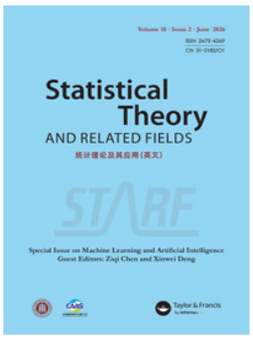

Taylor & Francis
Taylor & Francis Group

Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

On Kullback-Leibler divergence for medical diagnostics accuracy and cut-point selection criterion under tree or umbrella ordering

Hani M. Samawi, Marwan Alsharman & Jing Kersey

To cite this article: Hani M. Samawi, Marwan Alsharman & Jing Kersey (2026) On Kullback-Leibler divergence for medical diagnostics accuracy and cut-point selection criterion under tree or umbrella ordering, Statistical Theory and Related Fields, 10:2, 285-307, DOI: 10.1080/24754269.2026.2665851

To link to this article: https://doi.org/10.1080/24754269.2026.2665851

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 08 May 2026.

Submit your article to this journal

Article views: 94

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tstf20

STATISTICAL THEORY AND RELATED FIELDS
2026, VOL. 10, NO. 2, 285–307
https://doi.org/10.1080/24754269.2026.2665851

Taylor & Francis
by informa

OPEN ACCESS Check for updates

On Kullback-Leibler divergence for medical diagnostics accuracy and cut-point selection criterion under tree or umbrella ordering

Hani M. Samawi, Marwan Alsharman and Jing Kersey

Department of Biostatistics, Epidemiology, and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA, USA

ABSTRACT

Diagnostic testing typically involves two types of classification. Binary tests separate individuals into diseased or non-diseased groups, while multi-class methods, like tree or umbrella ordering, compare one class's biomarker levels to those of other classes. Kullback-Leibler divergence (KL), which measures the difference between two distributions, has been considered a valuable index for assessing the diagnostic performance of biomarkers. In this work, we derive and propose the total rule-in and rule-out Kullback-Leibler divergence (TTKL(c)) as a measure of accuracy, obtained by dichotomizing a continuous biomarker, and as an optimization criterion for cut-off point selection under tree or umbrella ordering. We have established a connection between the proposed TTKL(c) measure and the extended Youden index, which is the most used criterion for cut-off point selection. Additionally, we present both theoretical and numerical derivations for scenarios involving a single cut-off point under extended tree ordering. Graphically, KL divergence is represented through the information graph. Using simulation methods, we conducted a power study to compare the performance of our proposed methods with the extended Youden index under tree ordering, as well as the accuracy of optimal cut-off selection. This analysis provides insights into the effectiveness of TTKL(c) as a robust criterion for selecting cut-off points in multi-class diagnostic settings. A comprehensive data analysis of lung cancer data illustrates the proposed applications.

ARTICLE HISTORY
Received 23 July 2025
Revised 10 March 2026
Accepted 23 April 2026

KEYWORDS
ROC; cut-off point selection; diagnostics; Kullback-Leibler divergence AUC; Youden index; extended tree or umbrella ordering

1. Introduction

Medical diagnostic tests are typically binary, distinguishing between subjects with and without disease. For a continuous biomarker, a diagnostic cut-off point is necessary to classify a subject as diseased or non-diseased, aiding clinical decision-making. This cut-off point enables the biomarker to categorize patients into disease/positive or non-disease/negative groups. Sensitivity is the proportion of true positives, and specificity is the proportion of true negatives, given the disease status. Inferential procedures for sensitivity and specificity have been extensively researched across various study designs (Pepe, 2003).

CONTACT Hani M. Samawi hsamawi@georgiasouthern.edu Department of Biostatistics, Epidemiology, and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA 30460, USA

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

However, it is widely acknowledged that sensitivity and specificity may not fully capture a diagnostic test's clinical utility. Clinical utility refers to the practical value or usefulness of a medical test, procedure, intervention, or treatment in real-world clinical settings. It assesses whether a particular medical approach or intervention provides meaningful benefits in terms of diagnosis, treatment decisions, patient outcomes, or overall healthcare management (Knottnerus & Muris, 2003; Sackett et al., 1991).

In addition to binary and ordinal classifications, for multi-state diseases, certain diseases can be further divided into subclasses or subtypes that lack a clear stochastic (monotone) ordering between their measurements. Unlike the monotone ordinal ordering, biomarker measurements within one subclass may differ from those of other subtypes without exhibiting any discernible order. This approach allows for a more detailed and accurate representation of the disease states, as exemplified in the motivation example in Section 2. This is referred to as a tree or umbrella ordering. Specifically, let $Y_0, Y_1, Y_2, \ldots, Y_k$ denote biomarker values for non-diseased and subtypes of the disease subjects and $(Y_1, Y_2, \ldots, Y_k)$ means that there is no clearly defined order among disease classes (subtypes) $Y_1, Y_2, \ldots, Y_k$; then the tree or umbrella ordering is analytically defined as $Y_0 \prec (Y_1, Y_2, \ldots, Y_k)$ or $Y_0 \succ (Y_1, Y_2, \ldots, Y_k)$, respectively. For instance, lung cancer is classified into small-cell lung carcinomas (SCLC) and non-small-cell lung carcinomas (NSCLC), which are further divided into subtypes such as adenocarcinomas (AD), squamous cell carcinomas (SQ), pulmonary carcinoid (COID), or large-cell carcinomas. In this scenario, a straightforward tree-ordering pattern emerges within the disease subtypes, indicating distinct characteristics for each subtype.

In both tree and umbrella ordering, the main goal is to differentiate a specific class from the rest, which may comprise multiple sub-classes in any order. Traditionally, to assess a classifier's discriminatory ability under tree or umbrella ordering, it has been common to convert the problem into a two-class setting. This is done by comparing one class to a hypothetical combined class formed by merging all the other sub-classes without a specific order. As a result, the standard two-class ROC curve can be applied, and the resulting area under the curve (AUC) is referred to as the naive AUC (see Wang et al., 2016). Additionally, Wang et al. (2019) have introduced optimal cut-off point selection methods specifically tailored for tree or umbrella ordering. Recently, Samawi et al. (2023), extended the positive and negative predictive values (PPV and NPV) under tree or umbrella ordering. However, using these methods to determine the cut-off point may yield lower specificity and higher sensitivity for each subclass than using each subclass independently for cut-off determination. Using the Youden index under tree ordering presents a specific challenge. One concern is the possibility of obtaining negative values, which are neither logical nor desirable when determining a cut-off point. A negative Youden index implies that the biomarker's discriminatory ability to distinguish between disease subtypes is weak or even counterproductive. Consequently, using a cut-off point derived from a negative Youden index may yield an invalid or unreliable threshold.

The Kullback-Leibler (KL) divergence, also known as the information measure or relative entropy, quantifies the difference between two probability distributions. Lee (1999) explored the use of KL divergence as a metric for assessing the diagnostic performance of biomarkers. Specifically, Lee (1999) applied the KL divergence to discrete biomarkers, describing it as 'an abstract concept rooted in statistics and information theory'. However, Hughes (2013) later developed a diagrammatic representation of the KL divergence based on Lee (1999)'s work, referred to as an 'information graph', as noted by Benish (2002). Information graphs provide

a visual method for evaluating and comparing binary diagnostic tests, potentially increasing the practical appeal of KL divergence for clinicians.

In this paper, we extend the KL divergence for binary disease classification to extended tree ordering disease classification (TTKL). We show that TTKL for a continuous biomarker contains information about the sensitivity and the specificity of the diagnostic test and a reminder (loss of information due to dichotomizing the continuous marker at a specific threshold value, c). We propose using the total KL discrete version under extended tree ordering (TTKL), after dichotomizing a continuous marker, as an optimization criterion for cut-off point selection (optimizing the total before-test rule-in and rule-out) and linking the TTKL with some of the common receiver operating characteristic (ROC) measures under the extended tree or umbrella ordering. We investigate various applications of the TTKL divergence in medical diagnostics and demonstrate how it is illustrated through an information graph. We will present the relations theoretically and numerically in situations with one cut-off point (two categories).

We outline the paper's structure as follows. Section 2 provides the necessary background and justification for the proposed measures. In Section 3, we present the methodology for deriving the TTKL divergence under extended tree ordering. Section 4 is dedicated to the diagrammatic interpretation of TTKL divergence for continuous markers at a given cut-off point (c) concerning Bregman divergence. Section 5 presents and discusses the results of a simulation study that assesses the performance of the proposed measures. Section 6 presents an application of measures in a practical context, using lung cancer data as an example. Finally, Section 7 provides a comprehensive discussion of the study, including its limitations and implications.

2. Preliminaries

2.1. Motivating examples

Lung cancer provides a clear example of tree ordering in disease classification. The disease is categorized into two main types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). Within the NSCLC category, there are several subtypes, including adenocarcinoma (ADC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC), though no specific hierarchy exists among them. To differentiate NSCLC from healthy individuals or those with SCLC, a tree- or umbrella-ordering framework is required to provide a structured approach to classification (see Figure 1).

Another instance of disease classification with tree ordering can be observed in Colorectal Cancer (CRC), as discussed by Guinney et al. (2015). Through their research, they identified four consensus molecular subtypes (CMSs) with distinct characteristics, revealing significant interconnections among six independent classification systems. The first subtype, CMS1 (microsatellite instability immune), accounted for 14% of cases and exhibited hypermutation, microsatellite instability, and robust immune activation. The second subtype, CMS2 (canonical), accounted for 37% of cases and showed pronounced activation of the WNT and MYC signaling pathways, suggesting epithelial characteristics. CMS3 (metabolic), encompassing 13% of cases, displayed metabolic dysregulation alongside epithelial features. Lastly, CMS4 (mesenchymal) comprised 23% of cases and displayed prominent activation of the transforming growth factor-β pathway, stromal invasion, and angiogenesis. Within the disease, no stochastic ordering of subtype values was evident.

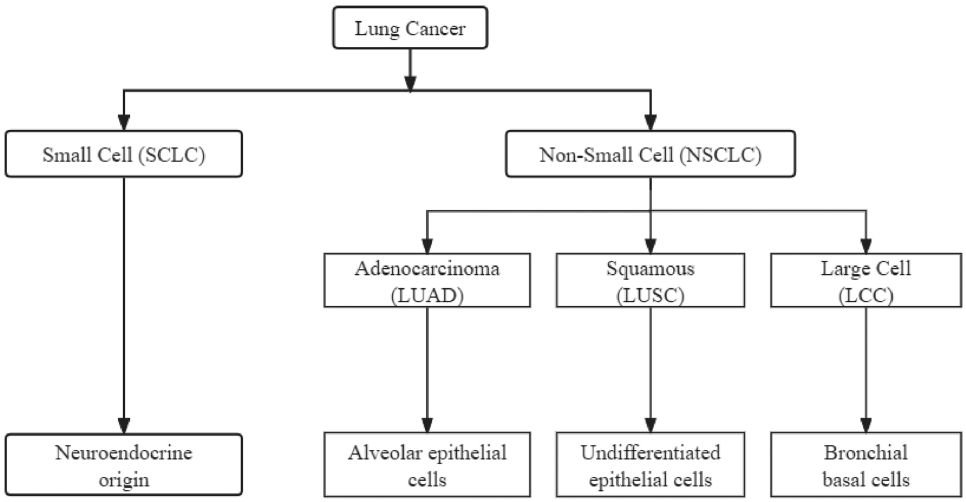

Figure 1. Subtypes of lung cancer disease.

2.2. Notation and definitions

Tree or umbrella ordering is a valuable concept in clinical practice, especially for molecular diagnostics of cancer subtypes. In oncology, a single cancer type often encompasses multiple subtypes without a specific hierarchical order, in contrast to conditions like Alzheimer's disease, where such an ordering is more clearly defined.

To analytically introduce tree or umbrella ordering, let $Y_0 \prec (Y_{11}, Y_{12}, \ldots, Y_{1k})$ or $Y_0 \succ (Y_{11}, Y_{12}, \ldots, Y_{1k})$, respectively, where $\succ$ or $\prec$ indicates stochastic ordering (see Figure 1), $Y_0, Y_{11}, Y_{12}, \ldots, Y_{1k}$ denote biomarker values for non-diseased and k subtypes of the diseased subjects and $(Y_{11}, Y_{12}, \ldots, Y_{1k})$ means that there is no clearly defined order among disease classes $(Y_{11}, Y_{12}, \ldots, Y_{1k})$. Without loss of generalization, we will focus on tree ordering, such that $Y_0 \prec (Y_{11}, Y_{12}, \ldots, Y_{1k})$.

