

Validação de Escores Prognósticos para Mortalidade de Pacientes com COVID-19 no Brasil

Lucas Emanuel F. Ramos

Orientadora; Profa. Magda Carvalho Pires (DEST/UFMG)

Co-orientadora; Profa. Milena Soriano Marcolino
(Departamento de Clínica Médica/UFMG)

26/03/2021

1. Introdução

- Ferramentas de predição de risco têm sido cada vez mais utilizadas no meio clínico para estimar e estratificar o risco de pacientes desenvolverem doenças ou mesmo de irem a óbito devido a complicações causadas por essas doenças.
- Em um cenário pandêmico como o provocado pelo novo coronavírus, um índice que identifique precocemente pacientes com alto risco de óbito devido à COVID-19 seria de extrema importância e urgência para direcionar as intervenções clínicas e terapêuticas.
- Para garantir a acurácia preditiva é crucial a validação da ferramenta, permitindo o seu uso e garantindo boa eficácia em outras populações.

1. Introdução

Objetivos

- 1 Apresentar técnicas para a validação de escores prognósticos
- 2 Aplicar os métodos para avaliar o escore $ABC_2 - SPH$ desenvolvido para estratificar o risco de óbito de pacientes com COVID-19 no Brasil
- 3 Estudar o desempenho do escore em outras populações

2.1. Metodologia: Tipos de Validação

- 1 **Validação Interna:** Divisão aleatória.
- 2 **Validação Temporal:** Divisão temporal.
- 3 **Validação Externa:** Novos dados de uma população de pacientes apropriada e similar (diferentes centros).

2.2. Metodologia: Desempenho Geral

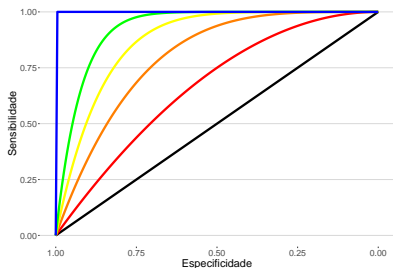
A distância entre os valores observados e preditos pelo modelo são comumente utilizadas para medir o desempenho geral. No caso em que a variável resposta é binária (óbito / alta) é utilizado o **Escore de Brier**:

$$BS = \frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2$$

$BS \approx 0$: Bom desempenho geral.

2.3. Metodologia: Discriminação

- Um bom modelo prognóstico deve discriminar os pacientes que possuem dos que não possuem o evento de interesse.
- Para o caso binário, a área abaixo da curva ROC é a medida mais utilizada.



Quanto mais próxima a curva ROC da curva azul ao lado, maior a área abaixo da curva (AUC) e melhor a discriminação do modelo prognóstico.

$AUC \approx 1$: Ótima discriminação.

Fig. 1: Diferentes curvas ROC.

2.4. Metodologia: Calibração

A calibração diz respeito à concordância entre os valores estimados e observados. O gráfico de calibração traça as probabilidades observadas e previstas.

A curva azul mostra a calibração logística e a curva preta é formada por suavização. Os pontos vermelhos são agrupamentos de observações.

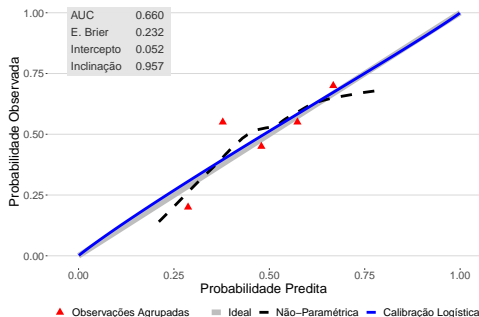


Fig. 2: Exemplo de um gráfico de calibração.

2.5. Metodologia: Intercepto e Inclinação de Cox

O método de Cox avalia a calibração ao ajustar uma regressão logística que explica o desfecho binário observado ($O = 1$) por meio das *log odds* das estimativas do modelo preditivo em estudo (E):

$$\text{logit}(P(O = 1)) = a + b * \text{logit}(E) \quad (1)$$

- a é o intercepto;
- b é a inclinação.

