

# The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy

M. Hassan Murad MD MPH, Lifeng Lin PhD, Haitao Chu MD PhD, Bashar Hasan MD, Reem A. Alsibai MD, Alzhras S. Abbas MBBS, Reem A. Mustafa MD MPH PhD, Zhen Wang PhD

■ Cite as: *CMAJ* 2023 July 17;195:E925-31. doi: 10.1503/cmaj.221802

## Abstract

**Background:** Sensitivity and specificity are characteristics of a diagnostic test and are not expected to change as the prevalence of the target condition changes. We sought to evaluate the association between prevalence and changes in sensitivity and specificity.

**Methods:** We retrieved data from meta-analyses of diagnostic test accuracy published in the Cochrane Database of Systematic Reviews (2003–2020). We used mixed-effects random-intercept linear regression models to evaluate the association between prevalence and logit-transformed sensitivity and specificity. The model evaluated all meta-analyses as nested within each systematic review.

**Results:** We analyzed 6909 diagnostic test accuracy studies from 552 meta-analyses that were included in 92 systematic reviews. For sensitivity, compared with the lowest quartile of prevalence, the second, third and fourth quartiles were associated with significantly higher odds of identifying a true positive case (odds ratio [OR] 1.17, 95% confidence interval [CI] 1.09–1.26; OR 1.32, 95% CI 1.23–1.41; OR 1.47, 95% CI 1.37–1.58; respectively). For specificity, compared with the lowest quartile of prevalence, the second, third and fourth quartiles were associated with significantly lower odds of identifying a true negative case (OR 0.74, 95% CI 0.69–0.80; OR 0.65,

95% CI 0.60–0.70; OR 0.47, 95% CI 0.44–0.51; respectively). Pooled regression coefficients from bivariate models conducted within each meta-analysis showed that prevalence was positively associated with sensitivity and negatively associated with specificity. Findings were consistent across subgroups.

**Interpretation:** In this large sample of diagnostic studies, higher prevalence was associated with higher estimated sensitivity and lower estimated specificity. Clinicians should consider the implications of disease prevalence and spectrum when interpreting the results from studies of diagnostic test accuracy.

Clinicians practising evidence-based medicine are familiar with the concepts of sensitivity and specificity, defined as the probability of a positive test given that the person has the target condition, and the probability of a negative test given that the person does not have the condition, respectively.<sup>1</sup> In practice, sensitivity and specificity are often treated as being independent from disease prevalence, defined as pre-test probability of disease or probability of the target condition in the study sample. This is often contrasted with positive and negative predictive values or post-test probabilities, which are highly dependent on prevalence and pre-test probability. Therefore, sensitivity and specificity are considered to be characteristics of the test as intrinsic accuracy measures, and independent from the characteristics of the population. The rationale for this assumption relates to the mathematical calculation of these measures from a classic 2 × 2 diagnostic table.

However, sensitivity and specificity have been found to change as disease prevalence changes. Leeflang and colleagues<sup>2</sup> provided 4 examples of individual diagnostic studies in which sensitivity and specificity changed as the prevalence changed, and Li and Fine<sup>3</sup> showed the same in 2 meta-analyses. In a separate study, Leeflang and colleagues<sup>4</sup> used data from 23 meta-analyses and evaluated the effects of prevalence on sensitivity and specificity using a bivariate random-effects model for each meta-analysis, with prevalence as a covariate.<sup>5,6</sup> The results suggested that specificity tended to be lower with a higher prevalence, whereas sensitivity changes did not follow a specific pattern.<sup>4</sup>

Overall, however, the number of studies that have evaluated the association between prevalence and sensitivity and specificity is small, although the issue is very important. Some of the reported changes in sensitivity and specificity reached

40 percentage points,<sup>4</sup> which can entirely change the impact of the test in management of patients and can lead to more false positives and negatives, under- and overdiagnosis, and possible harm to patients. Therefore, we sought to evaluate the relationship between prevalence and sensitivity and specificity in a large sample of studies of diagnostic test accuracy.

