

Transportability of Prognostic Markers: Rethinking Common Practices through a Sufficient-Component-Cause Perspective

Mohsen Sadatsafavi, Gavin Pereira, Wenjia Chen

Abstract

Transportability, the ability to maintain performance across populations, is a desirable property of markers of clinical outcomes. However, empirical findings indicate that markers often exhibit varying performances across populations. For prognostic markers that are advertised as predictive risk equations, oftentimes a form of updating is required when the equation is transported to populations with different disease prevalences. Here, we revisit transportability of prognostic markers through the lens of the foundational framework of sufficient component causes (SCC). We argue that transporting a marker “as is” implicitly assumes predictive values are transportable, whereas conventional prevalence adjustment shifts the locus of transportability to accuracy metrics (sensitivity and specificity). Using a minimalist SCC framework that decomposes risk prediction into its causal constituents, we show that both approaches rely on strong assumptions about the stability of cause distributions. A SCC framework instead invites making transparent assumptions about how different causes vary across populations, leading to different transportation methods. For example, in the absence of any external information other than disease prevalence, a cause-neutral perspective can assume all causes are responsible for change in prevalence, leading to a new form of marker transportation. Numerical experiments demonstrate that different transportability assumptions lead to varying degrees of information loss, depending on the distribution of causes. A SCC perspective challenges common assumptions and practices for marker transportability, and proposes transportation algorithms that reflect our knowledge or assumptions about how causes vary across populations.

Keywords: Prognosis; Predictive Value of Tests; Sensitivity and Specificity; Biomarkers; Etiology; Epidemiologic Methods

From Faculty of Pharmaceutical Sciences and Faculty of Medicine, Vancouver, Canada (MS); School of Population Health, WHO Collaborating Centre for Climate Change and Health Impact Assessment, and enAble Institute, Faculty of Health Sciences, Curtin University, Perth, Australia (GP); Faculty of Medicine, Universitas Negeri Malang, Indonesia (GP); Saw Swee Hock School of Public Health, National University of Singapore, Singapore (WC)

* Correspondence to Mohsen Sadatsafavi, Room 4110, Faculty of Pharmaceutical Sciences, 2405 Wesbrook Mall, Vancouver, British Columbia, V6T1Z3, Canada; email: mohsen.sadatsafavi@ubc.ca

Background

The underlying premise of reporting on the performance of biomarkers, tests, and prediction models (which we generally refer to as “markers”) is that performance metrics are transportable from one population to another. However, in practice, this premise more often than not fails to hold. Research has repeatedly shown the performance of markers can vary significantly across populations, especially when the prevalence of the outcome of interest differs⁴. For example, risk prediction models for cardiovascular disease have shown substantial degradation in performance when transported to a new population, leading to substantial risk of harm (defined as the net benefit of using the model being lower than not using it)⁵. The likelihood of harm was reduced when models were updated to account for the difference in prevalences between their source population and the target population. In another example, the performance of machine learning models for detecting pneumonia on chest X-rays substantially declined in data from settings not used to train the model⁶.

When transporting a marker for an outcome to a new population, some information about the characteristics of that population - generally referred to as case-mix- is often available. One of the most common information pieces is the outcome prevalence. For example, cancer registries often provide a good estimate of cancer risk in a population. This poses the question as to how such information can be used to revise our assessment of marker performance. For risk predictions models that return a quantitative estimate of outcome risk given patient characteristics, some form of model revision is often needed to correct for the under-estimation or over-estimation of risks in the new population⁷. When the model is a logistic regression, the most basic form of such updating involves modifying the intercept of the model to account for difference in prevalence between the source and target populations^{1,8,9}. Because in a logistic regression model, changing the intercept is equivalent to applying an odds-ratio to predicted risks, this method can be generalized to applying a correcting odds-ratio to the outputs of any risk prediction algorithm, including black-box (e.g., machine learning) models¹⁰.

The transportability of predictive information across populations is an active area of research in predictive analytics and machine learning. A recent scoping review categorized methods aimed at developing transportable marker, or making an existing marker transportable to a new setting, based on whether they require access to data from the target population, and whether they are purely data-driven or require contextual knowledge about associations¹¹. One common underlying framework in knowledge-driven approaches is causal graphs, particularly the pioneering work on selection diagrams by Pearl et al.¹². Generally, these methods aim to identify and remove predictors whose association with the outcome varies across populations^{13,14}. Causal diagrams have recently been used to study how common metrics of model performance for prognostic and diagnostic markers change across populations with difference case-mix¹⁵.

Causal graphs are not the only model for causation. Another is the sufficient component causes (SCC) model^{16,17}. SCC is a foundational model in that it establishes fully deterministic relationships between causes and effects (rather than a representation of statistical dependencies as in causal graphs). This framework has recently been applied to explore the biologic plausibility of different link functions for modeling binary outcomes, resulting in proposals for more transportable measures of association¹⁸ and more biology-aligned statistical models¹⁹.

In this paper, we use a parsimonious SCC model to study the most basic prediction setup: a binary factor that is used as a marker for the risk of a binary outcome. As a reference, we formulate this setup for prognostic markers, where adjustment for outcome prevalence seems to be a topical issue. This setup is the used to study marker transportability across populations, particularly as it relates to the variability in outcome prevalence. Our thesis is that by reducing the transportability problem to its basic constituents in this model, patterns will emerge that can provide insight into more complex scenarios.

A parsimonious SCC framework for prognostic markers

The SCC framework assumes the existence of sets of sufficient causes that bring about an event²⁰. Within each set, the causes are non-redundant (all elements with the set are required for the event to happen), but sets can act independently of each other²².

