Meta-Analysis of mutual information applied in EBM diagnostics

Athanasios Tsalatsanis[1], Iztok Hozo[2], and Benjamin Djulbegovic[3]

[1]University of South Florida
[2]Indiana University Northwest
[3]City of Hope National Medical Center

July 2, 2020

Abstract

Rationale Assessing the performance of diagnostic tests requires evaluation of the amount of diagnostic uncertainty the test reduces (i.e. 0% - useless test, 100% - perfect test). Statistical measures currently dominating the evidence-based medicine (EBM) field and particularly meta-analysis (e.g. sensitivity and specificity), cannot explicitly measure this uncertainty reduction. Mutual information (MI), an information theory statistic, is a more appropriate metric for evaluating diagnostic tests as it explicitly quantifies uncertainty and, therefore, facilitates natural interpretation of a test's value. In this paper, we propose the use of MI as a single measure to express diagnostic test performance and demonstrate how it can be used in meta-analysis of diagnostic test studies. Methods We use two cases from the literature to demonstrate the applicability of MI meta-analysis in assessing diagnostic performance. These cases are: 1) Meta-analysis of studies evaluating ultrasonography (US) to detect endometrial cancer and 2) meta-analysis of studies evaluating magnetic resonance angiography to detect arterial stenosis. Results Results produced by the MI meta-analyses are comparable to the results of meta-analyses based on traditionally used statistical measures. However, the results of MI are easier to understand as it relates directly to the extent of uncertainty a diagnostic test can reduce. For example, a US test diagnosing endometrial cancer is 40% specific and 94% sensitive. The combination of these values is difficult to interpret and may lead to inappropriate assessment (e.g. one could favour the test due to its high sensitivity, ignoring its low specificity). In terms of MI however, the test reduces diagnostic uncertainty by 10%, which is marginal and thus the test is clearly not very useful. Conclusions We have demonstrated the suitability of MI in assessing the performance of diagnostic tests, which can facilitate easier interpretation of the true utility of diagnostic tests.

Meta-Analysis of mutual information applied in EBM diagnostics

Athanasios Tsalatsanis[1], Iztok Hozo[2]and Benjamin Djulbegovic[3,4,5]

[1]Department of Internal Medicine, University of South Florida, Tampa, FL

[2]Department of Mathematics, Indiana University Northwest, Gary, IN

[3] Department of Supportive Care Medicine, City of Hope, 1500 East Duarte Rd, Duarte, CA

[4]Department of Hematology, City of Hope, 1500 East Duarte Rd, Duarte, CA;

[5]Evidence-based Analytics & Program for Comparative Effectiveness Research and Evidence-based Medicine, 1500 East Duarte Rd, Duarte, CA;

[*]Corresponding author

Email addresses:

AT: atsalats@usf.edu

IH: ihozo@iun.edu

BD: bdjulbegovic@coh.org

Abstract

Rationale

Assessing the performance of diagnostic tests requires evaluation of the amount of diagnostic uncertainty the test reduces (i.e. 0% - useless test, 100% - perfect test). Statistical measures currently dominating the evidence-based medicine (EBM) field and particularly meta-analysis (e.g. sensitivity and specificity), cannot explicitly measure this uncertainty reduction. Mutual information (MI), an information theory statistic, is a more appropriate metric for evaluating diagnostic tests as it explicitly quantifies uncertainty and, therefore, facilitates natural interpretation of a test's value. In this paper, we propose the use of MI as a single measure to express diagnostic test performance and demonstrate how it can be used in meta-analysis of diagnostic test studies.

Methods

We use two cases from the literature to demonstrate the applicability of MI meta-analysis in assessing diagnostic performance. These cases are: 1) Meta-analysis of studies evaluating ultrasonography (US) to detect endometrial cancer and 2) meta-analysis of studies evaluating magnetic resonance angiography to detect arterial stenosis.

Results

Results produced by the MI meta-analyses are comparable to the results of meta-analyses based on traditionally used statistical measures. However, the results of MI are easier to understand as it relates directly to the extent of uncertainty a diagnostic test can reduce. For example, a US test diagnosing endometrial cancer is 40% specific and 94% sensitive. The combination of these values is difficult to interpret and may lead to inappropriate assessment (e.g. one could favour the test due to its high sensitivity, ignoring its low specificity). In terms of MI however, the test reduces diagnostic uncertainty by 10%, which is marginal and thus the test is clearly not very useful.

Conclusions

We have demonstrated the suitability of MI in assessing the performance of diagnostic tests, which can facilitate easier interpretation of the true utility of diagnostic tests.

Introduction

It is widely acknowledged that the purpose of diagnostic testing is to reduce diagnostic uncertainty (e.g. by 0%, if the test is useless , or up to 100%, when the test is perfect) [1]. However, the current metrics of diagnostic performance [i.e. sensitivity (S), specificity (C), positive and negative likelihood ratios (LR+; LR-), diagnostic odds ratio (DOR), and area under curve (AUC)] cannot provide a direct assessment of the amount by which diagnostic uncertainty is reduced. Despite lacking this crucial clinical usefulness, these "traditional" diagnostic metrics are widely used as the preferred evidence-based medicine (EBM) diagnostic test measures[2,3].

