diagnostics

MDPI

Article

On the Dependence of the Critical Success Index (CSI) on Prevalence

Gashirai K. Mbizvo [1,2,3,*] and Andrew J. Larner [3]

[1] Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool L69 7BE, UK
[2] Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool L14 3PE, UK
[3] Cognitive Function Clinic, The Walton Centre NHS Foundation Trust, Liverpool L9 7LJ, UK; ajlarner241@aol.com
* Correspondence: gashirai.mbizvo@liverpool.ac.uk

Abstract: The critical success index (CSI) is an established metric used in meteorology to verify the accuracy of weather forecasts. It is defined as the ratio of hits to the sum of hits, false alarms, and misses. Translationally, CSI has gained popularity as a unitary outcome measure in various clinical situations where large numbers of true negatives may influence the interpretation of other, more traditional, outcome measures, such as specificity (Spec) and negative predictive value (NPV), or when unified interpretation of positive predictive value (PPV) and sensitivity (Sens) is needed. The derivation of CSI from measures including PPV has prompted questions as to whether and how CSI values may vary with disease prevalence (P), just as PPV estimates are dependent on P, and hence whether CSI values are generalizable between studies with differing prevalences. As no detailed study of the relation of CSI to prevalence has been undertaken hitherto, the dataset of a previously published test accuracy study of a cognitive screening instrument was interrogated to address this question. Three different methods were used to examine the change in CSI across a range of prevalences, using both the Bayes formula and equations directly relating CSI to Sens, PPV, P, and the test threshold (Q). These approaches showed that, as expected, CSI does vary with prevalence, but the dependence differs according to the method of calculation that is adopted. Bayesian rescaling of both Sens and PPV generates a concave curve, suggesting that CSI will be maximal at a particular prevalence, which may vary according to the particular dataset.

Keywords: Bayes formula; binary classification; critical success index; F measure; prevalence

Citation: Mbizvo, G.K.; Larner, A.J. On the Dependence of the Critical Success Index (CSI) on Prevalence. *Diagnostics* **2024**, *14*, 545. https://doi.org/10.3390/diagnostics14050545

Academic Editor: Francesco Sessa

Received: 26 January 2024
Revised: 28 February 2024
Accepted: 1 March 2024
Published: 5 March 2024

1. Introduction

The context of this paper is that many measures may be derived from the data cells in a 2 × 2 contingency table, which is used as the basis for evaluating any binary classification such as the outcome of a screening or diagnostic test accuracy study or a case-ascertainment algorithm [1]. Choosing the optimal measure(s) to describe the outcomes of a study may be dependent upon the nature of the available dataset.

For datasets with very large numbers of true negative (TN) outcomes in the base data, as seen for an example using routine epilepsy data [2–31], indices such as specificity (Spec), negative predictive value (NPV), and overall classification accuracy (Acc), which all feature TN values in both numerator and denominator, may be very high, indeed approaching values of 1. This is because the numbers of TN may approach the total number of observations (N), and hence swamp the values of the other cells of the 2 × 2 contingency table, namely, true positive (TP), false positive (FP), and false negative (FN).

This circumstance makes it difficult to rank the diagnostic accuracy of the corresponding case-ascertainment algorithms based on Spec, NPV, or Acc, as the figures are all similarly high [32]. In conditions such as dementia [33], motor neuron disease [34], and

epilepsy [35,36], systematic reviews of the diagnostic accuracy of routine data indicate that the original studies published have largely measured the positive predictive value (PPV) and sensitivity (Sens) without measuring Spec or NPV. This is because finding true negative cases in the community to verify an absent diagnostic code in a routine dataset is a challenge for researchers, who often only have permission to study populations that have been positively coded with the disease in question. Making a judgment on the optimal case-ascertainment algorithm for a particular condition based on either PPV or Sens is challenging because PPV and Sens tend to have an inverse relationship [37], so it is difficult to know which measure to prioritize to best indicate accuracy.

There are other examples in clinical medicine where large numbers of TN may complicate the interpretation of more traditional measures such as PPV and Sens, including National Institute for Clinical Excellence criteria for 2-week-wait suspected brain and CNS cancer referrals [38], polygenic hazard scores [39], and the evaluation of cognitive screening instruments [40]. Accordingly, as we have previously indicated, a metric is needed which eschews TN and combines PPV and Sens.

As we are not aware of such a metric currently in common use in medicine, we have proposed the use of the critical success index (CSI) for this purpose. This measure, which has been intermittently reinvented over the last century, has been variously known as the ratio of verification in the context of forecasting tornadoes in meteorology [41–50], and subsequently as the Jaccard index or similarity coefficient (J) [51], the threat score [52], the Tanimoto index [53], CSI [53,54], and most recently as F* [55].

In signal detection theory, CSI is defined as the ratio of hits to the sum of hits, false alarms, and misses [40,56]. In terms of the base data of the 2 × 2 contingency table [1]:

$$\mathrm{CSI} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$$

$$= \mathrm{TP}/(\mathrm{N} - \mathrm{TN})$$

CSI may also be expressed in terms of PPV and Sens [1]:

$$\mathrm{CSI} = 1/[(1/\mathrm{PPV}) + (1/\mathrm{Sens}) - 1] \quad (1)$$

We have demonstrated the advantages of using CSI to complement traditional diagnostic accuracy measures using real-word data in several conditions [32,40,57].