## Methods

We conducted a meta-epidemiological study using diagnostic meta-analyses published by the Cochrane Collaboration. This study follows the reporting guideline for meta-epidemiological methodology research.<sup>7</sup>

### Data source and study selection

We searched all reviews flagged as diagnostic test accuracy reviews in the Cochrane Database of Systematic Reviews published between January 2003 and January 2020. To be included, eligible meta-analyses had to report a classic 2 × 2 diagnostic table (i.e., true positive, false positive, false negative and true negative results) and include at least 4 studies. We did not restrict by patient population, type of diagnostic test, study location or condition studied. We excluded diagnostic studies that had a case-control design, in which disease prevalence is not reliably estimated and may not reflect the true population disease prevalence, severity or spectrum.

### Data extraction

We used the RCurl package in R to download data from the RM5 files of the systematic reviews from the Cochrane Database of Systematic Reviews. We converted these to comma-separated values files for analyses. In addition to 2 × 2 diagnostic tables, we extracted author names and publication time of the original studies of diagnostic test accuracy, as well as the systematic reviews. Pairs of reviewers independently and manually extracted the overall risk of bias across original studies, test type, study setting and target condition. We used the QUADAS-2 tool in evaluating risk of bias.<sup>8</sup> Reviewers also extracted whether the systematic reviews explicitly discussed spectrum bias or spectrum effect as a possible modifier of diagnostic accuracy, and if the reviews conducted subgroup analyses based on prevalence.

### Outcome

Our outcome of interest was the association between prevalence and sensitivity and specificity.

### Statistical analysis

In the main analysis, we fitted 2 separate models to evaluate the association between prevalence and logit-transformed sensitivity or logit-transformed specificity. These models were 3-level, mixed-effects, random-intercept linear regression models. Fixed effects included disease prevalence, categorized as quartiles based on the distribution within a meta-analysis ( $\leq 25$ th percentile as reference, 26th–50th percentile, 51st–75th percentile and  $> 75$ th percentile) or arbitrary prevalence cut-offs ( $< 25\%$ , 26%–50%, 51%–75%,  $> 75\%$ ), adjusting for the target condition category

(collapsed into 5 categories based on the discipline or specialty, namely internal medicine, surgery, oncology, obstetrics and gynecology, and neurology or psychiatry), type of diagnostic test (pathology and cytology, blood test, imaging, physical examination or symptomatology, urine or cerebrospinal fluid test and a category of multiple test types), study setting (inpatient, outpatient and mixed) and risk of bias (low, high and unclear). We included random effects for each meta-analysis and, then, nested within each systematic review. In a sensitivity analysis, we added the slope of prevalence as a random effect in the 3-level mixed effects model (Appendix 1, available at [www.cmaj.ca/lookup/doi/10.1503/cmaj.221802/tab-related-content](http://www.cmaj.ca/lookup/doi/10.1503/cmaj.221802/tab-related-content)).

In a secondary analysis, we applied a bivariate mixed-effects regression model to jointly estimate sensitivity and specificity for each individual meta-analysis,<sup>5</sup> and added the estimated prevalence (proportion of people with the target condition in each study) as a continuous covariate to the model. To estimate the direction and strength of the association between prevalence and logit-transformed sensitivity and specificity, we pooled regression coefficients of prevalence as estimated from each meta-analysis using the restricted maximum likelihood random-effects method. We conducted subgroup analyses based on category of target condition, type of diagnostic test, study setting and risk of bias.

We graphed trends of logit-transformed sensitivity and specificity against prevalence using a locally weighted scatterplot smoothing plot. We considered a 2-tailed  $p$  value of less than 0.05 statistically significant. We conducted all statistical analyses using Stata/SE, version 17.0 (StataCorp LLC). Analysis codes are listed in Appendix 1.