Consider a binary prognostic marker, such as the presence or absence of BRCA1 mutation, and a binary outcome, such as the occurrence of breast cancer. While BRCA1 mutation is associated with increased cancer risk²³, it is not a definitive marker: neither a negative BRCA1 mutation eliminates the risk of breast cancer, nor does its presence guarantee that cancer will occur. The fact that neither positive nor negative marker values are definitive indicates that there are at least two other mechanisms at play. On one hand, a BRCA1-positive individual must experience some other, key events that ultimately lead to cancer, explaining why not everyone with a BRCA1 mutation will develop cancer. On the other hand, cancer can also occur via pathways independent of the BRCA1 mutation, explaining why some breast cancer cases are BRCA1-negative.

We now formalize a minimal causal setup. We consider a binary prognostic marker T for a binary outcome D . For this marker to be informative but not definitive, at least two latent variables (or switches) must be present that can cause false-negative and false-positive responses. We model these switches as follows:

- The latent binary variable U represents universally required causes - for example factors that cause pre-cancers (*in-situ* cancer) to progress to malignancy. The absence of U is responsible for false-positive marker values.
- The latent binary variable V represents all alternative causes - for example pathways related to the effect of tobacco smoking, which increases the risk of breast cancer even among those without a BRCA1 mutation. The presence of V is responsible for false-negative marker values.

We assign the values of 1 and 0, respectively, to the ‘on’ and ‘off’ status of each of these switches. For the brevity of notations, by a simple character, we refer to the ‘on’ value, and by its dot-accented to its negated value (e.g., $\dot{U} = \neg U$ indicating $U = 0$).

The above setup results in the following outcome-generating process for prognostic markers:

$$D = (T \wedge U) \vee (V \wedge \dot{U}) = (T \vee V) \wedge U,$$

where $\wedge$ and $\vee$ are logical AND and OR, respectively.

Given that T , U , and V are all binary, a population is made up of 8 subgroups. As U and V are latent variables, the observed properties of the marker in a population is manifested in terms of $P(T, D)$, i.e., the two-by-two (contingency) table of marker by outcome status probabilities. We represent the contingency table by the sequence TP, FP, FN, TN , where $TP = P(TD = 1)$, $FP = P(T\dot{D} = 1)$, $FN = P(\dot{T}D = 1)$, $TN = P(\dot{T}\dot{D} = 1)$ are, respectively, true positive, false positive, false negative, and true negative probabilities.

Our interest is in the transportability of four fundamental performance characteristics of binary markers: positive predictive value ($PPV = P(D = 1|T = 1) = TP/(TP + FP)$), negative predictive value ($NPV = P(T = 1|D = 0) = TN/(FN + TN)$), sensitivity ($SE = P(T = 1|D = 1) = TP/(TP + FN)$), and specificity ($SP = P(T = 0|D = 0) = TN/(FP + TN)$), as a function of prevalence ($P(D = 1) = TP + FN$). By convention PPV and NPV are referred to as predictive values, and SE and SP as accuracy metrics.

Before proceeding, we note that one can create an equally parsimonious SCC setup by swapping the logical AND and OR in the above equation, resulting in the setup $D = (T \wedge V) \vee U$. However, this is mathematically symmetrical to our reference setup, as the complementary marker (whose positive and negative results are swapped) in this setup is a marker for not experiencing the outcome: $\dot{D} = (\dot{T} \vee \dot{V}) \wedge \dot{U}$. Because of this, any pattern we observe for PPV in the main formulation is also observed for NPV in this alternative formulation, albeit in the opposite direction, and vice versa. A similar symmetry exists between SE and SP in these setups. Table 1 shows these minimum SCC setups and the resulting contingency tables ($P(T, D)$).

Mapping between contingency table and cause probabilities

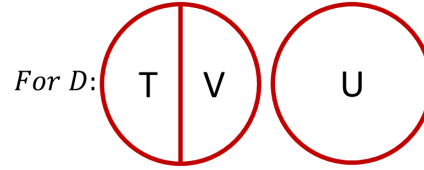
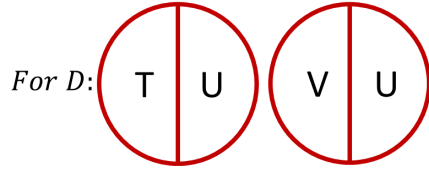
If all we know about marker performance in a populations is its contingency table, the full joint distribution of these causes among individuals in that population is unidentifiable, given its specification requires seven degrees of freedom, but the contingency table only offers three. However, any joint distribution can be replaced by independent marginal distributions for $P(T)$, $P(U)$, and $P(V)$ that would result in the same contingency table. Therefore, for studying contingency tables (and resulting marker performance metrics) for

Table 1: Minimal configurations for a prognostic for the reference (left) and its symmetrical (right) setup, and the resulting contingency tables. Top: logic equation; Middle: SCC diagram; Bottom: The resulting contingency table

Reference setup**Symmetrical setup**

$$D = (T \vee V) \wedge U$$

$$D = (T \wedge V) \vee U$$



	+	D	-
+	$TU\dot{V}$ TUV		$T\dot{U}\dot{V}$ $T\dot{U}V$
T			
-	$\dot{T}UV$		$\dot{T}\dot{U}\dot{V}$ $\dot{T}\dot{U}V$ $\dot{T}UV$

	+	D	-
+	$T\dot{U}\dot{V}$ $TU\dot{V}$ TUV		$T\dot{U}\dot{V}$
T			
-	$\dot{T}U\dot{V}$ $\dot{T}UV$		$\dot{T}\dot{U}\dot{V}$ $\dot{T}\dot{U}V$

Subgroups are defined as the combination of causes such that the product evaluates to 1. For example, $T\dot{U}\dot{V}$ is the subgroup where $T = 1$, $U = 0$, and $V = 1$.

a single population, it is sufficient to model the probability of these causes independently across individuals within that population. However, the prevalence of these causes across populations can be correlated, and it is this dependence structure that determines the transportability of performance metrics.

