ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Themed Section: Digital Health Technologies

Qini Curves for Potential Impact Assessment of Risk Predictive Models Informing Intervention Policies

Pierpaolo Palumbo, PhD

ABSTRACT

Objectives: Predictive models in medicine help make decisions about which individual to treat with a given therapeutic or preventive intervention. Before being tested in large field studies and recommended for clinical adoption, it is important to evaluate not only their statistical accuracy but also the impact they may have when used to inform health intervention policies. We aim to provide simple methods for the potential impact assessment of health intervention policies based on predictive models.

Methods: We propose an analytic framework based on Qini curves wherein prediction-based policies are analyzed on 2 impact endpoints: (1) the fraction of the population that would be selected for the intervention (coverage) and (2) the effect on the clinical outcomes of interest (disutility). The drivers of values are the disease prevalence, the predictive performance of the model, and the effectiveness of the intervention.

Results: We present simple formulas for calculating coverage and disutility from either observational or randomized controlled data. We illustrate possible value measures arising from geometrical properties on the Qini plane: delta coverage and disutility, number needed to treat, and integrated difference between Qini curves. We show the applicability of the Qini analysis by providing examples about the prevention of falls in older adults and prevention of secondary cardiovascular events with pioglitazone.

Conclusions: Coverage and disutility capture key value components of prediction-based policies. The method can be used for comparing models or tuning risk thresholds for managing trade-offs between conflicting objectives (eg, clinical benefits, side effects, and healthcare resources).

Keywords: impact, policy, population metrics, Qini curve, risk model.

VALUE HEALTH. 2026; 29(4):567–574

Highlights

- Predictive models are cornerstone digital technologies used to implement the paradigm of personalized medicine. Before being tested on large field studies and recommended for clinical adoption, it is important to estimate not only their statistical accuracy but also the consequences they could bring about if implemented on a population.
- Qini curves are gain-spend curves that are currently used for treatment benefit prediction models. In this study, we propose a methodology for deriving Qini curves on risk prediction models. The 2 endpoints are the fraction of the population that would be selected for the intervention (coverage) and the effect on the clinical outcome of interest (disutility).
- Qini curves can be used to appraise the impact of risk prediction models that inform health intervention policies before real-world implementation and tuning risk thresholds for managing trade-offs between conflicting objectives.

Introduction

Predictive health models are cornerstone elements of digital health technologies that help make a diagnosis of present conditions or a prognosis about future outcomes or may predict the responsiveness of different individuals to different interventions. In clinical practice, they are used to advise on therapeutic or preventive interventions to take on single individuals. At the same time, in public health, they support decision making about which population groups to target with such interventions and assist in preparing resource allocation plans.[1]

We distinguish predictive models into risk predictive models and treatment benefit predictive models. The former predicts the absolute risk of having or developing a given outcome of interest, whereas the latter predicts the benefit of an intervention over a given outcome at the single individual level. Treatment benefit models are also known as uplift models in business applications[2,3] and models for the heterogeneous treatment effect in the medical domain.[4-8]

The evidence-based evaluation process that accompanies a predictive model from its development to its recommendation for clinical uptake and real-world implementation has been described as staged into 3 phases,[9] which concern (1) predictive performance, (2) potential effect and usability, and (3) post-implementation impact. The first phase is addressed with internal and external validation studies.[10,11] In the second phase, the potential impact on health outcomes and healthcare resources can be estimated with decision analytic models,[12] whereas the third phase is conducted with comparative studies, such as before-after studies or clustered randomized controlled trials.[13-18] The early impact assessment of the second phase helps identify which models can progress along the evaluation process, reconciling the vast amount of models that reach the first phase with the high costs and complexity of the studies of the third phase.[9,17,19]

Different methods and metrics have been proposed to estimate the value of predictive models.[20-23] Qini curves are gain-spend curves that have been proposed to evaluate the value of treatment benefit models.[24-27] Because these models provide a prioritization score on subjects on whom to intervene, Qini curves plot the benefit of intervening on a subset of prioritized individuals (gain) as a function of the size of this subset (spend).

In this article, we aim to present Qini curves as a simple tool for the potential impact evaluation of risk predictive models before real-world implementation. We find relationships to express these curves in terms of 3 drivers of value: the prevalence of the targeted disease, the performance of the risk predictive model, and the effectiveness of the available intervention.[28] We highlight some geometrical properties of the Qini plane that are useful to appreciate the benefit of the predictive model. Finally, we present 2 examples of Qini curves for risk predictive models evaluated using observational and randomized data.

Methods

Problem Setup

We call Ω our target population. We define $X(\omega)$ as the observable features of an individual ω belonging to Ω, $w \in \{0;1\}$ as the intervention implemented on each individual (ie, to administer or not a given treatment) and $Y(\omega, w) \in \{0;1\}$ as a dichotomic health outcome for ω after receiving intervention w. Conventionally, we indicate with $Y = 1$ ($Y = 0$) the presence (respectively, absence) of the unpleasant disease condition.

We call f a risk prediction model for the outcome Y. Namely, $f(x, w)$ is an estimate of the probability for an individual ω with feature $x = X(\omega)$ to have outcome $y = 1$ after intervention w. We further call π the policy for intervention assignment. Namely, π is the function that assigns an individual ω with feature $x = X(\omega)$ to the intervention w: $w = \pi(x)$. We consider 4 policies:

1. $\pi_0(x) = 0$. Under this policy, w is 0 for all individuals (ie, none receives the treatment). We can refer to this as an intervention-on-none policy.
2. $\pi_1(x) = 1$. Under this policy, w is 1 for all individuals (ie, everyone receives the treatment). We can refer to this as the intervention-on-everyone policy.
3. $\pi_{r,\alpha}(x) \sim Bernoulli(\alpha)$. Under this policy, a fraction α of the individuals are selected for treatment $w = 1$, independent of their features x. This policy is a mixture of π_0 and π_1. We can refer to it as random policy.
4. $\pi_{f,t}(x) = I(f(x, 0) < t)$, in which $I()$ is the indicator function, which is 1 if its argument is true and 0 otherwise, and t is a threshold value over the estimated risk $f(x, 0)$. Under this policy, the treatment $w = 1$ is assigned to the individuals whose estimated risk exceeds a given threshold t, whereas all others do not receive the treatment.

