cases in two years, indicating the practicability of the ATT score, despite its methodological shortcomings. It is likely that the VetCOT data set will be source of further validation of the ATT or maybe the foundation for the development of another veterinary trauma score. The ATT score has been reported in at least seven research studies. The MGC5 assigns a score to variables of cortical, brainstem and motor function of dogs early after traumatic brain injury and modelled the overall result against survival to 48 hours. The retrospective data used originates from 38 dogs with acute head trauma admitted to a single referral centre (1999-99). Seven dogs died within 48 hours. Level of consciousness, motor activity and brainstem reflexes were assessed on first available neurological exam and given a score of 1 to 6 for each category, such that a total MGCS of 3 (worst) to 18 (best) ensued. Similar as was the case for the ATT, logistic regression was used to model the MGCS against mortality at 48 hours. No independent validation set has been used and calibration and discrimination was not determined. However, the score has been and is widely used in several research studies and is part of the VetCOT trauma database.

References


