

RESEARCH

Open Access



Visualizing a marker's degrees of necessity and of sufficiency in the predictiveness curve

Andreas Gleiss^{1*}

Abstract

Background The degrees to which a factor is necessary or sufficient for an event have been proposed as generalizations of attributable risk based on simple functions of unconditional and conditional event probabilities. Predictiveness curves show the risk for an event, as derived by a model with one or more predictors, depending on risk percentiles that represent the predictors' distribution in the underlying population.

Methods Connections between the degrees of necessity and of sufficiency and explained variation on the one hand and the predictiveness curve on the other hand are mathematically proved and exemplified using data of in-hospital death of Covid- 19 patients.

Results We show that the degrees of necessity and of sufficiency can be represented as proportions of areas easily identifiable in the plot of the predictiveness curve. In addition, we show that the proportion of explained variation, a common measure of predictiveness and relative importance of prognostic factors, is also closely connected to these areas.

Conclusion Our investigations demonstrate that the predictiveness curve extended by these new interpretations of areas provides a comprehensive evaluation of markers or sets of markers for prediction.

Trial registration Austrian Coronavirus Adaptive Clinical Trial (ACOVACT); ClinicalTrials.gov, identifier NCT04351724.

Keywords Predictiveness curve, Necessity, Sufficiency, Explained variation

Background

Predictiveness curves have been proposed to bring together, in a single intuitive graph, the view of the marker's effect on risk and the view of its classification performance [1]. While the effect can be quantified by, e.g., an odds ratio derived from a logistic regression model, the marker's importance is better described by measures of predictiveness. This contribution goes beyond known connections between predictiveness curve and explained variation [1] by showing that proportions of areas in the

plot of the predictiveness curve can be interpreted as the degrees of necessity and of sufficiency [2].

As an example, consider the study of 578 Covid- 19 patients in the early phase of the pandemic (year 2020, no virus mutations) [3]. The goal was to predict in-hospital mortality after day 4 by laboratory and other factors collected in clinical routine during the first four days in hospital. A cross-validated selection of factors resulted in age, fever on admission, and six variables derived from hematological parameters (mean and slope of lactate dehydrogenase, mean of Creatinine, mean and slope of C-reactive protein, slope of platelet count during the first four days).

Table 1 summarizes various statistical measures estimated from the data using a logistic regression model in SAS 9.4 (SAS Institute Inc., 2020). The odds ratio (OR) for the continuous age variable (1.10 per increase by one

*Correspondence:

Andreas Gleiss
andreas.gleiss@meduniwien.ac.at

¹ Center for Medical Data Science, Institute of Clinical Biometrics, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Table 1 Marginal (not shrunk) odds ratios (OR) with 95% confidence intervals (CI), apparent marginal proportion of explained variation (EV_{ind}), degree of necessity (DN_1 and DN_2) and of sufficiency (DS_1 and DS_2), Tjur's coefficient of determination (C_D), and standardized total gain (STG) for in-hospital death after day 4

Predictor	OR	CI	EV_{ind}	DN_1	DS_1	DN_2	DS_2	C_D	STG
Age (years)	1.10	1.08–1.13	0.23	0.76	0.28	0.70	0.22	0.22	0.53
Fever on admission	1.83	1.13–2.97	0.01	0.27	0.04	0.27	0.04	0.01	0.14
Hematological parameters ^a	-	-	0.30	0.67	0.40	0.62	0.29	0.29	0.53
Full model	-	-	0.49	0.88	0.50	0.85	0.40	0.48	0.73

^a up to day 4 (see text)

year) can hardly be compared to that of the dichotomous presence of fever on admission (OR = 1.83) or to the contribution of the set of six laboratory parameters. In contrast, the proportion of explained variation (EV) allows a clear ranking with EV equal to 0.30, 0.23 and 0.01 for laboratory parameters, age and fever, respectively. A closer look for the risk's dependence on age is presented in Fig. 1A: While the oldest patients observed a high risk for in-hospital death, this risk is well below the average risk of 14.7% for nearly two thirds of the patients in our sample.