Policy $\pi_{f,t}$ is the prediction-and-intervention (or test-and-treat) policy under evaluation, whereas the others serve as comparators and may represent the standard of care. The policy $\pi_{r,\alpha}$ may describe an observational regime in which the individuals are subject to different treatments according to criteria independent of their features X or may represent an experimental regime in which randomization guarantees that the treatment allocation is independent of X.

The prediction model f is often developed and evaluated using data collected under only 1 policy regime, which we assume to be π_0 for simplicity. We thus indicate $f(x, 0)$ as $f(x)$. Using data collected under π_0, we characterize f in terms of the true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates i achieves over the population. We call $Se = TP/(TP + FN)$ the sensitivity and $Sp = TN/(TN + FP)$ the specificity of the risk prediction model.

We characterize the effectiveness of the intervention over the population with its relative risk[29]:

$$RR = \frac{\sum_{\omega \in \Omega} Y(\omega, 1)}{\sum_{\omega \in \Omega} Y(\omega, 0)} \tag{1}$$

Similarly, we call RR_G the relative risk over a subgroup $G \subseteq \Omega$ of the population:

$$RR_G = \frac{\sum_{\omega \in G} Y(\omega, 1)}{\sum_{\omega \in G} Y(\omega, 0)} \tag{2}$$

We say that the intervention has uniform treatment effects when $RR_G = RR$ for all subgroups G of Ω. Otherwise, we state that it has heterogeneous treatment effects.

Finally, we choose 2 population metrics to evaluate the impact of a policy π: coverage and disutility. We define the coverage Co as follows:

$$Co(\pi) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \pi(X(\omega)) \tag{3}$$

It is the fraction of individuals receiving intervention $w = 1$. It provides information about the cost of its implementation and the number of possible side effects that may arise from its application. We define the disutility D as follows:

$$D(\pi) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} Y(\omega, \pi(X(\omega))) \tag{4}$$

It is the fraction of individuals developing the outcome $y = 1$. It measures the clinical effectiveness of the policy on the outcome Y. We call $D_0 = D(\pi_0) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} Y(\omega, 0) = TP + FN$ the health outcome rate (prevalence or incidence rate) under π_0. Plotting D as a function of Co makes the Qini curve.

All variable definitions are reported in Table 1.

Receiver Operating Characteristic (ROC) and Qini Curves

Given the formal framework presented in the section Problem setup, the fraction of individuals in the population that are assigned to the intervention (coverage of $\pi_{f,t}$) is given by the following (see Qini on observational data in Supplemental Materials):

$$Co(\pi_{f,t}) = D_0 Se + (1 - D_0)(1 - Sp) \tag{5}$$

The fraction of individuals affected by the disease after applying the intervention according to $\pi_{f,t}$ (disutility of $\pi_{f,t}$) is given by the following (see Qini on observational data in Supplemental Materials):

$$D(\pi_{f,t}) = D_0 - D_0(1 - RR_{G_t}) Se \tag{6}$$

in which RR_{G_t} is the relative risk over the group of individuals whose estimated risk exceeds the threshold t.

These equations describe the coverage and disutility of a prediction-and-intervention policy $\pi_{f,t}$ in terms of 3 drivers of value: the prevalence of the disease (D_0), the predictive

Table 1. Definitions.

Symbol	Definition
Population, intervention, and outcome	
Ω	Target population. $\|\Omega\|$ indicates its sample size (ie, the number of its individuals).
ω	Index individual of the population. $\omega \in \Omega$.
$X(\omega)$	Observable features of an individual ω.
w	Type of intervention implemented on each individual. We take it as dichotomic: $w \in \{0; 1\}$ (ie, to administer or not a given treatment).
$Y(\omega, w)$	Health outcome for ω after receiving intervention w. We take it as dichotomic: $Y(\omega,w) \in \{0; 1\}$. We indicate with $Y = 1$ $(Y = 0)$ the presence (respectively, absence) of the unpleasant disease condition.
RR	Relative risk of the intervention over the whole population: $RR = \frac{\Sigma_{\omega \epsilon \Omega} Y(\omega, 1)}{\Sigma_{\omega \epsilon \Omega} Y(\omega, 0)}$
RR_G	Relative risk of the intervention over a subgroup G: $RR = \frac{\Sigma_{\omega \epsilon G} Y(\omega, 1)}{\Sigma_{\omega \epsilon G} Y(\omega, 0)}$
RR_{Gt}	Relative risk of the intervention over the group of individuals whose estimated risk exceeds the threshold t.
Risk prediction model	
f	A risk prediction model for the outcome Y. Namely, $f(x, w)$ is an estimate of the probability for an individual ω with feature $x = X(\omega)$ to have outcome $y = 1$ after intervention w. For simplicity, we indicate $f(x, 0)$ as $f(x)$.
TP, TN, FP, FN	True-positive, true-negative, false-positive, and false-negative rates of $f(x)$.
Se, Sp	Sensitivity and specificity of $f(x)$.
AUC	Area under the receiver operating characteristic curve of $f(x)$. A measure of the discriminative ability of the prediction model across different threshold values.
Policy	
π	Policy for intervention assignment. Namely, π is the function that assigns an individual ω with feature $x = X(\omega)$ to the intervention w: $w = \pi(x)$.
π_0	Intervention-on-none policy. Under this policy, w is 0 for all individuals (ie, none receives the treatment: $\pi_0(x) = 0$)
π_1	Intervention-on-everyone policy. Under this policy, w is 1 for all individuals (ie, everyone receives the treatment: $\pi_1(x) = 1$).
$^{\pi}r,\alpha$	Random policy. Under this policy, a fraction α of the individuals are selected for treatment $w = 1$, independent of their features: $w = \pi_{r,\alpha}(x) \sim Bernoulli(\alpha)$, which indicates that w takes the value 1 with probability α (and value 0 with probability 1- α).
π_c	Generic comparator policy

