Risk of bias and limits of reporting in diagnostic accuracy studies for commercial point-of-care tests for respiratory pathogens

J.M. Hughes, C. Penney, S. Boyd and P. Daley

Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

Original Paper

Abstract

Commercial point-of-care (POC) diagnostic tests for Group A *Streptococcus*, *Streptococcus pneumoniae*, and influenza virus have large potential diagnostic and financial impact. Many published reports on test performance, often funded by diagnostics companies, are prone to bias. The Standards for Reporting of Diagnostic Accuracy (STARD 2015) are a protocol to encourage accurate, transparent reporting. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool evaluates risk of bias and transportability of results. We used these tools to evaluate diagnostic test accuracy studies of POC studies for three respiratory pathogens. For the 96 studies analysed, compliance was <25% for 14/34 STARD 2015 standards, and 3/7 QUADAS-2 domains showed a high risk of bias. All reports lacked reporting of at least one criterion. These biases should be considered in the interpretation of study results.

Key Findings

1. Commercial point-of-care diagnostic tests for respiratory pathogens have large potential financial impact.
2. Published reports on test performance, often funded by diagnostics companies, may be prone to bias.
3. We reviewed published studies and identified gaps in reporting and risk of bias.
4. These biases should be considered in the appropriate application of point-of-care testing.

Introduction

Group A *Streptococcus*, *Streptococcus pneumoniae*, and influenza virus are common respiratory pathogens in both community and hospital settings. Group A *Streptococcus* causes acute pharyngitis, rheumatic fever, rheumatic heart disease and acute post-streptococcal glomerulonephritis [1]. *S. pneumoniae* causes community-acquired pneumonia and meningitis, with significant economic burden [2]. Influenza, including pandemic strains such as H1N1 in 2009, causes morbidity and mortality despite vaccination, due to antigenic shift and an ageing population [3].

Point-of-care (POC) diagnostic testing is defined as testing conducted at the bedside by primary care physicians and nurses to provide rapid diagnostic information for clinical decision-making [4]. POC tests for respiratory infections such as Group A *Streptococcus*, *S. pneumoniae*, and influenza virus may be applied to respiratory specimens, or to urine [5–7]. These tests are often lateral flow devices containing paper impregnated with antibodies specific to antigens found in the pathogen of interest. When the device is exposed to a patient specimen containing the antigen of interest, there is a colour change indicating a positive test result without the use of costly laboratory equipment [8].

The clinical and economic impact of POC diagnostic testing is unknown, though POC testing for influenza has been shown to reduce the use of additional tests, decrease patient time in emergency departments, and decrease costs [5] It has also decreased overall antibiotic use and duration, and encouraged appropriate use of antivirals when positive [6]. POC testing for Group A *Streptococcus* may streamline management of patients with pharyngitis [7].

Considering potential beneficial cost impacts, many reports are published on POC test performance. Many published studies reporting these quality studies are funded by diagnostics companies, and are prone to biases in design or reporting, which may overestimate, or underestimate, the true performance of these tests. Biased estimates of test performance may lead to biased estimates of the impact of POC tests.

In 2003, the Standards for Reporting of Diagnostic Accuracy (STARD) Initiative published a standard list and flow diagram for studies of diagnostic accuracy, to encourage accurate, transparent reporting [9]. This was further updated in 2015 with expanded criteria [10]. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool for evidence-based

assessment of diagnostic accuracy in systematic reviews was developed in 2003 and revised in 2011 [11, 12]. QUADAS evaluates study design and identifies potential sources of bias using seven different criteria, as well as threats to transportability of research results.

In 2009, Fontela *et al.* reported on the quality of 90 studies of POC tests for HIV, tuberculosis and malaria using QUADAS and STARD, and demonstrated moderate- to low-quality and poor reporting [13]. Common sources of bias were inadequate description of withdrawals and reference test execution, index test and reference test review bias, and under-reporting of uninterpretable results. Less than 25% of the included studies reported methods for calculation and estimates of test reproducibility, adverse effects, estimates of diagnostic accuracy between subgroups, distribution of severity of disease or other diagnoses in study participants, number of eligible patients who did not participate in the study, blinding of the test readers, description of the team executing the test and management of indeterminate, invalid or outlier results [13]

In 2013, Jafari *et al.* investigated quality of studies of POC tests for *Treponema pallidum* ($n = 33$) using the same criteria, and demonstrated that five quality items remained unaddressed in 60% of papers, and clinical review bias, index test review bias, lack of reporting of uninterpretable results, reference test review bias and poor description of loss of patients were detected [14]. Wilczynski reviewed 240 studies from 2001 to 2005 and showed no improvement in completeness of reporting of following the initial publication of the STARD criteria [15].

Our objective was to explore completeness of reporting and limitations of applicability among studies reporting POC diagnostic testing performance for Group A *Streptococcus*, *S. pneumoniae*, and influenza virus.

Methods

Article search strategy

A PubMed search was conducted using defined search criteria ('Group A Streptococcus [explode]', 'Pneumococcus [explode]', 'Influenza [explode]'). Studies were screened based on title and keywords and abstract by author MH. Inclusion and exclusion criteria were then applied to full-text studies by MH and PD.