The next section summarizes predictiveness curves and their most important features and repeats definitions for measures of predictiveness. Further, the definitions and some important properties of the degree of necessity (DN) and of sufficiency (DS) are collected. "Results" section presents and illustrates novel connections between the predictiveness curve on the one hand, and EV, DN and DS on the other hand.

Methods

Predictiveness curves

Predictiveness curves have been proposed as graphical means to integrate the predictiveness of a marker with its performance as a classifier [1]. They show the risk for an event, as derived by a model with one or more predictors, depending on risk quantiles that represent the predictor's distribution in the underlying population. Figure 1A shows the risk of in-hospital death as predicted by age using a logistic regression model. In the *predictiveness curve* $R(q)$ in Fig. 1B the horizontal axis represents the quantiles q of the predictor distribution. More generally, risk predictions are ordered from lowest to highest values such that the horizontal axis shows quantiles of the risk distribution as predicted by the considered predictor [4].

Figure 1B also indicates the quantile q_0 where the predictiveness curve takes the value of the event prevalence, denoted by $P(D)$, the probability of the event of interest, D . At this point in the risk distribution, prediction conditional on the considered marker(s) equals unconditional risk

prediction. The horizontal line at $P(D)$ also represents prediction using an uninformative marker [1].

We denote the geometric areas between $R(q)$ and $P(D)$ as A_N (the part below $P(D)$) and A_S (the part above $P(D)$), respectively; they are shown as hatched areas in Fig. 2. Assuming a globally calibrated risk model where conditional risk averages to the unconditional risk in the population, A_N and A_S are of equal size [4, 5].

For our example of predicting in-hospital death of Covid-19 patients after day 4 by age, $P(D) = 0.147$, $q_0 = 0.64$ and $A_N = A_S = 0.066$. The predictiveness curve in Fig. 1B shows that 50% of the patients in our sample have a risk below 8% since $R(0.5) = 0.08$. On the other hand, starting from 30% defined as a risk threshold for in-hospital death, $R(0.81) = 0.3$ shows that 19% of the patients can be regarded to be at high risk.

Figure 3 in Steyerberg et al. [6] shows the predictiveness curve separately for patients without and for patients with the event. This enables to read off, for a given predicted risk level, the resulting specificity and 1 minus sensitivity, respectively, when the considered predictor is dichotomized at the value corresponding to this risk level. For our example data, the predictiveness curve stratified by event status is shown in Fig. 1C. Choosing a risk threshold of 14.7%, equal to unconditional risk, $P(D)$, the stratified curves in Fig. 1C cross the threshold at quantile $q_0^0 = 0.72$ and $q_0^1 = 0.19$, respectively, resulting in a specificity of 72% and a sensitivity of 81%.

Measures of predictiveness

Explained variation

The *indirect measure of explained variation* (EV_{ind}) quantifies the relative reduction in predictive inaccuracy when unconditional prediction is replaced by prediction conditional on a considered predictor [7]. EV equals the proportion of variation in the outcome that is explained by the considered predictor(s). It can therefore be used to quantify and compare the importance of prognostic factors [8]. By assuming a globally calibrated model this can be rewritten as the ratio of the variance