### Ethics approval

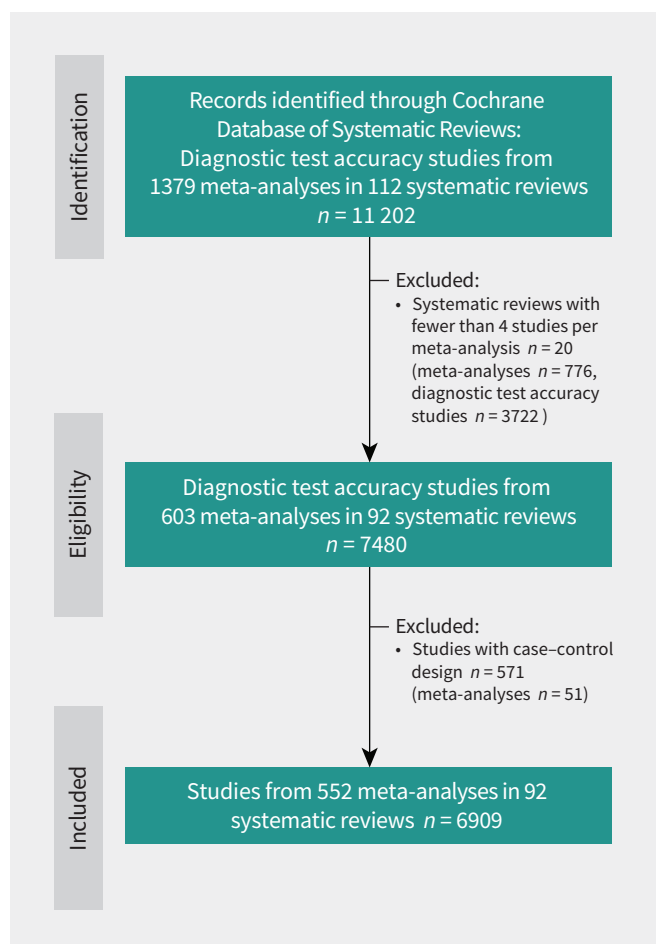
Ethics approval was not needed for this analysis.

## Results

We retrieved data from 112 systematic reviews of diagnostic test accuracy. After exclusions, we analyzed 6909 studies of diagnostic accuracy (i.e., 6909 2 × 2 tables) from 552 meta-analyses that were included in 92 systematic reviews (Figure 1). The individual diagnostic accuracy studies were published between 1961 and 2019, with a median sample size of 157 (interquartile range [IQR] 74–404) patients. The median number of original diagnostic accuracy studies within a meta-analysis was 17 (IQR 5–34) studies. Across the 552 meta-analyses, the median prevalence ranged from 0.07% to 94.90% (IQR 11.65%–41.15%). Only 14 (15.2%) of the 92 systematic reviews explicitly discussed spectrum bias or spectrum effect as a possible modifier of diagnostic accuracy, and only 7 (7.6%) conducted subgroup analyses based on prevalence.

### The association of prevalence with sensitivity and specificity

Figure 2 shows a positive association between sensitivity and prevalence, and Figure 3 shows a negative association between specificity and prevalence. In the main analysis, we fitted a



**Figure 1:** Study flow chart. A systematic review could include several meta-analyses. Individual diagnostic test accuracy studies refer to original studies included in each meta-analysis.

mixed-effects model with random effects for each meta-analysis nested within each systematic review. Table 1 shows the results stratified by prevalence quartiles and arbitrary prevalence cutoffs (25%, 50%, 75% and 100%). For sensitivity, compared with the lowest quartile of prevalence, the second, third and fourth quartiles were associated with significantly higher odds of identifying a true positive case (odds ratio [OR] 1.17, 95% confidence interval [CI] 1.09–1.26; OR 1.32, 95% CI 1.23–1.41; OR 1.47, 95% CI 1.37–1.58; respectively). For specificity, compared with the lowest quartile of prevalence, the second, third and fourth quartiles were associated with significantly lower odds of identifying a true negative case (OR 0.74, 95% CI 0.69–0.80; OR 0.65, 95% CI 0.60–0.70; OR 0.47, 95% CI 0.44–0.51; respectively). The sensitivity analyses in which prevalence was modelled as a random intercept and random slope showed similar findings (Appendix 1).