Continued in the next column

Table 1. Continued

Symbol	Definition
$^{\pi}f,t$	Prediction-and-intervention policy. Under this policy, the treatment $w = 1$ is assigned to the individuals whose estimated risk exceeds a given threshold t, whereas all others do not receive the treatment: $\pi_{f,t}(x) = I(f(x, 0) < t)$, in which I is the indicator function, which is 1 if its argument is true and 0 otherwise, and t is a threshold value over the estimated risk $f(x, 0)$.
$Co(\pi)$	Coverage of policy π. It is the fraction of individuals receiving intervention $w = 1$ under policy π. $Co(\pi) = \frac{1}{\|\Omega\|} \Sigma_{\omega \epsilon \Omega} \pi(X(\omega))$.
$D(\pi)$	Disutility of policy π. It is the fraction of individuals developing the outcome $y = 1$ under policy π. $D(\pi) = \frac{1}{\|\Omega\|} \Sigma_{\omega \epsilon \Omega} Y(\omega, \pi(X(\omega)))$.
D_0	Health outcome rate under $\pi_0 : D_0 = D(\pi_0) = \frac{1}{\|\Omega\|} \Sigma_{\omega \epsilon \Omega} Y(\omega, 0)$
$^{\Delta}(ft,c)^{Co}$, $^{\Delta}(ft,c)^{D}$	Differences in coverage and disutility between $\pi_{f,t}$ and a comparator policy π_c.
$^{\Delta}(ft,r)^{Co}$, $^{\Delta}(ft,r)^{D}$	Difference in coverage and disutility between $\pi_{f,t}$ and two random policies having respectively the same disutility and coverage of $\pi_{f,t}$.
$NNT(\pi)$	Number needed to treat under policy π (ie, the number of individuals who must be treated to gain a one-unit decrease in disutility).
$AUPEC(\pi)$	Area under the prescriptive effect curve. A measure of the effectiveness of integrating the prediction tool f with the available therapeutic or preventive intervention using policy π.
Example of Qini for fall risk screening (observational data)	
AGS/BGS	American Geriatrics Society/British Geriatrics Society (AGS/BGS) guidelines for falls prevention.
$NNT_{AGS/BGS,13.5s}$	Number needed to treat for the policy indicated by the American Geriatrics Society/British Geriatrics Society (AGS/BGS) guidelines using the Timed Up and Go (TUG) test for screening with a cutoff of 13.5 seconds.
$NNT_{\pi 1}$	Number needed to treat when providing the intervention to everybody, without any risk screening tool.
Example of Qini for secondary prevention of cardiovascular events with pioglitazone (randomized data)	
G_{00}, G_{10}, G_{01}, G_{11}	G_{ab} (with $a, b \in \{0; 1\}$) is the group of individuals randomized to treatment a by the random policy $\pi_{r,\alpha}$ and indicated to receive treatment b by the prediction tool f.
Y_{cv}, $Y_{fracture}$	Two health outcomes: cardiovascular events (Y_{cv}) and occurrence of bone fracture ($Y_{fracture}$) in a 5-year observation period.
D_{cv}, $D_{fracture}$	Disutilities calculated over the 2 health outcomes Y_{cv} and $Y_{fracture}$.

performance of the model (*Se* and *Sp*), and the effectiveness of the intervention (RR_{G_t}). Coverage is an affine function of sensitivity and specificity, whereas disutility is affine in sensitivity. An improvement in sensitivity brings an improvement in the disutility via the mediation of RR_{G_t}, with no impact if the intervention is not effective ($RR_{G_t} = 1$). Expressions for coverage and disutility also for the 3 comparator policies are provided in Appendix Table 1 in Supplemental Materials.

The relationship entailed by these formulas between the ROC curves and the Qini curves are graphically shown in Figure 1 using synthetic data (see Synthetic data in Supplemental Materials). The prediction-and-intervention policy $\pi_{f,t}$ is compared with an intervention-on-none policy (π_0), an intervention-on-everyone policy (π_1), and a random policy ($\pi_{r,\alpha}$). The random policy is represented as a straight line in both the ROC and Qini planes. It is parametrized by α and connects the points representing π_0 and π_1. The Qini curve shows that an improvement in the sensitivity determines an increased coverage and an improved (decreased) disutility, whereas an improvement in the specificity determines a reduced coverage. The same ROC curve corresponds to different Qini curves for different values of the intervention effectiveness *RR*.

Qini Analysis on Randomized Data

Although observational data can be advantageous for training risk prediction models because they are generally more abundant and representative of the target population, experimental data offer the opportunity to observe the characteristics of the individuals (X), the assigned intervention (w), and the clinical outcome (Y) without threats of confounding effects.