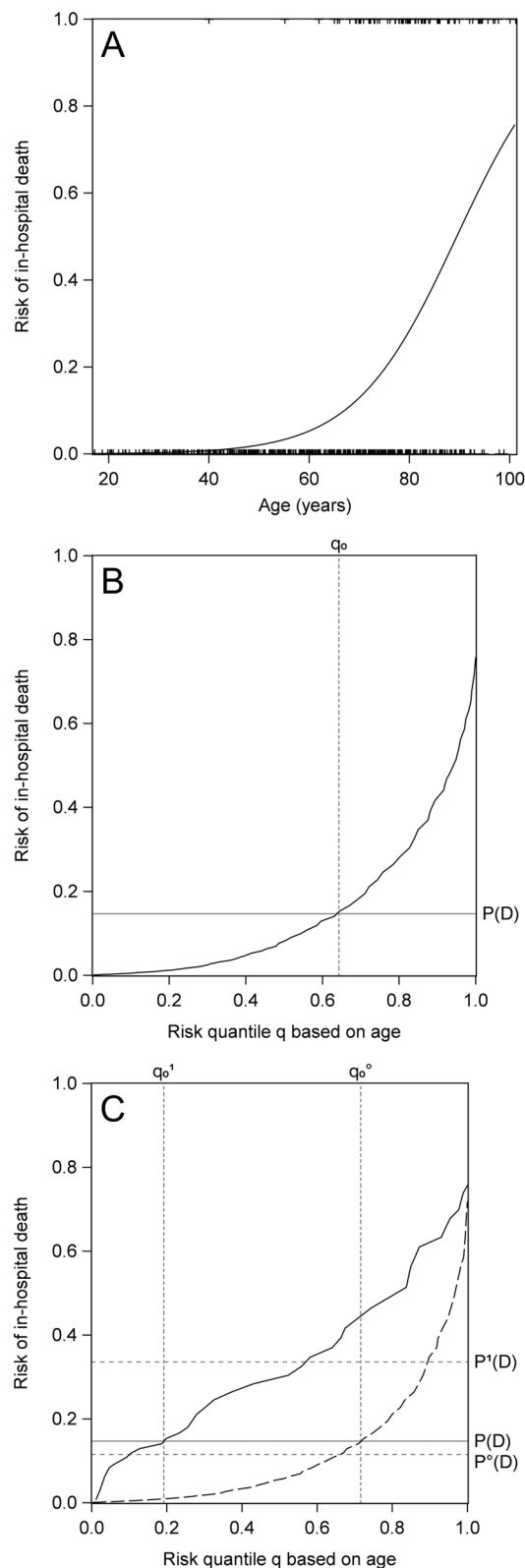


Fig. 1 Estimated logistic regression model for predicting in-hospital death after day 4 by age, ticks on bottom and top indicating non-events and events (**A**); smoothed predictiveness curve $R(q)$ for age on whole sample (**B**), restricted to events (black solid line) and to non-events (black dashed line) (**C**)

of conditional predictions to $P(D)[1 - P(D)]$ [2]. Thus, EV involves squared distances between $R(q)$ and $P(D)$.

Tjur proposed the *coefficient of discrimination* C_D , a measure of a ‘model’s ability to discriminate between successes and failures’, as the difference of $P^1(D)$ and $P^0(D)$, the expected conditional event probability among events and non-events, respectively [9]. For the example data, $P^1(D)$ and $P^0(D)$ are estimated by the average of the predicted event probabilities across all patients that died in hospital or did not die in hospital, respectively. For age as predictor of in-hospital mortality, $C_D = 0.22$ (Table 1) was based on an estimate of 0.331 for $P^1(D)$ and of 0.115 for $P^0(D)$ (see Fig. 1C). By noting that Tjur’s measure R_{mod}^2 exactly equals the plug-in estimator of EV_{ind} , we may conclude from Proposition 2 in [9] that $EV_{ind} = C_D$ (see also [10]). This asymptotic equality is also observed for our example data in Table 1.

Total gain

Bura & Gastwirth [5] were the first to propose the (absolute) area between predictiveness curve and $P(D)$ as a ‘measure of explanatory power’. They defined *total gain* (TG) as

$$TG = \int_0^1 |P(D) - R(q)| dq$$

Clearly, $TG = A_N + A_S$. Since the maximum value of TG is $2P(D)[1 - P(D)]$ [5], *standardized total gain* (STG) is defined as

$$STG = \frac{TG}{2P(D)[1 - P(D)]}$$

Since the denominator of STG only depends on $P(D)$, a comparison of STG between different prognostic factors within the same prediction task (and thus with the same event prevalence) is equivalent to a graphical comparison of TG. Gu & Pepe [11] have shown that STG equals the *Above average risk difference* (AARD) which is defined as the difference of q_0^0 and q_0^1 , the proportions of conditional predictions below $P(D)$ among non-events and events, respectively. AARD also equals Youden’s index evaluated at $P(D)$ [11].