Let F_0 and $F_{1i}(i = 1, 2, \ldots, k)$ denote the CDF of $Y_0, Y_{11}, Y_{12}, \ldots, Y_{1k}$. As proposed by Wang et al. (2016, 2019), sensitivity under the tree ordering is defined as $\mathrm{TSe}(c) = \Pr(\min(Y_{11}, Y_{12}, \ldots, Y_{1k}) > c) = \prod_{i=1}^{k}(1 - F_{1i}(c))$ and the specificity as $\mathrm{Sp}(c) = F_0(c)$, where c is the cut-off point of biomarker. However, for monotone (stochastic) ordering, we have $Y_0 \prec Y_1 \prec Y_2 \prec \cdots \prec Y_k$.

On the other hand, under extended tree ordering, we analytically have $(Y_{01}, Y_{02}, \ldots, Y_{0m}) \prec (Y_{11}, Y_{12}, \ldots, Y_{1k})$, where $(Y_{01}, Y_{02}, \ldots, Y_{0m})$ and $(Y_{11}, Y_{12}, \ldots, Y_{1k})$ indicate that there is no clearly defined order among non-disease $(Y_{01}, Y_{02}, \ldots, Y_{0m})$ and among disease classes $(Y_{11}, Y_{12}, \ldots, Y_{1k})$, respectively. Since all non-disease sub-classes are independent, then at a given cut-off point c, $\mathrm{TSp}(c) = \Pr(Y_{01}, Y_{02}, \ldots, Y_{0m} \leq c) = \Pr(\max(Y_{01}, Y_{02}, \ldots, Y_{0m}) \leq c) = \prod_{i=1}^{m} \Pr(Y_{0i} \leq c) = \prod_{i=1}^{m} F_{0i}(c)$, where F_{0i} is the underlying distribution function of the ith non-disease subclass Y_{0i} $(i = 1, 2, \ldots, m)$.

2.3. Preliminaries of Kullback-Leibler divergence for continuous marker

The Kullback-Leibler (KL) divergence quantifies the difference in distribution between two probability densities $f_1(x)$ and $f_0(x)$ distributed. Here $f_1(x)$ represents the probability density

function (PDF) for diseased patients, while $f_0(x)$ denotes the PDF for non-diseased patients. The discrimination information function $D(f_1, f_0)$ using (f_0) as the reference distribution, is defined as

$$D(f_1, f_0) = \int_{-\infty}^{\infty} f_1(x) \ln\left(\frac{f_1(x)}{f_0(x)}\right) \mathrm{d}x$$
$$= \int_{-\infty}^{\infty} f_1(x) \ln(f_1(x)) \, \mathrm{d}x - \int_{-\infty}^{\infty} f_1(x) \ln(f_0(x)) \, \mathrm{d}x, \quad (1)$$

as defined by Kullback and Leibler (1951).

It can be demonstrated that the Kullback-Leibler (KL) divergence $D(.,.) \geq 0$, and equality holds almost everywhere if and only if $f_1(x)$ and $f_0(x)$ are identical. Consequently, the KL divergence is always non-negative, with larger values indicating greater distinction between the two populations under analysis. It equals zero only when the distributions of diseased and non-diseased individuals are identical with respect to the diagnostic marker. Generally, $D(f_1, f_0) \neq D(f_0, f_1)$ does not exhibit symmetric properties. However, Hughes and Bhattacharya (2013) discussed the symmetry characteristics of bi-normal and bi-gamma Receiver Operating Characteristic (ROC) curves in the context of KL divergences. This measure can also be applied to discrete distributions by substituting integrals with summations. Nonetheless, as will be shown mathematically, categorizing a fundamentally continuous biomarker into discrete groups can reduce the information conveyed by the KL divergence $D(\cdot, \cdot)$.

In the case of discrete distributions, Lee (1999) defined $P_{\text{in}} = \mathrm{e}^{D(g_1, g_0)}$ and $P_{\text{out}} = \mathrm{e}^{D(g_0, g_1)}$ and showed that P_{in} is the ratio, for a randomly selected diseased subject, of the post-test disease odds to the pre-test disease odds. Whereas the P_{out} is the ratio of the pre-test disease odds to the post-test disease odds for a randomly selected disease-free subject. Hence, $P_{\text{in}}(P_{\text{out}})(\geq 1)$ measures the increase (decrease) in disease odds after a test for disease (control) subjects. In general, since both $D(f_1, f_0)$ and $D(f_0, f_1)$ are non-negative and are not bounded, Soofi et al. (1995), proposed a normalized transformation of the KL divergence given by the information distinguishability measure, which can be expressed as $\mathrm{ID}(f_1, f_0) = 1 - \exp[-D(f_1, f_0)]$. This ID measure is bounded between 0 and 1, with $\mathrm{ID} = 0$ indicating that the marker distributions are identical, whereas $\mathrm{ID} = 1$ indicates a complete separation of the marker distributions. This transformation yields a standardized, bounded measure that can be used to compare different biomarkers under study, such as the AUC and Youden index, especially for assessing the potential of before-test rule-in and rule-out.

In the next section, we will define the total KL under the tree ordering of discrete cases, denoted $\mathrm{TTKL}_{\text{discrete}}(c)$, and show that $\mathrm{TTKL}_{\text{discrete}}(c)$ measure is closely related to the sensitivity and specificity, as well as to the Youden index kernel and the natural logarithm of the diagnostic odds ratio under extended tree ordering. Thus, for continuous markers, we propose using $D(f_1, f_0)$ and $D(f_0, f_1)$ as an overall diagnostic accuracy measure for potential rule-in and rule-out. This is comparable to the application of sensitivity and specificity. Additionally, we propose using $\mathrm{TTKL}_{\text{discrete}}(c)$ as an objective function for selecting the optimal cut-off points to achieve the maximum information at the optimal cutoff for tests with high potential in both rule-in and rule-out situations. Like the Youden index, which corresponds to the sum of sensitivity and specificity, TTKL is the sum of the rule-in and rule-out measures of accuracy. This paper introduces a decision-oriented Kullback-Leibler divergence framework for diagnostic test evaluation and cut-off point selection. Unlike existing KL-based measures that emphasize global or symmetric divergence (e.g., Samawi et al., 2020;

Wang et al., 2019), the proposed approach explicitly decomposes diagnostic information into rule-in and rule-out components, reflecting clinical decision pathways. Furthermore, $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ uniquely leverages tree or umbrella ordering to collapse multi-class disease states into a clinically interpretable binary structure, enabling coherent threshold optimization across ordered outcomes. As a result, $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ provides cut-off points that balance sensitivity and specificity in a manner directly aligned with clinical relevance, demonstrating improved robustness and interpretability compared to AUC, Youden-, and existing KL-based methods.

3. Derivation of KL divergence as measures of accuracy under the extended tree ordering: relation with the common ROC indices

In this section, we will show how $D(f_1,f_0)$ and $D(f_0,f_1)$ are related to sensitivity and specificity, and hence to other indices such as the ETJ. Also, we will explain the rationale for using $\mathrm{TTKL}_{\mathrm{discrete}}$ divergence measures for cut-off point selection.

Using the extended tree ordering, for one cut-off point, let c be any diagnostic cut-off point. If the diseased individual has a value larger than c, then we define the stochastic classification matrix under extended tree ordering as follows:

$$P = \begin{matrix} D=0 \\ D=1 \end{matrix} \begin{pmatrix} \overset{T=0}{\mathrm{TSp}} & \overset{T=1}{1-\mathrm{TSp}} \\ 1-\mathrm{TSe} & \mathrm{TSe} \end{pmatrix} = \begin{pmatrix} \prod_{i=1}^{m} F_{0i}(c) & 1-\prod_{i=1}^{m} F_{0i}(c) \\ 1-\prod_{i=1}^{k}\left[1-F_{1i}(c)\right] & \prod_{i=1}^{k}\left[1-F_{1i}(c)\right] \end{pmatrix}.$$

Then, using Equation (1), we can define $D(f_1,f_0)$ under tree ordering as

$$\begin{aligned} D(f_1,f_0) &= \int_{-\infty}^{\infty} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x)}{\prod_{i=1}^{m} f_{0i}(x)}\right) \mathrm{d}x \\ &= \int_{c}^{\infty} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x)}{\prod_{i=1}^{m} f_{0i}(x)}\right) \mathrm{d}x + \int_{-\infty}^{c} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x)}{\prod_{i=1}^{m} f_{0i}(x)}\right) \mathrm{d}x \\ &= D_{\bar{c}}(f_1,f_0) + D_c(f_1,f_0). \end{aligned} \tag{2}$$

Taking $D_{\bar{c}}(f_1,f_0)$ and multiplying and dividing by $\frac{\prod_{i=1}^{k}[1-F_{1i}(c)]}{\prod_{i=1}^{k}[1-F_{1i}(c)]}$ inside the integral, we get

$$\begin{aligned} D_{\bar{c}}(f_1,f_0) &= \int_{c}^{\infty} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x)}{\prod_{i=1}^{k} f_{0i}(x)}\right) \mathrm{d}x \\ &= \int_{c}^{\infty} \frac{\prod_{i=1}^{k}[1-F_{1i}(c)]}{\prod_{i=1}^{k}[1-F_{1i}(c)]} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x) \cdot \frac{\prod_{i=1}^{k}[1-F_{1i}(c)]}{\prod_{i=1}^{k}[1-F_{1i}(c)]}}{\prod_{i=1}^{m} f_{0i}(x) \cdot \frac{1-\prod_{i=1}^{m} F_{0i}(c)}{1-\prod_{i=1}^{m} F_{0i}(c)}}\right) \mathrm{d}x \\ &= \prod_{i=1}^{k}[1-F_{1i}(c)] \int_{c}^{\infty} \frac{\prod_{i=1}^{k} f_{1i}(x)}{\prod_{i=1}^{k}[1-F_{1i}(c)]} \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x) \cdot \frac{\prod_{i=1}^{k}[1-F_{1i}(c)]}{\prod_{i=1}^{k}[1-F_{1i}(c)]}}{\prod_{i=1}^{m} f_{0i}(x) \cdot \frac{1-\prod_{i=1}^{m} F_{0i}(c)}{1-\prod_{i=1}^{m} F_{0i}(c)}}\right) \mathrm{d}x \end{aligned}$$

$$= \mathrm{TSe}(c)E_{f_{1\bar{c}}} \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x) \cdot \frac{\prod_{i=1}^{k}[1-F_{1i}(c)]}{\prod_{i=1}^{k}[1-F_{1i}(c)]}}{\prod_{i=1}^{m} f_{0i}(x) \cdot \frac{1-\prod_{i=1}^{m} F_{0i}(c)}{1-\prod_{i=1}^{m} F_{0i}(c)}}\right)$$

$$= \mathrm{TSe}(c)E_{f_{1\bar{c}}} \ln\left(\frac{f_{1\bar{c}}(X)}{f_{0\bar{c}}(X)}\right) + \mathrm{TSe}(c) \ln\left(\frac{\mathrm{TSe}(c)}{1-\mathrm{TSp}(c)}\right), \tag{3}$$

where $\{f_{1\bar{c}}\}$ is the truncated joint density $f_1 = \prod_{i=1}^{k} f_{1i}(x)$ on the interval (c, ∞) and $E_{f_{1\bar{c}}}(\ln(\frac{f_{1\bar{c}}(X)}{f_{0\bar{c}}(X)}))$ is the expected value of $\ln(\frac{f_{1\bar{c}}(X)}{f_{0\bar{c}}(X)})$ with respect to $\{f_{1\bar{c}}\}$. Similarly,

$$D_c(f_1, f_0) = \int_{-\infty}^{c} \prod_{i=1}^{k} f_{1i}(\mathbf{x}) \ln\left(\frac{\prod_{l=1}^{k} f_{1i}(x)}{\prod_{i=1}^{m} f_{0i}(x)}\right) \mathrm{d}x$$