Given data produced under the randomized policy $\pi_{r,\alpha}$, the coverage of $\pi_{f,t}$ can be estimated as follows (see Qini on randomized data in Supplemental Materials):

$$Co(\pi_{f,t}) = \frac{|G_{11}| + |G_{01}|}{|\Omega|} \tag{7}$$

in which G_{11} is the group of individuals who are assigned to the intervention ($w = 1$) by both the randomized and the prediction-based policies, G_{01} is the group of individuals who are in the control group according to $\pi_{r,\alpha}$ but would receive the intervention according to $\pi_{f,t}$, and $|\Omega|$ is the sample size of the randomized data. The disutility of $\pi_{f,t}$ can be estimated with inverse probability weighting as follows:

$$D(\pi_{f,t}) = \frac{1}{|\Omega|}\left\{\frac{1}{1-\alpha}\sum_{\omega \in G_{00}} Y(\omega, 0) + \frac{1}{\alpha}\sum_{\omega \in G_{11}} Y(\omega, 1)\right\} \tag{8}$$

in which G_{00} is the group of individuals who are not assigned to the intervention ($w = 0$) by either the randomized or the prediction-based policies, and α is the fraction of individuals randomized to the intervention by $\pi_{r,\alpha}$ (eg, $\alpha = 0.5$ for balanced cases and controls).

Value Measures

We highlight 3 approaches for appraising the value of a prediction-and-intervention policy in the Qini plane.

Delta coverage and disutility

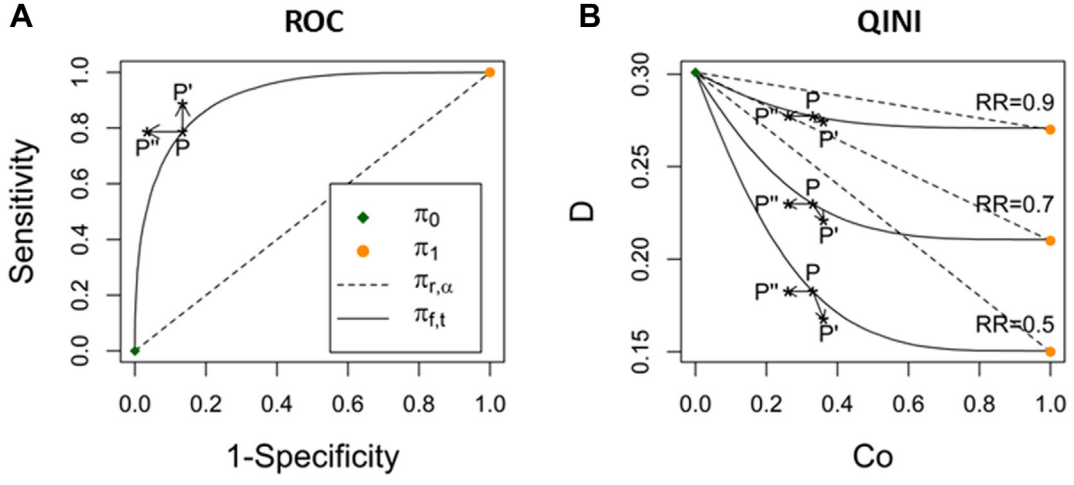
The value of an index prediction-and-intervention policy $\pi_{f,t}$ with respect to a comparator policy π_c may be appreciated by calculating the differences between the 2 in terms of coverage and disutility: $\Delta_{(ft,c)}D = D(\pi_c) - D(\pi_{f,t})$ and $\Delta_{(ft,c)}Co = Co(\pi_c) - Co(\pi_{f,t})$. When the comparator π_c represents the standard of care, $Co(\pi_c)$ and $D(\pi_c)$ are assumed to be known.

To evaluate the value of the prediction model f, we may take as comparator a random policy with either the same coverage or disutility of $\pi_{f,t}$. Closed-form formulas for such $\Delta_{(ft,r)}D$ and $\Delta_{(ft,r)}Co$ are provided in Delta coverage and disutility in Supplemental Materials.

Number needed to treat

A classical way to express the efficiency of an intervention is the number needed to treat (NNT) (ie, the number of individuals

Figure 1. ROC curves and Qini curves. One same ROC curve (panel A) is mapped into different Qini curves (panel B) depending on the effectiveness of the intervention (ie, its relative risk [RR]). The arrows show the effect of the predictive accuracy of the risk model on coverage (Co) and disutility (D). In particular, given a point P on the ROC and Qini curves, corresponding to a specific risk threshold, P′ and P″ represent improvements over P of 0.1 in sensitivity and specificity, respectively. Curves obtained with synthetic data (see Synthetic data in Supplemental Materials). π_0, intervention-on-none policy; π_1, intervention-on-everyone policy; $\pi_{(r,\alpha)}$, random policy; $\pi_{(f,t)}$, prediction-and-intervention policy.



who must be treated to gain a 1-unit decrease in disutility).[30] Because the abscissa indicates the number of individuals receiving the intervention and the ordinate indicates the disutility in the Qini plane, the NNT of a given policy π is the slope over the vertical of the segment connecting the points representing π and π_0 (Fig. 2). For the prediction-and-intervention policy $\pi_{f,t}$, the NNT is

$$NNT(\pi_{f,t}) = \frac{Co(\pi_{f,t}) - Co(\pi_0)}{Co(\pi_0) - Co(\pi_{f,t})} = \frac{TP(t) + FP(t)}{D_0 - RR_{G_t} TP(t) - FN(t)} = \frac{TP(t) + FP(t)}{TP(t)(1 - RR_{G_t})} \quad (9)$$