Screening paper eligibility

The inclusion criteria were original studies published in English in peer-reviewed journals, which included clinical specimens from human subjects, were published between 2004 and 2015, and which reported diagnostic performance of commercial POC tests for either Group A *Streptococcus*, *S. pneumoniae*, or influenza A or B. Studies were excluded if they did not include original data, if tests could not be performed outside of a microbiological laboratory, if full text was not available or duplicate reports.

Data abstraction

The data extracted from included studies were year of publication, continent of origin, journal name, commercial name of index test performed, reference standard test performed, the number of patients ongoing each test, stated industry involvement as defined by donation of test kits or statement of involvement and stated conflict of interest as defined by explicit mention in the report.

Each article was analysed by two of three authors (MH, SB or CP). Discrepancies in interpretation were resolved by a fourth author (PD).

Methodological applicability and bias assessment using QUADAS-2

Methodological applicability and risk of bias were assessed using QUADAS-2 items, and determined to be of low, high or unclear risk of bias for each item in the tool. These assessments refer to the risk of incorrect study conclusions based on study methods. See Table 1 for an explanation of the QUADAS-2 assessment criteria [10]. The proportion of high or unclear risk of bias was compared between studies of each organism using Pearson χ^2.

Reporting completeness assessment using STARD 2015

Reporting quality was assessed using the STARD 2015 checklist. Studies were evaluated based on the presence or absence of each criterion.

Statistical analysis was completed using SPSS (IBM® SPSS® Statistics Version 21, USA). All studies meeting inclusion criteria were analysed, and missing data were recorded as 'not reported'. STARD 2015 completeness scores were not combined into an overall score, as recommended by the original authors [10], but reported as counts of criteria achieved. Pearson χ^2 was used to compare the proportion of criteria achieved.

Table 1. QUADAS-2 criteria

| Item number	Criteria for low risk assessment
QUADAS1: Patient selection risk of bias	Patient enrolment strategy is specified and free of bias. A case–control design and inappropriate exclusions were avoided
QUADAS2: Patient selection applicability	There is no concern that the included patients do not match the review question
QUADAS3: Index test risk of bias	Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it pre-specified? The conduct or interpretation of the index test did not introduce bias
QUADAS4: Index test applicability	There is no concern that the index test, its conduct or interpretation differ from the review question
QUADAS5 Reference test risk of bias	The reference standard correctly classifies the target condition. The reference standard results are interpreted without knowledge of the results of the index test. The reference standard, its conduct or its interpretation have not introduced bias
QUADAS6: Reference test applicability	The target condition as defined by the reference standard matches the review question
QUADAS7: Flow and timing risk of bias	There is an appropriate interval between index test(s) and reference standard. All patients received the same reference standard. All patients included in the analysis and patient flow did not introduce bias
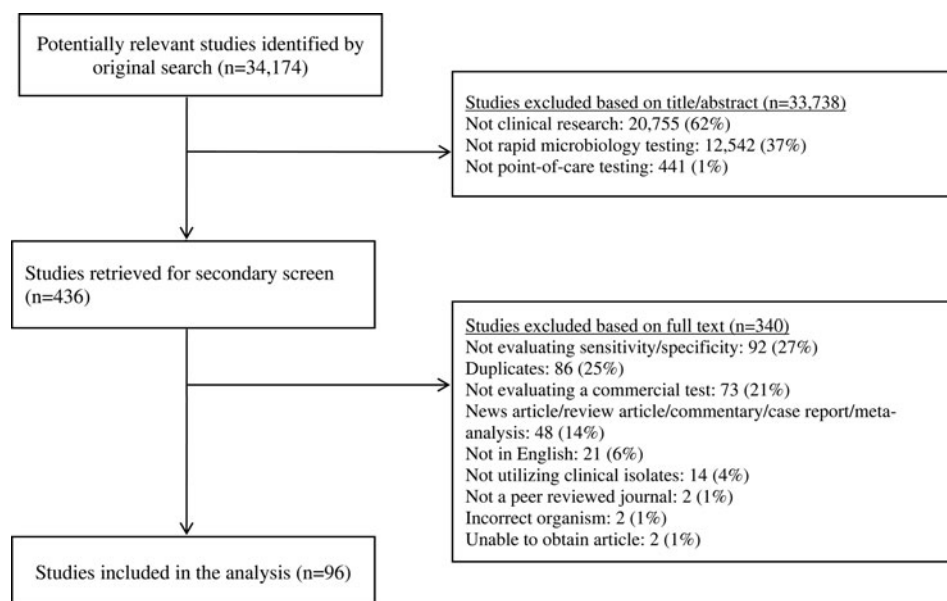
Fig. 1. Study selection.