It should be noted that CSI differs from, but is related to, another measure sometimes used for similar purposes of data analysis, which is variously called the Dice coefficient, the Sørensen-Dice coefficient, or the F measure [58–62], defined as:

$$\mathrm{F} = 2.\mathrm{TP}/(2.\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$$

$$= 2/[1/\mathrm{Sens} + 1/\mathrm{PPV}]$$

There is a monotonic relationship between CSI and F [63], such that:

$$\mathrm{F} = 2\mathrm{CSI}/(1 + \mathrm{CSI})$$

A question often raised about CSI concerns how its values relate to prevalence, P, the probability of a positive diagnosis. As we are not aware of any previous examination of this question, it merits further investigation. It is well-known that values of PPV vary with P, and thus are sensitive to class imbalance and may therefore not be generalizable between studies [63]. Since, as shown in Equation (1), CSI may be expressed in terms of PPV, a similar expectation will hold for CSI. Likewise, following from Equation (1), it may be asked whether CSI values track predominantly with Sens or PPV and whether this changes with P.

Here, we initially address two possible methods to illustrate the dependence of CSI on P, as previously suggested [57]: firstly, using the Bayes formula to recalculate PPV and then to recalculate CSI (hence, a two-step method); and secondly, using equations in which

CSI is expressed directly in terms of Sens, PPV, P, and the test threshold or probability of a positive test, denoted as Q. In addition, we introduce a third method in which Sens is also rescaled by using the Bayes formula to recalculate NPV and hence Sens. This then allows CSI values to be recalculated using both rescaled PPV and Sens.

2. Materials and Methods

2.1. Dataset

The public dataset from a published screening test accuracy study of a cognitive screening instrument [64], the Mini-Addenbrooke's Cognitive Examination (MACE) [65], was examined. In this study, MACE was administered to consecutive patient referrals (N = 755) to a dedicated cognitive disorders clinic located in a secondary care neuroscience center over the period of June 2014−December 2018 (inclusive). Other than those with a pre-existing diagnosis of dementia, there were no exclusion criteria. The diagnosis of dementia or mild cognitive impairment was made by the judgement of an experienced clinician using standard diagnostic criteria (DSM-IV; Petersen); in those without evidence of cognitive impairment, a diagnosis of subjective memory complaint (SMC) was made. MACE scores were not used to make criterion diagnoses in order to avoid review bias. Subjects gave informed consent, and the study was approved by the institute's committee on human research (Walton Centre for Neurology and Neurosurgery Approval: N 310).

In this cohort, 114 patients received a final criterial diagnosis (DSM-IV) of dementia (P = 0.151) [65]. The original analysis of the dataset established the optimal MACE cut-off for the diagnosis of dementia to be $\leq$20/30 (calculated from the maximal value for the Youden index), where TP = 104, FP = 188, FN = 10, and TN = 453. Hence, at this cut-off, Sens = 0.912, Spec = 0.707, PPV = 0.356, and NPV = 0.978.

From these base data, values of CSI across a range of P values (0.1 to 0.9, in 0.1 increments) were calculated using three different methods. As CSI is dependent on PPV and Sens (Equation (1)), it is appropriate to examine how its value changes with different methods of analysis, specifically how CSI changes with the change in PPV (Method 1), with the change in Sens (Method 2), and with the changes in both PPV and Sens (Method 3).

2.2. Method 1: CSI Recalculated via Bayes Formula for PPV

As Sens and Spec are relatively impervious to changes, in P, being strictly columnar ratios in the 2 × 2 contingency table, PPV may be recalculated for different values of P using the Bayes formula:

$$\text{PPV} = \text{Sens.P}/(\text{Sens.P}) + [(1 - \text{Spec}).\text{P}'] \quad (2)$$

where P′ = (1 − P). Using the base data (Sens = 0.912, Spec = 0.707), the values of PPV were calculated for P values ranging from 0.1 to 0.9.

The second step in this method used the recalculated PPV values at different prevalences to recalculate CSI values according to the relation to PPV and Sens (Equation (1)).

Thus, this approach requires the sequential application of Equations (1) and (2) to the base data. Results are displayed in a table and graphically.

2.3. Method 2: CSI Recalculated via Its Relation to Sens, PPV, P, and Q

The dependence of CSI on P, and the probability of a positive diagnosis, may be directly expressed in terms of Sens, PPV, P, and test threshold, the probability of a positive test (Q) [1]:

$$\text{CSI} = 1/[(\text{P} + \text{Q})/\text{Sens.P}] - 1 \quad (3)$$

$$\text{CSI} = 1/[(\text{P} + \text{Q})/\text{PPV.Q}] - 1 \quad (4)$$

Hence, the dependence of CSI on P may be addressed by calculating its value for different values of P at chosen values of Q. Q ranges from 0−1, where Q = 0 equates to a test threshold at which there are no positives (neither TP nor FP), and Q = 1 equates to a

threshold at which there are no negatives (neither TN nor FN). When Q = 0.5 in a balanced data set (P = 0.5), there are equal numbers of false positives and false negatives.

Using the base data (Sens = 0.912, PPV = 0.356), values of CSI were calculated for P values ranging from 0.1 to 0.9 to illustrate the dependence of CSI on P. Three conditions were examined: Q = 0.1 (very few false positives); Q = 0.5 (equal numbers of false positives and false negatives, if the dataset was balanced); and Q = 0.9 (very few false negatives).

Hence, this approach requires the application of either Equation (3) or Equation (4) to the base data. The results are displayed in tables and graphically.

2.4. Method 3: CSI Recalculated via Both Rescaled PPV and Sens

There is also a method to recalculate CSI using not only rescaled PPV, as in Method 1, but also rescaled Sens.

