Accuracy of package inserts that accompany SARS-CoV-2 rapid antigen tests

Bigio, Jacob; MacLean, Emily; Das, Rishav; Sulis, Giorgia; Kohli, Mikashmi; Berhane, Sarah; Dinnes, Jacqueline; Deeks, Jon; Brummer, Lucas; Denkinger, Claudia; Pai, Madhukar

DOI:
10.1016/S2666-5247(23)00222-7

Accuracy of package inserts of SARS-CoV-2 rapid antigen tests: a secondary analysis of manufacturer versus systematic review data

Jacob Bigio, Emily L-H MacLean, Rishav Das, Giorgia Sulis, Mikashmi Kohli, Sarah Berhane, Jacqueline Dinnes, Jonathan J Deeks, Lukas E Brümmer, Claudia M Denkinger, Madhukar Pai

Summary

Background Rapid antigen tests (RATs) were crucial during the COVID-19 pandemic. Information provided by the test manufacturer in product package inserts, also known as instructions for use (IFUs), is often the only data available to clinicians, public health professionals, and individuals on the diagnostic accuracy of these tests. We aimed to assess whether manufacturer IFU accuracy data aligned with evidence from independent research.

Methods We searched company websites for package inserts for RATs that were included in the July 2022 update of the Cochrane meta-analysis of SARS-CoV-2 RATs, which served as a benchmark for research evidence. We fitted bivariate hierarchical models to obtain absolute differences in sensitivity and specificity between IFU and Cochrane Review estimates for each test, as well as overall combined differences.

Findings We found 22 (100%) of 22 IFUs of the RATs included in the Cochrane Review. IFUs for 12 (55%) of 22 RATs reported statistically significantly higher sensitivity estimates than the Cochrane Review, and none reported lower estimates. The mean difference between IFU and Cochrane Review sensitivity estimates across tests was 12·0% (95% CI 7·5–16·6). IFUs in three (14%) of 22 diagnostic tests had significantly higher specificity estimates than the Cochrane Review and two (9%) of 22 had lower estimates. The mean difference between IFU and Cochrane Review specificity estimates across tests was 0·3% (95% CI 0·1–0·5). If 100 people with SARS-CoV-2 infection were tested with each of the tests in this study, on average 12 fewer people would be correctly diagnosed than is suggested by the package inserts.

Interpretation Health professionals and the public should be aware that package inserts for SARS-CoV-2 RATs might provide an overly optimistic picture of the sensitivity of a test. Regulatory bodies should strengthen their requirements for the reporting of diagnostic accuracy data in package inserts and policy makers should demand independent validation data for decision making.

Funding None.

Lancet Microbe 2023

Published Online October 13, 2023 https://doi.org/10.1016/S2666-5247(23)00222-7

Research Institute of the McGill University Health Centre, Montreal General Hospital, Montreal, QC, Canada (J Bigio MBBS, R Das MSc); **McGill International TB Centre, Montreal, QC, Canada** (J Bigio, R Das, Prof M Pai PhD); **Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada** (Prof M Pai); **Central Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia** (E L-H MacLean PhD); **School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada** (G Sulis PhD); **Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada** (G Sulis); **FIND, Geneva, Switzerland** (M Kohli PhD); **Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham, UK** (S Berhane PhD, J Dinnes PhD, Prof J J Deeks PhD); **NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK** (S Berhane, J Dinnes, Prof J J Deeks); **Division of Infectious Diseases and Tropical Medicine, Center of Infectious Diseases, Heidelberg University Hospital, Germany** (L E Brümmer BA, C M Denkinger PhD); **German Centre for Infection Research, Partner Site Heidelberg University Hospital, Heidelberg, Germany** (C M Denkinger)

Introduction

The COVID-19 pandemic led to a rapid increase in diagnostic testing, with billions of tests for SARS-CoV-2 performed worldwide since the start of 2020.[1] Among the most widely used diagnostic technologies are rapid antigen tests (RATs), which detect viral antigens using lateral flow immunoassays, typically in nasal or nasopharyngeal swabs. WHO recommends the use of RATs in a variety of settings, including for primary case detection in symptomatic individuals, asymptomatic individuals at high risk of COVID-19,[2] and as self-tests,[3] if they meet minimum performance requirements of at least 80% sensitivity and at least 97% specificity.[2]

Despite the prominent role RATs have in COVID-19 care and control efforts, the diagnostic accuracy (sensitivity and specificity) of RATs varies substantially between manufacturers, with many RATs failing to meet the WHO threshold.[4] Package inserts, also known as instructions for use (IFUs), of RATs frequently report manufacturer-produced diagnostic accuracy data alongside technical descriptions of how to use the test. In the absence of stringent regional regulatory agencies requiring independent diagnostic accuracy data, or in settings with limited access to scientific literature, the package insert information provided by the test manufacturer might be the only data source available to laboratory staff, clinicians, public health professionals, and individuals regarding the diagnostic accuracy of a particular antigen test.

There are few explicit requirements for the design or reporting of IFU diagnostic accuracy studies in applications for CE marking for in vitro diagnostics (IVDs;[5] which allows IVDs to be sold across the European Economic Area) or other forms of regional regulatory approval. Obtaining a CE mark involves a process of assessment through a notifying body, which determines whether a test complies with applicable

Correspondence to:
Dr Jacob Bigio, Research Institute of the McGill University Health Centre, Montreal General Hospital, Montreal, QC H3G 1A4, Canada
jacob.bigio@protonmail.com

Research in context

Evidence before this study

Optimism bias in package inserts has previously been reported for tuberculosis tests. We searched PubMed for full-text articles published from database inception to June 8, 2022, using the search terms ("instructions for use" OR "IFU" OR "package insert") AND ("Covid-19" OR "coronavirus disease 2019" OR "SARS-CoV-2"), with no restrictions on language. The search identified 31 full-text articles, none of which evaluated the diagnostic accuracy data provided in package inserts for COVID-19 diagnostics or compared these data with meta-analysed estimates.