Within a population, given specifying three independent Bernoulli distributions for these probabilities requires three degrees of freedom, this setup is identifiable and has a unique (1:1) mapping to a given contingency table. We use the shorthand notations $P_T = P(T = 1)$, $P_U = P(U = 1)$, and $P_V = P(V = 1)$. The 1:1 mapping can be formulated as follows:

- From $\{TP, FP, FN, TN\}$ to $\{P_T, P_U, P_V\}$:

$$P_T = TP + FP, \quad P_U = TP/P_T, \quad P_V = FN/[(1 - P_T)P_U]$$

- From $\{P_T, P_U, P_V\}$ to $\{TP, FP, FN, TN\}$:

$$TP = P_T P_U, \quad FP = P_T(1 - P_U), \quad FN = (1 - P_T)P_U P_V, \quad TN = 1 - TP - FP - FN.$$

As a numerical example, consider a population where $P_T = 0.25$, $P_U = 0.75$, and $P_V = 0.50$. The 8 subgroups are scattered into the four cells of the contingency table (see Table 1), which is presented in the top row of Table 2. This results in a prevalence of 0.469, SE of 0.400, SP of 0.882, PPV of 0.750, and NPV of 0.625. We will use this simple numerical example as a case study on various ways a marker can be transported.

Table 2: Numerical example of different marker transportation methodology

Description	$\{P_T, P_U, P_V\}$	Contingency table	SE	SP	PPV	NPV
Source population	$\{0.250, 0.750, 0.500\}$	$\{0.188, 0.062, 0.281, 0.469\}$	0.400	0.882	0.750	0.625
Target population	$\{0.333, 0.800, 0.667\}$	$\{0.267, 0.067, 0.356, 0.311\}$	0.429	0.824	0.800	0.467
By predictive values	$\{0.333, 0.750, 0.500\}$	$\{0.250, 0.083, 0.250, 0.417\}$	0.500	0.833	0.750	0.625
By accuracy	$\{0.293, 0.848, 0.623\}$	$\{0.249, 0.044, 0.373, 0.333\}$	0.400	0.882	0.848	0.472
Proportional odds	$\{0.349, 0.829, 0.617\}$	$\{0.290, 0.060, 0.333, 0.318\}$	0.465	0.841	0.829	0.489

SE: sensitivity; SP: specificity; PPV: positive predictive value; NPV: negative predictive value

Contingency table presents the sequence TP (true positive), FP (false positive), FN (false negative), TN (true negative)

Two common methods of transportability

Ultimately, using a marker for decision-making entails interpreting its predictive values - i.e., $P(D|T)$. Different transportation methods construct these predictive values by combining different pieces of information from the source and target populations.

Transportation by predicted values

Transportation by predictive values refers to the where predictive values from the source population are considered to be the same as in the target population. Given that $P(T, D) = P(T)P(D|T)$, this transportability means the contingency table in the target population is determined by $P(T)$ in that population, combined with predictive values from the source population. This approach is more common for prognostic markers, a typical example being risk scoring tools that directly return an estimate of $P(D = 1|T)$ for someone with marker value x . For a binary marker, this risk equation can be written as $P(D = 1|T) = (1 - NPV) + [PPV - (1 - NPV)]T$. In our numerical example, advertising prognostic information as a risk equation would result in $P(D = 1|T) = 0.375 + 0.375T$.

Transportation by accuracy

In transportation by accuracy, we assume that $P(T|D)$, i.e., the SE and SP, are equal between the source and target populations. In order to construct predictive probabilities, we apply the Bayes' rule:

$$P(D = 1|T) = \frac{P(D = 1)P(T|D = 1)}{P(T)} = \frac{P(D = 1)P(T|D = 1)}{P(D = 1)P(T|D = 1) + P(D = 0)P(T|D = 0)}.$$

Thus, under this transportability scheme we need access to prevalence in the target population to derive predictive values. This approach is the *modus operandi* for binary diagnostic markers, where the Bayes' rule is used to combine pre-test probability with marker accuracy estimates to derive the post-test probability of the outcome^{24,25}. Without any information that would distinguish the individual under evaluation, the pre-test probability is taken to be disease prevalence in the target population, which is equal to this mode of transportation.

Returning to our running example, imagine we are transporting the above-mentioned prognostic marker to a population where $P_T = 0.250$, $P_U = 0.750$, $P_V = 0.500$. Outcome prevalence in this population is 0.622 - 32.7% higher than in the source population. Transportation by predictive values preserves PPV and NPV in the target population, but will result in SE and SP of 0.5 and 0.833. Now, imagine we know the target prevalence. If we consider transportation by accuracy, we will arrive at the PPV and NPV values of, respectively, 0.750 and 0.625. The contingency table and marker performance metrics for the true population and those implied by the above-mentioned transportability methods are provided in Table 2.