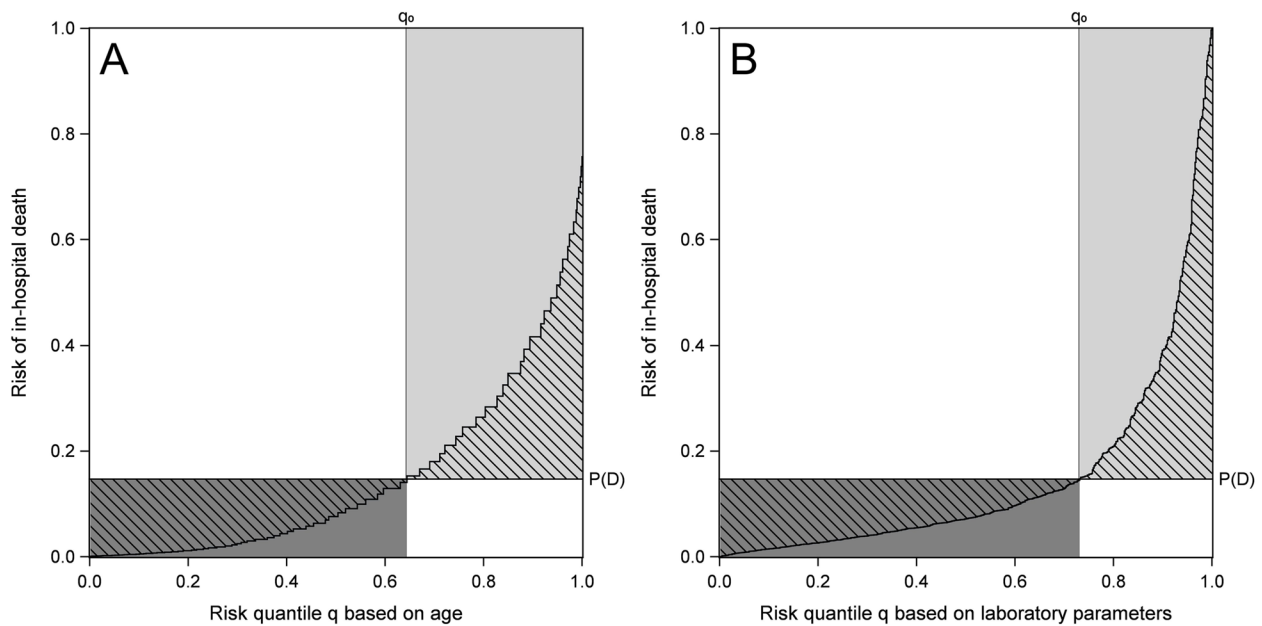


Fig. 2 Predictiveness curve with hatched areas A_N and A_S , for predicting in-hospital death after day 4 by age (A) and by laboratory parameters within first four days (B)