Table 2. All included studies

ID	First author	Year	Country of origin	Journal	Organism	QUADAS-2 risk of bias			STARD 2015 Standards met (/30)
						Low	High	Unclear	
1	Al-Najjar	2008	United Arab Emirates	*Int J Infect Dis*[1]	Group A Strep	5	0	2	10
2	Armengol	2004	USA	*J Pediatr*[2]	Group A Strep	4	1	2	17
4	Camurdan	2008	Turkey	*Int J Pediatr Otorhinolaryngol*[3]	Group A Strep	6	0	1	12
5	Cohen	2004	France	*Eur J Pediatr*[4]	Group A Strep	4	2	1	5
7	Felsenstein	2014	USA	*J Clin Microbiol*[5]	Group A Strep	6	0	1	22
8	Forward	2006	Canada	*Can J Infect Dis Med Microbiol*[6]	Group A Strep	7	0	0	12
10	Gurol	2010	Turkey	*Int J Pediatr Otorhinolaryngol*[3]	Group A Strep	5	1	1	15
11	Kim	2009	Korea	*Korean J Lab Med*[7]	Group A Strep	5	1	1	10
12	Kucuk	2014	Turkey	*Indian J Pediatr*[8]	Group A Strep	5	0	2	16
13	Lindbaek	2004	Norway	*Scand J Prim Health Care*[9]	Group A Strep	6	0	1	17
15	Noorbakhs	2011	Iran	*Iran J Microbiol*[10]	Group A Strep	5	1	1	12
16	Rimoin	2010	USA	*Int J Infect Dis*[1]	Group A Strep	5	0	2	18
17	Rogo	2010	USA	*Clin Pediatr*[11]	Group A Strep	5	0	2	8
18	Sarikaya	2010	Turkey	*Ear Nose Throat J*[12]	Group A Strep	5	0	2	14
19	Tanz	2009	USA	*J Pediatr*[2]	Group A Strep	6	0	1	18
20	Wright	2007	USA	*Mil Med*[13]	Group A Strep	7	0	0	6
21	Vakkila	2015	Finland	*J Clin Microbiol*[5]	Group A Strep	5	0	2	13
22	Andreo	2006	Spain	*Respir Med*[14]	*Streptococcus pneumoniae*	6	0	1	14

(*Continued*)

Table 2. (Continued.)

ID	First author	Year	Country of origin	Journal	Organism	QUADAS-2 risk of bias			STARD 2015
						Low	High	Unclear	Standards met (/30)
23	Briones	2006	Spain	Clin Vaccine Immunol[15]	S. pneumoniae	5	1	1	12
24	Ercis	2006	Turkey	Jpn J Infect Dis[16]	S. pneumoniae	5	1	1	13
25	Ishida	2004	Japan	J Infect Chemother[17]	S. pneumoniae	5	0	2	16
26	Charkaluk	2006	France	Diagn Microbiol Infect Dis[18]	S. pneumoniae	5	0	2	17
27	Genne	2005	Switzerland	Int J Infect Dis[1]	S. pneumoniae	5	0	2	17
28	Hohenthal	2008	Finland	Scand J Infect Dis[19]	S. pneumoniae	5	1	1	15
30	Lasocki	2006	France	Intensive Care Med	S. pneumoniae	7	0	0	22
31	Navarro	2004	Spain	J Clin Microbiol	S. pneumoniae	2	4	1	7
32	Perello	2011	Spain	Eur J Emerg Med[20]	S. pneumoniae	4	1	2	13
33	Roson	2004	Spain	Clin Infect Dis[21]	S. pneumoniae	4	2	1	21
34	Turner	2011	Thailand	BMC Infect Dis[22]	S. pneumoniae	4	2	1	17
35	Weatherall	2008	Australia	Emerg Med J[23]	S. pneumoniae	7	0	0	22
36	Zalacain	2014	Spain	Respirology	S. pneumoniae	6	0	1	16
38	Esposito	2004	Italy	Pediatr Infect Dis J[24]	S. pneumoniae	6	0	1	17
39	Monno	2013	Italy	J Microbiol Methods[25]	S. pneumoniae	4	2	1	9
40	Tzeng	2006	Taiwan	J Microbiol Immunol Infect[26]	S. pneumoniae	6	0	1	17
41	Vazquez	2004	Spain	Eur J Clin Microbiol Infect Dis[27]	S. pneumoniae	5	2	0	7
42	Al Johani	2011	Saudi Arabia	J Infect Public Health[28]	Influenza	7	0	0	14
43	Angoulvant	2011	France	Emerg Med J[23]	Influenza	7	0	0	14
44	Bell	2014	USA	J Clin Virol[29]	Influenza	4	0	3	19
45	Bellmann-Weiler	2010	Austria	Clin Microbiol Infect[30]	Influenza	4	1	2	7
46	Bhattacharya	2011	UK	Indian J Med Microbiol[31]	Influenza	4	1	2	11
47	Biggs	2010	USA	Emerg Med J[23]	Influenza	6	0	1	15
48	Bin Saeed	2014	Saudi Arabia	Saudi Med J[32]	Influenza	6	0	1	16
49	Boku	2013	Japan	Diagn Microbiol Infect Dis	Influenza	6	0	1	18
50	Boyanton	2014	USA	Diagn Microbiol Infect Dis	Influenza	3	1	3	11
51	Bruning	2014	The Netherlands	Diagn Microbiol Infect Dis	Influenza	4	1	2	11
52	Busson	2014	Belgium	Diagn Microbiol Infect Dis	Influenza	6	0	1	15
53	Chan	2012	Hong Kong, China	J Virol Methods[33]	Influenza	5	0	2	9
54	Cheng	2009	China	Diagn Microbiol Infect Dis	Influenza	6	0	1	20

(Continued)

Table 2. (*Continued.*)