$a \approx 0$ e $b \approx 1$: Ótima calibração.

2.5. Metodologia: Intercepto e Inclinação de Cox

É possível testar $H_0 : a = 0, b = 1$ a partir de um teste qui-quadrado que avalia se a diferença na soma dos quadrados dos resíduos dos modelos (1) e (2) é significativa.

$$\text{logit}(P(O = 1)) = \text{offset}(\text{logit}(E)) \quad (2)$$

2.6. Metodologia: Utilidade Clínica

- A utilidade clínica diz respeito à habilidade do modelo preditivo em auxiliar na tomada de decisão clínica;
- Geralmente são comparados os benefícios de se utilizar o modelo prognóstico ao invés das políticas padrão de tratamento a todos os paciente ou a nenhum.

Os benefícios são calculados para diferentes pontos de corte de classificação a partir da fórmula:

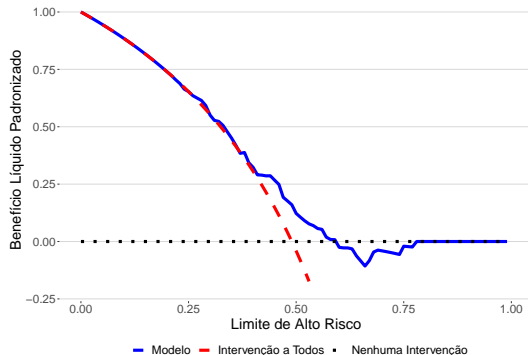
$$NetBenefit = NB = \frac{(TP - w * FP)}{n}$$

2.6. Metodologia: Utilidade Clínica

$$NetBenefit = NB = \frac{(TP - w * FP)}{n}$$

- TP são os verdadeiros positivos;
- FP são os falsos positivos;
- w é o peso relacionado ao ponto de corte p : $w = \frac{p}{1-p}$;
- n é o número total de pacientes.

2.6. Metodologia: Utilidade Clínica



Ao comparar curvas de decisão de diferentes modelos, o modelo de melhor ajuste terá sua curva acima das demais em toda a escala de possíveis pontos de corte.

Fig. 3: Curva de Decisão para dados simulados.

Neste exemplo, o modelo não apresenta grandes benefícios.

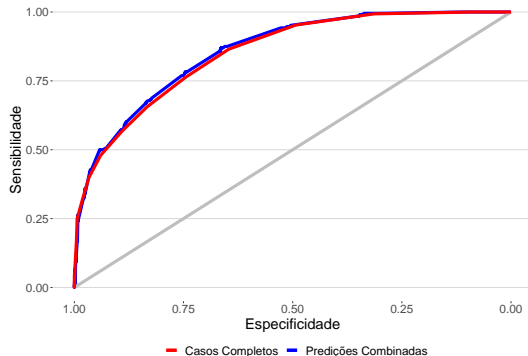
3. Resultados - Escore ABC_2 – SPH

- Base de derivação (3978 pacientes - 79,1%) - Incidência de 20,26%;
- Base de validação (1054 pacientes - 20,9%) - Incidência de 19,73%;
- Divisão das bases por data de admissão do paciente → Validação Temporal.
- As bases de dados foram imputadas independentemente (MICE - $m = 10$);
- Desfecho utilizado apenas na imputação da base de derivação.

3.1. Resultados - Abordagens de Validação

- 1 Avaliar o desempenho do escore em cada um dos bancos imputados e combinar as medidas de avaliação;
- 2 Combinar as predições dos $m = 10$ bancos imputados (média simples - *Rubin's Rules*) e obter medidas únicas de avaliação;
- 3 Avaliar escore prognóstico utilizando apenas os casos completos da base de validação (779 pacientes) - Incidência de 18,99%.

3.2. Resultados - Validação Temporal



Tanto a abordagem de combinar as predições quanto a de casos completos obtiveram ótima discriminação, apresentando AUC's de 85,9% e 85,2% respectivamente.

Fig. 4: Curvas ROC para as abordagens de combinar predições e casos completos.