$$= (1-\mathrm{TSe}(c))E_{f_{1c}}\left(\ln\left(\frac{f_{1c}(\mathbf{X})}{f_{0c}(\mathbf{X})}\right)\right) + (1-\mathrm{TSe}(c)) \ln\left(\frac{(1-\mathrm{TSe}(c))}{\mathrm{TSp}(c)}\right), \tag{4}$$

where $\{f_{1c}\}$ is the joint truncated $f_1 = \prod_{i=1}^{k} f_{1i}(x)$ density on the interval $(-\infty, c)$ and $E_{f_{1c}}\left(\ln\left(\frac{f_{1c}(X)}{f_{0c}(X)}\right)\right)$ is the expected value of $\ln\left(\frac{f_{1c}(X)}{f_{0c}(X)}\right)$ with respect to $\{f_{1c}\}$. Therefore, $D(f_1, f_0)$ and $D(f_0, f_1)$ are based on the expected log likelihood ratio of the diseased population relative to the non-diseased population and the expected log-likelihood ratio of the non-diseased population relative to the diseased population for a given marker. Besides, we can derive the following similarly:

$$D_{\bar{c}}(f_0, f_1) = \int_{c}^{\infty} \prod_{i=1}^{m} f_{0i}(x) \ln\left(\frac{\prod_{l=1}^{m} f_{0i}(x)}{\prod_{i=1}^{k} f_{1i}(x)}\right) \mathrm{d}x$$

$$= (1-\mathrm{TSp}(c))E_{f_{0\bar{c}}}\left(\ln\left(\frac{f_{0\bar{c}}(X)}{f_{1\bar{c}}(X)}\right)\right) + (1-\mathrm{TSp}(c)) \ln\left(\frac{(1-\mathrm{TSp}(c))}{\mathrm{TSe}(c)}\right) \tag{5}$$

and

$$D_c(f_0, f_1) = \int_{-\infty}^{c} \prod_{i=1}^{m} f_{0i}(x) \ln\left(\frac{\prod_{l=1}^{m} f_{0i}(x)}{\prod_{i=1}^{k} f_{1i}(x)}\right) \mathrm{d}x$$

$$= \mathrm{TSp}(c) \ln\left(\frac{\mathrm{TSp}(c)}{1-\mathrm{TSe}(c)}\right) + \mathrm{TSp}(c)E_{f_{0c}}\left(\ln\left(\frac{f_{0c}(X)}{f_{1c}(X)}\right)\right). \tag{6}$$

Finally, from (3) and (4) we have

$$D(f_1, f_0) = \mathrm{TSe}(c) \ln\left(\frac{\mathrm{TSe}(c)}{1-\mathrm{TSp}(c)}\right) + (1-\mathrm{TSe}(c)) \ln\left(\frac{1-\mathrm{TSe}(c)}{\mathrm{TSp}(c)}\right)$$

$$+ \mathrm{TSe}(c)E_{f_{1\bar{c}}}\left(\ln\left(\frac{f_{1\bar{c}}(X)}{f_{0\bar{c}}(X)}\right)\right) + (1-\mathrm{TSe}(c))E_{f_{1c}}\left(\ln\left(\frac{f_{1c}(X)}{f_{0c}(X)}\right)\right)$$

$$= D(\mathrm{TSe}(c), \mathrm{TSp}(c)) + \mathrm{TSe}(c)E_{f_{1\bar{c}}}\left(\ln\left(\frac{f_{1\bar{c}}(X)}{f_{0\bar{c}}(X)}\right)\right)$$

$$+ (1-\mathrm{TSe}(c))E_{f_{1c}}\left(\ln\left(\frac{f_{1c}(X)}{f_{0c}(X)}\right)\right) \tag{7}$$

and from (5) and (6) we have

$$\begin{aligned} D(f_0,f_1) &= (1-\mathrm{TSp}(c))\ln\left(\frac{1-\mathrm{TSp}(c)}{\mathrm{TSe}(c)}\right)+\mathrm{TSp}(c)\ln\left(\frac{\mathrm{TSp}(c)}{1-\mathrm{TSe}(c)}\right)\\ &\quad+(1-\mathrm{TSp}(c))E_{f_{0\bar{c}}}\left(\ln\left(\frac{f_{0\bar{c}}(X)}{f_{1\bar{c}}(X)}\right)\right)+\mathrm{TSp}(c)E_{f_{0c}}\left(\ln\left(\frac{f_{0c}(X)}{f_{1c}(X)}\right)\right)\\ &= D(\mathrm{TSp}(c),\mathrm{TSe}(c))+(1-\mathrm{TSp}(c))E_{f_{0\bar{c}}}\left(\ln\left(\frac{f_{0\bar{c}}(X)}{f_{1\bar{c}}(X)}\right)\right)\\ &\quad+\mathrm{TSp}(c)E_{f_{0c}}\left(\ln\left(\frac{f_{0c}(X)}{f_{1c}(X)}\right)\right). \end{aligned} \tag{8}$$

The total KL (TTKL) in the case of continuous markers for a given cut-off point (c) can be obtained by adding (7) and (8), and we have

$$\mathrm{TTKL} = [\mathrm{TSe}(c)+\mathrm{TSp}(c)-1]\,[\ln(\mathrm{TOR}(c))]+R(c) \tag{9}$$

where the remainder $R(c)$ is given by

$$\begin{aligned} R(c) &= \mathrm{TSe}(c)E_{f_{1\bar{c}}}\left(\ln\left(\frac{f_{1\bar{c}}(X)}{f_{0\bar{c}}(X)}\right)\right)+(1-\mathrm{TSe}(c))E_{f_{1c}}\left(\ln\left(\frac{f_{1c}(X)}{f_{0c}(X)}\right)\right)\\ &\quad+(1-\mathrm{TSp}(c))E_{f_{0\bar{c}}}\left(\ln\left(\frac{f_{0\bar{c}}(X)}{f_{1\bar{c}}(X)}\right)\right)+\mathrm{TSp}(c)E_{f_{0c}}\left(\ln\left(\frac{f_{0c}(X)}{f_{1c}(X)}\right)\right). \end{aligned} \tag{10}$$

When we dichotomize the diagnostic test at the cut-off point c, then as in Lee (1999) and Samawi et al. (2020), we have the discrete version of the two TTKL divergence measures:

$$D(\mathrm{TSe}(c),\mathrm{TSp}(c)) = \mathrm{TSe}(c)\ln\left(\frac{\mathrm{TSe}(c)}{1-\mathrm{TSp}(c)}\right)+(1-\mathrm{TSe}(c))\ln\left(\frac{1-\mathrm{TSe}(c)}{\mathrm{TSp}(c)}\right) \tag{11}$$

and

$$D(\mathrm{TSp}(c),\mathrm{TSe}(c)) = (1-\mathrm{TSp}(c))\ln\left(\frac{1-\mathrm{TSp}(c)}{\mathrm{TSe}(c)}\right)+\mathrm{TSp}(c)\ln\left(\frac{\mathrm{TSp}(c)}{1-\mathrm{TSe}(c)}\right). \tag{12}$$

We interpret the above measures as follows: A diagnostic test with a higher $D(f_1,f_0)$ value will, on average, increase the likelihood of a positive diagnosis among diseased subjects (true positive rate, TSe) relative to the likelihood of a false positive diagnosis among non-diseased subjects (false positive rate, 1-TSp). This indicates a higher potential for the disease to be ruled in. This 'rule-in' potential reflects the trade-off between sensitivity (true positive rate) and false positives (1-TSp), as well as false negatives (1-TSe) and specificity (true negative rate). Conversely, a diagnostic test with a greater $D(f_1,f_0)$ value will, on average, decrease the likelihood of a positive diagnosis among non-diseased subjects (false positives, 1-TSp) relative to its sensitivity (true positive rate), thereby enhancing its potential for ruling out the disease.

Moreover, from (9) and (10), the total sum of the two KL divergences of the discrete versions, denoted by $\mathrm{TTKL}_{\mathrm{discrete}}(c)$, is given by

$$\mathrm{TTKL}_{\mathrm{discrete}}(c) = \big[\mathrm{TSe}(c)+\mathrm{TSp}(c)-1\big]\left[\ln\left(\frac{\mathrm{TSe}(c)}{1-\mathrm{TSe}(c)}\right)+\ln\left(\frac{\mathrm{TSp}(c)}{1-\mathrm{TSp}(c)}\right)\right]$$

$$= \left[\mathrm{TSe}(c) + \mathrm{TSp}(c) - 1\right] \left[\ln\left(\mathrm{TOR}(c)\right)\right]. \tag{13}$$

Therefore, the continuous TTKL is a sum of the discrete parts $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ and a non-negative remainder $R(c)$, which is the loss of information from dichotomizing the continuous tests into a binary test at a selected diagnostic cut-off point. Also, this derivation demonstrates that the proposed measures of rule-in and rule-out by Lee (1999) are just the discrete part of KL measures when the continuous marker is dichotomized into two distinct categories.

3.1. Proposing using $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ for the optimal cut-off point selection (c) and its rationale

When selecting a threshold or cut-off point for an inherently continuous biomarker to create a binary disease status (non-diseased vs. diseased), some information will inevitably be lost due to the dichotomization. The choice of the objective function to estimate the optimal cut-off should align with the scientific or clinical purpose and the specific data context. For instance, if the goal is to enhance the ability to rule-in a disease, selecting the cut-off point that maximizes the $\mathrm{TTKL}(c)$ (total Kullback-Leibler divergence) would be more appropriate.

In particular, while AUC provides a global, threshold-free measure of discrimination and the (extended) Youden index identifies a cut-off by balancing sensitivity and specificity, neither directly reflects the clinical impact of a test result at the point of care. In contrast, $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ measures the total information gained from a biomarker at a specific decision threshold by jointly incorporating likelihood-based evidence from both positive and negative test results. From a clinician's perspective, this aligns directly with diagnostic reasoning, which centres on how test results modify the odds of disease before clinical action. We further added an explanation highlighting that maximizing $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ yields a cut-off that minimizes information loss due to dichotomization and explicitly accounts for both rule-in and rule-out performance via the diagnostic odds ratio. This provides a clinically interpretable advantage over the Youden index, particularly in tree-ordering settings where disease subtypes are heterogeneous and negative Youden values may arise.

Thus, we propose maximizing the $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ measure, which minimizes the remaining information loss, $R(c)$, associated with the cut-off point c. This approach offers an alternative optimization criterion for selecting the diagnostic cut-off point. Several numerical methods available in standard software, such as SAS and R, can be used to efficiently determine the optimal cut-off point. For example, one can find the maximum $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ by taking the first derivative with respect to c, setting it to zero to solve for the optimal c, and confirming the maximum by checking that the second derivative is negative at this point.

In practice, we can estimate the probability density functions (PDFs) and cumulative distribution functions (CDFs) for both diseased and non-diseased groups using empirical or kernel density estimators available in SAS and R. The $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ can then be numerically maximized using methods such as the Newton-Raphson method. Alternatively, the cut-off point can be determined by empirical search: evaluating the $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ for various values of c and selecting the $c = c_{kl}$ that yields the maximum $\mathrm{TTKL}_{\mathrm{discrete}}(c)$. While this latter approach may be somewhat slower than other numerical methods, it is still a viable option.

In addition, when $c = c_p$, where c_p is the optimal cut-off point obtained by the extended tree ordering Youden index, we find a link between the $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ measure and the extended tree ordering Youden index $\{\mathrm{ETJ}(c_p) = \mathrm{TSe}(c_p) + \mathrm{TSp}(c_p) - 1\}$ as $\mathrm{TTKL}(c_p) =$

$\mathrm{ETJ}(c_p)\ln(\mathrm{TOR}(c_p))$, where $\mathrm{TOR}(c_p)$ is the diagnostic odds ratio at the optimal cut-off point associated with the tree ordering Youden index.