ID	First author	Year	Country of origin	Journal	Organism	QUADAS-2 risk of bias			STARD 2015
						Low	High	Unclear	Standards met (/30)
55	Cheng	2011	China	*J Clin Virol*[29]	Influenza	5	0	2	10
56	Cho	2013	Republic of Korea	*J Virol Methods*[33]	Influenza	6	0	1	21
57	Choi	2011	Korea	*Yonsei Med J*[34]	Influenza	5	0	2	14
58	Choi	2010	South Korea	*J Clin Microbiol*	Influenza	4	0	3	10
59	Choi	2010	Korea	*Int J Microbiol Biotechnol*[35]	Influenza	5	0	2	11
60	Ciblak	2010	Turkey	*Scand J Infect Dis*[19]	Influenza	6	0	1	10
61	Cruz	2010	USA	*J Pediatr*[2]	Influenza	7	0	0	19
62	de la Tabla	2010	Spain	*Clin Microbiol Infect*[30]	Influenza	5	0	2	11
63	de Witte	2011	Belgium	*Eur J Clin Microbiol Infect Dis*[27]	Influenza	4	1	2	13
64	Drexler	2009	Germany	*Emerg Infect Dis*[36]	Influenza	4	2	1	8
65	Duman	2013	Turkey	*Pedriatr Emerg Care*[37]	Influenza	4	3	0	14
66	Dunn	2014	USA	*Diagn Microbiol Infect Dis*	Influenza	3	2	2	6
67	Claudia Fernandez	2010	United States	*Postgrad Med J*[38]	Influenza	5	0	2	9
68	Fuenzalida	2010	Spain	*Clin Microbiol Infect*[30]	Influenza	4	2	1	10
69	Ghebremedhin	2009	Germany	*J Med Microbiol*[39]	Influenza	7	0	0	14
70	Gimeno	2010	Spain	*Diagn Microbiol Infect Dis*	Influenza	5	0	2	7
71	Gimeno	2010	Spain	*Diagn Microbiol Infect Dis*	Influenza	5	2	0	8
72	Ginocchio	2009	USA	*J Clin Virol*[29]	Influenza	5	0	2	9
73	Gordon	2010	USA	*PLoS ONE*[40]	Influenza	5	0	2	16
74	Gordon	2009	USA	*PLoS ONE*[40]	Influenza	7	0	0	22
75	Hara	2013	Japan	*Diagn Microbiol Infect Dis*	Influenza	4	0	3	9
76	Hassan	2014	USA	*J Clin Microbiol*	Influenza	7	0	0	15
77	Hawkes	2010	Canada	*J Pediatr*[2]	Influenza	5	1	1	16
78	Herzum	2010	Germany	*Clin Chem Lab Med*[41]	Influenza	3	3	1	8
79	Karre	2010	USA	*J Clin Microbiol*	Influenza	5	0	2	3
80	Keitel	2011	Switzerland	*Eur J Pediatr*[42]	Influenza	6	0	1	13
81	Louie	2010	USA	*Emerg Infect Dis*[36]	Influenza	5	0	2	9
82	Miarka	2014	Poland	*Acta Biochim Pol*[43]	Influenza	3	1	3	12
83	Mitamura	2013	Japan	*J Infect Chemother*[17]	Influenza	6	0	1	13
84	Mitamura	2013	Japan	*J Virol Methods*[33]	Influenza	4	0	3	10
85	Nakao	2014	Japan	*Diagn Microbiol Infect Dis*	Influenza	4	1	2	11

Table 2. (Continued.)

| ID | First author | Year | Country of origin | Journal | Organism | QUADAS-2 risk of bias | | | STARD 2015 |
						Low	High	Unclear	Standards met (/30)
86	Nutter	2012	USA	*PLoS ONE*[40]	Influenza	4	1	2	13
87	Ozdemir	2012	Turkey	*J Int Med Res*[44]	Influenza	6	0	1	13
88	Poeppl	2011	Austria	*PLoS ONE*[40]	Influenza	6	0	1	19
89	Pongthanapisith	2011	Thailand	*J Infect*[45]	Influenza	5	1	1	11
90	Sandora	2009	USA	*Pediatr Infect Dis J*[24]	Influenza	5	0	2	14
91	Self	2012	USA	*Am J Emerg Med*[46]	Influenza	7	0	0	20
92	Steininger	2009	Austria	*Clin Microbiol Infect*[30]	Influenza	5	2	0	15
93	Stevenson	2010	USA	*J Clin Microbiol*	Influenza	5	2	0	9
94	Stripeli	2015	Greece	*Eur J Clin Microbiol Infect Dis*[27]	Influenza	6	1	0	17
95	Suntarattiwong	2010	Thailand	*Pediatr Infect Dis J*[24]	Influenza	6	1	0	10
96	Sutter	2012	USA	*J Med Virol*[47]	Influenza	3	4	0	10
97	Tai	2012	Taiwan	*J Formos Med Assoc*[48]	Influenza	7	0	0	17
98	Uyeki	2009	USA	*Clin Infect Dis*	Influenza	4	1	2	14
99	Vasoo	2009	USA	*Clin Infect Dis*	Influenza	3	4	0	9
100	Zazueta-Garcia	2014	Mexico	*J Infect Dev Ctries*[49]	Influenza	5	0	2	15
101	Zetti	2010	Malaysia	*Med J Malaysia*[50]	Influenza	4	2	1	10
102	Abu-Sabaah	2006	Saudi Arabia	*Br J Biomed Sc*[51]	Group A Strep	4	1	2	12