3.2. Resultados - Validação Temporal

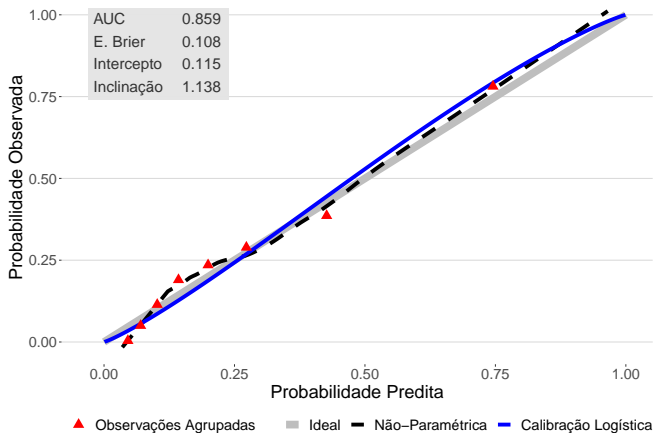


Fig. 5: Curvas de calibração ao combinar as previsões.

3.2. Resultados - Validação Temporal

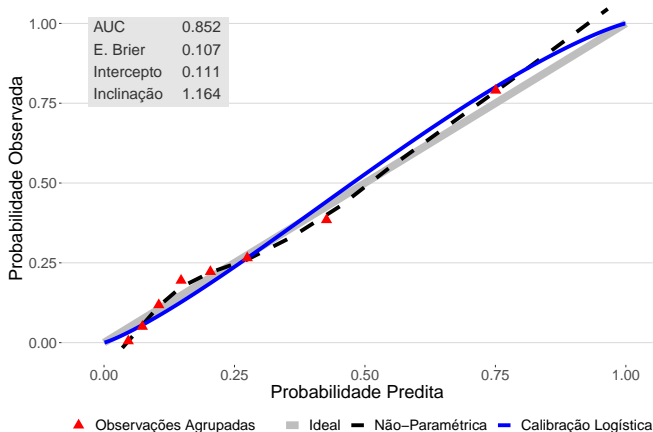


Fig. 6: Curvas de calibração ao considerar apenas os casos completos.

3.2. Resultados - Validação Temporal

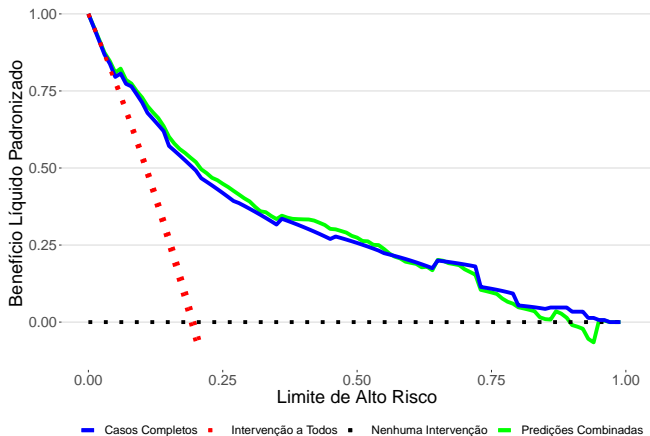


Fig. 7: Curvas de Decisão para as abordagens de combinar previsões e casos completos.

3.2. Resultados - Validação Temporal

Tabela 1: Métricas de Avaliação por Abordagem

	Bancos Imputados	Predições Combinadas	Casos Completos
AUC	85,69% (85,40%-85,92%)	85,86% (83,26%-88,47%)	85,2% (82,04%-88,36%)
Escore de Brier	0,109 (0,108-0,109)	0,108	0,107
Intercepto	0,102 (0,086-0,132)	0,115	0,111
Inclinação	1,127 (1,115-1,148)	1,138	1,164
P-valor (Int=0, Inc=1)	23,59% (15,61%-29,02%)	18,48%	16,99%

Nas métricas de avaliação por banco de validação imputado são exibidos: valor médio (mínimo - máximo). Para as predições combinadas e casos completos são exibidos os valores observados e intervalo de 95% de confiança para a AUC.