3.2. Proposed diagnostic measure of accuracy based on $\mathrm{TTKL}_{\mathrm{discrete}}(c)$

The $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ can be used to measure the overall accuracy of the total rule-in and rule-out performance. For one cut-off point of a continuous biomarker, we also propose using an overall measure of the accuracy of the total rule-in and rule-out performance, namely, the maximum of $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ as follows:

$$\mathrm{MTTKL}_{\mathrm{discrete}}(c_{kl}) = \max_{c} \mathrm{TTKL}_{\mathrm{discrete}}(c), \tag{14}$$

where c_{kl} are selected based on

$$\arg\max_{c}(\mathrm{TTKL}_{\mathrm{discrete}}) = \arg\max_{c}\{[\mathrm{TSe}(c) + \mathrm{TSp}(c) - 1][\ln(\mathrm{TOR}(c))]\}. \tag{15}$$

Note that, under mild regularity conditions, the $\mathrm{TTKL}_{\mathrm{discrete}}$ objective function is continuous over a compact parameter space, ensuring the existence of an optimal cut-point (Rudin, 1976). Moreover, when the distributions of disease and non-disease biomarkers are stochastically ordered, the objective admits a unique maximizer, establishing identifiability (Pepe, 2003). Let $\hat{c}_n$ denote the empirical maximizer of the estimated $\mathrm{TTKL}_{\mathrm{discrete}}$. Uniform convergence of the empirical objective implies $\hat{c}_n \xrightarrow{p} c$, guaranteeing the stability of the estimated cut-point. The curvature of the $\mathrm{TTKL}_{\mathrm{discrete}}$ objective further ensures local robustness to sampling variability, explaining the observed stability of the proposed method relative to competing criteria (van der Vaart, 1998).

3.3. Numerical illustrations

Section 3.3 is intended to provide numerical illustrations, implemented via Monte Carlo methods, that demonstrate how the proposed methodology in Sections 3.1 and 3.2 is applied in practice and to evaluate its exact performance under controlled distributional settings. Consequently, Table 1 presents numerical examples for populations with normally distributed non-diseased and diseased individuals. The sensitivity and specificity values at the optimal cut-off point, associated with the Youden index ($\mathrm{TSe}(c_p)$ and $\mathrm{TSp}(c_p)$), are shown, along with those associated with the maximum $\mathrm{TTKL}_{\mathrm{discrete}}(c)$. Additionally, the table lists values for rule-in and rule-out KL divergence measures ($\mathrm{ID}_{\mathrm{in}}(f_1, f_0)$, $\mathrm{ID}_{\mathrm{out}}(f_0, f_1)$) and the $\mathrm{TTKL}_{\mathrm{discrete}}$ for the two optimal cut-off points.

Across all parameter settings, sensitivity and specificity based on maximizing $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ demonstrate greater separation than those based on maximizing the Youden index under extended tree ordering, especially when the underlying distributions have different variances. When using the Youden index, sensitivity and specificity are more balanced, as it maximizes the sum of these two probabilities, thus better managing the trade-off between false positives and false negatives.

On the other hand, $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ divergence minimizes information loss by selecting the cut-off point that maximizes $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ for a continuous biomarker, but it gives less attention to balancing sensitivity and specificity. We suspect that this imbalance occurs because sensitivity and specificity are scaled by $\log(\mathrm{TOR}(c))$, which tends to amplify the more dominant probability.

Table 1. A numerical example of $\mathrm{TSe}(c_p), \mathrm{TSp}(c_p), \mathrm{TSe}(c_{kl}), \mathrm{TSp}(c_{kl})$, ID and $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ for normal distributions.

μ_{11} μ_{12} μ_{13}	μ_{01} μ_{02} μ_{03}	σ_{11} σ_{12} σ_{13}	σ_{01} σ_{02} σ_{03}	$\mathrm{TSe}(c_p)$	$\mathrm{TSp}(c_p)$	$\mathrm{TSe}(c_{kl})$	$\mathrm{TSp}(c_{kl})$	$\mathrm{ID}_{\mathrm{in}}(c_p)$ $\mathrm{ID}_{\mathrm{in}}(c_{kl})$	$\mathrm{ID}_{\mathrm{out}}(c_p)$ $\mathrm{ID}_{\mathrm{out}}(c_{kl})$	$\mathrm{TTKL}_{\mathrm{discrete}}(c_p)$ $\mathrm{TTKL}_{\mathrm{discrete}}(c_{kl})$
1.7	0.0	1.0	1.0	0.5373	0.5417	0.5360	0.5429	0.0124	0.0124	0.0250
1.8	0.01	1.0	1.0					0.0124	0.0124	0.0250
1.9	0.02	1.0	1.0							
1.8	0.0	1.0	1.0	0.5643	0.5675	0.5630	0.5688	0.0343	0.0343	0.0700
1.9	0.01	1.0	1.0					0.0343	0.0343	0.0700
2.0	0.02	1.0	1.0							
2.1	0.0	1.0	1.0	0.6220	0.6266	0.6195	0.6290	0.1186	0.1184	0.2523
2.0	0.01	1.0	1.0					0.1187	0.1185	0.2523
2.3	0.02	1.0	1.0							
2.4	0.0	1.0	1.0	0.7167	0.7193	0.7148	0.7212	0.3350	0.3345	0.8151
2.5	0.01	1.0	1.0					0.3353	0.3341	0.8151
2.7	0.02	1.0	1.0							
1.7	0.0	1.1	1.0	0.3759	0.6650	0.9711	0.0100	0.0037	0.0036	0.0073
1.8	0.01	1.1	1.0					0.0119	0.0084	0.0204
1.9	0.02	1.1	1.0							
1.8	0.0	1.0	1.2	0.5100	0.5060	0.9806	0.0082	0.0005	0.0005	0.0010
1.9	0.01	1.2	1.1					0.0056	0.0042	0.0100
2.0	0.02	1.2	1.1							
2.1	0.0	1.0	1.1	0.6521	0.5480	0.6886	0.4984	0.0777	0.0799	0.1641
2.0	0.01	1.0	1.1					0.0774	0.0818	0.1660
2.3	0.02	1.0	1.1							
2.4	0.0	1.0	1.2	0.6538	0.6496	0.6562	0.6472	0.1730	0.1733	0.3803
2.5	0.01	1.2	1.1					0.1729	0.1735	0.3803
2.7	0.02	1.2	1.1							

Table 1 also indicates that if a diagnostic test effectively discriminates between diseased and non-diseased individuals, more specific tests have a higher rule-in potential, whereas more sensitive tests perform better at ruling out disease. This supports the principles of 'rule-in, specific test' and 'rule-out, sensitive test'. As indicated by Samawi et al. (2020), the ID values ($D(\cdot,\cdot)$) flipped when we switch the disease and non-diseased distributions, and we refer to this as the symmetry of KL divergence, which has been proven to be valid under bi-normal and bi-gamma (Hughes & Bhattacharya, 2013).

Table 2 presents numerical examples in which the underlying distributions of the non-diseased and diseased populations are assumed to be gamma. Like the normal distribution cases, the sensitivity and specificity values based on $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ are further apart compared to those based on the Youden index. The pattern of rule-in and rule-out potential remains consistent with the observations from the normal distribution. Again, $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ divergence minimizes information loss by selecting the cut-off point that maximizes $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ for a continuous biomarker, but it gives less attention to balancing sensitivity and specificity.

Notably, under a bi-exponential distribution (a special case of bi-gamma), the ID values reversed when switching the diseased and non-diseased distributions, as observed in ROC measures such as sensitivity, specificity, and AUC (Samawi et al., 2020). Results from other distributions showed similar patterns.

Table 2. A numerical example of $\mathrm{TSe}(c_p), \mathrm{TSp}(c_p), \mathrm{TSe}(c_{kl}), \mathrm{TSp}(c_{kl})$, ID and $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ for gamma distributions.

β_{01} β_{02}	β_{11} β_{12}	α_{01} α_{02}	α_{11} α_{12}	$\mathrm{TSe}(c_p)$	$\mathrm{TSp}(c_p)$	$\mathrm{TSe}(c_{kl})$	$\mathrm{TSp}(c_{kl})$	$\mathrm{ID}_{\mathrm{in}}(c_p)$ $\mathrm{ID}_{\mathrm{in}}(c_{kl})$	$\mathrm{ID}_{\mathrm{out}}(c_p)$ $\mathrm{ID}_{\mathrm{out}}(c_{kl})$	$\mathrm{TTKL}_{\mathrm{discrete}}(c_p)$ $\mathrm{TTKL}_{\mathrm{discrete}}(c_{kl})$
1.00	0.70	3.00	5.00	0.6639	0.7008	0.6348	0.7286	0.2458	0.2415	0.5585
1.00	0.60	3.00	5.00					0.2487	0.2387	0.5600
1.00	0.70	3.00	5.00	0.7052	0.7782	0.6391	0.83567	0.4111	0.3927	1.0282
1.00	0.60	3.00	8.00					0.4316	0.3802	1.0432
1.00	0.70	4.00	6.00	0.6928	0.7606	0.6357	0.8117	0.3672	0.3528	0.8927
1.00	0.60	4.00	8.00					0.3822	0.3437	0.9026
1.00	0.50	3.00	5.00	0.7869	0.8377	0.7252	0.8881	0.6134	0.5895	1.8409
1.00	0.40	3.00	5.00					0.6441	0.5651	1.8657
1.00	0.30	3.00	5.00	0.9231	0.9495	0.8178	0.9903	0.9170	0.8935	4.7288
1.00	0.40	3.00	8.00					0.9639	0.8047	4.9537

4. The diagrammatic interpretation of TTKL divergence for continuous markers at a given cut-off point (c) concerning Bregman divergence

Hughes (2013) explored the diagrammatic interpretation of Kullback-Leibler (KL) divergence in the context of tree ordering for continuous markers, introducing the concept of an 'information graph' (Hughes, 2013). This graph is a visual representation used to assess and compare continuous markers at various cut-off points. Hughes (2013) demonstrated that constructing the information graph involves representing the discrete KL divergence as a discrete Bregman divergence (Bregman, 1967), a concept related to convex functions.

In this context, we will show that the total Kullback-Leibler (TTKL) divergence is a Bregman divergence associated with the negative Shannon entropy function (Shannon & Weaver, 1949) for continuous markers. This relationship holds for the discrete case at a specific cut-off point (c) under tree ordering. The negative Shannon entropy function for continuous variables in the context of tree or umbrella ordering is defined as follows:

$$I(f_1) = \int_{-\infty}^{\infty} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\prod_{i=1}^{k} f_{1i}(x)\right) \mathrm{d}x. \tag{16}$$

Using similar arguments as in (3), we have

$$I_{\bar{c}}(f_1) = \int_{c}^{\infty} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\prod_{i=1}^{k} f_{1i}(x)\right) \mathrm{d}x = \mathrm{TSe}(c) E_{f_{1\bar{c}}}\left(\ln\left(\prod_{i=1}^{k} f_{1i}(x)\right)\right) \tag{17}$$

and

$$I_{c}(f_1) = \int_{-\infty}^{c} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\prod_{i=1}^{k} f_{1i}(x)\right) \mathrm{d}x = (1 - \mathrm{TSe}(c)) E_{f_{1c}}\left(\ln\left(\prod_{i=1}^{k} f_{1i}(x)\right)\right), \tag{18}$$

and then

$$\begin{aligned} I(f_1) &= \mathrm{TSe}(c) E_{f_{1\bar{c}}}\left(\ln\left(\prod_{i=1}^{k} f_{1i}(x)\right)\right) + (1 - \mathrm{TSe}(c)) E_{f_{1c}}\left(\ln\left(\prod_{i=1}^{k} f_{1i}(x)\right)\right) \\ &= \mathrm{TSe}(c) \ln(\mathrm{TSe}(c)) + (1 - \mathrm{TSe}(c)) \ln(1 - \mathrm{TSe}(c)) \\ &\quad + \mathrm{TSe}(c) E_{f_{1\bar{c}}}\left(\ln\left(f_{1\bar{c}}(X)\right)\right) + (1 - \mathrm{TSe}(c)) E_{f_{1c}}\left(\ln\left(f_{1c}(X)\right)\right). \end{aligned} \tag{19}$$

Similarly,

$$I(f_0) = (1 - \mathrm{TSp}(c)) \ln\big(1 - \mathrm{TSp}(c)\big) + \mathrm{TSp}(c) \ln\big(\mathrm{TSp}(c)\big) + (1 - \mathrm{TSp}(c)) E_{f_{0c}}\left(\ln\left(f_{0c}(X)\right)\right) + \mathrm{TSp}(c) E_{f_{0c}}\left(\ln\left(f_{0c}(X)\right)\right). \tag{20}$$