Degrees of necessity and of sufficiency

A cause, represented by certain values of prognostic factors, is considered necessary for an event D if, without the cause, D cannot develop. It is considered sufficient for D if D is unavoidable in the presence of the cause. The degrees of necessity and of sufficiency, ranging from zero to one, have been developed as generalizations of attributable risk based on simple functions of the unconditional event probability $P(D)$ and probabilities $P(D|X)$ conditional on a predictor X [2]. There are two variants of DN and DS, the first giving more weight to extreme conditional probabilities:

$$DN_1 = \sqrt{E_{X<} \left\{ \frac{P(D) - P(D|X)}{P(D)} \right\}^2}$$

$$DS_1 = \sqrt{E_{X>} \left\{ \frac{P(D|X) - P(D)}{1 - P(D)} \right\}^2}$$

$$DN_2 = E_{X<} \left\{ \frac{P(D) - P(D|X)}{P(D)} \right\}$$

$$DS_2 = E_{X>} \left\{ \frac{P(D|X) - P(D)}{1 - P(D)} \right\}$$

where $E_{X<}(\cdot)$ and $E_{X>}(\cdot)$ denote expectation with respect to X conditional on $\{X : P(D|X) < P(D)\}$ and $\{X : P(D|X) > P(D)\}$, respectively. For an unfavorable

outcome, these two ranges of X values were termed ‘protective’ and ‘harmful’, respectively. Note that DN_2 thus equals the conditional expectation of attributable risk on the protective range [2].

The measures of variant 2 equal weighted differences between the event probability on the ‘harmful’ and the ‘protective’ range. In contrast, variant 1 also accounts for the variation of conditional probabilities within the ‘protective’ range for DN_1 and within the ‘harmful’ range for DS_1 [12]. Therefore, variant 1 is preferred. We have shown that $DN_1 \geq DN_2$ and $DS_1 \geq DS_2$ [12].

The product of DN_2 and DS_2 has been shown to provide a lower bound for the proportion of explained variation. The product of DN_1 and DS_1 is generally close to EV with an acceptable discrepancy for practical settings [12].

According to Table 1 none of the predictors or set of predictors up to day 4 is strongly sufficient for in-hospital death of Covid-19 patients after day 4. This means that on average, the involved independent variables in the specified model do not exhibit a strong ‘force of mortality’ in the ‘harmful’ range. Age is the most important single predictor explaining 23% of the variability of in-hospital mortality, with a high degree of necessity ($DN_1 = 0.76$) but low sufficiency ($DS_1 = 0.28$). The information derived from repeated measures of four routine laboratory parameters during the first four days explains 30% (with $DN_1 = 0.67$, $DS_1 = 0.40$). As is often the case, EV for the full model (0.49) is limited by the moderate sufficiency of all predictors ($DS_1 = 0.50$). In

contrast, on the necessity side ($DN_1 = 0.88$), patients with a prognostic index (obtained from the full model) in the ‘protective’ range, may expect an average decrease of in-hospital mortality by 88% relative to unconditional in-hospital mortality.

Dichotomous markers

In the case of a dichotomous predictor the predictiveness curve is a step function. For fever on admission in the example data the predictiveness curve jumps from the risk of in-hospital death predicted for patients without fever (10.7%) to the risk predicted for those with fever (18.0%) at the risk quantile $q_0 = 0.45$ (corresponding to the proportion of patients presenting without fever). Thus, A_N and A_S correspond to rectangles with an area of only 0.018. The degrees of necessity and sufficiency are low (Table 1). Note that $DN_1 = DN_2$ and $DS_1 = DS_2$ and $EV_{ind} = DN_1 \cdot DS_1 = DN_2 \cdot DS_2$ for a single dichotomous prognostic factor [12].

From Section 5 in reference [12] it directly follows that, for locally calibrated models, DN_2 and DS_2 (and by a similar argument STG) remain unchanged after dichotomization at the predictor’s average-risk value. In contrast, the preferred measures, DN_1 , DS_1 and EV_{ind} , decrease by an amount that depends on the conditional variances of the risk estimates [12]. These results are confirmed when age is dichotomized at 71 years (corresponding to a risk of $P(D) = 0.147$) in the data example: $DN_2 = 0.71$, $DS_2 = 0.22$, $STG = 0.54$ (cf. Table 1), while EV_{ind} decreases from 0.23 to 0.16, DN_1 from 0.76 to 0.71, and DS_1 from 0.28 to 0.22. This complements warnings against dichotomizing predictors regarding other statistical aspects [13].