Added value of this study

This is, to our knowledge, the first study to compare diagnostic accuracy estimates for SARS-CoV-2 rapid antigen tests (RATs) given by manufacturers in package inserts with summary meta-analysis estimates of independent research studies. We found that, on average, package inserts overestimated sensitivity substantially, but that there was wide variation in the degree of difference depending on the test.

Implications of all the available evidence

The quality of manufacturer-provided diagnostic accuracy data in package inserts of SARS-CoV-2 RATs is poor and frequently overestimates sensitivity. However, it is often the only data available to laboratory staff, clinicians, public health professionals, and individuals, probably leading to flawed clinical and public health decision making. The poor quality of data partially results from the absence of explicit requirements for the design or reporting of package insert diagnostic accuracy studies when manufacturers apply for regional or global regulatory approval, and mirrors similar findings from a study of tuberculosis diagnostics. Regulatory bodies should strengthen their requirements for the reporting of these data in package inserts, perhaps guided by the domains of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. Test users and policy makers should be aware that package inserts for COVID-19 RATs, in many cases, are providing an overly optimistic picture of the test's sensitivity and demand better quality evidence from independent research studies and meta-analyses.

legislation. However, the current process is one of notification, rather than of evaluation, and does not require the same level of investigation as for a drug or vaccine. Even when manufacturers apply for evaluation by the Global Fund Expert Review Panel for Diagnostic Products (ERPD)[6] or for WHO prequalification, for which guidance is provided for writing IFUs,[7] the requirements call for little more than a brief description of the study design, location, reference standard, and specimen type along with the results.

Clinical and public health decision makers, therefore, require an understanding of the quality, reliability, and applicability of the accuracy data presented in the package inserts, particularly as many of the RATs received emergency-use authorisations (EUAs), which did not follow the usual regulatory control mechanisms.

One way to assess the quality of data released by manufacturers is to compare sensitivity and specificity estimates in package inserts with test-specific estimates from peer-reviewed published research, synthesised in meta-analyses, an approach previously applied to tuberculosis diagnostics by Denkinger and colleagues.[8] They found what they described as optimism bias—ie, package inserts of tuberculosis diagnostic tests reported optimistic accuracy estimates when compared with estimates from published meta-analyses.

We compared the sensitivity and specificity estimates presented in package inserts for SARS-CoV-2 RATs with the estimates given for the same tests in a Cochrane Review[4] that provided summary meta-analysis estimates of the accuracy of SARS-CoV-2 RATs.

Methods

Study design

The Cochrane Review of SARS-CoV-2 RATs is a comprehensive systematic review and meta-analysis of available published and preprint diagnostic accuracy studies, which includes test-specific sensitivity and specificity estimates.[4] We chose this review for our secondary analysis as Cochrane Reviews are known for their methodological rigour and transparency of reporting[9,10] and it provides substantial details on the data sources used, the number of studies for each test, and other relevant parameters. Since its publication in July, 2022, it has been cited over 1000 times.

We used the Cochrane Review definition of RATs as antigen-detecting tests for current SARS-CoV-2 infection, which "have the capacity to be performed at the point of care or in a 'near-patient' testing role".[4] Only RATs included in the Cochrane Review update published on July 22, 2022,[4] which included diagnostic accuracy studies published up to March 8, 2021, were eligible for inclusion in our study. Future mention of the Cochrane Review in this Article refers specifically to this July 2022 update.

We searched company websites and other online sources up to Aug 9, 2022, for package inserts available in English of RATs included in the Cochrane Review. Some package inserts were additionally obtained via communication with contacts at FIND, the global alliance for diagnostics. RATs were included in our study when the package insert was available and contained sensitivity and specificity data compared with a PCR reference standard, with any type of sample and sample collection method, and study participants from any setting (including participants both

symptomatic or asymptomatic for COVID-19). Positive or negative percentage agreement with a PCR test was treated as sensitivity or specificity. If multiple versions of a package insert were available, data from the most recent version were used. RATs were excluded from the study if the package insert did not include both sensitivity and specificity data or if data were produced exclusively from contrived samples (in which a known amount of SARS-CoV-2 virus or its components is added to a swab from a healthy individual to imitate an infected sample). Test names are given as presented in the Cochrane Review for clarity.

No ethical approval was sought for our study as it involved no human participants and was a secondary analysis of publicly available data from test manufacturers and a published Cochrane Review.

Data collection

The following data were extracted from the package inserts: sensitivity and specificity estimates, with 95% CIs where available, numbers of true positives, true negatives, false positives and false negatives, sample size, sample collection method, and symptom status of patients. If the package insert presented data from multiple evaluations, the data were meta-analysed.

The following data were extracted from the Cochrane Review: sensitivity and specificity, with binomial exact 95% CIs, numbers of true positives, true negatives, false positives and false negatives, sample size, and number of studies meta-analysed.