The Bayes formula may be used to calculate different values of NPV across the range of P values:

$$NPV = Spec.P'/(Spec.P') + [(1 - Sens).P] \quad (5)$$

This allows for the recalculation of Sens at different P values using the equivalence shown by Kraemer, such that [66]:

$$(Sens - Q)/Q' = (NPV - P')/P$$

Rearranging this, values for Sens at a fixed Q may be calculated at variable P [1]:

$$Sens = [Q'.(NPV - P')/P] + Q \quad (6)$$

Hence, this approach requires the application of Equations (5) and (6) to the base data (Spec = 0.707; Q = 0.387 at optimal MACE cut-off of $\leq$20/30) to recalculate NPV and Sens, respectively.

With the rescaled Sens and the previously rescaled PPV (Table 1), it is then possible to recalculate CSI (Equation (1)). The results are displayed in a table and graphically.

Table 1. Values of PPV and CSI for dementia diagnosis at a fixed value of Q (MACE cut-off of $\leq$20/30) at various prevalence levels.

		MACE Cut-off $\leq$ 20/30 Sens = 0.912	
P	**P′**	**Recalculated PPV (from Equation (2))**	**Recalculated CSI (from Equation (1))**
0.1	0.9	0.257	0.251
0.2	0.8	0.437	0.420
0.3	0.7	0.571	0.542
0.4	0.6	0.675	0.634
0.5	0.5	0.757	0.705
0.6	0.4	0.824	0.763
0.7	0.3	0.879	0.810
0.8	0.2	0.926	0.850
0.9	0.1	0.966	0.883

3. Results

3.1. Method 1: CSI Recalculated via Bayes Formula for PPV

Using the Bayes formula (Equation (2)), both the recalculated values of PPV and CSI increased with increasing P (Table 1; Figure 1A). This confirms the expectation evident in

the Bayes formula that CSI, like PPV, is proportional to P in this formulation. This implies that the highest values of CSI will occur when P is high.