Moreover, under tree or umbrella ordering diagnostic test at the cut-off point (c), where $\mathrm{TSe}(c) = \prod_{i=1}^{k}(1 - F_{1i}(c))$ and $\mathrm{TSp}(c) = \prod_{i=1}^{m} F_{0i}(c)$, we have

$$I(\mathrm{TSe}(c)) = \mathrm{TSe}(c) \ln\big(\mathrm{TSe}(c)\big) + (1 - \mathrm{TSe}(c)) \ln\big(1 - \mathrm{TSe}(c)\big) \tag{21}$$

and

$$I(1 - \mathrm{TSp}(c)) = (1 - \mathrm{TSp}(c)) \ln\big(1 - \mathrm{TSp}(c)\big) + \mathrm{TSp}(c) \ln\big(\mathrm{TSp}(c)\big). \tag{22}$$

Generally, the negative of the Shannon Entropy Function can be written as $g(P) = P\ln(P) + (1 - P)\ln\big(1 - P\big)$ where $P = 1 - \mathrm{TSp}(c)$ or $P = \mathrm{TSe}(c)$ when the reference population is the non-diseased or the diseased population, respectively. In order to find the Bregman divergence, we need to find a tangent to the curve $g(P_1)$ with gradient $g'(P_1) = \ln(P_1) - \ln\big(1 - P_1\big) = \text{Slope}$, and an intercept $g(0) = g(P_1) - P_1 g'(P_1)$, drawn at a point P_1 (the reference point). Then, we can find the Bregman divergence, which is the distance between the tangent and the curve at a point P_0. Consequently, the Bregman divergence is given by

$$B(P_0, P_1) = g(P_0) - g(P_1) - (P_0 - P_1) g'(P_1) \tag{23}$$

and for the particular $g(P)$, we will show that this is a TKL divergence. In the case where a continuous variable is dichotomized at a cut-off point c, and from (20) and (23), we have

$$\begin{aligned} B(1 - \mathrm{TSp}(c), \mathrm{TSe}(c)) &= (1 - \mathrm{TSp}(c)) \ln\big(1 - \mathrm{TSp}(c)\big) + \mathrm{TSp}(c) \ln\big(\mathrm{TSp}(c)\big) \\ &\quad - \Big\{\mathrm{TSe}(c) \ln\big(\mathrm{TSe}(c)\big) + (1 - \mathrm{TSe}(c)) \ln\big(1 - \mathrm{TSe}(c)\big)\Big\} \\ &\quad - \Big\{(1 - \mathrm{TSp}(c)) - \mathrm{TSe}(c)\Big\}\Big[\ln\big(\mathrm{TSe}(c)\big) - \ln\big(1 - \mathrm{TSe}(c)\big)\Big] \\ &= (1 - \mathrm{TSp}(c)) \ln\left(\frac{1 - \mathrm{TSp}(c)}{\mathrm{TSe}(c)}\right) + \mathrm{TSp}(c) \ln\left(\frac{\mathrm{TSp}(c)}{1 - \mathrm{TSe}(c)}\right) \\ &= D\big(1 - \mathrm{TSp}(c), \mathrm{TSe}(c)\big). \end{aligned} \tag{24}$$

Finally, it can be shown that $B(\mathrm{TSe}(c), 1 - \mathrm{TSp}(c)) = D(\mathrm{TSe}(c), 1 - \mathrm{TSp}(c))$. Consequently, in the information graph, $\mathrm{TTKL}_{\mathrm{discrete}}(c)$ divergences are depicted by the vertical lines between the negative Shannon entropy curve and the two tangents to the curve at points $\mathrm{TSe}(c)$ and $1 - \mathrm{TSp}(c)$ (Hughes, 2013). This provides the theoretical foundation for constructing the information graph, as illustrated in Figure 3.

Thus, in the case of continuous distributions, Bregman divergence is given by

$$B(f_1, f_0) = \int_{-\infty}^{\infty} \left[T\left(\prod_{i=1}^{k} f_{1i}(x)\right) - T\left(\prod_{i=1}^{m} f_{0i}(x)\right) - T'\left(\prod_{i=1}^{m} f_{0i}(x)\right)\left(\prod_{i=1}^{k} f_{1i}(x) - \prod_{i=1}^{m} f_{0i}(x)\right)\right] \mathrm{d}x, \tag{25}$$

Table 3. Simulation scenarios for normal and gamma distributions with AUC under extended tree orderings (ETAUC, Wang et al., 2016).

Norman Distribution					
Scenarios	μ_{01}, μ_{02}	μ_{11}, μ_{12}	σ_{01}, σ_{02}	σ_{11}, σ_{12}	ETAUC
S1	0.00, 0.01	1.30, 1.50	1.00, 1.00	1.00, 1.00	0.5946
S2	0.00, 0.01	1.50, 1.70	1.00, 1.00	1.00, 1.00	0.6588
S3	0.00, 0.01	2.30, 2.70	1.00, 1.00	1.00, 1.00	0.8732
S4	0.00, 0.01	1.50, 1.70	1.10, 1.00	1.00, 1.20	0.6238
S5	0.00, 0.01	2.30, 2.70	1.00, 1.00	1.10, 1.10	0.8522
Gamma Distributions					
Scenarios	β_{01}, β_{02}	β_{11}, β_{12}	α_{01}, α_{02}	α_{11}, α_{12}	ETAUC
GS1	1.00, 1.00	0.70, 0.60	3.00, 3.00	4.00, 4.00	0.5774
GS2	1.00, 1.00	0.70, 0.60	3.00, 3.00	5.00, 5.00	0.7438
GS3	1.00, 1.00	0.70, 0.60	3.00, 3.00	6.00, 4.00	0.7157
GS4	1.00, 1.00	0.50, 0.40	3.00, 3.00	3.00, 3.00	0.5680
GS5	1.00, 1.00	0.30, 0.40	3.00, 3.00	5.00, 8.00	0.9820

where $T(t)$ is a strictly convex function on t. If we take $T(t) = t\ln(t)$, then with a little algebra, we have

$$B(f_1, f_0) = \int_{-\infty}^{\infty} \prod_{i=1}^{k} f_{1i}(x) \ln\left(\frac{\prod_{i=1}^{k} f_{1i}(x)}{\prod_{i=1}^{m} f_{0i}(x)}\right) \mathrm{d}x = D(f_1, f_0). \tag{26}$$

5. Simulation studies

To evaluate the performance of TTKL(c) and compare it with ETJ, we conducted a simulation study to assess the empirical power of these competing measures. We assume $m = 2$ and $k = 2$, meaning we have two subtypes of non-diseased patients and two subtypes of diseased patients. In our simulation, we generate data from normal and gamma distributions for non-disease and disease groups (subtypes). The parameters for these distributions are detailed in Tables 8 and 9, respectively, for power analysis, while those for the cut-off point simulations are presented in Table 3. We used equal sample sizes (20, 50, 100) in the simulations.

Scenario 1 assumes that the underlying distributions for the non-diseased and diseased groups are identical under the null hypothesis (H_0), implying that the biomarker cannot accurately distinguish between non-diseased and diseased individuals and is, therefore, medically irrelevant. Scenario 2 through 6 consider different underlying distributions for the non-diseased and diseased groups under the null and the alternative hypotheses, suggesting that the biomarker is clinically relevant and can effectively differentiate between these groups.

To determine the critical values of the tests under H_0, we simulated random samples 2000 times for each scenario, obtaining the 95% quantiles for the statistics (ETJ and TTKL(c)). We estimated the power using 2000 replications, rejecting the null hypothesis if the statistics estimated under the alternative hypotheses (H_a) exceeded their corresponding 95% quantiles under H_0. The sample sizes used in our simulation were $n = 20$, 50, and 100.

Additionally, we evaluated the optimal cut-point criteria using absolute relative bias (RBias), root-mean-square error (RMSE), total correct classification rate (TCCR), and maximum-minimum difference (MMDIF). The RBias of the optimal cut-off points was estimated from $N = 2000$ iterations, with its variance for each cut-off point captured by the

Table 4. Kernel estimation, Bias and RMSE to ETJ and $\text{TTKL}_{\text{discrete}}(c)$, for normal distributions.

	Sample size	ETJ			TTKL(c)		
Scenarios	n	$\hat{c}$	|Bias|	RMSE	$\hat{c}$	|Bias|	RMSE
S1	(20, 20, 20, 20)	0.7024	0.0021	1.2671	0.7681	0.0588	0.7781
	(50, 50, 50, 50)	0.7062	0.0017	0.6238	0.7180	0.0087	0.6254
	(100, 100, 100, 100)	0.6904	0.0141	0.3704	0.7166	0.0072	0.5025
S2	(20, 20, 20, 20)	0.8231	0.0202	0.5114	0.8298	0.0227	0.6804
	(50, 50, 50, 50)	0.8072	0.0043	0.3215	0.8100	0.0029	0.4968
	(100, 100, 100, 100)	0.8047	0.0017	0.1952	0.7973	0.0097	0.3918
S3	(20, 20, 20, 20)	1.2417	0.0031	0.1947	1.2709	0.0129	0.5248
	(50, 50, 50, 50)	1.2507	0.0059	0.1327	1.2541	0.0039	0.3521
	(100, 100, 100, 100)	1.2483	0.0035	0.0950	1.2766	0.0186	0.2630
S4	(20, 20, 20, 20)	0.8934	0.0566	1.0412	0.8804	0.0160	0.7394
	(50, 50, 50, 50)	0.8347	0.0022	0.4895	0.8939	0.0025	0.6035
	(100, 100, 100, 100)	0.8403	0.0034	0.2526	0.8983	0.0019	0.4651
S5	(20, 20, 20, 20)	1.2504	0.0108	0.2123	1.4127	0.0632	0.5574
	(50, 50, 50, 50)	1.2506	0.0110	0.1443	1.3906	0.0411	0.3839
	(100, 100, 100, 100)	1.2494	0.0098	0.1059	1.3788	0.0293	0.2840

Table 5. Kernel estimation, Bias and RMSE to ETJ and $\text{TTKL}_{\text{discrete}}(c)$, for gamma distributions.

	Sample size	ETJ			$\text{TTKL}_{\text{discrete}}(c)$		
Scenarios	n	$\hat{c}$	|Bias|	RMSE	$\hat{c}$	|Bias|	RMSE
GS1	(20, 20, 20, 20)	4.8527	0.5226	2.6148	4.6916	0.0821	2.2752
	(50, 50, 50, 50)	4.8849	0.5548	1.4022	5.0592	0.2855	2.0817
	(100, 100, 100, 100)	4.7945	0.4644	0.9015	5.2827	0.5090	1.6042
GS2	(20, 20, 20, 20)	5.0272	0.4299	0.9094	5.8991	1.1394	2.1193
	(50, 50, 50, 50)	4.9256	0.3284	0.5617	5.6972	0.9376	1.6839
	(100, 100, 100, 100)	4.8430	0.2458	0.4084	5.4500	0.6904	1.1699
GS3	(20, 20, 20, 20)	4.9973	0.3635	0.8208	6.0363	1.0626	2.1524
	(50, 50, 50, 50)	4.9334	0.2996	0.5744	5.7915	0.8179	1.6378
	(100, 100, 100, 100)	4.8986	0.2648	0.4541	5.5994	0.6258	1.2021
GS4	(20, 20, 20, 20)	5.4477	0.3997	1.9652	5.0543	1.1320	3.2389
	(50, 50, 50, 50)	5.4808	0.4329	1.0174	5.8713	0.3150	2.7436
	(100, 100, 100, 100)	5.4260	0.3780	0.8000	6.3557	0.1694	2.1114
GS5	(20, 20, 20, 20)	7.2808	0.0856	0.7671	9.6223	0.3031	1.8308
	(50, 50, 50, 50)	7.2407	0.0455	0.5243	9.5219	0.2027	1.7075
	(100, 100, 100, 100)	7.2279	0.0326	0.3869	9.5204	0.2012	1.6316

following formulas:

$$\text{RBias}(\widehat{T}) = \left[\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{T}_i - T\right)\Big/ T\right], \quad \text{Variance}(\widehat{T}) = \sum_{i=1}^{N}(\widehat{T}_i - T)^2\Big/(N-1).$$

Then, the RMSE is computed as $\text{RMSE} = \sqrt{\text{Bias}^2 + \text{Variance}(\widehat{T})}$, where $\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}\widehat{T}_i - T$. $\widehat{T}_i$ and T denote the estimated and true values of the cut-off point, respectively. The TCCR is computed as the summation of all the correct classification rates across all stages: $\text{TCCR} = \frac{1}{N}\sum_{i=1}^{N}[\text{TSp}_i(.) + \text{TSe}_i(.)]$.