As well as the overall summary sensitivity and specificity of each RAT, the Cochrane Review presented a sensitivity analysis of test sensitivity and specificity according to study "compliance with manufacturer IFUs".[4] Test evaluations were considered manufacturer-compliant when the study followed manufacturer instructions on three aspects of testing: sample type, use of viral transport medium, and timing between sample collection and testing. The data we used from the Cochrane Review were the sensitivity and specificity estimates from these IFU-compliant sensitivity analyses, as the comparison is fairest when diagnostic accuracy studies of a test were

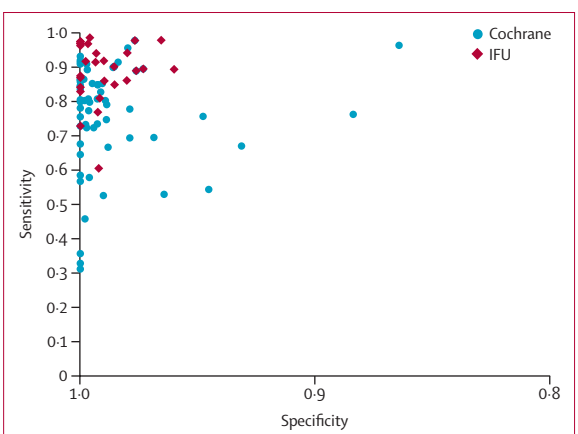

Figure 1: **Receiver operating characteristic plot displaying sensitivity and specificity estimates from the IFU and the Cochrane Review for the 22 included rapid antigen tests**
IFU=instructions for use.

	Number of evaluations; samples (cases)	Sensitivity (95% CI)	Specificity (95% CI)
AAZ, COVID-VIRO*			
Cochrane Review	2; 572 (239)	91·3 (78·2, 96·9)	94·0 (90·9 to 96·1)†
IFU	1; 226 (117)	96·6 (91·2 to 98·7)	100 (96·7 to 100)
Difference (95% CI), p value	NA	5·3 (–3·9 to 14·4), p=0·26	6·0 (3·5 to 8·6), p=0·0058‡
Abbott, BinaxNOW COVID-19 Ag card			
Cochrane Review	4; 2018 (358)	80·9 (67·6 to 89·6)	99·9 (99·5 to 100)
IFU	1; 460 (117)	84·6 (76·9 to 90·1)	98·5 (96·5 to 99·4)
Difference (95% CI), p value	NA	3·7 (–9·1 to 16·4), p=0·57	–1·3 (–2·6 to –0·06), p=0·040
Abbott, Panbio Covid-19 Ag			
Cochrane Review	11; 7718 (1397)	77·3 (68·7 to 84·0)	99·7 (99·5 to 99·8)
IFU	1; 585 (140)	91·4 (85·5 to 95·1)	99·8 (98·4 to 100)
Difference (95% CI), p value	NA	14·2 (5·2 to 23·1), p=0·0020	0·08 (–0·4 to 0·5), p=0·75
Access Bio, CareStart Covid-19 Ag			
Cochrane Review	1; 241 (69)	75·4 (63·9 to 84·1)§	94·8 (90·3 to 97·3)§
IFU	2; 272 (71)	90·1 (80·7 to 95·2)	99·5 (96·6 to 99·9)
Difference (95% CI), p value	NA	14·8 (2·5 to 27·1), p=0·019	4·7 (1·3 to 8·2), p=0·0070
Becton Dickinson, BD Veritor			
Cochrane Review	1; 1384 (116)	66·4 (57·0 to 74·9)	98·8 (98·1 to 99·3)
IFU	1; 226 (31)	83·9 (66·3 to 94·5)	100 (98·1 to 100)
Difference (95% CI), p value	NA	17·5 (2·0 to 33·0), p=0·077‡	1·2 (0·6 to 1·8), p=0·24‡
BIONOTE, NowCheck COVID-19 Ag			
Cochrane Review	2; 618 (181)	89·5 (84·1 to 93·2)	97·7 (95·8 to 98·8)
IFU	1; 400 (102)	89·2 (81·6 to 93·9)	97·3 (94·7 to 98·7)
Difference (95% CI), p value	NA	–0·3 (–7·8 to 7·2), p=0·94	–0·4 (–2·7 to 1·9), p=0·74
BIONOTE, NowCheck COVID-19 Ag (Nasal)			
Cochrane Review	1; 218 (79)	89·9 (81·0 to 95·5)	98·6 (94·9 to 99·8)
IFU	1; 218 (79)	89·9 (81·0 to 95·5)	98·6 (94·9 to 99·8)
Difference (95% CI), p value	NA	NA	NA
Coris Bioconcept, COVID-19 Ag Respi-Strip*			
Cochrane Review	3; 765 (408)	34·3 (29·9 to 39·1)	100 (99·0 to 100)†
IFU	3; 508 (183)	79·3 (59·9 to 90·8)	99·4 (97·8 to 99·9)†
Difference (95% CI), p value	NA	45·0 (28·8 to 61·1), p<0·0001	–0·6 (–1·5 to 0·2), p=0·23‡
Denka Co, QuickNavi COVID-19 Ag*			
Cochrane Review	2; 1633 (123)	84·2 (66·2 to 93·5)	100 (99·8 to 100)
IFU	2; 2048 (156)	81·3 (69·9 to 89·0)	100 (99·8 to 100)
Difference (95% CI), p value	NA	–2·9 (–19·3 to 13·4), p=0·73	NA
ECODiagnostica, COVID-19 Ag ECO			
Cochrane Review	1; 150 (55)	69·1 (55·2 to 80·9)	97·9 (92·6 to 99·7)
IFU	1; 426 (115)	96·5 (91·3 to 99·0)	99·7 (98·2 to 100)
Difference (95% CI), p value	NA	27·4 (14·8 to 40·1), p<0·0001‡	1·8 (–1·2 to 4·7), p=0·14‡