Meanwhile, there is a long tradition of quantifying diagnostic test performance in the field of information theory [4] . Although, conceptually speaking, the problems associated with medical diagnostic testing are similar to the problems faced in communication and information theory, for some reasons the field of EBM diagnostics has not embraced measures typically found in information theory.

One such measure, mutual information (MI) [5], used to evaluate association between two random variables, is considered the best metric to quantify diagnostic uncertainty and therefore test performance. [6] It has been used in a number of studies in medicine to explain the relationship between test results and disease states [7-14]. Yet it has been surpassingly missing from the EBM literature.

The most significant properties that establish MI as superior to traditional measures of diagnostic performance can be summarized as follows:

- MI quantifies the average amount of information that can be obtained about the value of a random variable (i.e. probability of disease before the diagnostic test) provided the value of another random variable is available (i.e. probability of disease after the diagnostic test) [15];
- MI quantifies the expected value of the amount of information a diagnostic test provides about the disease state, i.e. it takes into account all possible states that can be associated with the test results weighted by the likelihood of disease [16,17]. This number is particularly useful when comparing different diagnostic tests;
- MI summarizes test performance with a single meaningful number that corresponds to the average amount of information obtained by the diagnostic test and unlike the ROC it does not require a specified diagnostic cut-off point (threshold). The larger the MI value is, the greater the amount of diagnostic uncertainty reduced through the diagnostic test;
- MI can be applied to situations in which different test results are associated with different probabilities of disease[6,16];
- Unlike ROC and AUC, MI can be applied to a broad spectrum of testing situations ranging from the simple binary case (two test results and two disease states) to much more complicated situations in which a large number of test results (or a continuum of test results) are associated with multiple possible disease states[7-14];
- The maximal value of MI, formally referred to as channel capacity, can be used to identify the range of disease prevalence at which a diagnostic test is most useful;
- One way MI expresses information is in bits that range from 0 to infinity. In the simplest, binary case, where we are concerned if disease is present or not, the maximum number of bits is equal to 1[6];
- Finally, the relative expression of MI indicates the percentage of diagnostic information that can be reduced by a diagnostic test and it can range from 0% (a useless test) to 100% (a perfect test).

In this paper, we promote the notion that MI is a better measure for evaluation of diagnostic performance [8], both on theoretical and practical grounds. We extent the current work by explaining how MI can be meta-analyzed, and provide two illustrative examples of diagnostic test meta-analysis using MI.

Methods

Mutual information and diagnostic testing primer

Assume that a test (T) is used to examine whether a disease (D) is present in a group of N patients. For a diagnostic test, the values of specificity, sensitivity as well as the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results depend on whether the test turns out to be positive ($T+$), with probability t , or negative($T-$), and whether the disease is present($D+$), with probability p , or absent($D-$). To assist the reader, Table 1 summarizes the calculations of specificity, sensitivity, TP, TN, FP, and FN. Unabridged derivations are presented in the appendix.

The uncertainty of the state of disease *prior* to performing the diagnostic test is best expressed as entropy [4,15,18]:

$$H(D) = -(p\ p + (1-p)\ (1-p)),$$

where p is the probability of disease. The uncertainty due to the test is:

$$H(T) = -(\ \mathrm{t}\ t + (1-t)\ (1-t)),$$

where t is the probability of disease estimated by the diagnostic test T.

The MI is computed as:

$$I\left(D,T\right) = H\left(D\right) + H\left(T\right) - H\left(D,T\right).$$

where $H\left(D,T\right)$ is the joint entropy of disease and diagnostic test. MI can also be expressed in terms of the conditional entropy as well as the conditional probabilities of every test/disease outcome combination:

$$H\left(D|T\right) = H\left(D,T\right) - H\left(T\right)$$

Hence, the mutual information is also defined as:

$$I\left(D,T\right) = H\left(D\right) - H\left(D|T\right)$$

From the latter expression it is evident that MI explicitly describes the amount of diagnostic uncertainty that can be reduced by the diagnostic test. Clinically, it is particularly useful to express MI in relative terms, as it can indicate explicitly the percentage of diagnostic uncertainty a diagnostic test can reduce. Relative MI (RMI) is defined as:

$$I_R\left(D,T\right) = \frac{I\left(D,T\right)}{H\left(D\right)} = 1 - \frac{H(D|T)}{H(D)}$$

The quantity $\frac{H(D|T)}{H(D)}$, is the relative entropy associated with the test result (i.e. the percentage of uncertainty reduced by the test result).

Interpretation of uncertainty reduction

The amount of reduction of uncertainty defines the usefulness of a diagnostic test. Ideally, it would be defined in decision analytic context when the "useful" test is the one that affect our decisions and its downstream consequences. This, however, require case-specific decision modeling, which is not the focus of this paper. Alternatively, usefulness of a test can be defined according to magnitude of reduction of diagnostic uncertainty analogously to treatment effects as small, moderate or large[19]. Thus, we define small reduction of diagnostic uncertainty if it is less than 10%, moderate reduction between 20-30% and large reduction of diagnostic uncertainty if it exceeds 40-50%.