[1]*International Journal of Infectious Diseases*, [2]*Journal of Pediatrics*, [3]*International Journal of Pediatric Otorhinolaryngology*, [4]*European Journal of Pediatrics*, [5]*Journal of Clinical Microbiology*, [6]*Canadian Journal of Infectious Diseases and Medical Microbiology*, [7]*Korean Journal of Laboratory Medicine*, [8]*Indian Journal of Pediatrics*, [9]*Scandinavian Journal of Primary Health Care*, [10]*Iranian Journal of Microbiology*, [11]*Clinical Pediatrics*, [12]*Ear Nose & Throat Journal*, [13]*Military Medicine*, [15]*Clinical and Vaccine Immunology*, [16]*Japanese Journal of Infectious Disease*, [17]*Journal of Infection and Chemotherapy*, [18]*Diagnostic Microbiology and Infectious Disease*, [19]*Scandinavian Journal of Infectious Disease*, [20]*European Journal of Emergency Medicine*, [21]*Clinical Infectious Diseases*, [22]*BioMed Central Infectious Diseases*, [23]*Emergency Medicine Journal*, [24]*Pediatric Infectious Diseases Journal*, [25]*Journal of Microbiological Methods*, [26]*Journal of Microbiology, Immunology, and Infection*, [27]*European Journal of Clinical Microbiology and Infectious Disease*, [28]*Journal of Infection and Public Health*, [29]*Journal of Clinical Virology*, [30]*Clinical Microbiology and Infection*, [31]*Indian Journal of Medical Microbiology*, [32]*Saudi Medical Journal*, [33]*Journal of Virological Methods*, [34]*Yonsei Medical Journal*, [35]*Journal of Microbiology and Biotechnology*, [36]*Emerging Infectious Diseases*, [37]*Pediatric Emergency Care*, [38]*Postgraduate Medicine*, [39]*Journal of Medical Microbiology*, [40]*Public Library of Science*, [41]*Clinical Chemistry and Laboratory Medicine*, [42]*European Journal of Pediatrics*, [43]*Acta Biochimica Polonica*, [44]*The Journal of International Medical Research*, [45]*Journal of Infection*, [46]*American Journal of Emergency Medicine*, [47]*Journal of Medical Virology*, [48]*Journal of the Formosan Medical Association*, [49]*Journal of Infection in Developing Countries*, [50]*Medical Journal of Malaysia*, [51]*British Journal of Biomedical Science*.

Ethics

Because patient information was not analysed, ethics approval was not required.

Results

Article search, screening and data abstraction

The PubMed search identified 34 174 potential studies (Fig. 1). Of these, 33 738 were excluded by removal of duplicate studies, review of title, keyword and abstract, and 340 studies were further excluded by full-text review. Overall, 96 studies were included, including antigen tests from throat swabs for Group A *Streptococcus* (*n* = 18), antigen tests from urine for *S. pneumoniae* (*n* = 19) and antigen and molecular tests for influenza (*n* = 60) (Table 2).

Study characteristics are summarised in Table 3. The median number of patients included per study for both index and reference tests was 303. Most studies were reported from Europe

(42/96, 43.8%). Influenza tests were over-represented due to the 2009 pandemic year, with 37 studies published in 2009 out of 60 total studies on influenza tests (62%). Industry involvement was admitted or unclear in 65/96 studies (67.7%).

Methodological applicability and bias assessment using QUADAS-2

A summary of QUADAS-2 assessment can be found in Table 4, and a comparison of bias by QUADAS-2 domain is found in Figure 2.

Studies demonstrated a low risk of bias in three of the seven QUADAS-2 criteria, related to **applicability**, namely QUADAS2 (patient selection applicability), QUADAS4 (index test applicability) and QUADAS6 (reference test applicability). Studies demonstrated a high or unclear risk of bias in four of the seven QUADAS-2 criteria, related to **risk of bias**, namely QUADAS1 (patient selection risk of bias), QUADAS 3 (index test risk of

Table 3. Description of included studies (n = 96)

Characteristic	Frequency (%)
Disease	
Group A *Streptococcus*	18/96 (18.8%)
Streptococcus pneumoniae	18/96 (18.8%)
Influenza	60/96 (62.5%)
2009 Pandemic influenza	37/60 (61.7%)
Study continent of origin	
Asia	25
Australia and Oceania	1
Europe	42
North America	28
Median (range) patients per article for index test	302.5 (23–6114)
Median (range) patients per article for reference standard	302.5 (23–6114)
Year of publication	
2004	8 (8.3%)
2005	1 (1.0%)
2006	8 (8.3%)
2007	1 (1.0%)
2008	4 (4.2%)
2009	11 (11.5%)
2010	23 (24.0%)
2011	12 (12.5%)
2012	6 (6.3%)
2013	7 (7.3%)
2014	13 (13.5%)
2015	2 (2.1%)
Total number of journals	54
Industry involvement	
Yes	30 (31.3%)
No	31 (32.2%)
Unclear	35 (36.5%)
Conflict of interest stated	16 (16.7%)

