

Marcio Soares<sup>1</sup>, Dave A. Dongelmans<sup>2</sup>

## Por que não devemos usar o APACHE II como parâmetro para avaliação de desempenho e comparação?

*Why should we not use APACHE II for performance measurement and benchmarking?*

1. Departamento de Terapia Intensiva, Instituto D'Or de Pesquisa e Ensino, Rio de Janeiro (RJ), Brasil.

2. Departamento de Terapia Intensiva, Centro Médico Acadêmico, University of Amsterdam - Amsterdã, Holanda.

O *Acute Physiology and Chronic Health Evaluation* (APACHE) é o escore de avaliação de gravidade de doença mais utilizado nas unidades de terapia intensivas (UTI) de adultos. Os escores APACHE utilizam dados clínicos, fisiológicos e laboratoriais observados na admissão e durante as primeiras 24 horas de internação na UT para estimar a gravidade da doença aguda para um dado paciente, por meio de um escore em pontos e a respectiva probabilidade de óbito hospitalar. Embora os escores de gravidade de doença, inclusive o escore APACHE, não devam ser utilizados para guiar tomadas de decisão em pacientes individuais, eles são úteis para caracterizar pacientes em estudos clínicos, avaliar o desempenho da UTI e fins de comparação (*benchmarking*) em caso em que são necessárias correções quanto à mescla de casos.<sup>(1)</sup>

A primeira versão do APACHE data do início da década de 1980.<sup>(2)</sup> No entanto, o APACHE I era demasiadamente complexo e consumia muito tempo para poder ser usado na rotina da UTI. O escore APACHE II foi apresentado há mais de 35 anos, em 1985, e utilizava os dados de 5.815 pacientes internados entre 1979 e 1982 em 13 hospitais dos Estados Unidos.<sup>(3)</sup> O número de variáveis foi reduzido de 34 para 12, e foram fornecidos 50 grupos de diagnósticos à admissão para melhorar a precisão da predição do desfecho. O APACHE II foi rapidamente adotado pelas UTI em todo o mundo, e é hoje o escore mais utilizado em ensaios clínicos. O escore APACHE III foi publicado em 1991 com uso dos dados de 17.440 pacientes internados em 40 hospitais dos Estados Unidos.<sup>(4)</sup> Utilizaram-se estratégias de modelagem estatística mais sofisticadas, e se expandiram os números de grupos de diagnósticos à internação e de variáveis fisiológicas.

Mais ainda, foram fornecidas novas equações para predição de desfechos, além da mortalidade hospitalar. Durante a década de 1990, foram disponibilizadas versões atualizadas do APACHE III. Entretanto, apesar destas atualizações, as deteriorações do desempenho do modelo ao longo do tempo indicaram que seria necessário desenvolver uma nova equação.<sup>(5,6)</sup> Assim, o APACHE IV, que representa a versão mais recente dos escores APACHE foi introduzido em 2006.<sup>(5)</sup> Os investigadores utilizaram dados de mais de 110 mil internações em UTI de 45 hospitais, ainda restritos aos Estados Unidos. O número de grupos de diagnósticos à internação foi expandido para 116.

No entanto, por que os escores prognósticos devem ser regularmente atualizados? Não é surpreendente que o desempenho dos escores deteriore com o passar do tempo. Tal deterioração é invariavelmente caracterizada pela piora da discriminação (isto é, da capacidade de discriminar sobreviventes de não

**Conflitos de interesse:** Marcio Soares é sócio-fundador da Epimed Solutions, que comercializa o Epimed Monitor, um sistema de informação baseado na nuvem para gestão da qualidade e *benchmarking* de unidades de terapia intensiva. Dave A. Dongelmans é diretor-presidente da Dutch National Intensive Care Evaluation (NICE).

Submetido em 21 de fevereiro de 2017

Aceito em 22 de fevereiro de 2017

**Autor correspondente:**

Marcio Soares  
Instituto D'Or de Pesquisa e Ensino  
Rua Diniz Cordeiro, 30, 3º andar  
CEP: 22281-100 - Rio de Janeiro (RJ)  
E-mail: marciosoaarsms@gmail.com

**Editor responsável:** Jorge Ibrain Figueira Salluh

DOI: 10.5935/0103-507X.20170043

sobreviventes) e, mais importante, da calibração do modelo (isto é, piora da concordância entre o número observado e previsto de sobreviventes e não sobreviventes ao longo de todos os estratos de probabilidade de óbito), que podem ser atribuídas a uma série de razões<sup>(1,5)</sup> que incluem avanços na ciência médica; melhora do tratamento/controlado intensivo; melhora no controle dos pacientes e intervenções terapêuticas; modificação da mescla de casos (por exemplo, envelhecimento populacional e aumento do número de pacientes que vivem com comorbidades graves); e alterações nas políticas de admissão/alta/decisões de fim de vida.