In the secondary analysis (Table 2), the pooled regression coefficients of prevalence from bivariate regression models showed the direction and strength of the association between prevalence and sensitivity and specificity. In this analysis, data from 379 individual studies failed to converge. Thus, the analysis included 6530 diagnostic test accuracy studies. Among the 6530 studies, the target condition categories included internal

medicine ( $n = 2755$ , 42.2%), oncology ( $n = 1693$ , 25.9%), obstetrics and gynecology ( $n = 937$ , 14.3%), neurology or psychiatry ( $n = 578$ , 8.9%) and surgery ( $n = 567$ , 8.7%). The most common diagnostic test types were pathology and cytology ( $n = 1369$ , 21.0%), blood tests ( $n = 1310$ , 20.1%), imaging ( $n = 1334$ , 20.4%) and physical examination or symptomatology ( $n = 1174$ , 18.0%). The tests were conducted in outpatient settings ( $n = 1662$ , 25.5%), hospital settings ( $n = 929$ , 14.2%) and settings that were mixed or unclear ( $n = 3939$ , 60.3%). The risk of bias of the 6530 original studies was high in 1818 (27.8%) studies, low in 2841 (43.5%) studies and unclear in 1871 (28.7%) studies. The findings suggest a positive association with logit-transformed sensitivity (mean 0.92, standard error 0.10) and a negative association with logit-transformed specificity (mean  $-7.43$ , standard error 2.10). Findings were consistent across subgroups.

## Interpretation

We performed a meta-epidemiological analysis of 6909 studies of diagnostic tests and found that sensitivity is positively associated with prevalence, whereas specificity is negatively associated with prevalence. The direction of this change was also found in the secondary analyses in which a bivariate model incorporated prevalence as a covariate within each meta-analysis.

Our results are consistent with other studies that have attempted to evaluate whether there is an association between prevalence and sensitivity and specificity.<sup>2-4,9</sup> In terms of the direction of the association and its pattern, other studies suggested that specificity tended to be negatively associated with prevalence, as we observed, whereas sensitivity changes did not follow a specific pattern.<sup>2-4</sup> However, these studies were far smaller and may have been underpowered.

Although the definitions of sensitivity and specificity do not depend on prevalence, our results support the existence of such an association. Spectrum bias, also called spectrum effect, could partially explain this association.<sup>2,10,11</sup> A test may perform better when used to evaluate patients with more severe disease than it would in patients whose disease is less obvious or less advanced. Hence, if investigators choose clinically inappropriate populations when studying a diagnostic test, they can introduce spectrum bias, which may seriously affect the results to show that the test performs better than it actually does. Disease status is not truly binary; rather, a spectrum of continuous traits defines disease severity (e.g., serum glucose and a diagnosis of diabetes).<sup>9</sup> Patients with test values close to the test cut-off are more likely to be misclassified. This misclassification is correlated with population characteristics and prevalence.<sup>2,10,11</sup> Therefore, prevalence may be a surrogate for disease severity, and thus affects sensitivity and specificity. Brenner and Gefeller<sup>9</sup> evaluated the effect of a hypothetical continuous trait that categorizes people into diseased and not diseased, and found that the dependence of sensitivity on prevalence may be of similar magnitude to that of the positive predictive value. They expected that, as prevalence increases, sensitivity would increase and specificity would decrease, similar to our findings.<sup>9</sup>