Sensitivity and specificity

As the majority of diagnostic studies express diagnostic performance results in terms of sensitivity (S) and specificity (C), we show how MI relates to these measures.

$$I\left(D,T\right) = H\left(D\right) + H\left(T\right) - H\left(D,T\right) = Sp\left(\log_2\left(\frac{S\left(\left(1-S\right)p + C\left(1-p\right)\right)}{\left(1-S\right)\left(Sp + \left(1-C\right)\left(1-p\right)\right)}\right)\right) + C\left(1-p\right)\left(\log_2\left(\frac{C\left(Sp + \left(1-C\right)\left(1 \right.\right.}{\left(1-C\right)\left(\left(1-S\right)p + C \right.}\right.\right.$$

Meta-Analysis of entropy and mutual information

In most cases, decision-makers are not interested in evaluating the performance of a diagnostic test in a single study. Instead, they would like to know the totality of evidence generated in a series of studies evaluating the particular test. In such cases, a meta-analysis of summary statistics is employed.

Meta-analysis is initiated with the computation of a summary statistic for each study [20]. In our case, this summary statistic is the value of MI associated with the diagnostic test under investigation. The next step in meta-analysis is to compute the weighted average of MI, where the weights used are typically the inverse

of the MI variance, which is related to sample size [20]. According to Roulston [21], the variance of the entropy is given by

$$\mathrm{Var}\left(H\left(D\right)\right) = \left[\left(p + H\left(D\right)\right)^2 + \left(\left(1-p\right) + H\left(D\right)\right)^2\right] \cdot \frac{p\left(1-p\right)}{N}$$

which is valid for study sample size greater than 10.

Solving for the variance of MI we derive the expression:

$$\mathrm{Var}\left(I\left(D,T\right)\right) = \left(\left(p_{11}+p_{12}\right) + \left(p_{11}+p_{21}\right) - p_{11} + I\left(D,T\right)\right)^2 \left(\frac{p_{11}\left(1-p_{11}\right)}{N}\right) \backslash n + \left(\left(p_{11}+p_{12}\right) + \left(p_{12}+p_{22}\right) - p_{12} + I\left(D,T\right)\right.$$

See, table 1 for definitions of $p_{11},\ p_{12},\ p_{21}$, andp_{22}. Unabridged derivations are presented in the appendix. Numerical examples of these derivations are shown in Table 2.

Results

We present the application of MI meta-analysis based on two cases previously published in literature.

Case 1. Detection of endometrial cancer using endovaginal ultrasonography (US).

For this case, we used data presented in Deeks [22], originally published in Smith-Bindman et al [23]. The dataset is the result of a systematic review process on 35 papers presenting the diagnostic performance of endovaginal US in the detection of endometrial cancer. Evidence synthesis tables on test sensitivity and specificity are provided in Deeks [22].

Figure 1 displays the meta-analytic summary plots based on US studies. It includes the summary ROC curve, individual study estimate, and summary point estimate of the "traditional" measures of performance of endovaginal US in the detection of endometrial cancer. It is difficult to interpret, how "good" the test is, and in particular how much uncertainty the test reduced in each study where US was evaluated. For example, a US test diagnosing endometrial cancer is 40% specific and 94% sensitive. The combination of these values is difficult to interpret and may lead to inappropriate assessment (e.g. one could favour the test due to its high sensitivity, ignoring its low specificity). In terms of MI however, the test reduces diagnostic uncertainty by 10%, which is marginal and thus the test is clearly not very useful.

Figure 2 demonstrates meta-analysis of MI. We can clearly see that the US results provided only 0.05 (0.04 to 0.07) bits of information (recall, that the maximum amount information in the binary diagnostic case is 1). Although this gives us an estimate about overall diagnostic performance of US for diagnosis of endometrial cancer, what we really want to know is the amount of diagnostic uncertainty the US can possibly reduce (on scale 0 to 100%). This can be expressed by calculating RMI.

Figure 3a shows the performance of US expressed in terms of RMI. The information presented is much clearer: a decision-maker has much better understanding on how much diagnostic uncertainty was reduced in each study. The pooled estimate of the reduction in diagnostic uncertainty is 13% for pre-test probability of disease 14%. That is, US can reduce the uncertainty related to *endometrial cancer* by 13%. Figure 3b presents the sample size of each study.

Case 2: Contrast-enhanced magnetic resonance angiography (MRA) for arterial stenosis disease

In this case, we use meta-analysis data from the study of Menke and Larsen[24] summarizing evidence about how well MRA detects arterial stenosis. A total of 32 studies were included in the analysis.

Figure 4 depicts the meta-analytic summary plots. It includes the summary ROC curve, individual study estimate, and summary point estimate of the "traditional" measures of performance of MRA in diagnosis of arterial stenosis. As with figure 1, the interpretation of traditional statistics in terms of test performance

is difficult. For example, an MRA test diagnosing arterial stenosis is 96% specific and 78% sensitive. The combination of these values is difficult to interpret and may lead to inappropriate assessment. In terms of MI however, the test reduces diagnostic uncertainty by 49%, which indicates a clearly useful test.

Figure 5, demonstrates the meta-analysis of MI, in which it is shown that the information content of MRA in diagnosis of arterial stenosis is 0.53 (CI: 0.48, 0.57).