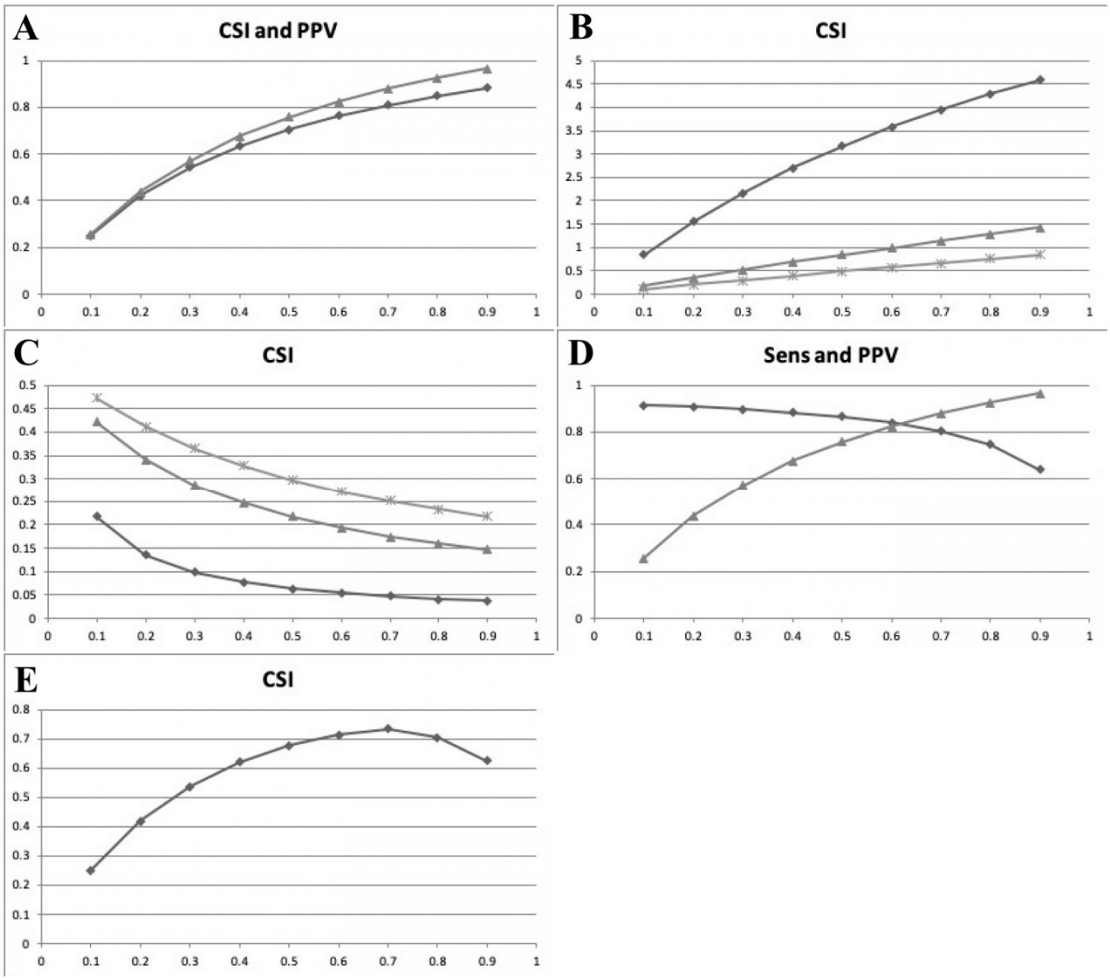

Figure 1. Panel of line graphs showing the study results. (**A**) Plot of CSI (♦) and PPV (▲) (y axis) for dementia diagnosis at fixed Q (Q = 0.387; MACE cut-off $\leq$ 20/30) versus prevalence P (x axis) calculated by sequential application of Equation (2) (Bayes formula) and Equation (1). (**B**) Plot of CSI (y axis) for dementia diagnosis at fixed Sens (0.912) and variable Q = 0.1 (♦), = 0.5 (▲), = 0.9 (*) versus prevalence P (*x* axis) calculated using Equation (3). (**C**) Plot of CSI (y axis) for dementia diagnosis at fixed PPV (0.356) and variable Q = 0.1 (♦), = 0.5 (▲), = 0.9 (*) versus prevalence P (x axis) calculated using Equation (4)**.** (**D**) Plot of Sens (♦) and PPV (▲) (*y* axis) for dementia diagnosis at fixed Q (Q = 0.387, MACE cut-off $\leq$ 20/30) versus prevalence P (*x* axis) calculated by application of Equations (2) and (6), respectively. (**E**) Plot of CSI (*y* axis) for dementia diagnosis at fixed Q (Q = 0.387, MACE cut-off $\leq$ 20/30) versus prevalence P (*x* axis), combining rescaled Sens and PPV (**D**).

3.2. Method 2: CSI Recalculated via Its Relation to Sens, PPV, P, and Q

Using Equation (3) (fixed Sens value), CSI increased with increasing P (Tables 2–4, 3rd column; Figure 1B). This implies that, with a fixed Sens, the highest values of CSI will occur when P is high.

Table 2. Values of CSI for dementia diagnosis at fixed values of Q = 0.1 and either Sens (0.912) or PPV (0.356) at various prevalence levels.

P	P + Q	CSI (Equation (3)) Sens = 0.912	CSI (Equation (4)) PPV = 0.356
0.1	0.2	0.838	0.217
0.2	0.3	1.55	0.135
0.3	0.4	2.16	0.098
0.4	0.5	2.70	0.077
0.5	0.6	3.17	0.063
0.6	0.7	3.58	0.054
0.7	0.8	3.95	0.047
0.8	0.9	4.28	0.041
0.9	1.0	4.58	0.037

Table 3. Values of CSI for dementia diagnosis at fixed values of Q = 0.5 and either Sens (0.912) or PPV (0.356) at various prevalence levels.

P	P + Q	CSI (Equation (3)) Sens = 0.912	CSI (Equation (4)) PPV = 0.356
0.1	0.6	0.179	0.421
0.2	0.7	0.352	0.341
0.3	0.8	0.520	0.286
0.4	0.9	0.682	0.247
0.5	1.0	0.838	0.217
0.6	1.1	0.990	0.193
0.7	1.2	1.14	0.174
0.8	1.3	1.28	0.159
0.9	1.4	1.42	0.146

Table 4. Values of CSI for dementia diagnosis at fixed values of Q = 0.9 and either Sens (0.912) or PPV (0.356) at various prevalence levels.

P	P + Q	CSI (Equation (3)) Sens = 0.912	CSI (Equation (4)) PPV = 0.356
0.1	1.0	0.100	0.473
0.2	1.1	0.199	0.412
0.3	1.2	0.295	0.365
0.4	1.3	0.390	0.328
0.5	1.4	0.483	0.297
0.6	1.5	0.574	0.272
0.7	1.6	0.664	0.251
0.8	1.7	0.752	0.233
0.9	1.8	0.838	0.217

Using Equation (4) (fixed PPV value), CSI decreased with increasing P (Tables 2–4, 4th column; Figure 1C). This implies that, with a fixed PPV, the highest value of CSI will occur when P is low.

3.3. Method 3: CSI Recalculated via Rescaled PPV and Sens

Using this method, neither PPV nor Sens is fixed, only Q. The rescaled values (Figure 1D) show Sens decreasing with increasing P (Table 5, column 4) and PPV increasing with increasing P (Table 5, column 3; and as per Table 1 and Figure 1A).

Table 5. Values of recalculated PPV (as per Table 1), Sens, and CSI for dementia diagnosis at various prevalence levels.

P	P′	Recalculated PPV (from Equation (2))	Recalculated Sens (from Equation (6))	Recalculated CSI (from Equation (1))
0.1	0.9	0.257	0.914	0.251
0.2	0.8	0.437	0.908	0.418
0.3	0.7	0.571	0.896	0.536
0.4	0.6	0.675	0.884	0.620
0.5	0.5	0.757	0.865	0.677
0.6	0.4	0.824	0.840	0.712
0.7	0.3	0.879	0.803	0.723
0.8	0.2	0.926	0.746	0.704
0.9	0.1	0.966	0.640	0.625

Combining these rescaled values as per Equation (1), CSI showed a concave curve when plotted against P (Table 5 column 5, Figure 1E). CSI values approximated PPV at low values of P (as in Figure 1A), and approximated Sens values at high values of P (compare Figure 1D,E).

4. Discussion

This study has shown that the dependence of CSI on P differs according to the method of calculation adopted.

Using either the Bayes formula method to rescale PPV (Equation (2)) or the direct method based on Sens (Equation (3)), the CSI values increased with increasing P. In these methods, the value of Sens was fixed, but the product (Sens.P) varied with P. Hence, the CSI values increased as P increased (Figure 1A,B).

In contrast, using the direct method based on PPV (Equation (4)), the CSI values decreased as P increased. With this method, the value of PPV was fixed, and hence, the product (PPV.Q) was also fixed for each of the three chosen values of Q (Tables 2–4, 4th column). Thus, the only changing variable in this method of calculation was (P + Q), which was inversely proportional to CSI (Equation (4)). This inverse relation was also expected on the basis of the observation that test Sens and PPV changed in opposite directions with the change in test cut-off [37]. This change in opposite directions was empirically observed in the previous analysis of the dataset used in this study [64].