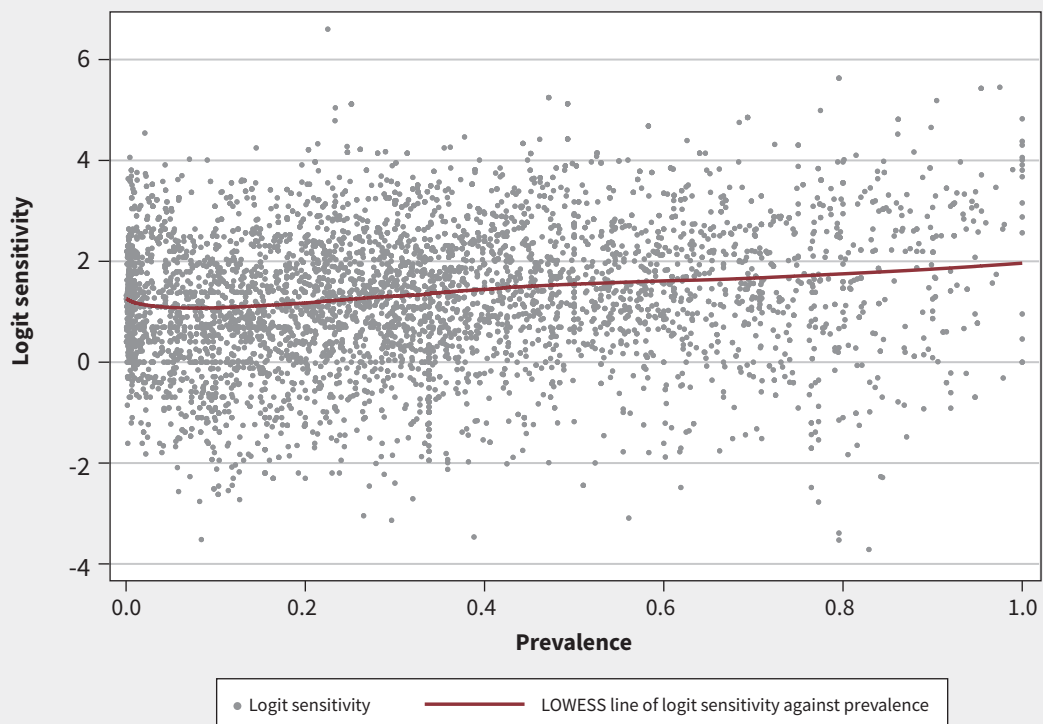


Figure 2: Locally weighted scatterplot smoothing (LOWESS) plot of logit-transformed sensitivity against prevalence, showing a positive association between prevalence and sensitivity.

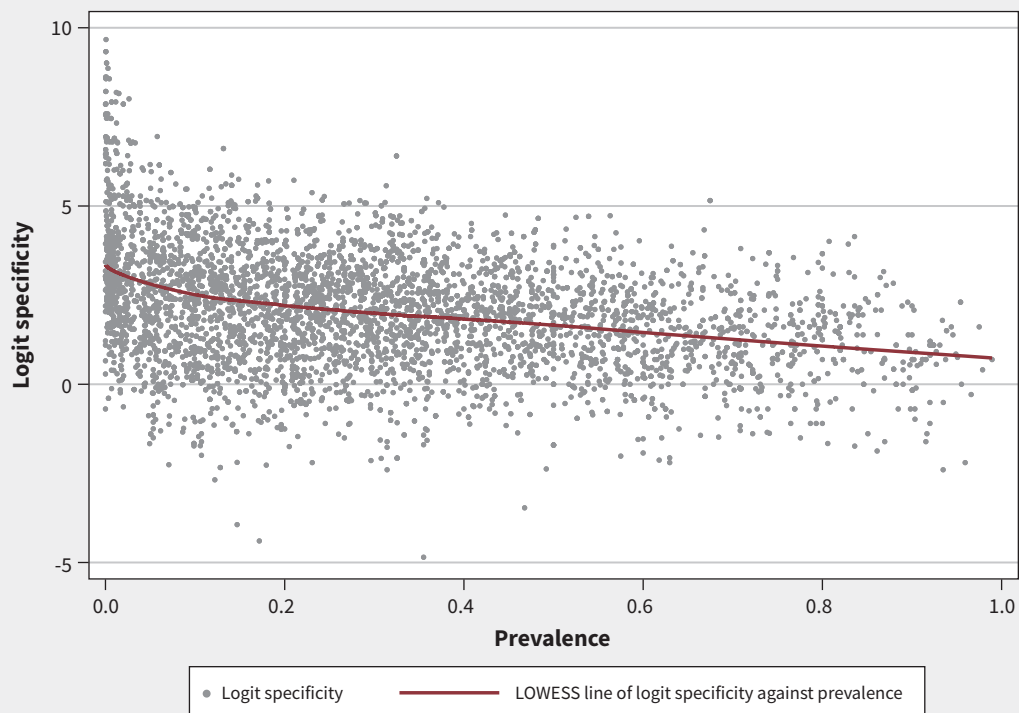


Figure 3: Locally weighted scatterplot smoothing (LOWESS) plot of logit-transformed specificity against prevalence, showing a negative association between prevalence and specificity.