Figure 6 depicts the RMI for reported by each study as well as the pool estimate, which is approximately 67% for pre-test probability of disease 25%. That is, the MRA reduces uncertainty related to arterial stenosis by 67%.

Discussion

Many authors have outlined a number of problems with the use of "traditional" measures of diagnostic performance[6,16,25-27]. These problems relate to the biases that plague studies evaluating diagnostic studies, and to the metrics themselves [28]. In this paper, we focus on the latter. In particular, we focus on the measurement of diagnostic accuracy as opposed to the impact of diagnostic tests on health outcomes, which depends on consideration of down-stream effects of testing such as the choice of treatment and will not be considered here.

With regards to diagnostic accuracy, it has been argued[6,8,29,30] that utilization of information theory, and particularly MI, has theoretical and practical advantages over the traditional measures at assessing the performance of a diagnostic test. Notably, MI and RMI can be used to explicitly quantify the amount of diagnostic uncertainty a test reduces. Such a direct measure can easily be used to evaluate test performance not only by trained researchers but also by any EBM literate practitioner. Here, we summarized the MI advantages over traditional measures and demonstrated how MI can be meta-analyzed using two cases from the literature.

The MI meta-analysis results presented in both cases show the superiority of MI and RMI over other metrics in conveying arguably the most useful clinical indicators of diagnostic test performance, namely the amount of diagnostic uncertainty reduced by the test. Clearly, consideration of other ethical and personal dilemmas is also involved in the administration of a diagnostic test. However, for the EBM community and the evidence synthesis practitioners , reduction of uncertainty is of outmost importance. In terms of derivation, MI is easily computed and meta-analyzed. In addition, although we have not emphasized it here, MI has particular advantages over other metrics when it comes to analysis of tests with continuous measurements such as PSA, blood pressure etc. Analysis of such tests with traditional metrics requires dichotomization of the test results discarding useful information [31]. On the other hand, MI can be computed both for discrete and continuous variables [32].

One limitation of MI is its reliance on prevalence, which even though represents theoretical advantages it introduces heterogeneity in meta-analysis. To solve this problem, we propose meta-analyzing RMI instead of MI, but at this time we know of no derivation of standard error for RMI. Further development in the field of research synthesis of diagnostic test performance may lie in the opportunity to develop robust meta-analytic techniques for RMI.

In summary, we believe that MI is the most meaningful measure for both decision makers and EMB researchers as it provides intuitive, easy to understand metrics that quantify diagnostic tests information content. We therefore, argue that the field of evidence-based diagnostics should adopt MI as its most useful metric.

References

1. Sox HC, Blatt MA, Higgins MC, Marton MC. Medical Decision Making. Boston: Butterworths; 1988.

2. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. Syst Rev 2013;2:82.

3. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working G. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889-97.

4. Shannon CE, Waever W. The mathematical theory of communication. Urbana: The University of Illinois Press; 1962.

5. Shannon C. A mathematical theory of communication, bell System technical Journal 27: 379-423 and 623–656. Mathematical Reviews (MathSciNet): MR10, 133e 1948.

6. Benish WA. Intuitive and axiomatic arguments for quantifying diagnostic test performance in units of information. Methods Inf Med 2009;48:552-7.

7. Somoza E, Mossman D. Comparing and optimizing diagnostic tests: an information-theoretical approach. Med Decis Making 1992;12:179-88.

8. Benish W. Mutual information as an index of diagnostic test performance. Methods of information in medicine 2003;42:260-4.

9. Mossman D, Somoza E. Diagnostic tests and information theory. J Neuropsychiatry Clin Neurosci 1992;4:95-8.

10. Somoza E, Soutullo-Esperon L, Mossman D. Evaluation and optimization of diagnostic tests using receiver operating characteristic analysis and information theory. International journal of bio-medical computing 1989;24:153-89.

11. Benish W. The use of information graphs to evaluate and compare diagnostic tests. Methods of information in medicine 2002;41:114-8.

12. Nelson GW, O'Brien SJ. Using mutual information to measure the impact of multiple genetic factors on AIDS. JAIDS Journal of Acquired Immune Deficiency Syndromes 2006;42:347-54.

13. Meyer CR, Boes JL, Kim B, et al. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. Medical image analysis 1997;1:195-206.

14. Diamond GA, Hirsch M, Forrester JS, et al. Application of information theory to clinical diagnostic testing. The electrocardiographic stress test. Circulation 1981;63:915-21.

15. Cover TM, Thomas JA. Elements of information theory: John Wiley & Sons; 2012.

16. Hughes G. Application of Information Theory to Epidemiology: American Phytopathological Society; 2012.

17. Hughes G, McRoberts N. The structure of diagnostic information. Australasian Plant Pathology 2014:1-20.

18. Djulbegovic B, Hozo I, Abdomerovic I, Hozo S. Diagnostic entropy as a function of therapeutic benefit/risk ratio. Med Hyoptheses 1995;45:503-9.

19. Djulbegovic B, Glasziou P, Chalmers I. The importance of randomised vs non-randomised trials. The Lancet 2019;394:634-5.

20. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. Systematic Reviews in Health Care: Meta-Analysis in Context, Second Edition 2001:285-312.

21. Roulston MS. Estimating the errors on measured entropy and mutual information. Physica D: Nonlinear Phenomena 1999;125:285-94.

22. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ 2001;323:157-62.

23. Smith-Bindman R, Kerlikowske K, Feldstein VA, et al. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. JAMA 1998;280:1510-7.

24. Menke J, Larsen J. Meta-analysis: Accuracy of contrast-enhanced magnetic resonance angiography for assessing steno-occlusions in peripheral arterial disease. Ann Intern Med 2010;153:325-34.

25. Knottnerus JA. The evidence base of clinical diagnosis. London: BMJ Books; 2002.

26. Hilden J. The area under the ROC curve and its competitors. Med Decis Making 1991;11:95-101.

27. Lee WC, Hsiao CK. Alternative summary indices for the receiver operating characteristic curve. Epidemiology 1996;7:605-11.

28. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD Statement for reporting of studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7-18.

29. Benish WA. Relative entropy as a measure of diagnostic information. Medical decision making 1999;19:202-6.

30. Wu Y, Alagoz O, Ayvaci MU, et al. A comprehensive methodology for determining the most informative mammographic features. Journal of digital imaging 2013;26:941-7.

31. Shapiro DE. The interpretation of diagnostic tests. Stat Methods Med Res 1999;8:113-34.

32. Ross BC. Mutual Information between Discrete and Continuous Data Sets. PloS one 2014;9:e87357.

Appendix - Unabridged derivations of MI, RMI and Var(MI)

Entropy is expressed as:

$$H\left(D\right) = -\left(P(D+)\ P(D+) + (1 - P(D+))\ (1 - P(D+))\right)$$

The uncertainty due to the diagnostic test is:

$$H\left(T\right) = -\left(\ P(D+|T+)\ P(D+|T+) + (1 - P(D+|T+))\ (1 - P(D+|T+))\right)$$

The mutual information is computed as:

$$I\left(D,T\right) = H\left(D\right) + H\left(T\right) - H\left(D,T\right) = H\left(D\right) - H\left(D|T\right)$$

The relative mutual information is computed as:

$$I_R\left(D,T\right) = \frac{I\left(D,T\right)}{H\left(D\right)} = 1 - \frac{H(D|T)}{H(D)}$$

In terms of sensitivity and specificity, mutual information is derived as:

$$I\left(D,T\right) = H\left(D\right)+H\left(T\right)-H\left(D,T\right) = P\left(T+|D+\right)P(D+)\left(\log_2\left(\frac{P\left(T+|D+\right)\left(\left(1-P\left(T+|D+\right)\right)P\left(D+\right)+P\left(T-|D-\right)\left(\right.\right.\right.$$

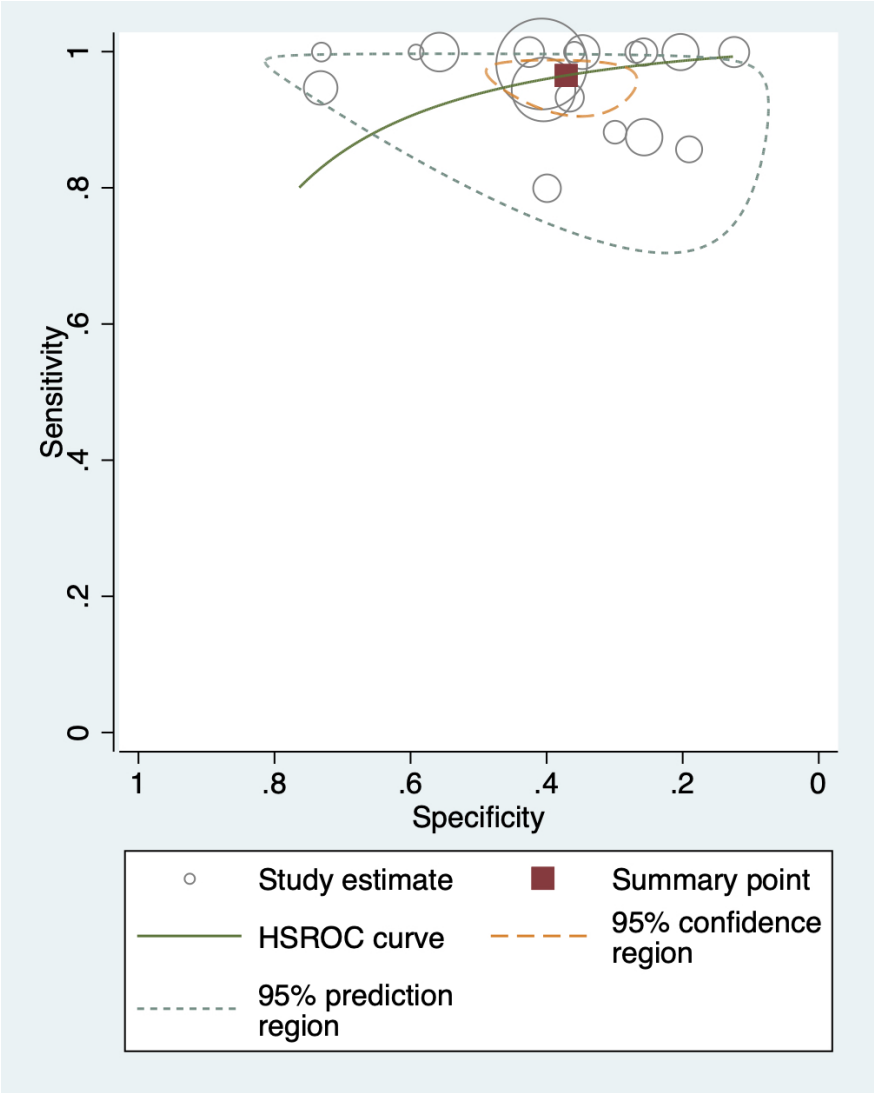
$$\left.\left.\left.\frac{}{\left(1-P\left(T+|D+\right)\right)\left(P\left(T+|D+\right)P\left(D+\right)+\left(1-P\left(T-|D-\right)\right.\right.\right.\right.$$