For the comparator π_1 (or equivalently any $\pi_{r,\alpha}$ with $\alpha < 0$),

$$NNT(\pi_1) = NNT(\pi_{r,\alpha}) = \frac{Co(\pi_1) - Co(\pi_0)}{Co(\pi_0) - Co(\pi_1)} = \frac{1}{D_0(1 - RR)} \quad (10)$$

Area under the prescriptive effect curve (AUPEC)

The third approach we propose is based on measuring the integrated difference between Qini curves. To appraise the value of $\pi_{f,t}$, we measure the area in the Qini plane between the curve representing $\pi_{f,t}$ and the curve representing the random intervention $\pi_{r,\alpha}$. Following Imai and Lingzhi Li,[24] we refer to this quantity as AUPEC. (Fig. 2). Because the area under the ROC curve (AUC) is a measure of the discriminative ability of the prediction model f across different threshold values,[31,32] AUPEC is a measure of the discriminative effectiveness of the prediction-and-intervention policy $\pi_{f,t}$, (ie, the effectiveness of integrating the prediction model f with the available therapeutic or preventive intervention).

Under the hypothesis of homogeneous treatment effects across population risk strata ($RR_{G_t} = RR$ for all t), area under the ROC curve and AUPEC are related by the following simple equation:

$$AUPEC = D_0(1 - D_0)(AUC - 0.5)(1 - RR) \quad (11)$$

This shows that the discriminative effectiveness is determined by the disease prevalence, the discriminative ability of the predictive model, and the effectiveness of the intervention. The proof is provided in AUPEC in Supplemental Materials.

Results

Qini Analysis on Observational Data

Here, we report an example of an impact analysis derived from Palumbo et al[33] that uses Qini curves obtained with observational data. The analyzed prediction-and-intervention policy is the program for fall prevention in community-dwelling older adults indicated by the American Geriatrics Society/British Geriatrics Society guidelines.[34] This policy consists of using a screening algorithm in combination with a multifactorial assessment and a tailored intervention. Namely, those who are screened as at high risk of falls are referred to a multifactorial assessment and a tailored preventive intervention for determining the main risk factors that are responsible for the increased fall risk and for intervening in those risk factors.

The prediction model f is the screening algorithm proposed by the American Geriatrics Society/British Geriatrics Society, which makes use of information about previous falls, fall injuries, and difficulties or abnormalities in gait and balance. The coverage Co is the fraction of older adults targeted with the multifactorial assessment and intervention, and the disutility D is the fraction of older adults experiencing at least 1 fall a year.

Given the low diffusion of fall preventive interventions, the treat-none policy π_0 well approximates the policy that generated the data and the standard of care, whereas $\pi_{f,t}$ is the policy to evaluate. The policy $\pi_{r,\alpha}$ mimics a preventive service that is accessed by individuals without risk-related criteria, as is the case for a service that is implemented heterogeneously over the national territory.

We used the InCHIANTI observational data[35] to derive the annual fraction of older adults experiencing at least 1 fall (D_0) and the accuracy of the screening algorithm in terms of *TP*, *TN*, *FP*, *FN*, *Se*, and *Sp*. The effectiveness of the multifactorial assessment and intervention was derived from the literature,[36] and we assumed that such effectiveness was homogeneous across population subgroups.[33]

The ROC plane (see Appendix Fig. 1 in Supplemental Materials) shows that the screening algorithm has sensitivity and specificity that range in the intervals of 35.1% to 43.3% and 79% to 84.3%, respectively, according to the threshold chosen on a test for detecting gait abnormalities (namely, the Timed-Up and Go [TUG] test).[34] The Qini curve (Fig. 3) shows that this preventive policy would target 18.4% to 24.2% (Co) of the older population with a multifactorial assessment and intervention, resulting in a fall rate of 12.8% to 13% fallers every year (D). The Qini curve further helps us to appraise the advantages of integrating the screening algorithm within the preventive policy with respect to others not based on predictive tools. More specifically, choosing the TUG test with a threshold equal to 13.5 s leads to more prevented cases ($\Delta_{(AGS/BGS13.5s,t)}D$ = 0.5% fallers a year), reduced

Figure 2. Value measures on Qini curves. The number needed to treat (NNT) of an index policy is represented by the slope over the vertical of the line connecting the intervention-on-none policy (π_0) to the index policy. The area under the prescriptive effect curve (AUPEC) is the area between the Qini curve of the random policy ($\pi_{(r,\alpha)}$) and the Qini curve of the index policy (in the figure, the prediction-and-intervention policy $\pi_{(f,t)}$). Curves obtained with synthetic data (see Synthetic data in Supplemental Materials). π_1, intervention-on-everyone policy.

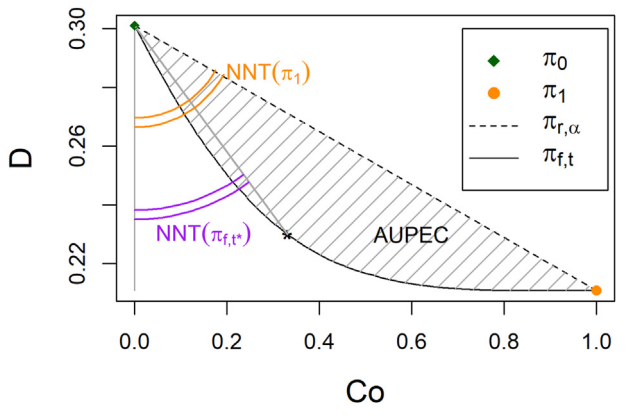


Figure 3. Qini curve for the prediction-based intervention for fall prevention of the American Geriatrics Society/British Geriatrics Society (AGS/BGS). The coverage (Co) is the fraction of the older population targeted with a multifactorial fall risk assessment and tailored intervention. The disutility (D) is the fraction of individuals who fall at least once a year. The fall risk model is a screening algorithm comprising different items, including the TUG test. The solid black line corresponds to the policy using the AGS/BGS screening algorithm for different thresholds on the TUG test. The asterisk corresponds to a threshold of 13.5 seconds (π(AGS/BGS,*13.5s)). The number needed to treat (NNT) for the intervention-on-everyone policy (π_1) is 32.2; the NNT for π(AGS/BGS,*13.5s) is 16.9. ΔCo and ΔD indicate the difference in coverage and disutility between π(AGS/BGS,*13.5s) and 2 random policies having, respectively, the same disutility and coverage of π(AGS/BGS,*13.5s) (horizontal and vertical gray solid lines).