**Table 1: Mixed-effects model incorporating all meta-analyses\***

Outcome	OR (95% CI)†	Range of prevalence,‡ %
<b>Stratification by prevalence quartiles§</b>		
Sensitivity		
1st quartile (≤ 25th percentile)	Ref.	0.00–80.00
2nd quartile (26th–50th percentile)	1.17 (1.09–1.26)	0.04–94.90
3rd quartile (51st–75th percentile)	1.32 (1.23–1.41)	0.06–100.00
4th quartile (> 75th percentile)	1.47 (1.37–1.58)	0.13–100.00
Specificity		
1st quartile (≤ 25th percentile)	Ref.	
2nd quartile (26th–50th percentile)	0.74 (0.69–0.80)	0.00–80.00
3rd quartile (51st–75th percentile)	0.65 (0.60–0.70)	0.04–94.90
4th quartile (> 75th percentile)	0.47 (0.44–0.51)	0.06–100.00
<b>Stratification by arbitrary cutoffs of prevalence</b>		
Sensitivity, %		
Prevalence ≤ 25	Ref.	0.00–25.00
Prevalence 26–50	1.38 (1.29–1.48)	25.06–50.00
Prevalence 51–75	1.60 (1.53–1.86)	50.25–75.00
Prevalence > 75	1.98 (1.72–2.28)	75.28–100.00
Specificity, %		
Prevalence ≤ 25	Ref.	0.00–25.00
Prevalence 26–50	0.65 (0.60–0.70)	25.06–50.00
Prevalence 51–75	0.42 (0.38–0.47)	50.25–75.00
Prevalence > 75	0.28 (0.24–0.33)	75.28–100.00

Note: CI = confidence interval, OR = odds ratio, Ref. = reference.

\*The models were 3-level, mixed-effects, random intercept linear regression models. Fixed effects included disease prevalence categorized as quartiles based on the distribution within a meta-analysis (≤ 25th percentile as reference, 26th–50th percentile, 51st–75th percentile and > 75 percentile) or arbitrary prevalence cutoffs (< 25%, 26%–50%, 51%–75%, > 75%), adjusting for the target condition category (collapsed into 5 categories based on the discipline or specialty: internal medicine, surgery, oncology, obstetrics and gynecology, and neurology or psychiatry), type of diagnostic test (pathology and cytology, blood test, imaging, physical exam or symptomatology, urine or cerebrospinal fluid test, and multiple test types), study setting (inpatient, outpatient and mixed), and risk of bias (low, high and unclear). Random effects were included for each meta-analysis and, then, nested within each systematic review.

†Odds ratio shows the changes of odds to correctly identify a true positive case for sensitivity and a true negative case for specificity, when prevalence increased from the reference category to a corresponding higher prevalence category.

‡Range of prevalence shows the prevalence range across all 552 meta-analyses.

§Prevalence quartile is identified within a meta-analysis.