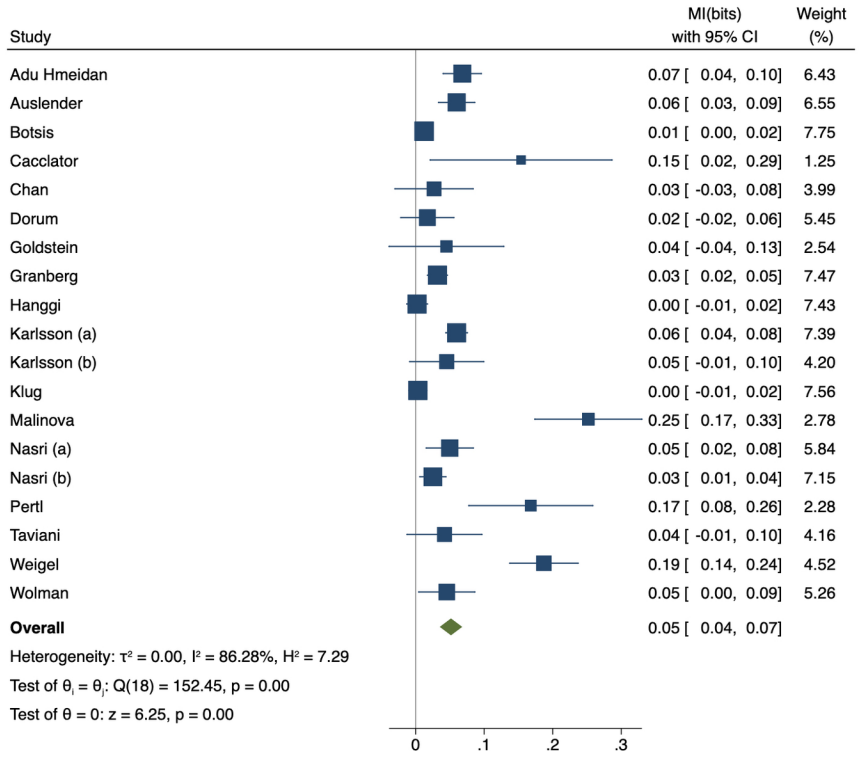
The variance of mutual information is computed as:

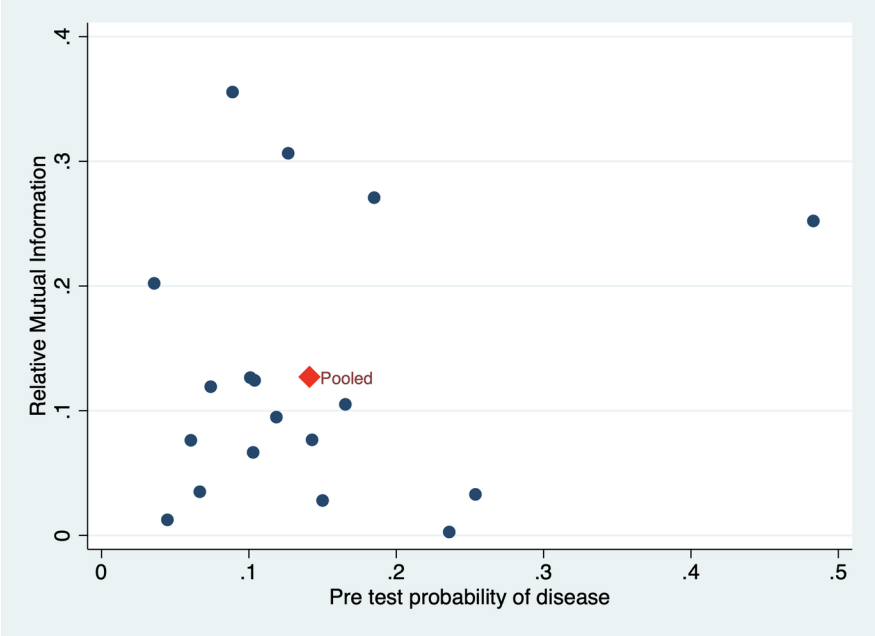
$$\mathrm{Var}\left(H\left(D\right)\right) = \left[\left(P(D+) + H\left(D\right)\right)^2 + \left(\left(1 - P(D+)\right) + H\left(D\right)\right)^2\right]\frac{P(D+)\left(1 - P(D+)\right)}{N}$$