Prevalence-adjustment for risk equations

A common way that prognostic markers are transported is in the format of a risk equation for $P(D|T)$, either explicitly, as is the case in regression-based prediction models, or implicitly, as in black-box (e.g., machine learning) models. A familiar modeling framework for binary outcomes is the logistic regression. There, the logit function ($\text{logit}(x) = \log(\frac{x}{1-x})$) is used as link function connecting marker value to outcome probability. Applying this function to both sides of the previous equation, we have

$$\text{logit}(P(D = 1|T)) = \text{logit}(P(D = 1)) + \log\left(\frac{P(T|D = 1)}{P(T|D = 0)}\right).$$

The last term on the right-hand side is the likelihood ratio (LR) of the marker between the diseased and non-diseased groups. The first term on the right-hand side is the logit of prevalence, which is not a function of marker value. These derivations indicate that the practice of prevalence-adjustment by the odds-ratio of prevalence between the source and target populations is equivalent to the application of Bayes' theorem. This approach changes the locus of transportability from predictive values to the LR. For binary markers, the LR is defined at two values. For $T = 0$ it is $(1 - SE)/SP$ (aka negative LR), and for $T = 1$ it is $SE/(1 - SP)$ (aka positive LR). For the marker to be transportable, both these LRs need to remain constant across populations, which will be the case if and only if both SE and SP remain constant.

Under what conditions are performance metrics transportable?

The above reasoning shows why adjustment for prevalence is generally expected to improve transportability, as the conventional wisdom is that SE and SP, being defined within the diseased and non-diseased groups, are less dependent on prevalence²⁶. While this might be intuitive as a general observation, under the SCC framework, none of these metrics are truly intrinsic. Rather, they emerge as properties of the distribution of causes, and their transportability depends on how this distribution varies across populations. One can create population-generating mechanisms where a given subset of these metrics remains constant while others vary. However, our parsimonious framework helps us examine the plausibility of such mechanisms.

Table 3 shows conditions for the distribution of causes across populations under which each of the four metrics remains constant (and therefore transportable) for a prognostic marker.

For the PPV to be transportable, the prevalence of universal causes should remain constant across populations. For NPV, transportability requires that the proportion in whom both the universal and alternative causes are present should remain constant. These conditions are both satisfied if P_U and P_V are stable across populations. In this scenario, the entirety of variation in outcome prevalence is attributable to variation in

Table 3: Conditions for Transportability of Metrics of Marker Performance (c indicates a constant value)

Metric	Transportability condition
PPV	$P_U = c$
NPV	$P_U P_V = c$
SE	$P_V(1 - P_T)/P_T = c$
SP	$P_T(1 - P_U)/[(1 - P_T)(1 - P_U P_V)] = c$

SE: sensitivity; SP: specificity; PPV: positive predictive value; NPV: negative predictive value

As an example of derivations, consider PPV. Its transportability means $P(D = 1|T = 1) = c$ (a constant). But $P(D = 1|T = 1) = (P(TUV = 1) + P(TUV = 1))/P(T = 1) = P(TU = 1)/P(T = 1) = P(U = 1|T = 1) = P(U = 1)$ (the last equality is based on our assumption of independence of cause distributions within a population).

T . In our breast cancer example, this would mean populations vary in breast cancer prevalence only because they differ in the BRCA1 mutation rate. For many outcomes, this assumption is unrealistic (e.g., many factors contribute to the variability in breast cancer risk²⁷). In comparison, conditions for the transportation by accuracy are more complicated. For SE, it requires the ratio of the probability of one cause (V) over the odds of another (T) to remain constant. For SP, it does not seem possible to specify a simple model on how causes should vary to ensure transportability. Overall, transportation by accuracy metrics for prognostic markers amounts to placing a very specific set of conditions on how causes vary across populations.

How do performance metrics vary by prevalence under different population-generating mechanisms?

In our simple causal framework, the degree of transportability of a marker by a given performance metric depends on how the causes vary across populations. Despite the simple setup, complex patterns arise due to the non-linearities in the interplay among causes. In this section we visualize how the relationship between performance metrics and outcome prevalence changes under various population-generating mechanisms. Taking the population of our numerical example as the baseline, we modeled the following scenarios: when populations vary only in one of causes; when they vary in two of the three causes; and when they vary in all three causes. When multiple causes vary, we assume they change by the same degree on the odds-ratio scale.

Results are provided in Figure 1. As explained above, under T -only variation, PPV and NPV remain transportable. Importantly, in all other conditions, PPV and NPV varied by prevalence. On the other hand, in none of the modeled scenarios did SE and SP remain unchanged. In fact, SE and SP could vary in either direction as a function of prevalence, reflecting the more complex conditions required for their transportability.