(Table continues on next page)

	Number of evaluations; samples (cases)	Sensitivity (95% CI)	Specificity (95% CI)
(Continued from previous page)			
Innova Medical Group, SARS-CoV-2 Ag			
Cochrane Review	1; 1676 (372)	57·5 (52·3 to 62·6)	99·6 (99·1 to 99·9)
IFU	1; 295 (75)	96·0 (88·8 to 99·2)	100 (98·3 to 100)
Difference (95% CI), p value	NA	38·5 (31·8 to 45·2), p<0·0001‡	0·4 (0·05 to 0·7), p=1·0‡
LumiraDx, SARS-CoV-2 Ag			
Cochrane Review	2; 741 (177)	91·2 (70·0 to 97·9)	98·6 (97·2 to 99·3)
IFU	2; 512 (123)	97·6 (92·7 to 99·2)	97·2 (95·0 to 98·4)
Difference (95% CI), p value	NA	6·4 (–5·9 to 18·7), p=0·31	–1·4 (–3·3 to 0·5), p=0·15
Mologic, COVID 19 Rapid Ag			
Cochrane Review	1; 650 (192)	90·6 (85·6 to 94·3)	100 (99·2 to 100)
IFU	1; 214 (113)	85·8 (78·0 to 91·7)	98·0 (93·0 to 99·8)
Difference (95% CI), p value	NA	–4·8 (–12·4 to 2·9), p=0·26‡	–2·0 (–4·7 to 0·7), p=0·032‡
Precision Biosensor, Exdia COVID-19 Ag			
Cochrane Review	1; 293 (90)	52·2 (41·4 to 62·9)	99·0 (96·5 to 99·9)
IFU	1; 99 (49)	93·9 (83·1 to 98·7)	98·0 (89·4 to 99·9)
Difference (95% CI), p value	NA	41·7 (29·3 to 54·0), p<0·0001‡	–1·0 (–5·1 to 3·1), p=0·49‡
Quidel, SOFIA SARS Antigen FIA‡			
Cochrane Review	3; 1000 (144)	76·4 (68·8 to 82·6)	99·5 (98·8 to 99·9)†
IFU	1; 209 (30)	96·7 (79·8 to 99·5)	100 (98·0 to 100)
Difference (95% CI), p value	NA	20·3 (10·8 to 29·7), p<0·0001	0·5 (0·01 to 0·9), p=1·0‡
RapiGEN, BIOCREDIT COVID-19 Ag			
Cochrane Review	2; 582 (195)	66·3 (52·9 to 77·5)	99·0 (97·3 to 99·6)
IFU	1; 161 (63)	85·7 (74·8 to 92·4)	99·0 (93·1 to 99·9)
Difference (95% CI), p value	NA	19·4 (4·2 to 34·6), p=0·013	0·01 (–2·2 to 2·2), p=0·99
SD Biosensor, STANDARD F COVID-19 Ag			
Cochrane Review	2; 1129 (159)	75·5 (68·2 to 81·5)	97·2 (96·0, 98·1)
IFU	1; 155 (55)	89·1 (77·8 to 95·0)	96·0 (89·8 to 98·5)
Difference (95% CI), p value	NA	13·6 (3·0 to 24·2), p=0·012	–1·2 (–5·2 to 2·8), p=0·55
SD Biosensor, STANDARD Q COVID-19 Ag (Nasal)*			
Cochrane Review	4; 621 (189)	85·2 (79·4 to 89·6)	99·3 (98·0 to 99·9)†
IFU	1; 503 (104)	97·1 (91·4 to 99·1)	100 (99·1 to 100)
Difference (95% CI), p value	NA	11·9 (5·9 to 17·9), p<0·0001	0·7 (–0·09 to 1·5), p=0·25*
SD Biosensor/Roche, STANDARD Q COVID-19 Ag			
Cochrane Review	15; 5116 (1197)	84·0 (79·2 to 87·9)	99·2 (98·8 to 99·4)
IFU	2; 1659 (153)	84·4 (74·6 to 90·9)	98·9 (98·3 to 99·3)
Difference (95% CI), p value	NA	0·3 (–8·8 to 9·5), p=0·94	–0·2 (–0·8 to 0·3), p=0·41
Shenzhen Bioeasy Biotech, 2019-nCoV Ag			
Cochrane Review	2; 855 (40)	72·5 (56·8 to 84·1)	92·5 (90·5 to 94·1)
IFU	1; 483 (362)	80·7 (76·3 to 84·4)	99·2 (94·4 to 99·9)
Difference (95% CI), p value	NA	8·2 (–6·3 to 22·6), p=0·27	6·7 (4·2 to 9·1), p<0·0001

(Table continues on next page)

performed according to manufacturer instructions. Tests for which the Cochrane Review did not include IFU-compliant estimates for both sensitivity and specificity were excluded.

Further sensitivity analyses in the Cochrane Review separated IFU-compliant data for symptomatic and asymptomatic participants.[4] We a priori assumed that the IFUs would give higher estimates for sensitivity than the Cochrane Review based on similar work by members of our team on tuberculosis diagnostic IFUs.[8] Therefore, for our analysis of sensitivity data, we used Cochrane Review IFU-compliant summary estimates for symptomatic participants, as we assumed that these estimates would be higher than the estimates for both symptomatic and asymptomatic participants combined and so represented the most conservative option for our analysis. For specificity, we also used combined Cochrane Review IFU-compliant summary estimates for symptomatic participants.