In addition, the maximum-minimum difference (MMDIF), which is proposed by Wang et al. (2019) to measure the balance of correct classification rates among disease stages, and it is computed as $\text{MMDIF} = \frac{1}{N}\sum_{i=1}^{N}\frac{\max(\text{TSp}_i(.),\text{TSe}_i(.)) - \min(\text{TSp}_i(.),\text{TSe}_i(.))}{\min(\text{TSp}_i(.),\text{TSe}_i(.))}$.

A smaller MMDIF shows a more balanced measure in selecting optimal cut-points.

Tables 4-7 present the simulation studies for the cut-off point selection and provide the results of the simulation for Kernel estimation, Bias, and RMSE to ETJ and $\text{TTKL}_{\text{discrete}}(c)$, as well as TCCR, MMDIF for normal and gamma distributions. In general, $\text{TTKL}_{\text{discrete}}(c)$

Table 6. Simulation results for TCCR, MMDIF to ETJ and $\text{TTKL}_{\text{discrete}}(c)$, for normal distributions.

	Sample size	ETJ						$\text{TTKL}_{\text{discrete}}(c)$					
Scenarios	n	TP0	TP1	TCCR	MMDIF	ETJ	TTKL	TP0	TP1	TCCR	MMDIF	ETJ	TTKL
S1	(20, 20, 20, 20)	0.5712	0.5582	1.1294	0.0232	0.1294	0.1236	0.5701	0.5282	1.0983	0.0792	0.0983	0.1523
	(50, 50, 50, 50)	0.5597	0.5579	1.1176	0.0034	0.1176	0.0813	0.5600	0.5480	1.1080	0.0218	0.1080	0.0891
	(100, 100, 100, 100)	0.5565	0.5614	1.1179	0.0089	0.1179	0.0714	0.5639	0.5498	1.1136	0.0256	0.1136	0.0754
S2	(20, 20, 20, 20)	0.6070	0.5976	1.2046	0.0157	0.2046	0.2408	0.6056	0.5818	1.1874	0.0409	0.1874	0.2711
	(50, 50, 50, 50)	0.6030	0.5975	1.2005	0.0092	0.2005	0.1949	0.5987	0.5934	1.1921	0.0089	0.1921	0.2051
	(100, 100, 100, 100)	0.6043	0.6006	1.2049	0.0061	0.2049	0.1877	0.5979	0.5992	1.1970	0.0022	0.1970	0.1904
S3	(20, 20, 20, 20)	0.7642	0.7656	1.5299	0.0019	0.5299	1.3705	0.7539	0.7377	1.4915	0.0220	0.4915	1.5307
	(50, 50, 50, 50)	0.7715	0.7659	1.5374	0.0073	0.5374	1.3422	0.7646	0.7547	1.5193	0.0132	0.5193	1.3813
	(100, 100, 100, 100)	0.7760	0.7707	1.5467	0.0069	0.5467	1.3707	0.7799	0.7569	1.5368	0.0305	0.5368	1.3919
S4	(20, 20, 20, 20)	0.6145	0.5462	1.1607	0.1250	0.1607	0.1663	0.6058	0.5402	1.1459	0.1215	0.1459	0.2031
	(50, 50, 50, 50)	0.6047	0.5532	1.1579	0.0930	0.1579	0.1295	0.6122	0.5323	1.1445	0.1502	0.1445	0.1318
	(100, 100, 100, 100)	0.6021	0.5536	1.1557	0.0877	0.1557	0.1136	0.6203	0.5328	1.1531	0.1643	0.1531	0.1220
S5	(20, 20, 20, 20)	0.7680	0.7253	1.4932	0.0588	0.4932	1.1803	0.7928	0.6608	1.4536	0.1999	0.4536	1.3345
	(50, 50, 50, 50)	0.7728	0.7268	1.4996	0.0632	0.4996	1.1490	0.8038	0.6751	1.4789	0.1907	0.4789	1.2093
	(100, 100, 100, 100)	0.7753	0.7320	1.5073	0.0590	0.5073	1.1656	0.8092	0.6831	1.4923	0.1846	0.4923	1.1836

Table 7. Simulation results for TCCR, MMDIF to ETJ and $\text{TTKL}_{\text{discrete}}(c)$, for gamma distributions.

	Sample size	ETJ						$\text{TTKL}_{\text{discrete}}(c)$					
Scenarios	n	TP0	TP1	TCCR	MMDIF	ETJ	TTKL	TP0	TP1	TCCR	MMDIF	ETJ	TTKL
GS1	(20, 20, 20, 20)	0.7234	0.4065	1.1299	0.7794	0.1299	0.1396	0.6297	0.4553	1.0850	0.3831	0.0850	0.1767
	(50, 50, 50, 50)	0.7223	0.3959	1.1182	0.8243	0.1182	0.0909	0.6955	0.3954	1.0909	0.7591	0.0909	0.1058
	(100, 100, 100, 100)	0.7119	0.4005	1.1125	0.7775	0.1125	0.0718	0.7549	0.3465	1.1014	1.1784	0.1014	0.0837
GS2	(20, 20, 20, 20)	0.7528	0.5760	1.3287	0.3069	0.3287	0.5606	0.8081	0.4733	1.2813	0.7075	0.2813	0.5773
	(50, 50, 50, 50)	0.7419	0.5880	1.3299	0.2618	0.3299	0.5069	0.8079	0.4862	1.2941	0.6618	0.2941	0.5200
	(100, 100, 100, 100)	0.7314	0.6031	1.3345	0.2126	0.3345	0.4967	0.7997	0.5139	1.3136	0.5562	0.3136	0.5142
GS3	(20, 20, 20, 20)	0.7502	0.5501	1.3003	0.3636	0.3003	0.4813	0.8133	0.4333	1.2466	0.8769	0.2466	0.4808
	(50, 50, 50, 50)	0.7407	0.5569	1.2976	0.3300	0.2976	0.4186	0.8167	0.4500	1.2667	0.8148	0.2667	0.4498
	(100, 100, 100, 100)	0.7386	0.5625	1.3011	0.3131	0.3011	0.4101	0.8148	0.4645	1.2794	0.7541	0.2794	0.4349
GS4	(20, 20, 20, 20)	0.8007	0.3487	1.1493	1.2966	0.1493	0.1842	0.6554	0.4274	1.0828	0.5334	0.0828	0.2522
	(50, 50, 50, 50)	0.8093	0.3302	1.1395	1.4510	0.1395	0.1331	0.7564	0.3361	1.0925	1.2506	0.0925	0.1691
	(100, 100, 100, 100)	0.8064	0.3278	1.1342	1.4603	0.1342	0.1105	0.8398	0.2642	1.1040	2.1794	0.1040	0.1379
GS5	(20, 20, 20, 20)	0.9546	0.8691	1.8236	0.0984	0.8236	4.4347	0.9856	0.7460	1.7316	0.3211	0.7316	3.5803
	(50, 50, 50, 50)	0.9498	0.8802	1.8300	0.0791	0.8300	4.2508	0.9842	0.7606	1.7448	0.2941	0.7448	4.1766
	(100, 100, 100, 100)	0.9493	0.8892	1.8385	0.0677	0.8385	4.2847	0.9848	0.7651	1.7500	0.2871	0.7500	4.3973

provided larger cut-off points and less information loss than ETJ. Therefore, it will provide higher specificity and greater rule-in ability, which is more clinically beneficial in the case of a new pandemic or disease. Also, cut-off points have slightly larger biases and RMSEs in most cases. On the other hand, as expected, ETJ has a slightly larger TCCR and smaller MMDIF in most cases.

Table 8 presents the power analysis results for the normal distributions underlying the non-disease and disease populations, while Table 9 presents the results for Gamma distributions. Tables 8 and 9 show that both diagnostic accuracy measures, ETJ and $\text{TTKL}_{\text{discrete}}(c)$, provide a close estimation to the nominal value of the tests (0.05) in Scenario 1. On the other hand, Scenarios 2-6 in Tables 8 and 9 show that $\text{TTKL}_{\text{discrete}}(c)$ is at least as good as ETJ in some mean shift of the underlying distribution from the null to alternative hypotheses.

In general, as the distributions under H_0 and H_a become increasingly distinct, the power of the tests increases. In addition, the power of the tests increases with sample size. Thus, our power analysis shows that $\text{TTKL}_{\text{discrete}}(c)$ can capture some differences among the disease groups as well as ETJ. Finally, the choice of the right diagnostic measure depends not only on its power to discriminate but also on the purpose for which the underlying biomarker is being used. Our proposed $\text{TTKL}_{\text{discrete}}(c)$ measure is recommended as a cut-off point selection

Table 8. Power analysis for the normal underlying distribution.

Distribution for H_0	Distribution for H_0	Distribution for H_a	Distribution for H_a	Sample size	Power of the tests	
Non-disease	Disease	Non-disease	Disease	n	ETJ	TTKL(c)
$N(0, 1.0)$	$N(0, 1.0)$	$N(0, 1.0)$	$N(0, 1.0)$	20, 20, 20, 20	0.0660	0.0515
$N(0.01, 1.0)$	$N(0.01, 1.0)$	$N(0.01, 1.0)$	$N(0.01, 1.0)$	50, 50, 50, 50	0.0555	0.0465
				100, 100, 100, 100	0.0480	0.0485
Scenario #1						
$N(0, 1.0)$	$N(1, 1.0)$	$N(0, 1.0)$	$N(1.9, 1.0)$	20, 20, 20, 20	0.9465	0.6620
$N(0.01, 1.0)$	$N(1.1, 1.0)$	$N(0.01, 1.0)$	$N(2.0, 1.0)$	50, 50, 50, 50	1.0000	0.9690
				100, 100, 100, 100	1.0000	1.0000
Scenario #2						
$N(0, 1.0)$	$N(1, 1.0)$	$N(0, 1.0)$	$N(1.9, 1.2)$	20, 20, 20, 20	0.8810	0.5410
$N(0.01, 1.0)$	$N(1.1, 1.0)$	$N(0.01, 1.0)$	$N(2.0, 1.0)$	50, 50, 50, 50	0.9985	0.9005
				100, 100, 100, 100	1.0000	0.9995
Scenario #3						
$N(0, 1.0)$	$N(1, 1.1)$	$N(0, 1.0)$	$N(2.5, 1.0)$	20, 20, 20, 20	1.0000	0.9520
$N(0.01, 1.2)$	$N(1.1, 1.2)$	$N(0.01, 1.0)$	$N(2.3, 1.0)$	50, 50, 50, 50	1.0000	1.0000
				100, 100, 100, 100	1.0000	1.0000
Scenario #4						
$N(0, 1.0)$	$N(1, 1.1)$	$N(0, 1.0)$	$N(2.5, 1.0)$	20, 20, 20, 20	0.9990	0.8750
$N(0.01, 1.2)$	$N(1.1, 1.2)$	$N(0.01, 1.0)$	$N(2.3, 1.3)$	50, 50, 50, 50	1.0000	0.9985
				100, 100, 100, 100	1.0000	1.0000
Scenario #5						
$N(0, 1.0)$	$N(1, 1.1)$	$N(0, 1.0)$	$N(2.9, 1.0)$	20, 20, 20, 20	1.0000	0.9920
$N(0.01, 1.2)$	$N(1.1, 1.2)$	$N(0.01, 1.0)$	$N(2.7, 1.3)$	50, 50, 50, 50	1.0000	1.0000
				100, 100, 100, 100	1.0000	1.0000
Scenario #6						

Table 9. Power analysis for the Gamma underlying distribution.