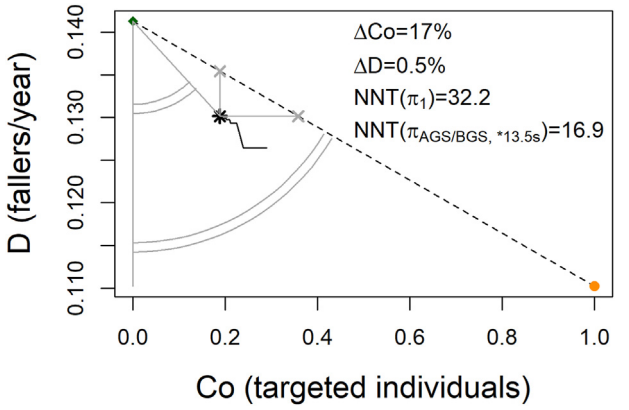
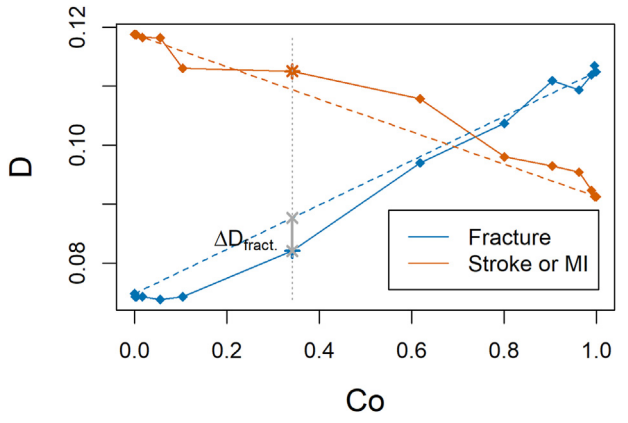


Figure 4. Qini curves for prevention of secondary cardiovascular events with pioglitazone on individuals at low risk of bone fracture. The coverage (Co) represents the fraction of the population (nondiabetic patients with a recent history of cardiovascular events and with insulin resistance) targeted with pioglitazone for prevention of secondary cardiovascular events. The disutility (D) represents the fraction of individuals who incur a secondary stroke or myocardial infarction (MI) (orange curves) or bone fracture (blue curves) in a 5-year observation period. The prediction-and-intervention policy (solid lines) consists of treating with pioglitazone those at low risk of bone fracture, according to the bone fracture risk model by Viscoli et al.[38] The asterisk and the dotted vertical line indicate the cutoff for the lowest tertile for the risk of fracture. The random policy is represented with dashed lines. As coverage with pioglitazone increases, secondary cardiovascular events drop from 11.9% to 9.1%, whereas bone fractures rise from 7.5% to 11.2%. The prediction-and-intervention policy produces a reduction of bone fractures with respect to a random policy with the same coverage (8.2% vs 8.8%, $\Delta D_{(fract.)}$ = 0.6%).



coverage ($\Delta_{(AGS/BGS13.5s,r)}Co$ = 17% individuals), which is connected to reduced costs for prevention, and an increased efficiency of the intervention ($NNT_{AGS/BGS,13.5s}$ =16.9 vs. NNT_{π_1} = 32.2).

Qini Analysis on Randomized Data

We show an example of the Qini methodology for risk models using randomized data considering the Insulin Resistance Intervention After Stroke study, a randomized controlled trial that aimed to test the efficacy of an insulin sensitizer, pioglitazone, for secondary prevention of cardiovascular events.[37] In the Insulin Resistance Intervention After Stroke study, they included 3876 participants with a recent history of ischemic stroke or transient ischemic attack, with insulin resistance, and without diabetes (target population Ω): 1939 participants were assigned to pioglitazone and 1937 to placebo. The study demonstrated that pioglitazone (intervention w) is effective in reducing the risk of recurring cardiovascular events after 5 years but at the expense of causing side effects. Of these, the most concerning were bone fractures. For this reason, Viscoli et al[38] developed a predictive model (f) for the 5-year risk of bone fractures and proposed to supply pioglitazone therapy only to those at low risk (prediction-and-intervention policy $\pi_{f,t}$).

We consider 2 health outcomes: the occurrence of cardiovascular events (stroke or myocardial infarction, Y_{cv}) and the occurrence of bone fracture ($Y_{fracture}$) in a 5-year observation period. Consistently, we evaluate the policy $\pi_{f,t}$ looking at two Qini curves for the 2 disutilities D_{cv} and $D_{fracture}$.