Results

In this section, we show that areas resulting from the predictiveness curve plot are closely connected to EV, DN and DS. Further connections between various measures of prediction performance (including total gain) and their asymptotic properties for hypothesis testing have been studied by Pepe et al. [14].

DN and DS and the predictiveness curve

In Appendix 1, we show that DN_2 equals the area A_N between $P(D)$ and the predictiveness curve for quantiles corresponding to the ‘protective’ range, relative to the total area below $P(D)$ in the same range, which is equal to $P(D) \cdot q_0$:

$$DN_2 = \frac{A_N}{P(D)q_0}$$

Analogously, DS_2 is equal to the area A_S between $P(D)$ and the predictiveness curve on the ‘harmful’ range, relative to the total area above $P(D)$ in the same range, which is equal to $(1 - P(D)) \cdot (1 - q_0)$:

$$DS_2 = \frac{A_S}{(1 - P(D))(1 - q_0)}$$

Thus, instead of standardizing $TG = A_N + A_S$, A_N and A_S are each standardized to their respective maximum possible value. Note, however, that these maxima are conditional on q_0 and do not necessarily correspond to a calibrated step function. In contrast, the denominator of STG results from the global calibration constraint leading to a calibrated step function with a step from 0 to 1 at $q_0 = 1 - P(D)$ [5], in which case $DN_2 = DS_2$.

Figure 2 shows the numerators of DN_2 and DS_2 , A_N and A_S , respectively, as hatched areas and their denominators as shaded areas. For age $A_N = A_S = 0.0665$ and for the laboratory parameters $A_N = A_S = 0.0660$. This results in nearly identical values for TG as well as for STG . However, due to different sizes of the shaded areas for the two predictor sets the values for DN , DS and EV differ considerably between these predictors (Table 1). Since $DN_1 \geq DN_2$ and $DS_1 \geq DS_2$ these relative areas in the predictiveness plot represent lower bounds for the preferred variant 1 of the degrees of necessity and of sufficiency.

Explained variation and the predictiveness curve

In Appendix 2 we show that STG is an upper bound for EV_{ind} . In our example, however, this bound is quite weak (cf. Table 1). Thus, we conclude that $DN_2 \cdot DS_2 \leq EV_{ind} \leq STG$. As explained above, equality for the lower bound holds with a single dichotomous prognostic factor, while for the upper bound with a perfectly separating 0–1 step function at $q_0 = 1 - P(D)$. Consequently, the proportion of variation in the outcome that is explained by a prognostic marker or a set of markers can be intuitively gauged from the areas A_N and A_S produced by the predictiveness curve.

Graphical comparison of predictiveness curves per se between predictors is often hampered by the fact that different predictors imply different values for q_0 (see, e.g., Fig. 2 age vs. laboratory parameters; or Fig. 2 in [1]). In contrast, EV, DN and DS give quantifications of relevant aspects of the predictiveness curve that incorporate q_0 . This allows for an easy comparison between predictors, which can be enhanced by a bootstrap-based statistical test [15].

As mentioned in the previous section, DN_2 and DS_2 only quantify the variation of conditional probabilities between the mean predictions on the ‘protective’ and the ‘harmful’ range. The same would also hold for any

standardization of TG (including STG and a standardization by the sum of the denominators of DN_2 and DS_2). In contrast, EV_{ind} , DN_1 and DS_1 also account for the variation of conditional probabilities *within* the ‘protective’ range and the ‘harmful’ range [12], making them more sensitive to the shape of the predictiveness curve.

In Fig. 3 of Steyerberg et al. [6] a horizontal line at $P(D)$ is shown in a graph of the predictiveness curve stratified by event status. This could be enhanced by horizontal lines at $P^1(D)$ among events and at $P^0(D)$ among non-events, respectively. Their vertical distance thus is an estimate of C_D . For our example data, this is shown in Fig. 1C. Similarly, this plot is enhanced by vertical lines at q_0^1 for events and at q_0^0 for non-events, with their horizontal distance giving an estimate of $AARD$.