Distribution for H_0	Distribution for H_0	Distribution for H_a	Distribution for H_a	Sample size	Power of the tests	
Non-disease	Disease	Non-disease	Disease	n	ETJ	TTKL(c)
$G(3.0, 1.00)$	$G(3.0, 1.00)$	$G(3.0, 1.0)$	$G(3.0, 1.00)$	20, 20, 20, 20	0.0502	0.0490
$G(3.1, 1.00)$	$G(3.1, 1.00)$	$G(3.1, 1.0)$	$G(3.1, 1.00)$	50, 50, 50, 50	0.0511	0.0465
				100, 100, 100, 100	0.0488	0.0462
Scenario #1						
$G(3.0, 1.00)$	$G(4.0, 0.80)$	$G(3.0, 1.00)$	$G(5.0, 0.80)$	20, 20, 20, 20	0.5898	0.2107
$G(3.1, 1.00)$	$G(4.1, 0.70)$	$G(3.1, 1.00)$	$G(5.1, 0.70)$	50, 50, 50, 50	0.8961	0.4408
				100, 100, 100, 100	0.9445	0.7185
Scenario #2						
$G(3.0, 1.00)$	$G(4.0, 0.80)$	$G(3.0, 1.00)$	$G(4.0, 0.55)$	20, 20, 20, 20	0.7983	0.464
$G(3.1, 1.00)$	$G(4.1, 0.70)$	$G(3.1, 1.00)$	$G(4.1, 0.50)$	50, 50, 50, 50	0.9891	0.8273
				100, 100, 100, 100	0.9999	0.9825
Scenario #3						
$G(3.0, 1.00)$	$G(4.0, 0.80)$	$G(3.0, 1.0)$	$G(5.0, 0.60)$	20, 20, 20, 20	0.9379	0.6434
$G(3.1, 1.00)$	$G(4.1, 0.70)$	$G(3.1, 1.0)$	$G(5.1, 0.65)$	50, 50, 50, 50	0.9999	0.9683
				100, 100, 100, 100	1.0000	0.9999
Scenario #4						
$G(3.0, 1.00)$	$G(4.0, 0.80)$	$G(3.0, 1.00)$	$G(6.0, 0.60)$	20, 20, 20, 20	0.9996	0.9682
$G(3.1, 1.00)$	$G(4.1, 0.70)$	$G(3.1, 1.00)$	$G(6.1, 0.65)$	50, 50, 50, 50	1.0000	1.0000
				100, 100, 100, 100	1.0000	1.0000
Scenario #5						
$G(3.0, 1.00)$	$G(4.0, 0.80)$	$G(3.0, 1.0)$	$G(7.0, 0.65)$	20, 20, 20, 20	1.0000	0.9951
$G(3.1, 1.00)$	$G(4.1, 0.70)$	$G(3.1, 1.0)$	$G(6.1, 0.60)$	50, 50, 50, 50	1.0000	1.0000
				100, 100, 100, 100	1.0000	1.0000
Scenario #6						

criterion for several situations, when the purpose of the diagnostics is to test before ruling in and ruling out patients.

6. Illustration using the lung cancer subtypes study

We utilized data from a complete cohort comprising 203 patient samples, which included 139 lung adenocarcinomas (AD) that had 12 suspected metastases of extrapulmonary origin,

Table 10. Descriptions of the biomarkers, optimal cut-off points under tree-ordering.

Cold Standard classification	Biomarker	Mean	SD	Cut-off point (c) based on ETJ	ETJ	Cut-off point (c) Based on $TTKL_{discrete}(c)$	$TTKL_{discrete}(c)$
NL	Tetranectin*	0.0014	0.0006	0.0031	0.7153	0.00374	8.2460
AD		0.0151	0.0225				
COID		0.0097	0.0054				
SCLC		0.0166	0.0089				
SQ		0.0201	0.0575				
NL	Ficolin*	0.0012	0.0007	0.0030	0.8263	0.0038	10.0707
AD		0.0148	0.0093				
COID		0.0162	0.0105				
SCLC		0.0149	0.0069				
SQ		0.0290	0.0185				
NL	Achaete-scute	−7.4418	16.8061	−25.6764	0.0033	−0.8756	0.0354
AD		18.4704	60.4319				
COID		24.6353	27.6975				
SCLC		15.1000	26.7297				
SQ		13.3529	39.8	339			
NL	Insulinoma	−14.4394	18.6747	−32.7157	0.1021	−4.0783	0.2813
AD		4.4778	28.6397				
COID		349.0500	182.1132				
SCLC		710.1633	381.7100				
SQ		−10.5800	28.1101				
NL	Gastrin-releasing peptide*	−0.1148	0.2348	−0.0228	0.0834	0.0673	1.2564
AD		0.0420	0.2521				
COID		0.0360	0.0838				
SCLC		0.0127	0.0250				
SQ		0.2492	1.7868				
NL	CD55*	0.0026	0.0009	0.0038	0.3688	0.0064	2.0687
AD		0.0069	0.0070				
COID		0.0246	0.0171				
SCLC		0.0163	0.0096				
SQ		0.0281	0.0515				

Note: *The reciprocal of the original values is used in the analysis.

21 squamous (SQ) cell carcinoma cases, 20 pulmonary carcinoid (COID) tumors, and 6 small-cell lung cancers (SCLC), along with 17 normal lung (NL) samples to demonstrate our approach of tree ordering. More information on the clinical data for adenocarcinoma patients and details regarding the tumor and control samples can be found in SampleData.xls (Primary Dataset and Dataset A sheets), and in the supporting information published on the PNAS website (www.pnas.org) and at www.genome.wi.mit.edu/MPR/lung; see Bhattacharjee et al. (2001). We considered several biomarkers but chose seven that had better classification ability than others in the data (ficolin, tetranectin, insulinoma, achaete, gastrin, chromogranin, CD55, and thymidylate). Descriptive statistics and the cut-off point based on the ETJ of the six selected biomarkers are presented in Table 10.

As per the World Health Organization (WHO), lung cancer accounts for 11.6% of all cancer cases globally, making it the most commonly diagnosed cancer. In the United States, lung cancer is the second most commonly diagnosed cancer, following breast cancer in women and prostate cancer in men, and is responsible for the highest number of cancer-related deaths. It is worth noting that smoking is strongly linked to the prevalence of lung cancer and quitting smoking and adopting other measures to limit exposure to tobacco smoke can significantly reduce the risk of developing lung cancer.

Figure 2 shows plots of the empirical distribution of the six biomarkers of lung cancer data. Clearly, the graphs of the disease biomarkers for the disease subtypes show that the assumption of extended tree ordering is satisfied.

Table 10 shows that Ficolin has the highest information value among the biomarkers analysed, followed by Tetranectin and CD55. Furthermore, as shown in Table 11, all three biomarkers exhibit a higher capacity to rule in patients, regardless of the method used

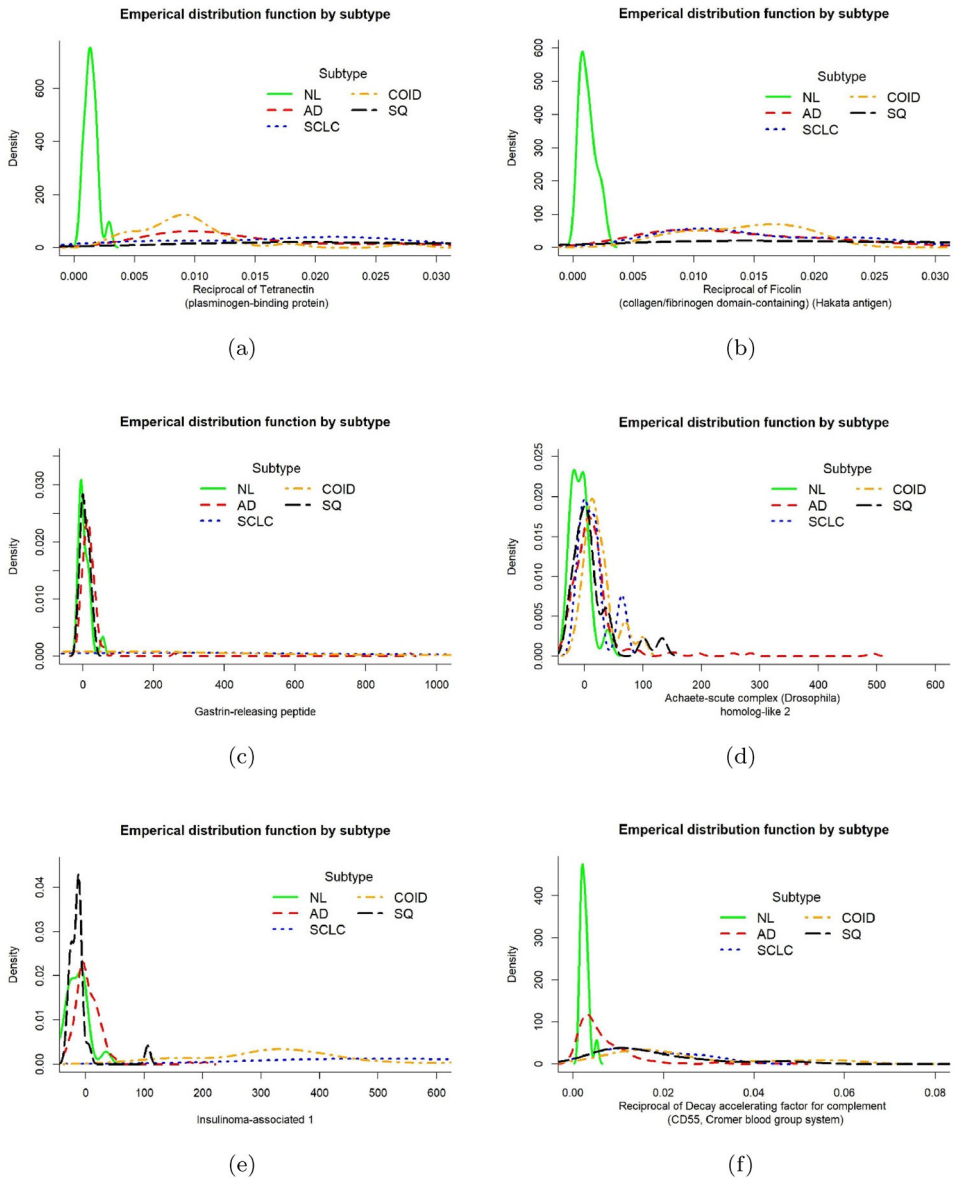

Figure 2. Plots of the empirical distribution of the six biomarkers of lung cancer data.

to select the optimal cut-off point. Conversely, the other three biomarkers (Achaete-scute, Insulinoma, and Gastrin-releasing) display a greater ability to rule out patients.

Additionally, Table 12 presents bootstrap 95% confidence intervals for ETJ and TTKL(c) across various cut-off point selection methods. Notably, the use of ETJ for optimal cut-off point selection results in a shorter 95% confidence interval for both ETJ and TTKL(c), indicating a more precise estimation.

Furthermore, $\text{TTKL}_{\text{discrete}}(c)$ remains non-negative and interpretable in these scenarios, offering a stable criterion for cut-off selection. $\text{TTKL}_{\text{discrete}}(c)$-based cut-offs could lead to

Table 11. Estimation of the proposed measures for the seven selected biomarkers and their bootstrap (standard error).

	Using ETJ cut-off point					Using $TTKL_{discrete}(c)$ cut-off point				
Biomarker	TSp	TSe	ID_{in}	ID_{out}	$TTKL_{discrete}(c)$	TSp	TSe	ID_{in}	ID_{out}	$TTKL_{discrete}(c)$
Tetranectin *	0.9877	0.7276	0.9270	0.7054	3.8392	0.9999	0.6773	0.0992	0.6772	8.2460
	(0.013)	(0.060)	(0.014)	(0.0679)	(3.8142)	(0.0005)	(0.0555)	(0.0066)	(0.0528)	(2.9479)
Ficolin *	0.9909	0.8354	0.9692	0.8241	5.2182	0.9999	0.7913	0.9998	0.7913	10.0707
	(0.048)	(0.095)	(0.0315)	(0.0831)	(1.3788)	(0.0000)	(0.0413)	(0.0157)	(0.0528)	(2.4074)
Achaete-scute	0.1353	0.8861	0.0005	0.0005	0.0010	0.6886	0.2282	0.0169	0.0181	0.0354
	(0.040)	(0.000)	(0.0000)	(0.0001)	(0.0000)	(0.0606)	(0.0641)	(0.0016)	(0.0592)	(0.1904)
Insulinoma	0.1712	0.9309	0.0444	0.0573	0.1085	0.7302	0.0784	0.1110	0.1508	0.2813
	(0.061)	(0.140)	(0.0092)	(0.1513)	(0.2078)	(0.0867)	(0.0651)	(0.0803)	(0.1792)	(0.2740)
Gastrin-releasing peptide*	0.6026	0.4808	0.0142	0.0140	0.0280	0.7758	0.0011	0.2189	0.6355	1.2564
	(0.095)	(0.095)	(0.0632)	(0.1691)	(0.1553)	(0.0789)	(0.0661)	(0.1803)	(0.2124)	(1.0526)
CD55*	0.9074	0.4616	0.3691	0.2774	0.1361	0.9998	0.2629	0.8288	0.2620	2.0687
	(0.061)	(0.068)	(0.2068)	(0.1160)	(0.6829)	(0.0004)	(0.0467)	(0.1147)	(0.0466)	(2.3871)

Table 12. Bootstrap 95% confidence intervals for the proposed measures of the seven selected biomarkers.