The Qini curves (Fig. 4) show that by increasing the coverage of pioglitazone from 0% to 100% in the eligible population, the 5-year risk of cardiovascular events (stroke and myocardial infarction) drops from 11.9% to 9.1%, whereas the risk of bone fractures rises from 7.5% to 11.2%. The prediction model is used to spare high-risk individuals from receiving a therapy that could exacerbate their fracture risk (ROC curve; see Appendix Fig. 2 in Supplemental Materials). For example, identifying those in the lowest risk tertile t entails reduced cardiovascular events with respect to the treat-none policy ($D_{cv}(\pi_{f,t})$ = 11.3% vs $D_{cv}(\pi_0)$ = 11.9%) but also reduced fractures with respect to the treat-everyone policy ($D_{fracture}(\pi_{f,t})$ = 8.2% vs $D_{fracture}(\pi_1)$ = 11.2%) and a random policy with the same coverage (8.2% vs 8.8%, $\Delta_{(ft,r)}D_{fracture}$ = 0.6%).

Discussion

The Qini analysis that we have presented in this article is a method for estimating the impact of a risk predictive model used in combination with an intervention strategy before real-world implementation. Although Qini curves are currently used for treatment benefit predictive models, we have presented a methodology for deriving these curves also for risk predictive models, which are more common in the literature and in clinical practice.

The risk predictive model was intended as either diagnostic or prognostic, and accordingly, the intervention was thought to be therapeutic or preventive.

We chose coverage and disutility to capture essential impact dimensions at a population level[9] and analyze the trade-offs inside the problem that the prediction-and-intervention policy aims to optimize. In the first example that we have presented, the prediction-and-intervention policy was intended to prevent falls (ie, reduce the disutility) while limiting the preventive intervention only to the subgroup of high-fall-risk individuals (ie, limiting the coverage). The disutility was directly defined as the fall incidence, whereas the coverage determines the healthcare resources that are needed to implement the preventive intervention.[39,40] In the second example, the 2 conflicting objectives were reducing the risk of secondary cardiovascular events while keeping bone fractures to a minimum. Accordingly, we defined 2 disutilities for the 2 clinical outcomes. We used the coverage (the fraction of individuals treated with pioglitazone) to highlight the trade-off between the 2 disutilities and show the advantages of the predictive model for fractures. Taken together, coverage and disutility form a plane that mirrors the cost-disutility plane used in health technology assessment.[41]

Although coverage and disutility are the endpoints of the Qini analysis, the inputs are 3 drivers of value, namely, the prevalence of the targeted disease, the performance of the predictive model, and the effectiveness of the available intervention.[28] Inputs and outputs are linked with a simple analytical model that supports theory-based evaluation.[42]

Input drivers of value can be derived from different data sources. In the first example, we derived the predictive performance of the risk predictive model and the disease incidence under the no intervention policy from observational data. The effectiveness of the intervention was taken from the literature and was used for the Qini analysis under the assumption of homogeneous treatment effects. In the second example, the experimental data were sufficient to run the analysis without additional information sources or assumptions on homogeneous treatment effects.

From a graphical point of view, Qini curves aim to parallel ROC curves, the latter being used for assessing the predictive performance of a risk predictive model, whereas the former is used for estimating its potential impact. In the Qini plane, we highlighted 3 possible measures of value based on linear distances (ΔCo and ΔD), angles (NNT), and areas (AUPEC). ΔCo and ΔD are often used in health technology assessment, NNT is a classical population metric to express the efficiency of an intervention, and AUPEC is a summary indicator of the efficiency of a prediction-and-intervention policy over different risk thresholds.[24,26]

We used closed-form algebraic formulas, graphical methods, and minimal external sources of data other than those used for developing the predictive model itself. Given the high number of risk predictive models and the low number of those that progress along the evaluation phases, the technical simplicity of performing Qini analyses is intended to induce more predictive models to be evaluated for their potential impact at an early stage.

For the sake of simplicity, we did not explicitly model many factors that characterize the complexity of interventions based on predictive models. For example, our analysis does not explicitly consider issues of usability of the predictive model, transferability of evidence results from studies in clinical practice, or other sociotechnical translational issues that deserve attention in later stages of the evaluation process.[9,43,44] Remarkably, costs were also not included explicitly but can be easily estimated downstream using coverage and disutility estimates. Finally, we focused on 3 elementary policies as comparators: the intervention-on-none, the intervention-on-everyone, and the random policies. However, according to the decision problem at hand, it may be necessary to identify and operationalize more complex standard of care pathways.[45]

Conclusions

Qini analysis can be used for evaluating the potential impact of a risk predictive model used in combination with a therapeutic or preventive intervention. It can be performed using simple closed-form formulas and focuses on population impact measures. Its simplicity aims at encouraging early impact assessment of predictive models in medicine.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2025.01.024.

Article and Author Information

Accepted for Publication: January 28, 2025

Published Online: March 8, 2025

doi: https://doi.org/10.1016/j.jval.2025.01.024

Author Affiliations: Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi"-DEI, University of Bologna, Bologna, Italy (Palumbo).

Correspondence: Pierpaolo Palumbo, PhD, Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi"-DEI, University of Bologna, Via Saragozza, 8, 40123 Bologna, Italy. Email: pierpaolo.palumbo@unibo.it

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This research was cofunded by the Italian Complementary National Plan PNC-I.1 "Research initiatives for innovative technologies and pathways in the health and welfare sector" D.D. 931 of 06/06/2022, Digital Lifelong Prevention (DARE) initiative, code PNC0000002, CUP: (B53C22006450001).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The author is grateful to Prof Lorenzo Chiari and the anonymous reviewers for providing comments on the manuscript. The Insulin Resistance Intervention after Stroke (PI Walter N. Kernan, supported by grant U01NS044876 from the National Institute of Neurological Disorders and Stroke) data and analyses presented in this manuscript are based on the research files downloaded from the Archived Clinical Research website of the National Institute of Neurologic Disease and Stroke.