3.2. Resultados - Validação Temporal

Tabela 2: Grupos de Risco do Escore ABC_2 – SPH por Base de Validação

Grupo de Risco	Prob. Preditá	Bases Imputadas		Casos Completos	
		Pacientes	Óbitos	Pacientes	Óbitos
Baixo (0-1)	0%-5,9%	290	1 (0.3%)	199	1 (0.5%)
Intermediário (2-4)	6%-14,9%	394	47 (11.9%)	306	34 (11.1%)
Alto (5-8)	15%-49,9%	252	73 (29%)	194	54 (27.8%)
Muito Alto (9-20)	50%-100%	118	87 (73.7%)	80	59 (73.8%)
Geral	-	1054	208 (19.7%)	779	148 (19%)

Para as bases imputadas foi obtido o escore mais frequente (moda) do paciente.

3.3. Resultados - Validação Externa

- A validação temporal verifica a adequação do escore $ABC_2 - SPH$ para os pacientes brasileiros;
- Para a generalização para outras populações é necessária uma validação externa.

474 pacientes admitidos com COVID-19 no Hospital Universitário Vall d'Hebron, em Barcelona, Espanha, entre março e maio de 2020 foram selecionados para a aplicação e validação do escore $ABC_2 - SPH$.

3.3. Resultados - Validação Externa

- Todos estes pacientes contém as informações necessárias para o cálculo do escore (Casos Completos);
- A amostra contém 82 óbitos (incidência de 17,3%);
- Os critérios de inclusão e exclusão de pacientes foram os mesmos adotados para a construção do escore.

3.3. Resultados - Validação Externa

Ao aplicar o escore $ABC_2 - SPH$ nos pacientes do Hospital Universitário Vall d'Hebron, obteve-se uma ótima discriminação com AUC de 89,9%.

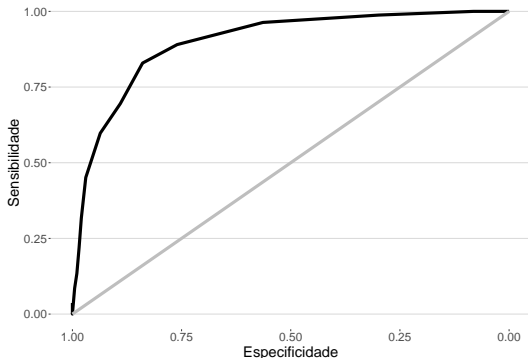


Fig. 8: Curva ROC obtida com os pacientes espanhóis.

3.3. Resultados - Validação Externa

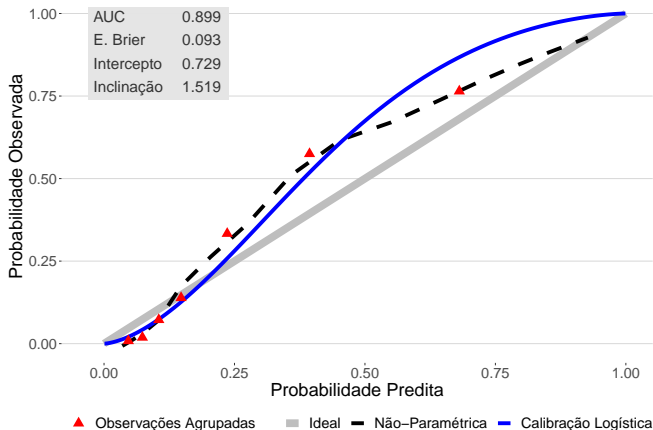


Fig. 9: Curvas de calibração obtidas na validação externa.
P-valor para $H_0 : a = 0, b = 1$ foi 0,001

3.3. Resultados - Validação Externa

Tabela 3: Grupos de Risco do Escore $ABC_2 - SPH$ nas Bases de Validação Temporal Imputadas e Validação Externa

Grupo de Risco	Prob. Preditada	Bases Imputadas		Validação Externa	
		Pacientes	Óbitos	Pacientes	Óbitos
Baixo (0-1)	0%-5,9%	290	1 (0.3%)	118	1 (0.008%)
Intermediário (2-4)	6%-14,9%	394	47 (11.9%)	225	13 (5.8%)
Alto (5-8)	15%-49,9%	252	73 (29%)	97	42 (43.3%)
Muito Alto (9-20)	50%-100%	118	87 (73.7%)	34	26 (76.5%)
Geral	-	1054	208 (19.7%)	474	82 (17.3%)

Para as bases imputadas de derivação foi obtido o escore mais frequente (moda) do paciente.