Outcomes

The primary outcome of this study was a calculation of the differences in sensitivity and specificity estimates provided by the IFU and the Cochrane Review for each SARS-CoV-2 RAT, as well as overall combined values across tests.

Statistical analysis

Analyses were performed using Stata/SE, version 17.0. To compare the estimates of sensitivity and specificity for each RAT we fitted bivariate hierarchical models[11,12] via the meqrlogit command. We compared studies from the Cochrane Review with those from the IFU by including an indicator variable in the random-effects logistic regression models. To reproduce the summary sensitivity and specificity estimates from the Cochrane Review, the same bivariate models were fitted, but with additional parameters added to model the IFU data. While modelling the latter, due to the very small number of evaluations, models were simplified firstly by assuming no correlation between sensitivity and specificity estimates, and secondly by setting near-zero variance estimates of the random effects to zero.[13] Where there was only one IFU evaluation, random effects were also set to zero. We obtained absolute differences in sensitivity or specificity and corresponding p values and 95% CIs post estimation by using the model parameters and the nlcom command in Stata. In instances where only one study was available per test or when tests were being directly compared following summing of counts of the 2×2 tables, we performed test comparison by calculating point estimates and 95% CIs for a risk difference.

In addition to the outcomes according to each RAT, overall difference in summary sensitivity and specificity between the package inserts and Cochrane Review was calculated. A single bivariate model that included a

covariate term (indicating IFU or Cochrane) was fitted to the entire data. Zero values in the difference indicate agreement of the estimates in the IFU with the pooled data in the Cochrane Review. Positive differences indicate that the estimates in the IFU were higher than those in the Cochrane Review and negative differences indicate that estimates in the IFU were lower than those in the Cochrane Review. The differences show the percentage difference of COVID-19 cases claimed as RAT positives for sensitivity, and of COVID-19 non-cases claimed as RAT negatives for specificity, when comparing IFU and Cochrane Review.

Role of the funding source

There was no funding source for this study.

Results

This secondary analysis was conducted between Nov 23, 2021, and May 5, 2023. Of the 22 RATs for which the Cochrane Review presented IFU-compliant sensitivity and specificity data (of 50 RATs analysed in the Cochrane Review), IFUs could be retrieved for all 22 (100%). All 22 included IFUs used PCR reference standards. Ten (45%) IFUs presented data from symptomatic patients only, eight (36%) did not specify the symptom status of the patients, two (9%) had both symptomatic and asymptomatic patients, and two (9%) presented multiple evaluations where some evaluations did not specify the symptom status and others had both symptomatic and asymptomatic patients.

17 (77%) of the diagnostic tests have received EUAs from the US Food and Drug Administration[14] or are included on the EU common list of COVID-19 antigen tests,[15] whereas five (23%) received approval from neither agency. Full manufacturer details and information on regulatory approval are given in the appendix (pp 1–2). Although the results of the quality assessment in the Cochrane Review were not separated by diagnostic test, 142 (93%) of 152 included studies had high risk of bias in at least one domain of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[16] A receiver operating characteristic plot displaying sensitivity and specificity estimates from the IFUs and the Cochrane Review for the 22 included RATs is shown in figure 1.

All studies included in the Cochrane Review used PCR reference standards to diagnose cases. The median sample size of PCR-diagnosed cases per RAT was 109 (IQR 71–140) in IFUs and 168 (90–239) in the Cochrane Review. The number of studies pooled for test-specific Cochrane Review sensitivity estimates ranged from two to 15. In 12 (55%) of 22 RATs, the IFU estimates for sensitivity were statistically significantly higher at the 5% level than the Cochrane Review estimates (table). In ten (45%) of 22, there was no statistically significant difference between the IFU and Cochrane Review estimates. No IFU estimated statistically significantly lower sensitivity than the Cochrane Review. The mean difference between IFU and Cochrane Review sensitivity estimates across tests was 12·0% (95% CI 7·5–16·6).

All studies included in the Cochrane Review used PCR reference standards or pre-pandemic respiratory samples to determine SARS-CoV-2 infection negativity. The median sample size of PCR negatives per RAT was 211 (IQR 109–389) in IFUs and 448 (203–1268) in the Cochrane Review. The number of studies pooled for test-specific Cochrane Review specificity estimates ranged from two to 15. In three (14%) of 22 RATs, the IFU estimate for specificity was statistically significantly higher than the Cochrane Review estimate (table). In two (9%) of 22 RATs, the IFU estimate was statistically significantly lower than the Cochrane Review estimate (table). In 17 (77%) of 22 RATs, there was no statistically significant difference between the IFU and Cochrane Review estimates. The mean difference between IFU and Cochrane Review specificity estimates across tests was 0·3% (95% CI 0·08–0·5).

IFU and Cochrane Review sensitivity and specificity estimates, with sample sizes and percentage differences, are shown in the table. The percentages differences arranged in descending order by sensitivity difference are shown in figure 2. A forest plot showing sensitivity and specificity data from all studies included in the IFUs and Cochrane Review arranged by test name is shown in the appendix (p 3).