REFERENCES

1. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* 2nd ed. Berlin, Germany: Springer; 2019.
2. Radcliffe NJ. Using control groups to target on predicted lift: building and assessing uplift model. *Direct Mark Anal J.* 2007;1(3):14–21.

3. Ascarza E. Retention futility: targeting high-risk customers might be ineffective. *J Mark Res*. 2018;55(1):80–98.
4. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *Proc Natl Acad Sci*. 2016;113(27):7353–7360.
5. Rekkas A, Paulus JK, Raman G, et al. Predictive approaches to heterogeneous treatment effects: a scoping review. *BMC Med Res Methodol*. 2020;20(1):1–12.
6. Wang G, Heagerty PJ, Dahabreh IJ. Using effect scores to characterize heterogeneity of treatment effects. *JAMA*. 2024;331(14):1225–1226.
7. Kent DM, Steyerberg E, Van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245.
8. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;113(523):1228–1242.
9. Khalifa M, Magrabi F, Gallego B. Developing a framework for evidence-based grading and assessment of predictive tools for clinical decision support. *BMC Med Inform Decis Mak*. 2019;191:1–17.
10. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338(June):1432–1435.
11. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLOS Med*. 2013;10(2):e1001381.
12. Briggs ADM, Wolstenholme J, Blakely T, Scarborough P. Choosing an epidemiological model structure for the economic evaluation of non-communicable disease public health interventions. *Popul Health Metr*. 2016;14:17.
13. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338(june):1487–1490.
14. Schaafsma JD, van der Graaf Y, Rinkel GJE, Buskens E. Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. *J Clin Epidemiol*. 2009;62(12):1248–1252.
15. Wallace E, Smith SM, Perera-Salazar R, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). *BMC Med Inform Decis Mak*. 2011;11(1):62.
16. van Giessen A, Peters J, Wilcher B, et al. Systematic review of health economic impact evaluations of risk prediction models: stop developing, start evaluating. *Value Heal*. 2017;20(4):718–726.
17. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res*. 2018;21:1–11.
18. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. 2020;26(9):1351–1363.
19. Wallace E, Uijen MJM, Clyne B, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. *BMJ Open*. 2016;6(3):e009957.
20. Peirce CS. The numerical measure of the success predictions. *Science*. 1884;4(93):453–454.
21. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–574.
22. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med*. 1975;293(5):229–234.
23. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. *Med Decis Mak*. 1988;8(4):279–289.
24. Imai K, Li ML. Experimental evaluation of individualized treatment rules. *J Am Stat Assoc*. 2023;118(541):242–256.
25. Rößler J, Schoder D. Bridging the gap: a systematic benchmarking of uplift modeling and heterogeneous treatment effects methods. *J Interact Mark*. 2022;57(4):629–650.
26. Yadlowsky S, Fleming S, Shah N, Brunskill E, Wager S. Evaluating treatment prioritization rules via rank-weighted average treatment effects. *J Am Stat Assoc*. 2024;0(0):1–24.
27. Sverdrup E, Wu H, Athey S, Wager S. Qini curves for multi-armed treatment rules. https://arxiv.org/abs/2306.11979v3; Published online June 21, 2023. Accessed July 6, 2024.
28. Soares MO, Walker S, Palmer SJ, Sculpher MJ. Establishing the value of diagnostic and prognostic tests in health technology assessment. *Med Decis Mak*. 2018;38(4):495–508.
29. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd edt. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
30. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728–1733.
31. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. *Biostatistics*. 2005;6(2):227–239.
32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–138.
33. Palumbo P, Becker C, Bandinelli S, Chiari L. Simulating the effects of a clinical guidelines screening algorithm for fall risk in community dwelling older adults. *Aging Clin Exp Res*. 2019;31(8):1069–1076.
34. Panel on prevention of falls in older persons, American Geriatrics Society and British Geriatrics Society. Prevention of Falls in Older Persons: AGS/BGS Clinical Practice Guideline. http://geriatricscareonline.org/toc/updated-american-geriatrics-societybritish-geriatrics-society-clinical-practice-guideline-for-prevention-of-falls-in-older-persons-and-recommendations/CL014; Published 2011. Accessed December 31, 2024.
35. Ferrucci L, Bandinelli S, Benvenuti E, et al. Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J Am Geriatr Soc*. 2000;48(12):1618–1625.
36. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. 2012;12(9):CD007146.
37. Kernan WN, Viscoli CM, Furie KL, et al. Pioglitazone after ischemic stroke or transient ischemic attack. *N Engl J Med*. 2016;374(14):1321–1331.
38. Viscoli CM, Kent DM, Conwit R, et al. Scoring system to optimize pioglitazone therapy after stroke based on fracture risk. *Stroke*. 2019;50(1):95–100.
39. McLean K, Day L, Dalton A. Economic evaluation of a group-based exercise program for falls prevention among the older community-dwelling population. *BMC Geriatr*. 2015;15:33.
40. Wu S, Keeler EB, Rubenstein LZ, Maglione MA, Shekelle PG. A cost-effectiveness analysis of a proposed national falls prevention program. *Clin Geriatr Med*. 2010;26(4):751–766.
41. Eckermann S, Briggs A, Willan AR. Health technology assessment in the cost-disutility plane. *Med Decis Mak*. 2008;28(2):172–181.
42. White H. Theory-based impact evaluation: principles and practice. *J Dev Eff*. 2009;1(3):271–284.
43. Ferrante Di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686.
44. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015;350:h1258.
45. Graziadio S, Gregg E, Allen AJ, et al. Is the comparator in your diagnostic cost-effectiveness model "standard of care"? Recommendations from literature reviews and expert interviews on how to identify and operationalize It. *Value Heal*. 2024;27(5):585–597.