Discussion

We conclude with some remarks:

Predictiveness curves as well as measures like EV , DN and DS assume a representative sample as their basis. Furthermore, only calibrated models should be compared based on predictiveness curves [16].

In the presentation of the theory and also of the data example, we used apparent values of EV , DN and DS , but also of $R(q)$ (cf. [4]). Since these measures and graphs have been derived on the same data that were used for estimating conditional probabilities bootstrap techniques could be applied to correct for over-optimism.

In the case of non-monotone conditional probabilities, the given definition for $R(q)$ ensures a monotone predictiveness curve by ordering the horizontal axis of the plot according to $P(D|X)$ instead of X [4]. Since the definitions of the ‘protective’ and the ‘harmful’ range do not depend on any ordering, the relations presented in this contribution still hold.

Predictiveness curves as well as EV , DN and DS do not depend on the model type used for estimating $P(D|X)$. All relations presented here are also valid for estimation algorithms other than the logistic regression model.

Several of the methods presented in this contribution have been extended to survival data, see [2, 7, 17].

SAS and R code for highlighting the areas A_N and A_S in the predictiveness curve have been added to the ‘Nec-Suff’ repository at www.github.com/agleiss.

Conclusions

This contribution offers an interpretation for the lower and the upper part of the total area between predictiveness curve and the horizontal line representing event prevalence as the degree of necessity and the degree of sufficiency, respectively. In addition, we have shown that certain areas and distances in the plot of the predictiveness curve are connected to measures of

predictiveness. Thus, the intuitive approach to compare these features between predictiveness curves for different prognostic factors or sets of factors results in a choice of factors according to their importance.

Appendix 1

Denote the predictor’s cumulative distribution function by $F(\cdot)$. By using $x = F^{-1}(q)$ for a given quantile q , $R(q) = P[D|X = F^{-1}(q)]$ [4]. Since by definition $q_0 = \int_{P(D) > P(D|X)} dF(x)$ we have

$$\begin{aligned} DN_2 &= E \left[\frac{P(D) - P(D|X)}{P(D)} \middle| P(D) > P(D|X) \right] \\ &= \frac{1}{q_0} \int_{P(D) > P(D|X)} \frac{P(D) - P(D|X)}{P(D)} dF(x) \\ &= \frac{1}{P(D)q_0} \int_0^{q_0} [P(D) - P(D|F^{-1}(q))] dq \\ &= \frac{A_N}{P(D)q_0} \end{aligned}$$

using the substitution $q = F(x)$.

Appendix 2

Gu & Pepe [11] show that

$$EV_{ind} = \int_0^1 \{P[P(D|X) > r|D] - P[P(D|X) > r|\neg D]\} dr$$

with $\neg D$ denoting the counter-event to D , and that the integrand is maximized at $r = P(D)$. Since $AARD$ exactly equals the integrand at $r = P(D)$ and $STG = AARD$ according to [11], this proves that $EV_{ind} \leq STG$.

Abbreviations

AARD	Above average risk difference
DN	Degree of necessity
DS	Degree of sufficiency
EV	Explained variation
OR	Odds ratio
STG	Standardized total gain
TG	Total gain

Acknowledgements

The author thanks Stefan Heber for the possibility to use the data published in Heber et al. [3] as an illustrative example. Furthermore, comments by Michael Schemper on a previous version of the manuscript are gratefully acknowledged.

Authors’ contributions

Not applicable (one author only).

Funding

This work has not been supported by any funding.

Data availability

The dataset analyzed during the current study is not publicly available but is available from the corresponding author of reference [3] on reasonable request.