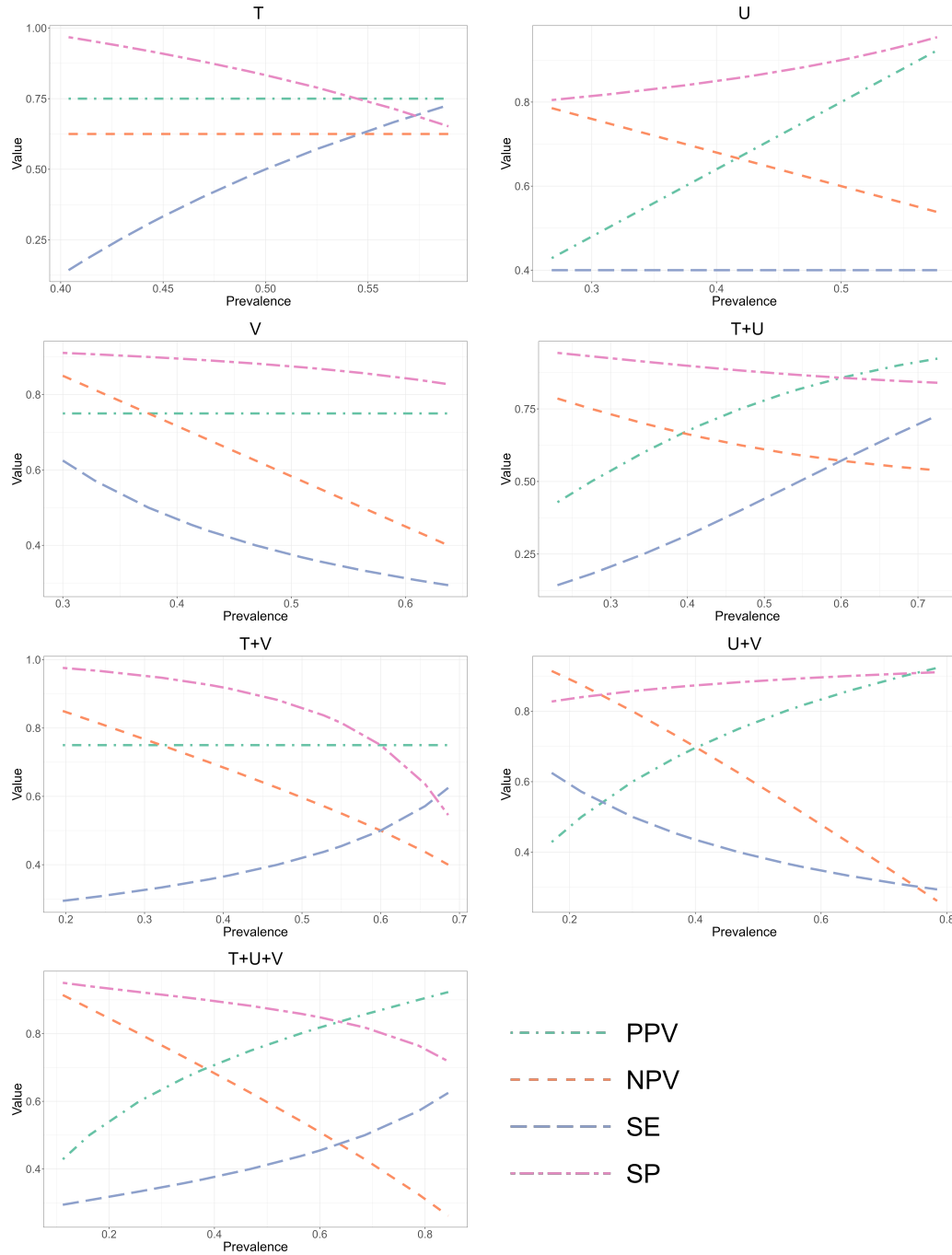
Other forms of marker transportation under the SCC framework

This framework enables us to express our beliefs on how causal pathways vary across populations, resulting in algorithms that are transparent and explicit in their underlying assumptions. Consider investigating a rare ancestral mutation as a prognostic marker for cancer that is known to have minimal variations across populations. In this case, it is reasonable to assume that P_T remains essentially constant across populations. When transporting this marker, our focus would therefore be on scenarios where differences in $P(U, V)$ derive variations in cancer prevalence. The resulting transportation method will be different from transportation by predictive values or accuracy metrics. Indeed, both these transportation methods implicitly require that an increase in outcome prevalence be accompanied by change in marker positivity.

Proportional-odds transportability method

Methods based on causal graphs provide specific solutions for marker transportability if causal associations between specific factors that determine the marker and outcome values are known. But, what if the only

Figure 1: Relationship between prevalence and marker performance metrics under various population-generating mechanisms for a prognostic marker



SE: sensitivity; SP: specificity; PPV: positive predictive value; NPV: negative predictive value

The base setup is $P_T = 0.25$, $P_U = 0.75$, $P_V = 0.5$. Changes across populations are modeled by applying varying odds-ratios to the components that vary (labeled on top of each panel).

information we have is differences in outcome prevalence across populations? Our parsimonious framework enables an overall assessment of the plausibility of the assumptions under various transportability methods even under such a general case. For transportability by predictive values, one can question as to why only one cause (T , the one we happen to be measuring) should be responsible for change in prevalence. Transportability by accuracy is also questionable due to unclear assumptions it places on the distribution of causes. For this general case, it might be more reasonable to take a neutral yet transparent stance about the variations in causes, for example that all causes move by the same extent to cause change in prevalence. As an implementation of this approach, instead of applying a correcting odds-ratio to predicted values, one can solve for a common odds-ratio that, when applied to P_T , P_U , and P_V in the source population, results in a new population that matches the outcome prevalence in the target population. This “proportional odds” assumption is a new, distinct transportability approach that would generate different updated values of marker performance metrics compared with both conventional methods. Details of solving for this common odds-ratio are provided in the Appendix.

In our numerical example (last row of Table 2), we need to apply an odds-ratio of 1.612 to the three causes to match the target population’s outcome prevalence. The contingency table under this method of transportability will be $\{0.290, 0.060, 0.333, 0.318\}$. This in turn results in SE of 0.465 and SP of 0.841. The PPV and NPV are, respectively, 0.829 and 0.489.

Information loss under different transportability assumptions

Ultimately, any assumption about how causes vary across populations will be a simplified version of a complex reality. The discrepancy between marker performance under a given transportability assumption and the true marker performance results in loss of prognostic information. We conducted brief simulation studies to explore such information loss under simple population-generating scenarios. The discrepancy between an assumed versus the true marker performance can be measured in different ways, each focusing on certain aspects of performance (e.g., discrimination, calibration, or prediction error). A more foundational approach is to measure information loss by the Kullback-Leibler divergence (D_{KL}), an information-theoretic measure of the discrepancy between a true distribution and a candidate distribution²⁸. D_{KL} quantifies the additional number of bits required to encode information from the true distribution using the candidate distribution, rather than using the true distribution itself²⁹. In our case, these distributions are the true contingency table ($P(T, D)$) in the target population versus the one implied by a given transportability algorithm.