Using the third method, in which both PPV and Sens were rescaled via the Bayes formula, the relationship between CSI and P was shown to be a concave curve. This suggests that CSI is maximal at a particular prevalence, which may vary according to the particular dataset under examination. It was previously shown, using the same dataset, that another unitary measure based on Sens and PPV, the F measure (the harmonic mean of Sens and PPV), showed a concave curve when plotted against P, with a maximum value at P = 0.7, but falling away at both higher and lower values of P. The finding of maximal CSI at P = 0.7 in this dataset was previously predicted, since CSI and F share a monotonic relationship [1,63]. The findings suggest that, at least in this cohort, CSI values follow PPV at low values of P and follow Sens at high values of P, but this needs further investigation in other patient cohorts.

This concave relationship is simply a reflection of the fact that CSI is dependent on both P and Q, as per Equations (3) and (4). In other words, this reflects the known trade-off

relationship between PPV and Sens, where one decreases as the other increases [37]. Just as paired outcome measures may be dependent on either P (PPV, NPV, and their complements) or Q (Sens, Spec, and their complements), unitary measures are also often functions of both P and Q. This is the case not only for CSI, but also for the F measure, Youden index (Y), predictive summary index (PSI), Matthews' correlation coefficient (MCC), and the harmonic mean of Y and PSI (HMYPSI) (Table 6). All showed concave relationships to P in this dataset [1].

Table 6. Summary of dependence of unitary measures on P and Q.

Unitary Measure	Dependence on P and Q
Critical success index (CSI)	$CSI = 1/[(P + Q)/Sens.P] - 1$ $CSI = 1/[(P + Q)/PPV.Q] - 1$
F measure (F)	$F = 2.Sens.P/(Q + P)$ $F = 2.PPV.Q/(Q + P)$
Youden index (Y)	$Y = (Sens - Q)/P'$ $Y = (Spec - Q')/P$ $Y = (Q - Q^2/P - P^2).PSI$
Predictive summary index (PSI)	$PSI = (PPV - P)/Q'$ $PSI = (NPV - P')/Q$ $PSI = (P - P^2/Q - Q^2).Y$
Matthews' correlation coefficient (MCC)	$MCC = \sqrt{(P - P^2/Q - Q^2).Y}$ $MCC = \sqrt{(Q - Q^2/P - P^2).PSI}$
Harmonic mean of Y and PSI (HMYPSI)	$HMYPSI = 2/(1/Y).[(1 + (Q - Q^2)/(P - P^2)]$ $HMYPSI= 2/(1/PSI).[(P - P^2)/(Q - Q^2) + 1]$

The major strength of this study is that it is, to our knowledge, the first to address the dependence of CSI on prevalence. A limitation is that it is based on a dataset of a single diagnostic test accuracy study. Future studies may examine other larger datasets, including those from different sources, such as case-ascertainment algorithms.

Hence, in conclusion, we suggest that there is no simple answer to the question of how CSI is dependent on P, other than that it is, and this depends on the method of calculation chosen to examine the relationship. In real-world situations, the dependence of CSI on P is not, and cannot be, independent of Q. Thus, conclusions based on outcome values of CSI (and indeed F measure) may be dataset-specific and not easily translated or generalized to other situations, as is recognized to necessarily be the case for PPV. Moreover, pragmatically, this is also the case for Sens since, although it is algebraically unrelated to P as a strictly columnar ratio in the 2 × 2 contingency table, it varies according to the heterogeneity of clinical populations (ditto Spec) [67], as is implied in the dependence of the Youden index on P (Table 6).

Author Contributions: Conceptualization, A.J.L.; methodology, G.K.M. and A.J.L.; formal analysis, G.K.M. and A.J.L.; writing—original draft, A.J.L.; writing—review and editing, G.K.M. and A.J.L.; supervision, A.J.L. All authors have read and agreed to the published version of the manuscript.

Funding: G.K.M. is supported by an NIHR Clinical Lectureship (CL-2022-07-002). The funders played no role in the design or conduct of this work. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Walton Centre for Neurology and Neurosurgery (protocol code N 310).

Informed Consent Statement: Not applicable.