The design of diagnostic studies and enrolment procedures could lead to increased spectrum effect. For example, test accuracy studies that enroll patients from subspecialty clinics may include higher prevalence and sicker patients than when the test is conducted in a primary care clinic because of referral bias.<sup>2</sup> The association between accuracy and prevalence may have other explanations. For example, Leeflang<sup>2</sup> suggests that an inadequate or imperfect reference standard will underestimate accuracy, but this effect may decline at a higher prevalence. The observed association between test accuracy and prevalence may also be related to the fact that it was evaluated within a meta-analysis and therefore reflects sample sensitivity and specificity, whereas the population sensitivity and specificity remain theoretically independent.<sup>3</sup>

For evidence-based practitioners, the current findings suggest that when they apply evidence about diagnostic accuracy of

a test, they should compare the prevalence of disease in their context with that in available studies. A very different prevalence may indicate a different disease spectrum and uncertainty about the sensitivity and specificity. This uncertainty also affects the likelihood ratios, which are derived from sensitivity and specificity and may have a stronger association with prevalence.<sup>9</sup>

For researchers conducting original studies on diagnostic accuracy, special attention should be paid to the spectrum of the disease in their sample so that it is consistent with the spectrum of disease among people who will receive the test in practice. In addition, when the spectrum or prevalence of disease is highly variable, researchers should plan stratified analyses by disease severity and spectrum.

For researchers conducting systematic reviews of diagnostic test studies, our results suggest the need to consider the effect of prevalence on diagnostic accuracy measures. They may consider

**Table 2: Pooled regression coefficients of prevalence from bivariate regression models**

Outcome	No. of original diagnostic test studies (no. of meta-analyses)	Pooled regression coefficient of prevalence on logit-transformed sensitivity, mean $\pm$ SE	Pooled regression coefficient of prevalence on logit-transformed specificity, mean $\pm$ SE
Overall*	6530 (520)	0.92 $\pm$ 0.10	-7.43 $\pm$ 2.10
Target condition category			
Internal medicine	2755 (202)	0.72 $\pm$ 0.16	-7.93 $\pm$ 4.48
Oncology	1693 (135)	1.78 $\pm$ 0.26	-6.93 $\pm$ 1.51
Obstetrics and gynecology	937 (76)	1.02 $\pm$ 0.43	-16.64 $\pm$ 7.40
Surgery	567 (39)	0.59 $\pm$ 0.40	-1.09 $\pm$ 1.21
Neurology or psychiatry	578 (68)	0.24 $\pm$ 0.22	-0.34 $\pm$ 1.54
Type of tests			
Pathology and cytology	1369 (82)	1.55 $\pm$ 0.29	-9.94 $\pm$ 2.42
Blood tests	1310 (95)	0.56 $\pm$ 0.23	-17.40 $\pm$ 9.75
Imaging	1334 (119)	1.21 $\pm$ 0.21	-1.40 $\pm$ 0.47
Physical examination or symptomatology	1174 (121)	0.64 $\pm$ 0.19	-1.94 $\pm$ 0.54
Multiple types	1265 (88)	0.94 $\pm$ 0.26	-12.18 $\pm$ 5.97
Urine or CSF	78 (15)	-0.30 $\pm$ 0.49	2.88 $\pm$ 1.66
Study settings			
Inpatients	929 (70)	0.08 $\pm$ 0.28	-19.31 $\pm$ 12.87
Outpatients	1662 (148)	1.11 $\pm$ 0.34	-11.74 $\pm$ 3.91
Mixed or unclear	3939 (80)	0.98 $\pm$ 0.12	-2.64 $\pm$ 0.66
Risk of bias			
Low	2841 (209)	1.28 $\pm$ 0.21	-9.81 $\pm$ 2.82
Unclear	1871 (146)	0.97 $\pm$ 0.20	-1.67 $\pm$ 0.96
High	1818 (165)	0.45 $\pm$ 0.15	-10.78 $\pm$ 6.19

Note: CSF = cerebrospinal fluid, SE = standard error.

\*32 meta-analyses with 379 diagnostic test studies from the primary analyses were not included as they failed to converge.

conducting subgroup analyses by prevalence (e.g., low, medium, high), although these exploratory analyses can be affected by chance findings and errors related to the arbitrary categorizing of prevalence into discrete categories. Another approach that may avoid some of these limitations is to jointly model disease prevalence with diagnostic test sensitivity and specificity, such as by using trivariate generalized linear mixed models.<sup>12-14</sup> Such models can directly estimate the correlations between diagnostic accuracy measures with prevalence on a logit-transformed scale, but require a large number of studies (> 10) to achieve model convergence.