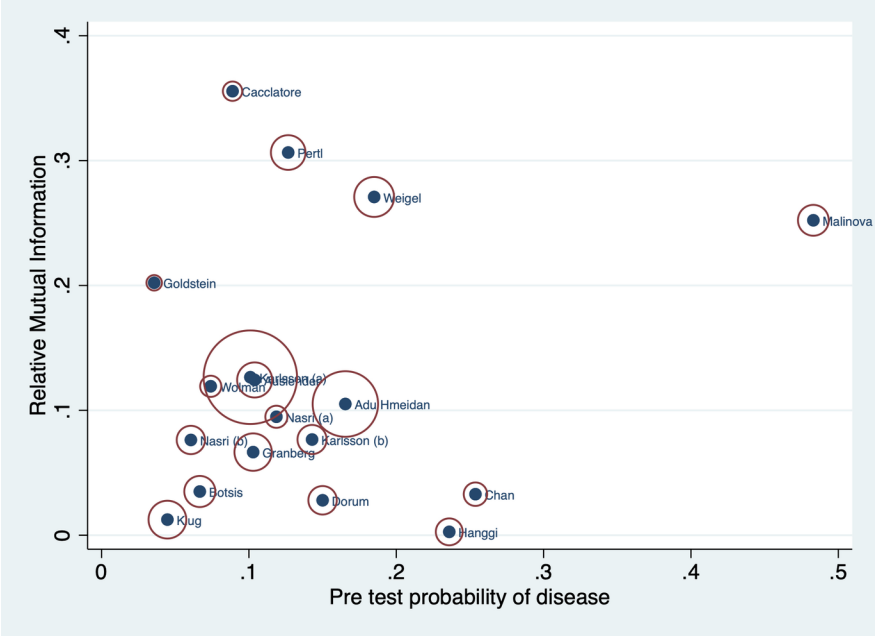
and:

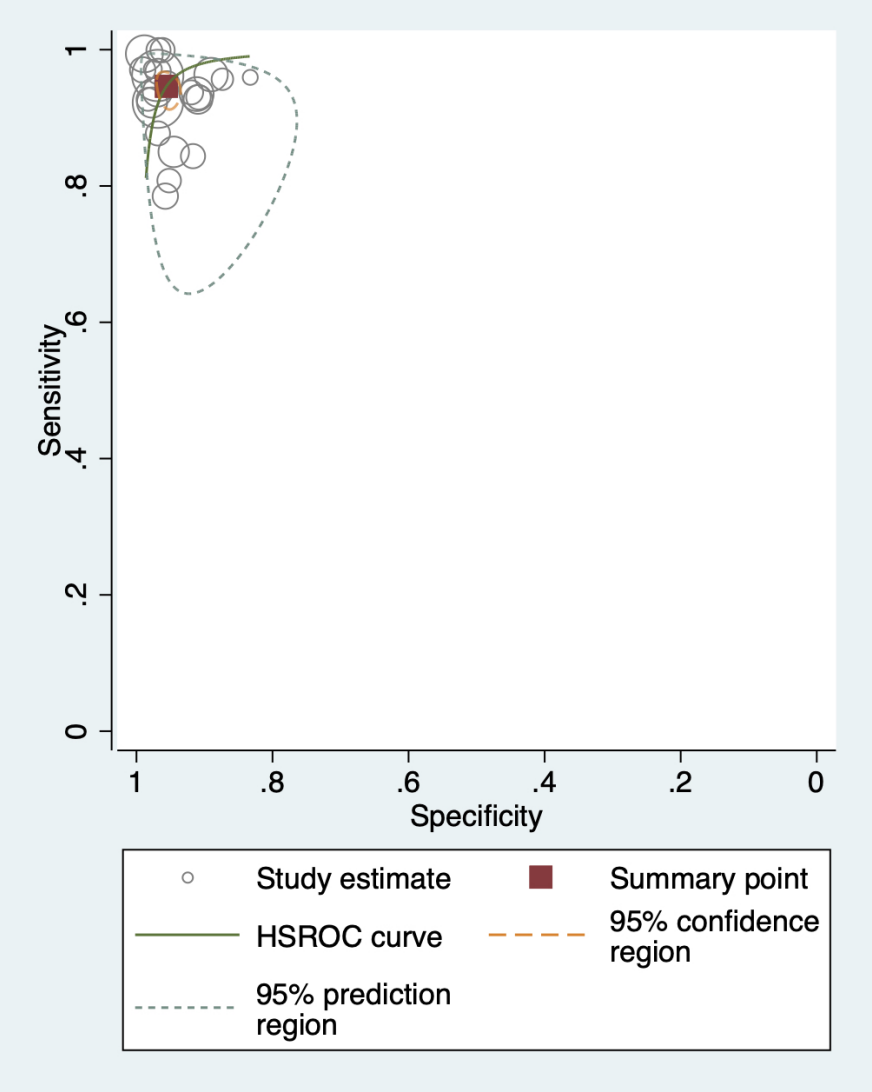
$$\mathrm{Var}\left(I\left(D,T\right)\right) = \left(\left(P\left(T+|D+\right)P(D+) + \left(1-P\left(T+|D+\right)\right)P(D+)\right) + \left(P\left(T+|D+\right)P(D+) + \left(1-P\left(T-|D-\right)\right)\left(1-P\left(D+\right)\right)\right.\right.$$