See Online for appendix

	Number of evaluations; samples (cases)	Sensitivity (95% CI)	Specificity (95% CI)
(Continued from previous page)			
Siemens, CLINITEST Rapid COVID-19 Ag*			
Cochrane Review	1; 178 (91)	80·2 (70·8 to 87·2)§	100 (95·8 to 100)
IFU	2; 1102 (228)	97·8 (94·8 to 99·1)	99·7 (99·0 to 99·9)†
Difference (95% CI), p value	NA	17·6 (9·2 to 26·0), p<0·0001	−0·3 (−0·7 to 0·04), p=1·0‡
Sugentech, SGTI-flex COVID-19 Ag			
Cochrane Review	1; 106 (78)	52·6 (40·9 to 64·0)	96·4 (81·7 to 99·9)
IFU	1; 183 (83)	91·6 (83·4 to 96·5)	99·0 (94·6 to 100)
Difference (95% CI), p value	NA	39·0 (26·4 to 51·6), p<0·0001‡	2·6 (−4·6 to 9·7), p=0·39‡
All tests combined			
Cochrane Review	63; 28264 (5949)	79·0 (75·1 to 82·4)	99·0 (98·8 to 99·1)
IFU	29; 10944 (2549)	91·0 (87·9 to 93·4)	99·3 (99·1 to 99·4)
Difference (95% CI), p value	NA	12·0 (7·5 to 16·6), p<0·0001	0·3 (0·08 to 0·5), p=0·0090

IFU=instructions for use. NA=not applicable. *Sensitivity and specificity pooled separately. †2×2 data combined before calculating estimates. ‡Derived using Stata's csi command to calculate point estimates and CIs for the risk difference. §Method used for deriving CIs (bivariate model) differs from the original Cochrane Review (exact/Clopper-Pearson binomial CIs).

Table 1: **Comparison of sensitivity and specificity estimates in package inserts and Cochrane Review for antigen tests for SARS-CoV-2 infection**

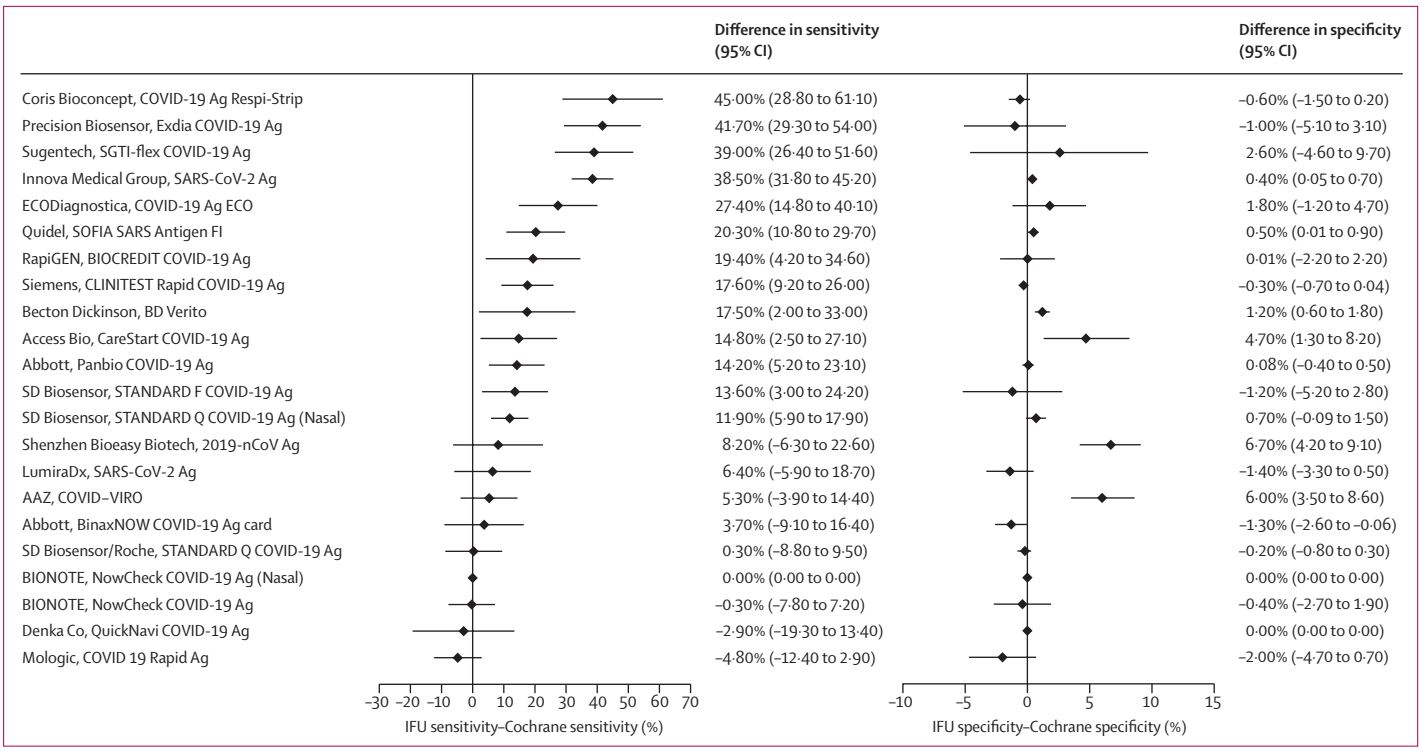

Figure 2: **Differences in sensitivity and specificity estimates between IFUs and Cochrane Review for the 22 included rapid antigen tests**
IFU=instructions for use.

Discussion

If 100 people with SARS-CoV-2 infection were tested with each of the RATs in this study, the Cochrane Review estimates suggest that, on average, 12 fewer people would be correctly diagnosed than is suggested by estimates from the package inserts. If 100 people with SARS-CoV-2 infection were tested with the RAT with the most optimistic package insert estimate in our dataset, 45 fewer people would be correctly diagnosed than is suggested by that package insert estimate. These findings have important implications for public health. Results from RATs are frequently used to determine whether someone should begin, or continue, isolating and to contribute data to local and national surveillance systems, and package inserts often provide the only data available to laboratory staff, clinicians, public health professionals, and individuals on the accuracy of a test. Many of these RATs are also sold as self-tests, in line with WHO guidance.[3] The poor quality of these data is, therefore, likely to have led to flawed clinical, public health, and individual decision making, with impacts on individuals using the tests and those around them, as well as on the overall control of the spread of COVID-19.

