

MedicalBiostatistics.com

APACHE scores

Among many severity scoring systems, APACHE scores are the most commonly used and deserve some understanding. The basic premise in these scores is that worst physiological derangement noted during first 24 hours after admission in an intensive care unit (ICU) more or less determines the chance of hospital survival as these define organ insufficiency. This implies, though, that treatment and care are not of much consequence as they are nearly same in hospitals across the United States where this system was evolved.

APACHE-I was proposed in the year 1981 and was surprisingly accurate in predicting mortality in patients in a variety of ICUs. An exception noted later was the patients requiring coronary by-pass graft where the physiological derangement was high but mortality was low. APACHE-I considered 34 routinely collected physiological measurements in most US hospitals and required no extra efforts. Each of these measurements was assigned weight according to the severity of derangement. For example, if serum pH value is either <7.15 or ≥ 7.70 , the weight is +4 as both are considered equally grave, whereas normal value between 7.33 and 7.49 has weight zero as this is no derangement¹. Slightly higher pH value between 7.50 and 7.59 was assigned weight +1 but on the lower side the value between 7.25 and 7.32 was assigned weight +2 as this is considered relatively more harmful. Sum of these weights for 34 measurements was APACHE score. Higher the score more is the chance of death. However, this was found too complex for adoption.

APACHE-II is the simplified version of APACHE-I and included only 12 physiological measurements. But it added points ranging from 0 for age <45 years to 6 for age ≥ 75 years, and previous history (5 points for nonoperative or postoperative emergency patients and 2 points for elective postoperative patients in the past). Maximum possible score is 71 although in practice none exceeds 55. A score of 40 or more has been seen to be strongly associated with hospital death. These scores were woven into a logistic regression with mortality as the outcome using data from a large number of ICUs across the United States. The equation derived was

$$\ln\left(\frac{R}{1-R}\right) = -3.517 + (0.146 * \text{APACHE-II score}) + (0.603, \text{ only if postoperative surgery}) + (\text{diagnostic category weight}),$$

where R is the predicted risk of death, and diagnostic category weight was separately derived for 50 disease groups. For example, this weight for asthma/allergies is -2.108 and for cardiogenic shock is $+0.393$. Negative weight implies that the risk of mortality is less and positive weight implies that the risk is more. These weights are available in Knaus et al. (1985).

Subsequently, APACHE-III appeared between 1991 and 2002 in several different versions. This included 17 physiological variables, adjustment of location and length of stay before ICU admission, and used splines for statistical modeling. The last version of APACHE-III covered 96 disease groups. APACHE-IV appeared in 2006 and included 116 disease groups. This revised the prediction equation, used five new predictors, extended splines, and made prior length of stay continuous in terms of minutes and not just in integer days. The details of APACHE-IV have been described by Zimmerman and Kramer.²

In short, APACHE-III and IV are more complex and only marginally increase the predictive accuracy over APACHE-II. Thus many still prefer APACHE-II. The percentage of ICU patients correctly classified into survive/death by APACHE II was observed as 85.5% in U.S. hospitals and the area under ROC curve was 0.863.¹ Note that correct prediction is not as high as the hype around this scoring system. Moreover, not many studies are available that can guide us regarding use of APACHE in developing countries where ICU care and mortality could be very different.

This scoring system is applicable to critical cases wherein survival is known not to exceed 80%. If I am naïve and use this as predictivity for survival without using any scoring system, I would be right in at least 80% cases. Thus a scoring system such as APACHE-II adds just about 5% to the accuracy of prediction. You may like to examine if it is worth taking and using more than 12 physiological measurements for a gain of paltry 5%.

Notwithstanding the problems just enumerated, APACHE scoring is still useful in many setups. You can legitimately compare severity of cases admitted in intensive care unit (ICU) of one hospital with that in another hospital, or in two or more groups such as severity in people in different occupations. Similarly, if a regimen is effective in 72% cases with APACHE-II score between 20 and 24, and another is effective in 78% with the same score, you are confident of 6% difference in efficacy. Also, if average APACHE-II score in critical cases admitted in a hospital during 2005-2009 is 17 and the average rises to 21 in cases admitted during 2010-2014, it would be legitimate to say that the cases admitted later are more severe. Actual utility of APACHE scores lie in this kind of comparison rather than in predicting survival.

References

1. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Crit Care Med* 1985; 13:818-829.
2. Zimmerman JE, Kramer AA. Outcome prediction in critical care: the Acute Physiology and Chronic Health Evaluation models. *Curr Opin Crit Care* 2008; 14:491-497.