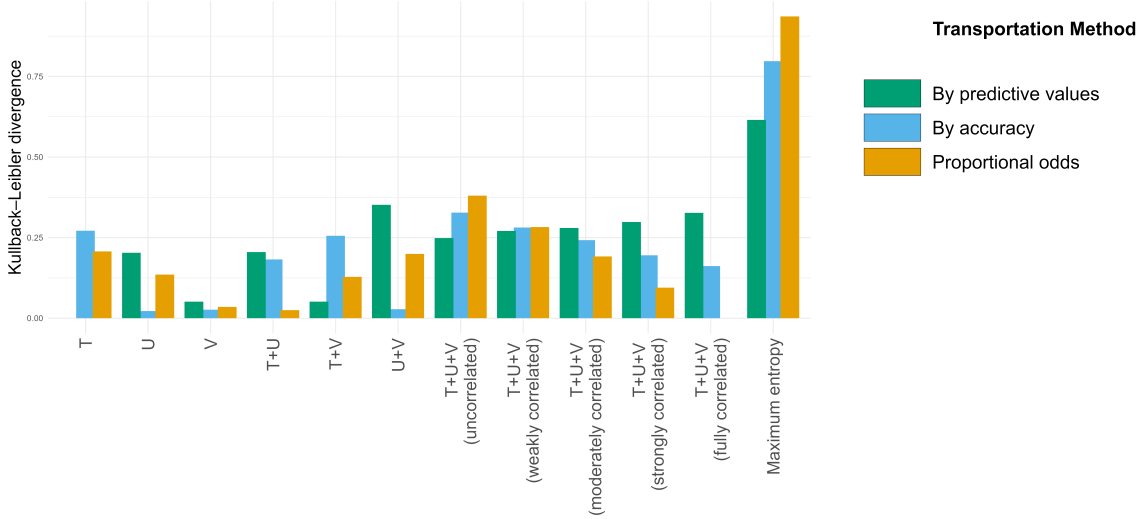
We modeled population-generating scenarios similar to those presented in Figure 1. Because in many realistic settings all three causes are likely to vary, we explored this setup with more depth, covering scenarios where causes had positive correlation of varying strengths. We also modeled a ‘maximum entropy’ scenario where the three causes have independent uniform distributions; albeit unrealistic, this scenario represents maximum theoretical randomness, and thus places an upper limit on information loss³⁰. For each scenario, we simulated 10,000 random pairs of source and target populations. For each pair, we applied the following transportation strategies: by accuracy, by predictive values, and via the proportional-odds assumption explained previously.

Results are provided in Figure 2. Overall, they show that no transportability method is universally better than others. They confirm that if the outcome prevalence variation is entirely due to T , PPV and NPV are fully transportable ($D_{KL} = 0$). However, under other scenarios, transportation by predictive values resulted in information loss, sometimes substantially. For transportability by accuracy, in none of the scenarios was the loss zero, indicating that none of the modeled scenarios were compatible with stable SE and SP - again, due to the complex requirements for the stability of these metrics. Notably, for the scenarios where the three causes changed together, the new, proportional-odds assumption performed better than both conventional transportability methods.

Discussion

We constructed a parsimonious causal framework to study the transportability of prognostic markers. Several observations from this theoretical exploration deserve reiterating. First, the common practice of transporting prognostic markers by predictive values relies on the strong assumption that the prevalences of universal and alternative causes remain stable across populations. We showed that the conventional prevalence-adjustment

Figure 2: Information loss associated with different methods for transporting a marker under various population-generating scenarios



Kullback-Leibler divergence (D_{KL}) = $\sum_{i=1}^4 P(a_i) \log_2(P(b_i)/P(a_i))$ with a_1, a_2, a_3, a_4 being the four cells of the contingency table from the correct $P(T, D)$, and b_i s the corresponding ones from $P(T, D)$ constructed given the transportability method. Aside in the maximum entropy scenario (whose description provided in text), the variable component in each scenario had a standard logit-normal distribution e.g., $\text{logit}(T) \propto \text{Normal}(0, 1)$ for scenarios where T was variable.

No correlation, weak, moderate, strong, and full correlation correspond to correlation coefficients of, respectively, 0, 0.25, 0.50, 0.75, and 1 for logit-transformed probabilities.

method is the exact implementation of Bayes' rule, thus changing the locus of transportability to accuracy (sensitivity and specificity). Nevertheless, no easily explainable or biologically plausible mechanism is likely to generate fully transportable accuracy metrics. In our explored population-generating scenarios, transporting a prognostic marker by accuracy was not universally better than transporting by predictive values. However, in scenarios where causes were positively correlated, which is likely to be common, transporting by accuracy reduced information loss compared with transportability by predictive values. However, in the same scenarios, the new 'proportional odds' method of transportation performed better.

Our numerical results were related to a selected ways the causes vary across populations. Still, they should be sufficient to question some common assumptions and practices, including the practice of advertising prognostic risk questions without emphasizing their dependence on case-mix and outcome prevalence, and considering sensitivity and specificity as intrinsic properties of markers. While these insights are from studying binary markers, the core findings can be extended to continuous markers and multi-variable risk equations. Transforming such markers 'as is' is equal to assuming that stratum-specific predictive values ($P(D|T)$) remain constant. This is equal to attributing variations in prevalence entirely to variations in P_T . Consider a multi-variable risk score such as the QRISK3 for cardiovascular diseases³¹. Even though this model includes up to 22 predictors, is it plausible that other causes of cardiovascular diseases, including environmental exposures, lifestyle choices, access to and quality of preventive care, and genetic risk factors, are the same across populations? This assumption is needed for claiming that QRISK3 predictions, developed using primary care UK data, are transportable to other settings. On the other hand, conventional prevalence-adjustment would indicate that stratum-specific likelihood ratios would be transportable (i.e., $P(T = x|D = 1)/P(T = x|D = 0) = c$ for all x). This assumption will hold if the distributions of predicted risks within outcome status strata remain the same. This condition places very specific constraints on how the causes of diseases should vary across populations.