	Using ETJ cut-off point				Using $TTKL_{discrete}(c)$ cut-off point			
	ETJ		$TTKL_{discrete}(c)$		ETJ		$TTKL_{discrete}(c)$	
Biomarker	Lower	Upper	Lower	Upper	Lower (ETJ)	Upper (ETJ)	Lower	Upper
Tetranectin *	0.2375	0.4371	0.5734	9.2065	0.2370	0.4221	2.1762	12.0589
Ficolin *	0.3122	0.4760	1.1107	6.5197	0.3015	0.4772	3.3557	12.1088
Achaete-scute	0.0000	0.1606	0.0000	0.2886	0.0000	0.1162	0.0100	0.3968
Insulinoma	0.0000	0.2604	0.0019	0.7254	0.0000	0.0631	0.0060	0.9560
Gastrin-releasing*	0.0000	0.0857	0.0004	0.4798	−0.3559	−0.0837	0.3854	4.0931
CD55*	0.0385	0.4021	0.0654	2.3542	0.1856	0.3697	1.2468	9.2670

different clinical recommendations compared with those derived from AUC or the Youden index, depending on whether confirmatory diagnosis or screening is prioritized.

Another important finding is that using the cut-off point based on maximizing $TTKL_{discrete}(c)$, rather than other methods, such as maximizing $ETJ(c)$, consistently minimizes the loss of information when dichotomizing a continuous biomarker. For instance, in the case of marker Ficolin, we compare the information derived from two cut-off points: one by maximizing ETJ, and the other by maximizing $TTKL_{discrete}(c)$, as illustrated in Figure 2.

In Figure 3, the black dashed line represents Shannon entropy as a function of p (probability ranging from 0 to 1). The tangent lines are drawn at $p = 1 - \text{TSp}$ and $p = \text{TSe}$ at the cut-off points selected by $TTKL_{discrete}(c)$ or $ETJ(c)$, using the control or diseased population as the reference, respectively. The lengths of these dashed lines indicate the information content of the biomarker for diagnosis at the selected cut-off points. Specifically, the blue lines show $KL_{in} = 3.4722$ and $KL_{out} = 1.7375$ for the ETJ-based cut-off point, while the red lines show $TTKL_{in} = 6.6927$ and $TTKL_{out} = 1.5662$ for the $TTKL_{discrete}(c)$-based cut-off point.

Clearly, the cut-off point selected by $TTKL_{discrete}(c)$ provides more information, as evidenced by the greater lengths of the two dashed red lines ($6.627 + 1.5662 > 3.4722 + 1.7375$). This highlights the larger rule-in potential of the $TTKL_{discrete}(c)$-based cut-off compared to the ETJ-based cut-off for the same biomarker Ficolin∗. Specifically, a patient diagnosed as positive using the TTKL divergence cut-off point is approximately 25.041 times more likely to be a true positive ($\exp[6.6927 - 3.4722] = 25.041$). This makes the $TTKL_{discrete}(c)$ cut-off point provide a more specific test ($\text{TSp} = 0.9999$) with fewer false positives compared to the ETJ cut-off point. However, a non-diseased individual diagnosed as negative using

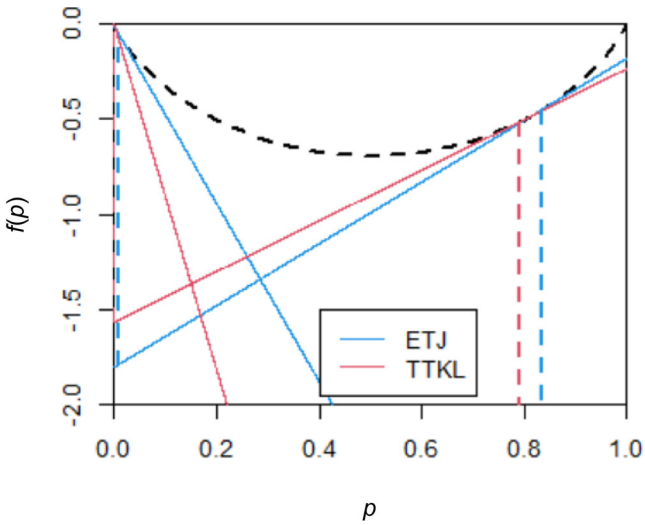

Figure 3. Information graph of selected marker (Ficolin∗). For ETJ, the cut-off point is at 0.003 with TSe = 0.8354 and TSp = 0.9909; $KL_{in} = 3.4722$, $KL_{out} = 1.7375$. For TTKL(c), the cut-off point is at 0.0038 with TSe = 0.9999 and TSp = 0.7913; $TTKL_{in} = 6.6927$, $TTKL_{out} = 1.5662$.

the $TTKL_{discrete}(c)$ cut-off point is about 16% less likely to be a true negative compared to the ETJ-based diagnosis ($\exp[1.5662 - 1.7375] = 0.84$). This indicates a lower sensitivity and more false negatives for the $TTKL_{discrete}(c)$ cut-off point. Thus, different cut-off point selection methods can lead to significantly different diagnostic performances. Unlike ETJ, which balances TSe and TSp, the $TTKL_{discrete}(c)$ criteria emphasize that marker Ficolin∗ has a greater overall rule-in potential and provides a more specific diagnosis at the cut-off point. Nonetheless, in terms of overall diagnostic accuracy, the sums based on $TTKL_{discrete}(c)$ and ETJ are quite similar, even though $TTKL_{discrete}(c)$ optimization is not aimed at maximizing the total true classifications.

7. Discussion

In this study, we examine the applications of Kullback-Leibler (KL) divergences in medical diagnostics, particularly their utility as an overall measure of pre-test rule-in and rule-out potential and as an optimization criterion for cut-off point selection under an extended tree or umbrella ordering ($TTKL_{discrete}(c)$). This finding is noteworthy because $TTKL_{discrete}(c)$ divergence measures provide a new perspective for diagnostic threshold selection by minimizing the information loss associated with categorizing a continuous biomarker for diagnostic purposes.

Typically, overall diagnostic measures such as $TTKL_{discrete}(c)$ divergence and the area under curve (AUC) are used in Phase I exploratory diagnostic accuracy studies to find potential biomarkers, which are generally intended for use as screening tests rather than confirmatory diagnostic tests. Our results showed that diagnostic tests capable of distinguishing diseased subtypes from non-diseased subtypes exhibit higher rule-in potential when they are more specific and better rule-out potential when they are more sensitive. This supports the established principle that specific tests are more effective for ruling in a condition, while sensitive tests are more effective for ruling out a condition.

We also linked $\text{TTKL}_{\text{discrete}}(c)$ divergence with common receiver operating characteristic (ROC) measures, providing both analytical and numerical relationships for scenarios involving a single cut-off point under extended tree ordering. Additionally, we discussed the graphical interpretation of $\text{TTKL}_{\text{discrete}}(c)$ divergence, known as the information graph, and illustrated this with real data.

Numerical examples using normal and gamma-distributed populations showed that sensitivity and specificity values based on maximizing $\text{TTKL}_{\text{discrete}}(c)$ are more disparate than using the Youden index. The Youden index maximizes the sum of sensitivity and specificity, achieving a more balanced trade-off between false positives and false negatives. In contrast, $\text{TTKL}_{\text{discrete}}(c)$ divergence minimizes information loss when categorizing a continuous biomarker but places less emphasis on balancing sensitivity and specificity. This imbalance likely arises because sensitivity and specificity are scaled by $\log(\text{OR}(c))$, which amplifies the dominant probability. However, the proper choice of diagnostic measure depends not only on its discriminative power but also on the intended use of the underlying biomarker. Therefore, we recommend using the $\text{TTKL}_{\text{discrete}}(c)$ measure for diagnostics focussed on pre-test rule-in and rule-out assessments.

In conclusion, $\text{TTKL}_{\text{discrete}}(c)$ divergence offers a valuable tool for optimizing diagnostic thresholds and enhancing the accuracy of medical diagnostics by minimizing information loss and providing robust measures of diagnostic performance.

Author contributions

CRediT: **Hani M. Samawi:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing; **Marwan Alsharman:** Data curation, Formal analysis, Investigation, Software, Validation, Writing – review & editing; **Jing Kersey:** Conceptualization, Data curation, Formal analysis, Investigation, Software, Visualization, Writing – review & editing

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Hani M. Samawi http://orcid.org/0000-0003-3817-2249

References

Benish, W. (2002). The use of information graphs to evaluate and compare diagnostic tests. *Methods of Information in Medicine*, *41*(2), 114–118. https://doi.org/10.1055/s-0038-1634294

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., E. J. Mark, Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., & Meyerson, M. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences*, *98*(24), 13790–13795. https://doi.org/10.1073/pnas.191502998

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, *7*(3), 200–217. https://doi.org/10.1016/0041-5553(67)90040-7

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., Melo, F. D. S. E., Missiaglia, E., Ramay, H., Barras, D., … Homicsko, K. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11), 1350–1356. https://doi.org/10.1038/nm.3967

Hughes, G. (2013). Information graphs for epidemiological applications of the Kullback-Leibler divergence. *Methods of Information in Medicine*, *53*(1), IV–VI.

Hughes, G., & Bhattacharya, B. (2013). Symmetry properties of bi-normal and bi-gamma receiver operating characteristic curves are described by Kullback-Leibler divergences. *Entropy*, *15*(4), 1342–1356. https://doi.org/10.3390/e15041342

Knottnerus, J. A., & Muris, J. W. (2003). Assessment of the accuracy of diagnostic tests: The cross-sectional study. *Journal of Clinical Epidemiology*, *56*(11), 1118–1128. https://doi.org/10.1016/S0895-4356(03)00206-3

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Lee, W. C. (1999). Selecting diagnostic tests for ruling out or ruling in disease: The use of the Kullback-Leibler distance. *International Journal of Epidemiology*, *28*(3), 521–525. https://doi.org/10.1093/ije/28.3.521

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press.

Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). McGraw-Hill Book.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine*. Little, Brown and Co.

Samawi, H., Alsharman, M., Keko, M., & Kersey, J. (2023). Post-test diagnostic accuracy measures under tree ordering of disease classes. *Statistics in Medicine*, *42*(28), 5135–5159. https://doi.org/10.1002/sim.v42.28

Samawi, H., Yin, J., Zhang, X., Yu, L., Rochani, H., Vogel, R., & Mo, C. (2020). Kullback-Leibler divergence for medical diagnostics accuracy and cut-point selection criterion: How it is related to the Youden index. *Journal of Applied Bioinformatics & Computational Biology*, *9*(2), 1–10. https://doi.org/10.37532/jabcb.2020.9(2).168

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Soofi, E. S., Ebrahimi, N., & Habibullah, M. (1995). Information distinguishability with application to the analysis of failure data. *Journal of the American Statistical Association*, *90*(430), 657–668. https://doi.org/10.1080/01621459.1995.10476560

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Wang, D., Attwooda, K., & Tiana, L. (2016). Receiver operating characteristic analysis under tree orderings of disease classes. *Statistics in Medicine*, *35*(11), 1907–1926. https://doi.org/10.1002/sim.v35.11

Wang, D., Feng, Y., Attwood, K., & Tian, L. (2019). Optimal threshold selection methods under tree or umbrella ordering. *Journal of Biopharmaceutical Statistics*, *29*(1), 98–114. https://doi.org/10.1080/10543406.2018.1489410