More broadly, the results of our analysis echo those of the similar study of tuberculosis diagnostic tests conducted by Denkinger and colleagues,[8] which also found that IFUs substantially overestimate sensitivity in comparison with published meta-analyses, although the authors did not compute a pooled estimate of the overestimation. Further work looking at package inserts for tests for other diseases is required to determine whether tests for tuberculosis and SARS-CoV-2 infection are outliers, or whether this problem is common to many diseases. Nevertheless, our results suggest that at the start of a new pandemic, or on the introduction of a new diagnostic test or class of tests for a particular condition, test users should be aware that package inserts might be providing an overly optimistic picture of the sensitivity of a test. Unfortunately, the wide variation in the quality of package insert data means that some tests will be unfairly maligned by this suspicion—45% of tests in our study had sensitivity estimates that were not statistically different in the IFU and Cochrane Review.

None of the IFUs presented sensitivity estimates that were significantly lower than the Cochrane Review. There are clear reasons why manufacturers might produce optimistic diagnostic accuracy estimates. In a competitive market, it is commercially advantageous to advertise high accuracy and there are few explicit requirements for the design and reporting of IFU diagnostic accuracy studies in applications for CE marking, Global Fund ERPD Review,[6] or WHO prequalification.[7] IFU diagnostic accuracy studies tend to have small sample sizes, are often based in a single setting in a single country, and might use a case-control design, which is well known to overestimate test accuracy,[17] since people with advanced

disease are often compared with healthy people. Given the range of biases described in the QUADAS-2 tool[16]—including around patient selection, conduct of the index test and reference standard, and flow and timing of patients through the study—the information provided in even the most explicit IFU we reviewed in this study was inadequate to judge the quality of the data.

These findings mirror those of a 2021 study examining package inserts for SARS-CoV-2 RATs included on the approved list of the German Federal Institute for Drugs and Medical Devices.[18] Two-thirds of package inserts reviewed in that work contained no information on study design, whereas more than 80% contained no clinical information, such as whether patients were symptomatic or asymptomatic. The same study noted that RATs included on the approved list showed substantial variability in manufacturer claims of test performance, although the authors did not compare the package insert data with independent studies or meta-analyses of the diagnostic accuracy of the same RATs. Professional users of RATs should be critical of the quality of data provided to them by manufacturers, particularly as similar limitations regarding the reliability and accuracy of these data can be expected for other commercial rapid tests in the future.

Given that package insert data is often the only source of information available to users regarding the diagnostic accuracy of a test, regulatory bodies should strengthen their requirements for the reporting of these data in IFUs when companies apply for WHO prequalification, Global Fund ERPD review, or medical device approval from national regulatory bodies such as the Food and Drug Administration in the USA. For example, companies could be required to report more details on the study design in the package insert, perhaps guided by the domains of the QUADAS-2 tool. In many cases, this information already exists and is shared with regulatory authorities, but is not currently included in IFUs. Care should be taken to avoid making these requirements onerous for companies, however, as rapid innovation in diagnostic testing for SARS-CoV-2 at the start of the pandemic required companies to take financial risks and produced widespread benefits. Given that many of the RATs are also marketed as self-tests, regulatory bodies should additionally require manufacturers to include extra package insert information written for the general public, as all IFUs reviewed in this study used highly technical language. For example, a lay summary could explain what sensitivity means in terms of the number of people out of a thousand with SARS-CoV-2 infection who would incorrectly test negative using the test.

Many available RATs for SARS-CoV-2 have not been evaluated in independent clinical studies and it might, therefore, be incumbent on the user—whether at national, regional, or even institutional level—to conduct validation studies before roll-out. The conduct of high-quality independent diagnostic accuracy studies should be encouraged in general and might be particularly important for low-income and middle-income settings, as these are rarely the site for such studies and might be under-represented in subsequent meta-analyses. The GRADE approach provides a framework for guideline development relating to diagnostic tests.[19] Finally, manufacturers should be encouraged to update the diagnostic accuracy data they include in newer versions of product IFUs to incorporate data from independent evaluations or meta-analyses.

Our study had four main limitations. First, although Cochrane Reviews are recognised as the highest international standard of systematic reviews,[20] the systematic review process inherently relies on the availability and methodological quality of relevant published and preprint studies. Many of the studies included in the Cochrane Review we used had methodological limitations.[4] Therefore, although the Cochrane Review estimates can be seen as the best available evidence, they are not perfect. Second, nine (41%) of 22 Cochrane Review values for sensitivity and specificity were based on only one study and were not actually meta-analyses. However, in these cases with only one Cochrane Review study, the ratios of average sample sizes in IFU:Cochrane Review were 0·74 for sensitivity and 0·58 for specificity, so the average sample sizes were larger in the Cochrane Review. Third, although it is difficult to quantify, some of the difference in IFU and Cochrane Review estimates might be due to manufacturer diagnostic accuracy studies being performed on patients infected with original SARS-CoV-2 strains, whereas Cochrane Review estimates might include studies performed on patients infected with later variants. However, this possibility does not change the conclusion that package inserts frequently provide diagnostic accuracy data that do not correspond with real-world performance. Finally, to produce a fair comparison with the data provided by the manufacturers in the package inserts, we used IFU-compliant data from the Cochrane Review. However, in practice, RATs are often used in ways that do not follow the manufacturer's instructions—only 43% of studies included in the Cochrane Review were compliant with the IFU, with a further 19% not providing enough information to judge. If tests are used without following manufacturer instructions, the difference between accuracy in practice and accuracy reported by manufacturers might differ from what we have shown in ways that are difficult to quantify.