Table 4. QUADAS-2 results

Risk of Bias	Low	High	Unclear
	Overall (n = 96, %)		
QUADAS1: Patient selection risk of bias	39 (40.6)	15 (15.6)	42 (43.8)
QUADAS2: Patient selection applicability	85 (88.5)	7 (7.3)	4 (4.2)
QUADAS3: Index test risk of bias	37 (38.5)	14 (14.6)	45 (46.9)
QUADAS4: Index test applicability	94 (97.9)	2 (2.1)	0
QUADAS5: Reference test risk of bias	69 (71.9)	7 (7.3)	20 (20.8)
QUADAS6: Reference test applicability	94 (97.9)	2 (2.1)	0
QUADAS7: Flow and timing risk of bias	70 (72.9)	17 (17.7)	9 (9.4)
	Group A *Streptococcus* (n = 19)		
QUADAS1: Patient selection risk of bias	9 (50.0)	1 (5.3)	8 (42.1)
QUADAS2: Patient selection applicability	18 (100)	0	0
QUADAS3: Index test risk of bias	8 (42.1)	1 (5.3)	9 (50.0)
QUADAS4: Index test applicability	18 (100)	0	0
QUADAS5: Reference test risk of bias	18 (100)	0	0
QUADAS6: Reference test applicability	10 (52.6)	2 (10.5)	6 (31.6)
QUADAS7: Flow and timing risk of bias	14 (73.7)	3 (15.8)	1 (5.3)
	Streptococcus pneumoniae (n = 19)		
QUADAS1: Patient selection risk of bias	10 (52.6)	5 (26.3)	3 (15.8)
QUADAS2: Patient selection applicability	15 (78.9)	3 (15.8)	0
QUADAS3: Index test risk of bias	7 (38.9)	0	11 (57.9)
QUADAS4: Index test applicability	17 (89.5)	1 (5.3)	0
QUADAS5: Reference test risk of bias	17 (89.5)	1 (5.3)	0
QUADAS6: Reference test applicability	10 (52.6)	3 (15.8)	5 (26.3)
QUADAS7: Flow and timing risk of bias	15 (78.9)	3 (15.8)	0
	Influenza (n = 60)		
QUADAS1: Patient selection risk of bias	20 (33.3)	9 (15.0)	31 (51.7)
QUADAS2: Patient selection applicability	52 (86.7)	4 (6.7)	4 (6.7)
QUADAS3: Index test risk of bias	22 (36.7)	13 (21.7)	25 (41.7)

(*Continued*)

bias), QUADAS5 (reference test risk of bias) and QUADAS 7 (flow and timing risk of bias). For QUADAS1, 15 studies (15.6%) demonstrated a high risk of bias, and 42 studies (43.8%) demonstrated an unclear risk of bias. For QUADAS3, 14 studies (14.6%) demonstrated a high risk of bias, and 45 studies (46.9%) demonstrated an unclear risk of bias. For QUADAS5, seven studies (7.3%) demonstrated a high risk of bias, and 20 studies (20.8%) demonstrate an unclear risk of bias. For QUADAS7, 17 studies (17.7%) demonstrated a high risk of bias and nine studies (9.4%) demonstrated an unclear risk of bias.

The proportion of studies with a high or unclear risk of bias for each QUADAS criterion was statistically similar between studies by an organism, except QUADAS5 (reference test risk of bias), which demonstrated a higher proportion of a high or unclear risk of bias

Table 4. (*Continued.*)

Risk of Bias	Low	High	Unclear
QUADAS4: Index test applicability	59 (98.3)	1 (1.7)	0
QUADAS5: Reference test risk of bias	59 (98.3)	1 (1.7)	0
QUADAS6: Reference test applicability	49 (81.7)	2 (2.1)	9 (15.0)
QUADAS7: Flow and timing risk of bias	41 (68.3)	11 (18.3)	8 (13.3)

among studies of influenza (81.7%) compared with studies of Group A *Streptococcus* (55.65) and *S. pneumoniae* (55.6%) (*P* = 0.022).

Quality of reporting assessment using STARD 2015

A summary of STARD 2015 assessment by an organism can be found in Table 5. The mean STARD compliance among all studies was 11/30 criteria (S.D. 4.3). Only two studies (2.1%) explicitly stated compliance with the STARD reporting guidelines. Mean STARD compliance between studies by an organism was similar (Group A *Streptococcus* 11.7 criteria, *S. pneumoniae* 13.4 criteria, influenza 10.0 criteria) (*P* = 0.38). Twenty criteria had low inclusion (<50% of studies included the criterion) (criteria 1, 2, 4, 5, 6, 9, 10, 12, 13, 14, 15, 16, 18, 20, 21, 22, 24, 25, 26, 28).

Discussion

We identified four QUADAS-2 criteria associated with a risk of bias that were at a high risk or unclear risk of bias among included studies. Studies performed better among criteria associated with applicability.

Patient selection risk of bias (QUADAS1)

To minimise the risk of bias in patient selection, a study should have appropriate methods of recruiting patients, such as either consecutive

patients who present for care over time, or a random selection from a larger subset of patients. Included patients should be of heterogeneous composition, because exclusion of complex patients may overestimate test accuracy and compromise external validity of the study, whereas exclusion of healthy patients could underestimate test accuracy. Furthermore, case–control designs in which patients with obvious disease are selected as cases, and patients without obvious disease are selected as controls, could overestimate test accuracy through selection bias. We observed a large proportion of studies (42 studies, 43.8%) in which the assessment of risk of bias in patient selection was unclear, meaning that reports did not contain adequate information to reassure the reader that the risk of bias was low. Without this information, it is difficult to make inferences on the validity of the study.