Declarations**Ethics approval and consent to participate**

The studies involving human participants were reviewed and approved by the Ethical commission of the Medical University of Vienna (EK1315/2020). The patients/participants provided their written informed consent to participate in this study. The study was performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 November 2024 Accepted: 27 March 2025

Published online: 23 April 2025

References

1. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol*. 2008;167(3):362–8. <https://doi.org/10.1093/aje/kwm305>.
2. Gleiss A, Schemper M. Quantifying degrees of necessity and of sufficiency in cause-effect relationships with dichotomous and survival outcomes. *Stat Med*. 2019;38:4733–48. <https://doi.org/10.1002/sim.8331>.
3. Heber S, Pereyra D, Schrottmaier WC, Kammerer K, Santol J, Rumpf B, et al. A Model Predicting Mortality of Hospitalized Covid-19 Patients Four Days After Admission: Development, Internal and Temporal-External Validation. *Front Cell Infect Microbiol*. 2022;11:795026. <https://doi.org/10.3389/fcimb.2021.795026>. PMID: 35141170.
4. Huang Y, Sullivan Pepe M, Feng Z. Evaluating the predictiveness of a continuous marker. *Biometrics*. 2007;63(4):1181–8. <https://doi.org/10.1111/j.1541-0420.2007.00814.x>.
5. Bura E, Gastwirth JL. The Binary Regression Quantile Plot: Assessing the Importance of Predictors in Binary Regression Visually. *Biom J*. 2001;43(1):5–21.
6. Steyerberg EW, Veldter MM, Leening MJ, Postmus D, D'Agostino RB Sr, Van Calster B, Pencina MJ. Graphical assessment of incremental value of novel markers in prediction models: From statistical to decision analytical perspectives. *Biom J*. 2015;57(4):556–70. <https://doi.org/10.1002/bimj.201300260>.
7. Schemper M. Predictive accuracy and explained variation. *Stat Med*. 2003;22:2299–308. <https://doi.org/10.1002/sim.1486>.
8. Gleiss A, Zeillinger R, Braicu EI, Trillsch F, Vergote I, Schemper M. Statistical controversies in clinical research: the importance of importance. *Ann Oncol*. 2016;27:1185–9. <https://doi.org/10.1093/annonc/mdw159>.
9. Tjor T. Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. *Am Stat*. 2009;63:366–72. <https://doi.org/10.1198/tast.2009.08210>.
10. Pepe MS, Feng Z, Gu JW. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al. *Stat Med*. 2008;27(2):173–81. <https://doi.org/10.1002/sim.2991>.
11. Gu W, Pepe M. Measures to summarize and compare the predictive capacity of markers. *Int J Biostat*. 2009;5(1):Article 27. <https://doi.org/10.2202/1557-4679.1188>.
12. Gleiss A, Henderson R, Schemper M. Degrees of necessity and of sufficiency: Further results and extensions, with an application to Covid-19 mortality in Austria. *Stat Med*. 2021;40:3352–66. <https://doi.org/10.1002/sim.8961>.
13. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127–41. <https://doi.org/10.1002/sim.2331>.
14. Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Stat Med*. 2013;32(9):1467–82. <https://doi.org/10.1002/sim.5727>.
15. Heinze G, Schemper M. Comparing the importance of prognostic factors in Cox and logistic regression using SAS. *Comput Methods Progr Biomed*. 2003;71(2):155–63. [https://doi.org/10.1016/S0169-2607\(02\)00077-9](https://doi.org/10.1016/S0169-2607(02)00077-9).
16. Casalicchio G, Bischl B, Boulesteix AL, Schmid M. The Residual-Based Predictiveness Curve: A Visual Tool to Assess the Performance of Prediction Models. *Biometrics*. 2016;72:392–401. <https://doi.org/10.1111/biom.12455>.
17. Choodari-Oskoei B, Royston P, Parmar MK. The extension of total gain (TG) statistic in survival models: properties and applications. *BMC Med Res Methodol*. 2015;15:50. <https://doi.org/10.1186/s12874-015-0042-x>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.