We focused on prognostic markers given the recent debates on transportability of prognostic information and the merit of prevalence-adjustment to improve transportability. The application of this framework for diagnostic markers is also important and deserves its own airing. For such markers, the path of causality is from the disease to the test. As such, an equivalent setup to our reference setup for a diagnostic marker

would be $T = (D \vee V) \wedge U$. From this setup, it is immediately obvious that transportation by accuracy (the default approach for diagnostic markers) requires that universal and alternative causes (U and V) remain stable across populations. This strong requirement can explain why sensitivity and specificity tend to vary across populations²⁻⁴. How the SCC framework can inform transportability of diagnostic markers needs to be pursued in future work.

How can these findings inform practice? we question the contemporary practice of advertising risk prediction models for outcome risk as transportable. If reliable, context-specific information is available on the distribution of predictors and their relationships with the outcome and with each other, methods based on causal graphs can be used to build transportable models or to design tailored transportation strategies¹³. In contrast, when only general information, such as outcome prevalence or test positivity rates, are available, the SCC framework can be used to formulate transportability algorithms that utilize such broad information. Instead of fixed transportability rules, this framework offers a foundation for algorithms that reflect our assumptions on how underlying causes vary, rather than keeping performance metrics, which are emergent properties of such causes, constant.

We conclude this paper by providing a few areas for further inquiry. Whether SCC-informed transportation algorithms such as the proportional-odds method actually improve transportability can be tested in empirical studies. Our arguments were based on precise knowledge of outcome prevalence in the target population. To what extent using a noisy estimate of prevalence will help or harm transportability needs to be investigated. The issue of uncertain prevalence estimate is also applicable to conventional method. Empirical studies on conventional prevalence adjustment also estimated the prevalence from the sample that was subsequently used to assess marker performance - effectively taking its value as known^{5,32}. Further, our explorations were for when performance metrics are derived from a single source population. Without knowing how populations differ from each other, the choice of transportation method will require assumptions on the distribution of causes. On the other hand, when contingency tables from multiple populations are at hand, this choice can be learned from the data. This results in SCC-based model specifications for meta-analysis of marker performance studies, where the estimands are the parameters that govern the joint distribution of T , U , and V . This will provide an alternative to modeling the joint distribution of prevalence, sensitivity, and specificity (or prevalence and predictive values)^{33,34}. The more explanatory nature of SCC-based specification might provide better fit to the data. In addition to pooled estimates of performance metrics and their predictive distribution for a new population, this approach provides an overall estimate of the degree by which universal and alternative causes are responsible for between-population variations, which might be of secondary interest.

Appendix: Transportation by the proportional-odds assumption

Let $\{TP, FP, FN, TN\}$ be the elements of the two-by-two contingency table. We would like to transport this marker to a new setting where prevalence is π^* . Our goal is to construct a predicted contingency table for the target population, defined by $\{TP^*, FP^*, FN^*, TN^*\}$.

Step 1. Map from $\{TP, FP, FN, TN\}$ to $\{P_T, P_U, P_V\}$ (see text)

Step 2. Find the odds-ratio x such that when applied to P_T, P_U, P_V , results in updated probabilities P_T^*, P_U^*, P_V^* that correspond to the desired prevalence. Note that prevalence is equal to $P_U^*(P_T^* + P_V^* - P_T^*P_V^*)$.

Given that applying an odds-ratio x to probability p can be written as as the function $\frac{px}{1-p+px}$, we have

$$P_T^*(x) = \frac{P_T^*x}{1-P_T^*+P_T^*x} \text{ (similar for } P_U \text{ and } P_V).$$

This, we should solve for x in $f(x) = \pi^*$ where

$$f(x) = P_U^*(x)[P_T^*(x) + P_V^*(x) - P_T^*(x)P_V^*(x)].$$

We consider non-degenerate scenarios where $0 < P_T < 1$, $0 < P_U < 1$, and $0 < P_V < 1$. Given that $f(x)$ is continuous, $f(0) = 0$, and $\lim_{x \rightarrow \infty} f(x) = 1$, there is always at least one real solution for x . Further, given that

- $\frac{dP_T^*(x)}{dx} > 0$, $\frac{dP_U^*(x)}{dx} > 0$, and $\frac{dP_V^*(x)}{dx} > 0$, and
- $\frac{\partial f(x)}{\partial P_T^*(x)} > 0$, $\frac{\partial f(x)}{\partial P_U^*(x)} > 0$, $\frac{\partial f(x)}{\partial P_V^*(x)} > 0$,

$\frac{df(x)}{dx} > 0$ for all x . As such, this solution is unique. This solution can be found using univariate root finding methods (e.g., *uniroot()* in R).

Of note, re-arranging the terms reveals that x can also be expressed as the real root of the cubic equation $ax^3 + bx^2 + cx + d = 0$ with

$$\begin{aligned} a &= (1 - \pi^*)P_TP_UP_V, \\ b &= (3\pi^* - 2)P_TP_UP_V - \pi^*[P_TP_U + P_TP_V + P_UP_V] + P_TP_U + P_UP_V, \\ c &= -\pi^*[3P_TP_UP_V - 2P_TP_U - 2P_TP_V - 2P_UP_V + P_T + P_U + P_V], \\ d &= -\pi^*(1 - P_T)(1 - P_U)(1 - P_V); \end{aligned}$$

which can be solved using standard methods (e.g., *polyroot()* in R).