Conduct or interpretation of index test (QUADAS3)

If index tests are interpreted with knowledge of the results of reference tests, index test interpretation may be biased towards overestimation of test accuracy. To prevent this bias, operators performing the test must perform the index test prior to the reference test, or at least must be adequately blinded to the reference test results. Reports must adequately describe the methods of blinding, such as performing the index test and reference test in two different laboratories, or relabelling and changing the order of specimens to protect the blind. A further cause of bias in conduct or interpretation of the index test is the timing in which the threshold of detection is determined. If the threshold of detection is selected based on analysing the completed study results, this may overestimate test accuracy. Threshold of detection should be defined prior to collecting data, and this must be stated in the report. We observed 46 studies (46.9%) in which information was not provided to allow the reader to assess that the index test was performed without bias (unclear risk of bias).

Reference test risk of bias (QUADAS5)

Bias may be introduced in the reference test if the reference test is unlikely to correctly classify the condition, or if the reference test
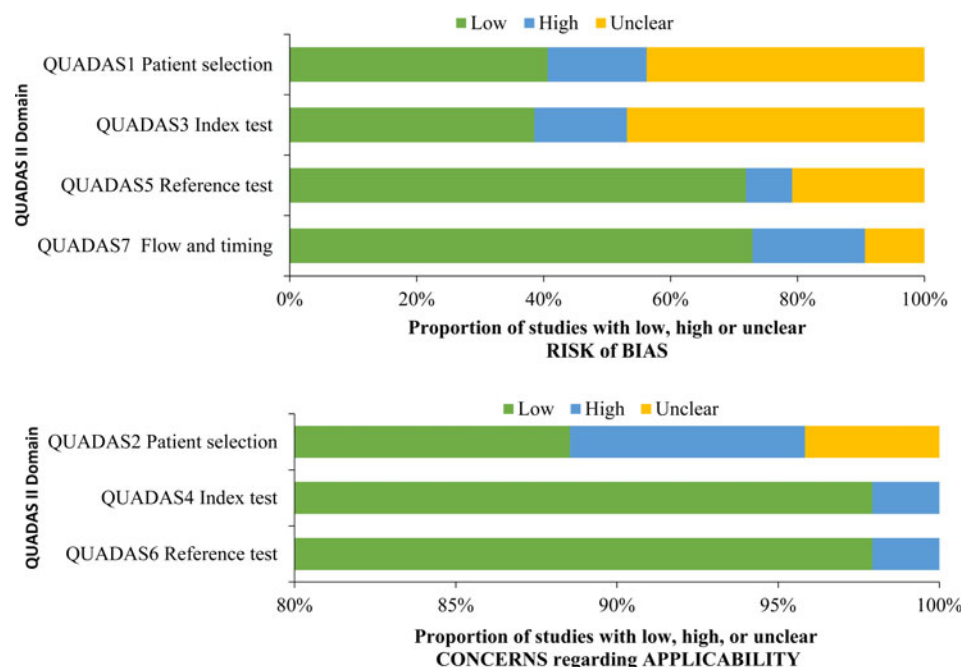

Fig. 2. QUADAS criteria among all studies (*n* = 96).

Table 5. STARD 2015 criteria with 25% or less inclusion

STARD 2015 criteria	Per cent inclusion
STARD9: METHODS Whether participants formed a consecutive, random or convenience series	25
STARD12B: METHODS Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	20.8
STARD21A: RESULTS Distribution of severity of disease in those with the target condition	18.8
STARD22: RESULTS Time interval and any clinical interventions between index test and reference standard	18.8
STARD17: METHODS Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	18.8
STARD13B: METHODS Whether clinical information and index test results were available to the assessors of the reference standard	17.7
STARD21B: RESULTS Distribution of alternative diagnoses in those without the target condition	16.7
STARD13A: METHODS Whether clinical information and reference standard results were available to the performers/readers of the index test	14.6
STARD19: RESULTS Flow of participants, using a diagram	11.5
STARD15: METHODS How indeterminate index test or reference standard results were handled	11.5
STARD28: OTHER Registration number and name of registry	6.3
STARD18: METHODS Intended sample size and how it was determined	3.1
STARD16: METHODS How missing data on the index test and reference standard were handled	1
STARD25: RESULTS Any adverse events from performing the index test or the reference standard	0

is interpreted with knowledge of the results of the index test. We observed 20 studies (20.8%) in which the risk of bias in this criterion was unclear. Reference tests for respiratory infections include culture and nucleic acid amplification tests. Culture tests may be influenced by the presence of normal respiratory flora, or collection after antibiotic treatment. POC tests for influenza were statistically more biased than other tests in this criterion.

Risk of bias pertaining to flow and timing (QUADAS7)

Patient flow must be well documented, including exclusions and missing data, as patients who are not included in the final analysis may differ significantly from those included. An acute respiratory infectious disease evolves quickly, and index and reference tests must be collected simultaneously. Any delay between index and reference test collection

may cause the natural disease progression to change the results and thus the performance of the delayed test. We observed 17 studies (17.7%) with a high risk of bias in this criterion.

Fourteen of the 30 STARD 2015 criteria demonstrated compliance among <25% of the reports. Criterion 14 (definition and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory) may be less relevant to qualitative reference standard tests. Criteria 15 and 16 (whether clinical information and reference standard results were available to the performers of the index test, whether clinical information and index test were available to the assessors of the reference standard) address blinding, which is not commonly included in the diagnostic study designs, although expected in the treatment trial designs. Criteria 18 and 19 (how indeterminate index test or reference standard results were handled, how missing data on the index test or reference standard were handled) is particularly important in the analysis, since indeterminate results should be considered missing data, and missing data may bias conclusions.