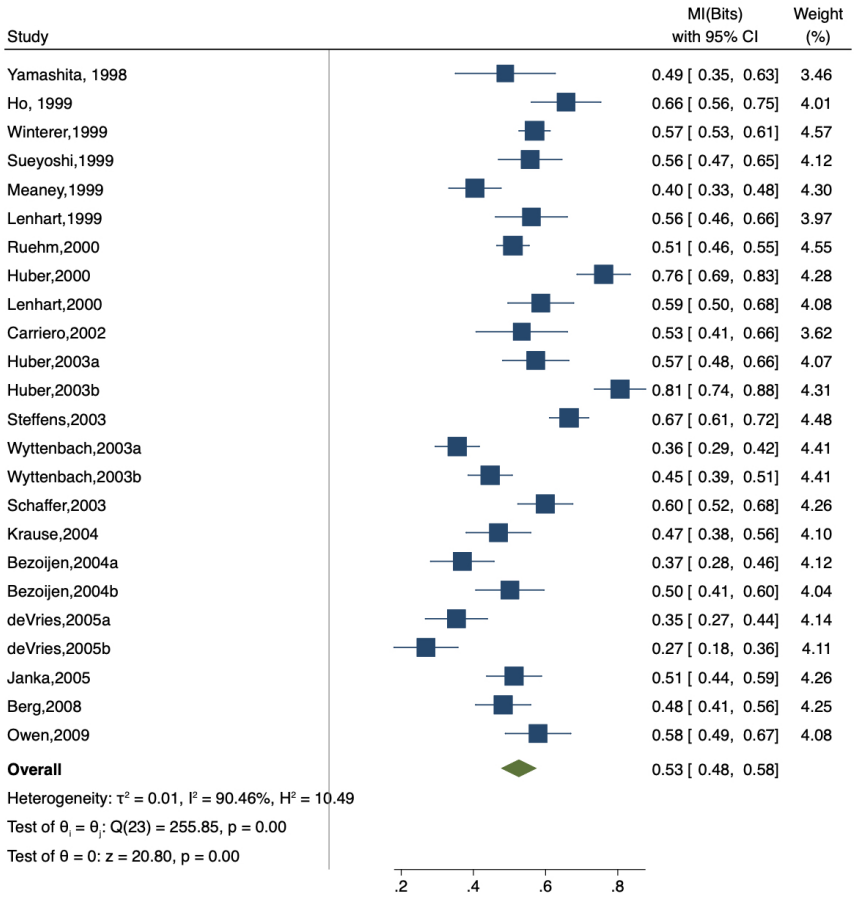


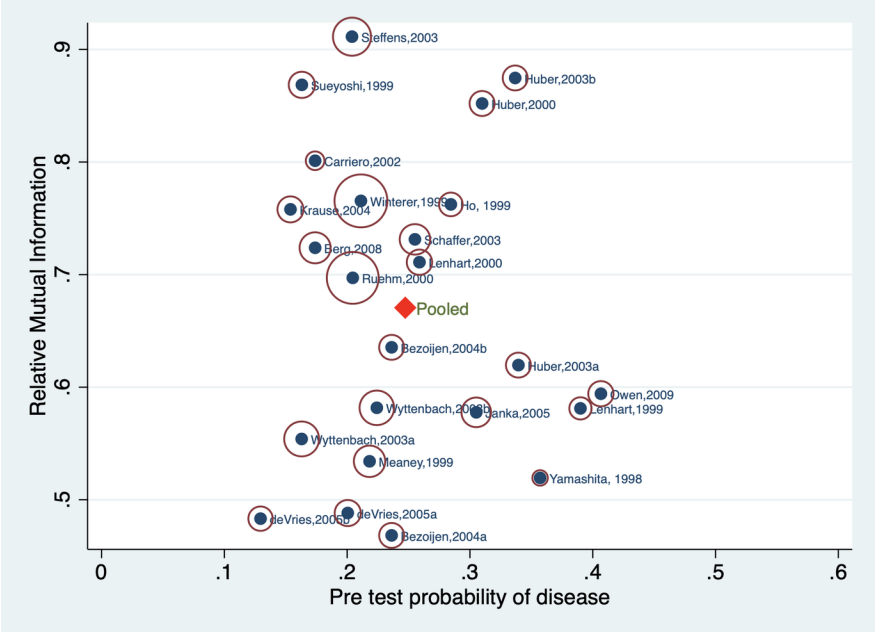












Hosted file

`Table 1.docx` available at https://authorea.com/users/339139/articles/465391-meta-analysis-of-mutual-information-applied-in-ebm-diagnostics

Hosted file

`Table 2.docx` available at https://authorea.com/users/339139/articles/465391-meta-analysis-of-mutual-information-applied-in-ebm-diagnostics