Data Availability Statement: Anonymized base data in this secondary analysis are available for public use from the authors of the original study on a CCBY basis [64].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Larner, A.J. *The 2 × 2 Matrix: Contingency, Confusion and the Metrics of Binary Classification*, 2nd ed.; Springer: London, UK, 2024.
2. Aaberg, K.M.; Gunnes, N.; Bakken, I.J.; Lund Soraas, C.; Berntsen, A.; Magnus, P.; Lossius, M.I.; Stoltenberg, C.; Chin, R.; Suren, P. Incidence and Prevalence of Childhood Epilepsy: A Nationwide Cohort Study. *Pediatrics* **2017**, *139*, e20163908. [CrossRef]
3. Bellini, I.; Policardo, L.; Zaccara, G.; Palumbo, P.; Rosati, E.; Torre, E.; Francesconi, P. Identification of prevalent patients with epilepsy using administrative data: The Tuscany experience. *Neurol. Sci.* **2017**, *38*, 571–577. [CrossRef]
4. Chen, C.C.; Chen, L.S.; Yen, M.F.; Chen, H.H.; Liou, H.H. Geographic variation in the age- and gender-specific prevalence and incidence of epilepsy: Analysis of Taiwanese National Health Insurance-based data. *Epilepsia* **2012**, *53*, 283–290. [CrossRef]
5. Christensen, J.; Vestergaard, M.; Olsen, J.; Sidenius, P. Validation of epilepsy diagnoses in the Danish National Hospital Register. *Epilepsy Res.* **2007**, *75*, 162–170. [CrossRef]
6. Coulter, A.; Brown, S.; Daniels, A. Computer held chronic disease registers in general practice: A validation study. *J. Epidemiol. Community Health* **1989**, *43*, 25–28. [CrossRef]
7. De Jesus-Alvelo, I.; Labovitz, D. How Reliable Are the ICD9-CM Billing Codes in the Administrative Data to Estimate the Risk of Seizures and Epilepsy after Stroke? *Neurology* **2013**, *80*. [CrossRef]
8. Engeland, A.; Bjorge, T.; Daltveit, A.K.; Vollset, S.E.; Furu, K. Validation of disease registration in pregnant women in the Medical Birth Registry of Norway. *Acta Obstet. Gynecol. Scand.* **2009**, *88*, 1083–1089. [CrossRef]
9. Foebel, A.D.; Hirdes, J.P.; Heckman, G.A.; Kergoat, M.J.; Patten, S.; Marrie, R.A. Diagnostic data for neurological conditions in interRAI assessments in home care, nursing home and mental health care settings: A validity study. *BMC Health Serv. Res.* **2013**, *13*, 457. [CrossRef]
10. Fonferko-Shadrach, B.; Lacey, A.S.; White, C.P.; Powell, H.W.R.; Sawhney, I.M.S.; Lyons, R.A.; Smith, P.E.M.; Kerr, M.P.; Rees, M.I.; Pickrell, W.O. Validating epilepsy diagnoses in routinely collected data. *Seizure* **2017**, *52*, 195–198. [CrossRef]
11. Franchi, C.; Giussani, G.; Messina, P.; Montesano, M.; Romi, S.; Nobili, A.; Fortino, I.; Bortolotti, A.; Merlino, L.; Beghi, E.; et al. Validation of healthcare administrative data for the diagnosis of epilepsy. *J. Epidemiol. Community Health* **2013**, *67*, 1019–1024. [CrossRef]
12. Frost, F.J.; Hurley, J.S.; Petersen, H.V.; Gunter, M.J.; Gause, D. A comparison of two methods for estimating the health care costs of epilepsy. *Epilepsia* **2000**, *41*, 1020–1026. [CrossRef]
13. Holden, E.W.; Grossman, E.; Nguyen, H.T.; Gunter, M.J.; Grebosky, B.; Von Worley, A.; Nelson, L.; Robinson, S.; Thurman, D.J. Developing a computer algorithm to identify epilepsy cases in managed care organizations. *Dis. Manag.* **2005**, *8*, 1–14. [CrossRef]
14. Jette, N.; Reid, A.Y.; Quan, H.; Hill, M.D.; Wiebe, S. How accurate is ICD coding for epilepsy? *Epilepsia* **2010**, *51*, 62–69. [CrossRef]
15. Lee, S.Y.; Chung, S.E.; Kim, D.W.; Eun, S.H.; Kang, H.C.; Cho, Y.W.; Do Yi, S.; Kim, H.D.; Jung, K.Y.; Cheong, H.K.; et al. Estimating the Prevalence of Treated Epilepsy Using Administrative Health Data and Its Validity: ESSENCE Study. *J. Clin. Neurol.* **2016**, *12*, 434–440. [CrossRef]
16. Marrie, R.A.; Yu, B.N.; Leung, S.; Elliott, L.; Caetano, P.; Warren, S.; Wolfson, C.; Patten, S.B.; Svenson, L.W.; Tremlett, H.; et al. The utility of administrative data for surveillance of comorbidity in multiple sclerosis: A validation study. *Neuroepidemiology* **2013**, *40*, 85–92. [CrossRef]
17. Meeraus, W.H.; Petersen, I.; Chin, R.F.; Knott, F.; Gilbert, R. Childhood epilepsy recorded in primary care in the UK. *Arch. Dis. Child* **2013**, *98*, 195–202. [CrossRef]
18. Moura, L.M.; Price, M.; Cole, A.J.; Hoch, D.B.; Hsu, J. Accuracy of claims-based algorithms for epilepsy research: Revealing the unseen performance of claims-based studies. *Epilepsia* **2017**, *58*, 683–691. [CrossRef]
19. Parko, K.; Thurman, D.J. Prevalence of epilepsy and seizures in the Navajo Nation 1998–2002. *Epilepsia* **2009**, *50*, 2180–2185. [CrossRef]
20. Pina-Garza, J.E.; Vekeman, F.; Cheng, W.; Tuttle, E.; Giguere-Duval, P.; Oganisian, A.; Damron, J.; Duh, M.S.; Shen, V.; Isojarvi, J.; et al. Development of a claims-based classifier to identify lennox-gastaut syndrome. *Neurology* **2015**, *84*. [CrossRef]
21. Pugh, M.J.; Parko, K. Research using archival health care data: Let the buyer beware. *Epilepsia* **2015**, *56*, 321–322. [CrossRef]
22. Pugh, M.J.; Van Cott, A.C.; Cramer, J.A.; Knoefel, J.E.; Amuan, M.E.; Tabares, J.; Ramsay, R.E.; Berlowitz, D.R.; Treatment In Geriatric Epilepsy Research, t. Trends in antiepileptic drug prescribing for older patients with new-onset epilepsy: 2000–2004. *Neurology* **2008**, *70*, 2171–2178. [CrossRef] [PubMed]
23. Rehman, R.; Everhart, A.; Frontera, A.T.; Kelly, P.R.; Lopez, M.; Riley, D.; Sajan, S.; Schooff, D.M.; Tran, T.T.; Husain, A.M. Implementation of an established algorithm and modifications for the identification of epilepsy patients in the veterans health administration. *Epilepsy Res.* **2016**, *127*, 284–290. [CrossRef] [PubMed]
24. Reid, A.Y.; St Germaine-Smith, C.; Liu, M.; Sadiq, S.; Quan, H.; Wiebe, S.; Faris, P.; Dean, S.; Jette, N. Development and validation of a case definition for epilepsy for use with administrative health data. *Epilepsy Res.* **2012**, *102*, 173–179. [CrossRef] [PubMed]
25. Shackleton, D.P.; Westendorp, R.G.; Kasteleijn-Nolst Trenite, D.G.; de Boer, A.; Herings, R.M. Dispensing epilepsy medication: A method of determining the frequency of symptomatic individuals with seizures. *J. Clin. Epidemiol.* **1997**, *50*, 1061–1068. [CrossRef]