3.3. Resultados - Validação Externa

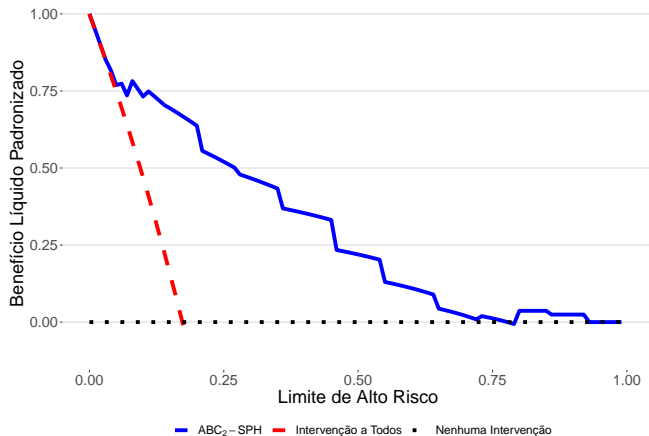


Fig. 10: Curva de Decisão considerando os casos completos de Barcelona.

4. Discussão

- A validação temporal pôde constar com segurança a adequação do escore $ABC_2 - SPH$ tanto em termos de discriminação quanto de calibração e utilidade clínica para os pacientes brasileiros;
- Quanto à aplicação do escore em outras populações, a validação externa indicou excelente discriminação e boa utilidade clínica, entretanto, uma má calibração;
- Os grupos de risco construídos discriminam bem os pacientes espanhóis com relação ao risco de óbito, o que pode ser suficiente para sua aplicação, mas para melhores resultados uma recalibração do escore $ABC_2 - SPH$ pode ser realizada para adequá-lo à nova população.

4. Discussão

Possíveis fontes de viés na validação externa:

- 1 Amostra pequena (474 pacientes - 82 óbitos);
- 2 Pacientes de apenas um centro médico;
- 3 Diferentes políticas de internação: mediana do tempo de internação dos pacientes no coorte brasileiro é de 7 dias, enquanto que para os pacientes espanhóis considerados a mediana é de 21 dias.

5. Conclusões

- 1 O escore $ABC_2 - SPH$ pôde ser validado para estratificar o risco de óbitos de pacientes com COVID-19 em hospitais públicos e privados do Brasil;
- 2 Quanto sua aplicação em pacientes espanhóis, o escore apresentou bom desempenho em discriminar os pacientes em grupos de risco de óbito, entretanto, é recomendada sua recalibração para readequar as probabilidades preditas pelo escore de acordo com os percentuais de óbitos observados.

5. Conclusões

Futuros estudos:

- Recalibração do escore $ABC_2 - SPH$ para o coorte de Barcelona;
- Aplicação e validação do escore em outras populações, preferencialmente com maiores amostras e pacientes de diversos centros médicos (públicos e privados).

6. Referências

- STEYERBERG, Ewout W. et al. **Clinical prediction models**. Cham: Springer International Publishing, 2019.
- ALTMAN, Douglas G. et al. **Prognosis and prognostic research: validating a prognostic model**. Bmj, v. 338, 2009.
- VAN CALSTER, Ben et al. **Calibration: the Achilles heel of predictive analytics**. BMC medicine, v. 17, n. 1, p. 1-7, 2019.
- RUBIN, Donald B. **An overview of multiple imputation**. In: Proceedings of the survey research methods section of the American statistical association. 1988. p. 79-84.
- MARCOLINO, Milena S. et al. **ABC2-SPH risk score for in-hospital mortality in COVID-19 patients: development, external validation and comparison with other available scores**. medRxiv, 2021. Preprint at <https://doi.org/10.1101/2021.02.01.21250306>