In summary, our study found that the diagnostic accuracy data reported in most of the package inserts for SARS-CoV-2 RATs we identified overestimated sensitivity in comparison with Cochrane Review meta-analyses of the same tests. As these data are frequently used to inform clinical, public health, and personal decision making, particularly in the early stages of a pandemic when few other sources of information are available, test users should be aware that package inserts often provide

an overly optimistic picture of the sensitivity of a test. Where possible, data from independent validation studies should be used for policy and decision making. Regulatory bodies should strengthen their requirements for the reporting of data in IFUs.

Contributors

JB and MP conceptualised the study. JB, EL-HM, RD, GS, and SB curated the data. JB, EL-HM, and SB did the formal analysis. JB, EL-HM, RD, GS, and SB did the investigation. JB, EL-HM, GS, MK, SB, JD, JJD, and MP were responsible for the study methods. JB was responsible for project administration. JB and JD sourced resources. JB, EL-HM, and SB ran the software. MP supervised the study. JB, EL-HM, RD, and GS validated the data. JB and SB did the original data visualisation. JB wrote the original draft of the Article. JB, EL-HM, RD, GS, MK, SB, JD, JJD, LEB, CMD, and MP reviewed and edited drafts of the Article. JB and RD accessed and verified all the data in this study. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

JB has received consulting fees from the New Diagnostics Working Group of the Stop TB Partnership. SB, JD, and JJD are supported by the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre. This Article presents independent research supported by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham National Health Services (NHS) Foundation Trust and the University of Birmingham. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. CMD has received research grants from the US National Institutes of Health, German Ministry of Education and Research, German Alliance for Global Health Research, United States Agency for International Development, FIND, German Center for Infection Research, UNAIDS, WHO, and Roche. CMD also declares a payment from Roche Diagnostics that she accepted as German law requires a manufacturer to pay for the use of data for regulatory purposes. Data was generated as part of an independent evaluation by CMD and team. CMD is an academic editor of *PLoS Medicine* and sits on the WHO technical advisory group on tuberculosis diagnostics. MP holds a Canada Research Chair Award from the Canadian Institutes of Health Research and serves as an adviser to the following non-profit agencies in global health: Bill & Melinda Gates Foundation; Foundation for Innovative New Diagnostics, WHO, and Stop TB Partnership. All other authors declare no competing interests.

Data sharing

Cochrane Review data are publicly available at https://www.cochranelibrary.com/cdsr/doi/10·1002/14651858.CD013705.pub3/full. Package inserts and our Stata dataset and code are available at https://www.doi.org/10.17632/73shkb5kr2.1.

References

1 Our World in Data. Total COVID-19 tests. 2022. https://ourworldindata.org/grapher/full-list-total-tests-for-covid-19?country= (accessed July 21, 2022).

2 WHO. Antigen-detection in the diagnosis of SARS-CoV-2 infection: interim guidance. Geneva: World Health Organization, 2021.

3 WHO. Use of SARS-CoV-2 antigen-detection rapid diagnostic tests for COVID-19 self-testing: interim guidance. Geneva: World Health Organization, 2022.

4 Dinnes J, Sharma P, Berhane S, et al. Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Systematic Rev* 2022; **7:** CD013705.

5 Grifa RA, Pozzoli G. Performance evaluation of in vitro diagnostic medical devices: methodology and differences compared to studies on other medical devices. *Microchem J* 2018; **136:** 279–82.

6 The Global Fund. Diagnostic product questionnaire for product evaluation by the Global Fund Expert Review Panel for Diagnostic Products. Version 6. 2017. https://www.theglobalfund.org/media/5845/psm_expertreviewpanelproductevaluationfordiagnosticproducts_questionnaire_en.docx (accessed Oct 12, 2022).

7 WHO. Technical guidance series for WHO prequalification – diagnostic assessment. Designing instructions for use for in vitro diagnostic medical devices. Geneva: World Health Organization, 2017.

8 Denkinger CM, Grenier J, Minion J, Pai M. Promise versus reality: optimism bias in package inserts for tuberculosis diagnostics. *J Clin Microbiol* 2012; **50:** 2455–61.

9 Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; **149:** 889–97.

10 Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; **4:** e78.

11 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:** 1331–32, author reply 1332–33.

12 Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; **58:** 982–90.

13 Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017; **26:** 1896–911.

14 Food and Drug Administration. In vitro diagnostics EUAS - antigen diagnostic tests for SARS-CoV-2. 2022. https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas-antigen-diagnostic-tests-sars-cov-2 (accessed Oct 12, 2022).

15 European Commission Directorate-General for Health and Food Safety. EU Common list of COVID-19 antigen tests. 2022. https://health.ec.europa.eu/system/files/2022–07/covid-19_eu-common-list-antigen-tests_en.pdf (accessed Oct 11, 2022).

16 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.

17 Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; **282:** 1061–66.

18 Özcürümez M, Katsounas A, Holdenrieder S, von Meyer A, Renz H, Wölfel R. Assessment of SARS-CoV-2 rapid antigen tests. 2021; **45:** 143–8.

19 Schünemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008; **336:** 1106–10.

20 Cipriani A, Furukawa TA, Barbui C. What is a Cochrane review? *Epidemiol Psychiatr Sci* 2011; **20:** 231–33.