26. Syvertsen, M.; Nakken, K.O.; Edland, A.; Hansen, G.; Hellum, M.K.; Koht, J. Prevalence and etiology of epilepsy in a Norwegian county-A population based study. *Epilepsia* **2015**, *56*, 699–706. [CrossRef]
27. Tan, M.; Wilson, I.; Braganza, V.; Ignatiadis, S.; Boston, R.; Sundararajan, V.; Cook, M.J.; D'Souza, W.J. Development and validation of an epidemiologic case definition of epilepsy for use with routinely collected Australian health data. *Epilepsy Behav.* **2015**, *51*, 65–72. [CrossRef] [PubMed]
28. Tu, K.; Wang, M.; Jaakkimainen, R.L.; Butt, D.; Ivers, N.M.; Young, J.; Green, D.; Jette, N. Assessing the validity of using administrative data to identify patients with epilepsy. *Epilepsia* **2014**, *55*, 335–343. [CrossRef] [PubMed]
29. Wassenaar, M.; Carpay, J.A.; Sander, J.W.; Thijs, R.D. Validity of health insurance data to identify people with epilepsy. *Epilepsy Res.* **2018**, *139*, 102–106. [CrossRef]
30. Williamson, T.; Green, M.E.; Birtwhistle, R.; Khan, S.; Garies, S.; Wong, S.T.; Natarajan, N.; Manca, D.; Drummond, N. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann. Fam. Med.* **2014**, *12*, 367–372. [CrossRef]
31. Pickrell, W.O.; Lacey, A.S.; Bodger, O.G.; Demmler, J.C.; Thomas, R.H.; Lyons, R.A.; Smith, P.E.; Rees, M.I.; Kerr, M.P. Epilepsy and deprivation, a data linkage study. *Epilepsia* **2015**, *56*, 585–591. [CrossRef]
32. Mbizvo, G.K.; Bennett, K.H.; Simpson, C.R.; Duncan, S.E.; Chin, R.F.M.; Larner, A.J. Using Critical Success Index or Gilbert Skill score as composite measures of positive predictive value and sensitivity in diagnostic accuracy studies: Weather forecasting informing epilepsy research. *Epilepsia* **2023**, *64*, 1466–1468. [CrossRef] [PubMed]
33. Wilkinson, T.; Ly, A.; Schnier, C.; Rannikmae, K.; Bush, K.; Brayne, C.; Quinn, T.J.; Sudlow, C.L.M.; Group, U.K.B.N.O.; Dementias Platform, U.K. Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimer's Dement.* **2018**, *14*, 1038–1051. [CrossRef] [PubMed]
34. Horrocks, S.; Wilkinson, T.; Schnier, C.; Ly, A.; Woodfield, R.; Rannikmae, K.; Quinn, T.J.; Sudlow, C.L. Accuracy of routinely-collected healthcare data for identifying motor neurone disease cases: A systematic review. *PLoS ONE* **2017**, *12*, e0172639. [CrossRef] [PubMed]
35. Kee, V.R.; Gilchrist, B.; Granner, M.A.; Sarrazin, N.R.; Carnahan, R.M. A systematic review of validated methods for identifying seizures, convulsions, or epilepsy using administrative and claims data. *Pharmacoepidemiol. Drug Saf.* **2012**, *21*, 183–193. [CrossRef] [PubMed]
36. Mbizvo, G.K.; Bennett, K.H.; Schnier, C.; Simpson, C.R.; Duncan, S.E.; Chin, R.F.M. The accuracy of using administrative healthcare data to identify epilepsy cases: A systematic review of validation studies. *Epilepsia* **2020**, *61*, 1319–1335. [CrossRef]
37. Wang, H.; Wang, B.; Zhang, X.; Feng, C. Relations among sensitivity, specificity and predictive values of medical tests based on biomarkers. *Gen. Psychiatr.* **2021**, *34*, e100453. [CrossRef] [PubMed]
38. Mbizvo, G.K.; Larner, A.J. Isolated headache is not a reliable indicator for brain cancer. *Clin. Med.* **2022**, *22*, 92–93. [CrossRef] [PubMed]
39. Mbizvo, G.K.; Larner, A.J. Re: Realistic expectations are key to realising the benefits of polygenic scores. *BMJ* **2023**. [CrossRef]
40. Larner, A.J. Assessing cognitive screening instruments with the critical success index. *Prog. Neurol. Psychiatry* **2021**, *25*. *in press*. [CrossRef]
41. Chae, S.; Shin, J.; Kwon, S.; Lee, S.; Kang, S.; Lee, D. PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. *Sci. Rep.* **2021**, *11*, 11952. [CrossRef]
42. Kim, R.S.; Moon, Y.J.; Gopalswamy, N.; Park, Y.D.; Kim, Y.H. Two-step forecast of geomagnetic storm using coronal mass ejection and solar wind condition. *Space Weather* **2014**, *12*, 246–256. [CrossRef]
43. Pavlovic, R.; Chen, J.; Anderson, K.; Moran, M.D.; Beaulieu, P.A.; Davignon, D.; Cousineau, S. The FireWork air quality forecast system with near-real-time biomass burning emissions: Recent developments and evaluation of performance for the 2015 North American wildfire season. *J. Air Waste Manag.* **2016**, *66*, 819–841. [CrossRef] [PubMed]
44. El Jarroudi, M.; Kouadio, L.; Delfosse, P.; Tychon, B. Brown rust disease control in winter wheat: I. Exploring an approach for disease progression based on night weather conditions. *Environ. Sci. Pollut. Res.* **2014**, *21*, 4797–4808. [CrossRef] [PubMed]
45. Shin, D.; Kim, J.H. A New Application of Unsupervised Learning to Nighttime Sea Fog Detection. *Asia-Pac. J. Atmos. Sci.* **2018**, *54*, 527–544. [CrossRef] [PubMed]
46. Skinner, P.S.; Wheatley, D.M.; Knopfmeier, K.H.; Reinhart, A.E.; Choate, J.J.; Jones, T.A.; Creager, G.J.; Dowell, D.C.; Alexander, C.R.; Ladwig, T.T.; et al. Object-Based Verification of a Prototype Warn-on-Forecast System. *Weather Forecast.* **2018**, *33*, 1225–1250. [CrossRef]
47. Nguyen, P.; Ombadi, M.; Gorooh, V.A.; Shearer, E.J.; Sadeghi, M.; Sorooshian, S.; Hsu, K.L.; Bolvin, D.; Ralph, M.F. PERSIANN Dynamic Infrared-Rain Rate (PDIR-Now): A Near-Real-Time, Quasi-Global Satellite Precipitation Dataset. *J. Hydrometeorol.* **2020**, *21*, 2893–2906. [CrossRef] [PubMed]
48. Jing, J.R.; Li, Q.; Peng, X. MLC-LSTM: Exploiting the Spatiotemporal Correlation between Multi-Level Weather Radar Echoes for Echo Sequence Extrapolation. *Sensors* **2019**, *19*, 3988. [CrossRef] [PubMed]
49. Gilbert, G.K. Finley's tornado predictions. *Am. Meteorol. J.* **1884**, *1*, 166–172.
50. World Meteorological Organization. *Forecast Verification for the African Severe Weather Forecasting Demonstration Projects*; No. 1132; World Meteorological Organization: Geneva, Switzerland, 2014.
51. Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytol.* **1912**, *11*, 37–50. [CrossRef]