Como a sobrevivência de muitas condições que necessitam de tratamento intensivo melhorou nas décadas recentes, à medida que o tempo passa, os escores prognósticos tendem a superestimar a probabilidade de óbito, assim resultando em taxas de mortalidade padronizadas (TMP) mais baixas. O relato dos dados de qualidade e desempenho das UTI disseminam-se rapidamente em todo o mundo. Em muitos países, os sistemas de escore têm sido utilizados para fins de *benchmarking* entre diferentes unidades. No entanto, são necessários estudos de validação antes que se utilizem estes instrumentos em uma região ou país específico. Ao longo das últimas décadas, uma série de estudos que utilizaram bases de dados contemporâneos observou que o APACHE II é impreciso para avaliação do desempenho e *benchmarking*.<sup>(7,8)</sup> Consequentemente, diversas iniciativas de melhoria da qualidade substituíram este escore por versões atualizadas (APACHE III e APACHE IV), ou por outro índice de gravidade de doença.<sup>(7,9-13)</sup> Como alternativa, alguns países tentaram desenvolver equações customizadas dos escores APACHE II para superar o mau desempenho da equação original.<sup>(14)</sup> Entretanto, estas práticas foram abandonadas há mais de uma década, em função tanto da disponibilização de versões mais recentes do APACHE quanto de outros escores, ou por conta da decisão de desenvolver instrumentos mais ajustados ao local. Um bom exemplo do uso de ambas as estratégias é o *Case Mix Programme of the Intensive Care National Audit & Research Centre* (ICNARC) do Reino Unido. Na metade da década de 1990, disponibilizaram-se versões adaptadas do APACHE II com uso de coeficientes específicos do Reino Unido, com recalibrações frequentes.<sup>(9,14)</sup> Em 2007, o modelo ICNARC foi publicado e apresentou precisão maior que a de outros escores, inclusive os escores APACHE II e III, tornando-se o escore padrão para fins de

avaliação do desempenho e *benchmarking* no Reino Unido.<sup>(9,14)</sup> Há alguns anos, o ICNARC tomou a decisão de não mais fazer recalibrações do APACHE II. O registro holandês de UTI (*National Intensive Care Evaluation - NICE*) recebe dados de 85 UTI da Holanda.<sup>(10)</sup> Desde o ano passado, o NICE deixou de utilizar o APACHE II nos relatórios devolutivos, em razão de sua má adaptação à situação holandesa. Mesmo depois de recalibrado, o desempenho do modelo permanecia muito baixo. Nos atuais relatórios devolutivos, são utilizados os modelos APACHE IV e *Simplified Acute Physiology Score* (SAPS) II. No sítio público de internet do NICE, os dados das UTI individuais, assim como os dados agregados são apresentados com base no modelo APACHE IV em vários grupos de pacientes. Uma vantagem importante do modelo APACHE IV é que ele pode também ser utilizado em pacientes de cirurgia cardíaca. Isto permite uma comparação corrigida conforme a mescla de casos neste grupo de pacientes. Deve ser dito que, em razão da baixa ocorrência de mortalidade nesta população de pacientes, é difícil encontrar diferenças estatisticamente significantes entre os centros.

Em resumo, a despeito da miríade de argumentos utilizados por alguns médicos e administradores (Tabela 1), o escore APACHE II não pode ser recomendado para avaliações de desempenho e *benchmarking*. Para isto, devem ser utilizadas versões atualizadas de escores de gravidade, apropriadamente validadas para o país ou região.

### Suporte financeiro

O Dr. Soares é parcialmente suportado por verbas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

**Tabela 1** - Argumentos frequentemente utilizados para continuar a usar o escore *Acute Physiology and Chronic Health Evaluation II*

- Familiaridade com a compreensão da representatividade de um escore para um dado paciente
- Familiaridade com o uso do escore
- Permitir comparações de TMA ao longo do tempo
- Permitir comparações em termos de severidade da doença entre novos e antigos estudos clínicos
- Como o escore superestima a mortalidade, TMA baixas podem ser facilmente demonstradas
- Simplesmente inércia

TMA - taxa de mortalidade ajustada.

## REFERÊNCIAS

1. Salluh JI, Soares M. ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care*. 2014;20(5):557-65.
2. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9(8):591-7.
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-29.
4. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100(6):1619-36.
5. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34(5):1297-310.
6. Zimmerman JE, Kramer AA. Outcome prediction in critical care: the Acute Physiology and Chronic Health Evaluation models. *Curr Opin Crit Care*. 2008;14(5):491-7.
7. Harrison DA, Lone NI, Haddow C, MacGillivray M, Khan A, Cook B, et al. External validation of the Intensive Care National Audit & Research Centre (ICNARC) risk prediction model in critical care units in Scotland. *BMC Anesthesiol*. 2014;14:116.
8. Brinkman S, Bakhshi-Raiez F, Abu-Hanna A, de Jonge E, Bosman RJ, Peelen L, et al. External validation of Acute Physiology and Chronic Health Evaluation IV in Dutch intensive care units and comparison with Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score II. *J Crit Care*. 2011;26(1):105.e11-8.
9. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K. A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model. *Crit Care Med*. 2007;35(4):1091-8.
10. van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data Resource Profile: the Dutch National Intensive Care Evaluation (NICE) Registry of Admissions to Adult Intensive Care Units. *Int J Epidemiol*. 2015;44(6):1850-1850h.
11. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). *Crit Care Med*. 2007;35(3):827-35.
12. Ferrando-Vivas P, Jones A, Rowan KM, Harrison DA. Development and validation of the new ICNARC model for prediction of acute hospital mortality in adult critical care. *J Crit Care*. 2017;38:335-9.
13. Associação de Medicina Intensiva Brasileira (AMIB). Comissão de Defesa do Exercício Profissional. Regulamento Técnico para Funcionamento de Unidades de Terapia Intensiva - AMIB. São Paulo: AMIB; 2009. [Internet]. [citado 2017 Jun 14]. Disponível em: <http://www.amib.org.br/fileadmin/RecomendacoesAMIB.pdf>
14. Harrison DA, Rowan KM. Outcome prediction in critical care: the ICNARC model. *Curr Opin Crit Care*. 2008;14(5):506-12.