Criterion 20 (any analysis of variability in diagnostic accuracy distinguishing pre-specified from exploratory) addresses test repeatability, which was generally not performed in the diagnostic studies we examined. Criterion 21 (intended sample size and how it was determined) reflects appropriate study planning to determine statistical power. Criterion 22 (flow of participants, using a diagram) accounts for patients lost to follow-up or excluded, which represent missing data. Criteria 24 and 25 (distribution of severity of disease in those with the target condition, distribution of alternate diagnoses in those without the target condition) address generalisability assessment. Criterion 26 (time interval and any clinical interventions between index test and reference standard) allows the reader to assess the risk that the patient's condition changed between index and reference test application. Criterion 29 (any adverse events from performing the index test or reference standard) was not reported by a single study. It is not generally suspected that the application of a diagnostic test should cause patient harm, however it is possible. Criterion 32 (registration number and name of registry) indicates transparency in design and reporting.

Many of the STARD 2015 reporting shortfalls previously noted [13, 14] are still lacking in this study, including lack of reporting of blinding for index and reference test execution, reporting of indeterminate and missing data, flow of participants and reporting of withdrawals, and distribution of severity of disease and alternative diagnoses [13, 14]. Therefore, despite updated criteria, there continues to be weak reporting of POC diagnostic studies for respiratory pathogens. This may be due to perceived lack of utility of reporting, or purposeful omission to enhance the perceived impact of the results. Continued weak adherence to STARD 2015 over included years from 2004 to 2015 enhances the findings by Wilczynski, which noted no change in quality, despite the publication of STARD [15].

This study reports on the studies of diagnostic accuracy for respiratory infections and indicates that there are many areas of reporting that fail to meet the outlined criteria. This has never been reported among tests for these pathogens. More stringent reporting requirements from journals, including reporting of the STARD 2015 criteria flowcharts in the methodology sections, may enhance the quality of published works.

Strengths of this study include reporting on POC tests in current use for common respiratory pathogens, and a comprehensive review of literature using broad search terms, which is believed to include all potential studies in the specified period at the time of

data abstraction. However, some studies may not have been detected by the search terms. Additional limitations of the study include restriction of data abstraction to studies in English only, as well as the subjective nature of interpretation of the QUADAS-2 criteria.

Future research may include quality assessments of reports of conventional diagnostic tests in microbiology, for which fewer and older studies are available.

Author contributions. MH: preliminary literature search, primary screen, secondary screen and data abstraction, data analysis, creation and amendment of manuscript. SB: data abstraction. CP: data abstraction. PD: review of data discrepancies, amendment and review of manuscript, and data analysis.

Declaration of Interest. None.

References

1. **Carapetis JR, *et al.*** (2005) The global burden of group A streptococcal diseases. *Lancet Infect Diseases* **5**(11), 685–694.
2. **Drijkoningen JJ and Rohde GG** (2014) Pneumococcal infection in adults: burden of disease. *Clinical Microbioly and Infection* **20**(Suppl. 5), 45–51.
3. **Molinari NA, *et al.*** (2007) The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* **25**(27), 5086–5096.
4. **Blyth CC, Booy R and Dwyer DE** (2011) Point of care testing: diagnosis outside the virology laboratory. *Methods in Molecular Biology* **665**, 415–433.
5. **Abanses JC, *et al.*** (2006) Impact of rapid influenza testing at triage on management of febrile infants and young children. *Pediatric Emergency Care* **22**(3), 145–149.
6. **Noyola DE and Demmler GJ** (2000) Effect of rapid diagnosis on management of influenza A infections. *Pediatric Infectious Diseases Journal* **19**(4), 303–307.
7. **Lean WL, *et al.*** (2014) Rapid diagnostic tests for group A streptococcal pharyngitis: a meta-analysis. *Pediatrics* **134**(4), 771–781.
8. **Pai NP, *et al.*** (2012) Point-of-care testing for infectious diseases: diversity, complexity, and barriers in low- and middle-income countries. *Public Library of Science Medicine* **9**(9), e1001306.
9. **Bossuyt PM, *et al.*** (2003) Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. *American Journal of Clinical Pathology* **119**(1), 18–22.
10. **Cohen JF, *et al.*** (2016) STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *British Medical Journal Open* **6**(11), e012799.
11. **Whiting P, *et al.*** (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BioMed Central Medical Research Methodoly* **3**, 25.
12. **Whiting PF, *et al.*** (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* **155**(8), 529–536.
13. **Fontela PS, *et al.*** (2009) Quality and reporting of diagnostic accuracy studies in TB, HIV and malaria: evaluation using QUADAS and STARD standards. *Public Library of Science One* **4**(11), e7753.
14. **Jafari Y, *et al.*** (2013) Are *Treponema pallidum* specific rapid and point-of-care tests for syphilis accurate enough for screening in resource limited settings? Evidence from a meta-analysis. *Public Library of Science One* **8**(2), e54695.
15. **Wilczynski NL** (2008) Quality of reporting of diagnostic accuracy studies: no change since STARD statement publication – before-and-after study. *Radiology* **248**(3), 817–823.