52. Palmer, W.C.; Allen, R.A. *Note on the Accuracy of Forecasts Concerning the Rain Problem*; U.S. Weather Bureau manuscript: Washington, DC, USA, 1949.
53. Donaldson, R.J.; Dyer, R.M.; Kraus, M.J. An objective evaluator of techniques for predicting severe weather events. In *Preprints, 9th Conference on Severe Local Storms. Norman, Oklahoma*; National Technical Information Service: Springfield, VI, USA, 1975; pp. 312–326.
54. Schaefer, J.T. The critical success index as an indicator of warning skill. *Weather Forecast.* **1990**, *5*, 570–575. [CrossRef]
55. Hand, D.J.; Christen, P.; Kirielle, N. F*: An interpretable transformation of the F-measure. *Mach. Learn.* **2021**, *110*, 451–456. [CrossRef]
56. Space Weather Prediction Center. *Forecast Verification Glossary*; National Oceanic and Atmospheric Administration: Boulder, CO, USA, 2022.
57. Mbizvo, G.K.; Simpson, C.R.; Duncan, S.E.; Chin, R.F.M.; Larner, A.J. Critical success index or F measure to validate the accuracy of administrative healthcare data identifying epilepsy in deceased adults in Scotland. *Epilepsy Res.* **2024**, *199*, 107275. [CrossRef]
58. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
59. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* **1948**, *5*, 1–34.
60. Van Rijsbergen, C.J. Foundation of evaluation. *J. Doc.* **1974**, *30*, 365–373. [CrossRef]
61. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
62. Powers, D.M.W. What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes. *arXiv* **2015**, arXiv:1503.06410. [CrossRef]
63. Jolliffe, I.T. The Dice co-efficient: A neglected verification performance measure for deterministic forecasts of binary events. *Meteorol. Appl.* **2016**, *23*, 89–90. [CrossRef]
64. Larner, A.J. MACE for Diagnosis of Dementia and MCI: Examining Cut-Offs and Predictive Values. *Diagnostics* **2019**, *9*, 51. [CrossRef]
65. Hsieh, S.; McGrory, S.; Leslie, F.; Dawson, K.; Ahmed, S.; Butler, C.R.; Rowe, J.B.; Mioshi, E.; Hodges, J.R. The Mini-Addenbrooke's Cognitive Examination: A new assessment tool for dementia. *Dement. Geriatr. Cogn. Disord.* **2015**, *39*, 1–11. [CrossRef]
66. Kraemer, H.C. *Evaluating Medical Tests: Objective and Quantitative Guidelines*; Sage Publications: New York, NY, USA, 1992.
67. Leeflang, M.M.; Rutjes, A.W.; Reitsma, J.B.; Hooft, L.; Bossuyt, P.M. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* **2013**, *185*, E537–E544. [CrossRef] [PubMed]