3. Map the resulting $\{P_T^*, P_U^*, P_V^*\}$ to $\{TP^*, FP^*, FN^*, TN^*\}$ (see text)

References

1. Morise AP, Diamond GA, Detrano R, et al. The effect of disease-prevalence adjustments on the accuracy of a logistic prediction model. *Medical Decision Making* 1996; 16: 133–142.
2. Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: Implications for evidence-based diagnosis. *Journal of Clinical Epidemiology* 2009; 62: 5–12.
3. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Statistics in Medicine* 1997; 16: 981–991.
4. Murad MH, Lin L, Chu H, et al. The association of sensitivity and specificity with disease prevalence: Analysis of 6909 studies of diagnostic test accuracy. *Canadian Medical Association Journal* 2023; 195: E925–E931.
5. Gulati G, Upshaw J, Wessler BS, et al. Generalizability of Cardiovascular Disease Clinical Prediction Models: 158 Independent External Validations of 104 Unique Models. *Circulation Cardiovascular Quality and Outcomes* 2022; 15: e008487.
6. Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine* 2018; 15: e1002683.
7. Steyerberg EW. Updating for a new setting. In: Steyerberg EW (ed). Cham: Springer International Publishing, pp. 399–429.
8. Janssen KJM, Moons KGM, Kalkman CJ, et al. Updating methods improved the performance of a clinical prediction model in new patients. *Journal of Clinical Epidemiology* 2008; 61: 76–86.
9. Sadatsafavi M, Tavakoli H, Safari A. Marginal Versus Conditional Odds Ratios When Updating Risk Prediction Models. *Epidemiology* 2022; 33: 555–558.
10. Meijerink LM, Dunias ZS, Leeuwenberg AM, et al. Updating methods for artificial intelligence-based clinical prediction models: A scoping review. *Journal of Clinical Epidemiology* 2025; 178: 111636.
11. Ploddi K, Sperrin M, Martin GP, et al. Scoping review of methodology for aiding generalisability and transportability of clinical prediction models. Epub ahead of print 2024. DOI: 10.48550/ARXIV.2412.04275.
12. Pearl J, Bareinboim E. Transportability of causal and statistical relations: A formal approach. *2011 IEEE 11th International Conference on Data Mining Workshops* 2011; 540–547.
13. Subbaswamy A, Schulam P, Saria S. Preventing failures due to dataset shift: Learning predictive models that transport. Epub ahead of print 2018. DOI: 10.48550/ARXIV.1812.04597.
14. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics*. Epub ahead of print 19 November 2019. DOI: 10.1093/biostatistics/kxz041.
15. Amsterdam WAC van. A causal viewpoint on prediction model performance under changes in case-mix: Discrimination and calibration respond differently for prognosis and diagnosis predictions. Epub ahead of print 2024. DOI: 10.48550/ARXIV.2409.01444.
16. Rothman KJ. Causes. *American Journal of Epidemiology* 1976; 104: 587–592.
17. Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. *American Journal of Public Health* 2005; 95: S144–S150.
18. Van Der Laan MJ, Hubbard A, Jewell NP. Estimation of Treatment Effects in Randomized Trials With Non-Compliance and a Dichotomous Outcome. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 2007; 69: 463–482.
19. Daniel RM, Farewell DM, Huitfeldt A. ‘Does God toss logistic coins?’ and other questions that motivate regression by composition. *Journal of the Royal Statistical Society Series A: Statistics in Society* 2024; 187: 636–655.
20. Flanders WD. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *European Journal of Epidemiology* 2006; 21: 847–853.

21. Kezios KL, Hayes-Larson E. Sufficient component cause simulations: An underutilized epidemiologic teaching tool. *Frontiers in Epidemiology*; 3. Epub ahead of print 10 November 2023. DOI: 10.3389/fepid.2023.1282809.
22. Suzuki E, Yamamoto E, Tsuda T. Identification of operating mediation and mechanism in the sufficient-component cause framework. *European Journal of Epidemiology* 2011; 26: 347–357.
23. Paul A. The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. *Frontiers in Bioscience* 2014; 19: 605.
24. Bours MJL. Bayes’ rule in diagnosis. *Journal of Clinical Epidemiology* 2021; 131: 158–160.
25. Johnson KM. Erratum to: Using bayes’ rule in diagnostic testing: A graphical explanation. *Diagnosis* 2018; 5: 89–89.
26. Altman DG, Bland JM. Statistics notes: Diagnostic tests 2: Predictive values. *BMJ* 1994; 309: 102–102.
27. Hortobagyi GN, Garza Salazar J de la, Pritchard K, et al. The global breast cancer burden: Variations in epidemiology and survival. *Clinical Breast Cancer* 2005; 6: 391–401.
28. Lee WC. Selecting diagnostic tests for ruling out or ruling in disease: The use of the kullback-leibler distance. *International Journal of Epidemiology* 1999; 28: 521–525.
29. Joyce JM. Kullback-leibler divergence. In: *International encyclopedia of statistical science*. Springer Berlin Heidelberg, pp. 720–722.
30. Thomas M. Cover, Thomas JA. *Elements of information theory*. 2nd ed. Nashville, TN: John Wiley & Sons, 2006.
31. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ* 2017; j2099.
32. Ho JK, Safari A, Adibi A, et al. Generalizability of risk stratification algorithms for exacerbations in COPD. *Chest* 2023; 163: 790–798.
33. Chu H, Nie L, Cole SR, et al. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection. *Statistics in Medicine* 2009; 28: 2384–2399.
34. Chu H, Guo H, Zhou Y. Bivariate random effects meta-analysis of diagnostic studies using generalized linear mixed models. *Medical Decision Making* 2009; 30: 499–508.