### Limitations

Although the large number of studies of diagnostic test accuracy included in this analysis represents a strength, it also made granular evaluation at a study level infeasible for some factors. We were able to address only risk of bias, target condition, study setting and diagnostic test type. Other causes of variability include the stringency or misclassification of the gold-standard test or the threshold used to categorize study participants as positive for the target condition; these may have contributed to the observed association

with prevalence. Spectrum bias or effect was not explored by most of the systematic reviews evaluated in this study; hence, we did not have sufficient information to further explore this issue. We also acknowledge the heterogeneity across the studies, which we addressed by clustering the regression analysis within each meta-analysis, given that the authors had decided that the studies were sufficiently similar to include in the same meta-analysis. Random measurement error and variability in prevalence may lead to regression dilution bias, although such bias would likely drive the estimated association toward the null.<sup>15</sup> Despite our effort in excluding duplicate and overlapping studies or participants, this possibility still exists, which is a limitation of meta-epidemiological research in which the study is the unit of analysis.

### Conclusion

We found that the estimated sensitivity and specificity of diagnostic tests are associated with the prevalence of the target condition. Prevalence could be a surrogate of disease spectrum and should be considered when interpreting the results of studies that evaluate diagnostic test accuracy.



## References

1. Rosner B. *Fundamentals of biostatistics*. 8th ed. Boston: Brooks Cole, Cengage Learning; 2015.
2. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* 2009;62:5-12.
3. Li J, Fine JP. Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis. *Biostatistics* 2011;12:710-22.
4. Leeflang MM, Rutjes AW, Reitsma JB, et al. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* 2013;185:E537-44.
5. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.
6. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59:1331-2, author reply 32-3.
7. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med* 2017;22:139-42.
8. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
9. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med* 1997;16:981-91.
10. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598-602.
11. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926-30.
12. Chen Y, Liu Y, Ning J, et al. A hybrid model for combining case-control and cohort studies in systematic reviews of diagnostic tests. *J R Stat Soc Ser C Appl Stat* 2015;64:469-89.
13. Chu H, Nie L, Cole SR, et al. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Stat Med* 2009;28:2384-99.
14. Ma X, Nie L, Cole SR, et al. Statistical methods for multivariate meta-analysis of diagnostic tests: An overview and tutorial. *Stat Methods Med Res* 2016;25:1596-619.
15. Hutcheon JA, Chiolerio A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010;340:c2289.

**Competing interests:** Haitao Chu is employed by Pfizer and owns stock in the company. No other competing interests were declared.

This article has been peer reviewed.

**Affiliations:** Evidence-Based Practice Center (Murad, Hasan, Alsibai, Abbas, Wang), Mayo Clinic, Rochester, Minn.; Department of Epidemiology and Biostatistics (Lin), University of Arizona, Tucson, Ariz.; Statistical Research and Data Science Center (Chu), Pfizer, New York, NY; Department of Internal Medicine (Mustafa), University of Kansas Health System, Kansas City, Kan.

**Contributors:** M. Hassan Murad and Zhen Wang contributed to the conception and design of the work. Bashar Hasan, Reem Alsibai and Alzhras Abbas contributed to data acquisition; Lifeng Lin, Haitao Chu, Bashar Hasan, Reem Alsibai, Alzhras Abbas and Zhen Wang contributed to data analysis. All of the authors contributed to interpretation of data. M. Hassan Murad and Zhen Wang drafted the manuscript. All of the authors revised it critically for important intellectual content, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

**Content licence:** This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**Funding:** Lifeng Lin is supported by the United States National Institute of Mental Health (R03 MH128727) and the US National Library of Medicine (R01 LM012982).

**Data sharing:** All data are publicly available. The RCurl package of R can be used to download data from the .rm5 files of the systematic reviews from the Cochrane Database of Systematic Reviews.

**Accepted:** June 1, 2023

**Correspondence to:** M. Hassan Murad,  
Murad.mohammad@mayo.edu