Ariella P Dale, Christian Marchello and Mark H Ebell

Clinical gestalt to diagnose pneumonia, sinusitis, and pharyngitis:

a meta-analysis

Abstract

Background
The overall clinical impression ('clinical gestalt') is widely used for diagnosis but its accuracy has not been systematically studied.

Aim
To determine the accuracy of clinical gestalt for the diagnosis of community-acquired pneumonia (CAP), acute rhinosinusitis (ARS), acute bacterial rhinosinusitis (ABRS), and streptococcal pharyngitis, and to contrast it with the accuracy of clinical decision rules (CDRs).

Design and setting
Systematic review and meta-analysis of outpatient diagnostic accuracy studies in ambulatory care.

Method
PubMed and Google were searched for studies in outpatients that reported sufficient data to calculate accuracy of the overall clinical impression and that used the same reference standard. Study quality was assessed using Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2), and measures of accuracy calculated using bivariate meta-analysis.

Results
The authors identified 16 studies that met the inclusion criteria. The summary estimates for the positive (LR+) and negative likelihood ratios (LR–) were LR+ 7.7, 95% confidence interval (CI) = 4.8 to 11.5, and LR– 0.54, 95% CI = 0.42 to 0.65 for CAP in adults, LR+ 2.7, 95% CI = 1.1 to 4.3 and LR– 0.63, 95% CI = 0.20 to 0.98 for CAP in children, LR+ 3.0, 95% CI = 2.1 to 4.4 and LR– 0.37, 95% CI = 0.29 to 0.46 for ARS in adults, LR+ 3.9, 95% CI = 2.4 to 5.9 and LR– 0.33, 95% CI = 0.20 to 0.50 for ABRS in adults, and LR+ 2.1, 95% CI = 1.6 to 2.8 and LR– 0.47, 95% CI = 0.36 to 0.60 for streptococcal pharyngitis in adults and children. The diagnostic odds ratios were highest for CAP in adults (14.2, 95% CI = 9.0 to 21.0), ARS in adults (8.3, 95% CI = 4.9 to 13.1), and ABRS in adults (13.0, 95% CI = 5.0 to 27.0), as were the C-statistics (0.80, 0.77, and 0.84 respectively).

Conclusion
The accuracy of the overall clinical impression compares favourably with the accuracy of CDRs. Studies of diagnostic accuracy should routinely include the overall clinical impression in addition to individual signs and symptoms, and research is needed to optimise its teaching.

Keywords
acute rhinosinusitis; community-acquired pneumonia; diagnosis; evidence-based medicine; overall clinical impression; pharyngitis.

INTRODUCTION

The overall clinical impression, also called 'clinical gestalt', is an intuitive approach to decision making used by physicians to make clinical diagnoses. It takes into account multiple signs and symptoms without necessarily using an analytic approach such as a point score or algorithm, and is an inductive approach based on pattern recognition rather than a hypotheticodeductive approach. Some studies have shown that inductive pattern-recognition strategies may be more widely used and more successful than hypotheticodeductive strategies.[1–3] However, proponents of evidence-based practice encourage the use of clinical decision rules (CDRs) for diagnosis, as do practice guidelines. CDRs use a formal approach such as multivariate analysis or recursive partitioning to identify signs, symptoms, and point-of-care tests that are the best independent predictors of a diagnosis or clinical outcome. They are then typically converted to a simple point score or algorithm such as the Ottawa Ankle Rules for ankle injury,[4] or the Wells rule to diagnose pulmonary embolism.[5] The goal of CDRs is to improve the efficiency and accuracy of clinical diagnosis and thereby reduce unnecessary testing.[6]

However, CDRs may be cumbersome to access and use at the point of care. As a result, CDRs are only infrequently used in real-world clinical practice.[7] Instead, clinicians rely on their overall clinical impression. As the overall clinical impression can incorporate additional variables not included in the CDR, it has the potential of being more accurate. For example, while a clinical rule may categorise a patient as being at low risk for group A beta-haemolytic streptococcal (GABHS) pharyngitis, knowing that a sibling was diagnosed with GABHS pharyngitis the week before could be an important factor.

For acute respiratory tract infections, CDRs have been developed to diagnose GABHS pharyngitis,[8,9] acute rhinosinusitis (ARS) and acute bacterial rhinosinusitis (ABRS),[10] and community-acquired pneumonia (CAP).[11] In this study, the authors performed a systematic review of the accuracy of the overall clinical impression for GABHS pharyngitis, ARS, and CAP, which has not been systematically studied before, and evaluated how its accuracy compared with that of CDRs for the same conditions.

METHOD

Search

For this systematic review, PubMed was searched for published studies using a search strategy (available from the authors), combining synonyms for overall clinical impression, the clinical diagnosis, and ambulatory care. The reference lists of all included studies were also searched to identify studies not captured by the PubMed search strategy. In addition,

AP Dale, PhD, MPH, healthcare-associated infections surveillance data coordinator, Colorado Department of Public Health and Environment, Denver, US. **C Marchello**, PhD, MSc, fellow, University of Otago, Dunedin, New Zealand. **MH Ebell**, MD, MSc, professor, Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, Georgia, US.

Address for correspondence
Mark Ebell, 125 Miller Hall, University of Georgia Health Sciences Campus, Athens, GA 30602, US.

Email: ebell@uga.edu

Submitted: 24 November 2018; **Editor's response:** 19 December 2018; **final acceptance:** 4 January 2019.

This is the full-length article (published online 18 Jun 2019) of an abridged version published in print. Cite this version as: **Br J Gen Pract 2019; DOI: https://doi.org/10.3399/bjgp19X704297**

How this fits in

It is known that the overall clinical impression is widely used in clinical practice but has not been systematically studied. This study showed that in adults the overall clinical impression had good accuracy for the diagnosis of community-acquired pneumonia, for acute rhinosinusitis, and for acute bacterial rhinosinusitis. It had moderate accuracy for diagnosis of streptococcal pharyngitis and for pneumonia in children. In each case, the accuracy of the overall clinical impression was similar to or better than that for a clinical decision rule for the same conditions. Thus, the overall clinical impression has good accuracy and is an important diagnostic tool that is deserving of further study and quantification.

published systematic reviews of the clinical diagnosis of GABHS pharyngitis, CAP, and ARS or ABRS were searched for additional studies,[12–16] as were the first 50 results returned by a Google search of '<disease> diagnosis clinical impression' for each disease. The search was not restricted by language, country, or date of publication.

Inclusion and exclusion criteria

The present research was limited to prospective studies that reported diagnostic data regarding the accuracy of the overall clinical impression (clinical gestalt) to diagnose CAP, ARS, ABRS, or acute GABHS pharyngitis. ARS was defined as abnormal imaging, and ABRS as abnormal culture of antral puncture fluid. Studies were limited to the ambulatory-care setting (outpatient clinic, urgent care, or emergency department [ED]) as hospital-acquired and ventilator-associated pneumonia are separate clinical entities. All patients must have received the same acceptable reference standard: chest radiograph (CXR), lung ultrasound, or computed tomography (CT) for pneumonia; imaging or antral puncture fluid analysis for ARS; and throat culture for GABHS pharyngitis. The authors excluded studies of nosocomial infections, infections in immunocompromised persons, or studies of the diagnosis of bacteraemia or sepsis. The authors included studies of both children and adults. Studies of ARS using inspection of antral puncture fluid or bacterial culture as the reference standard were classified as also diagnosing ABRS.

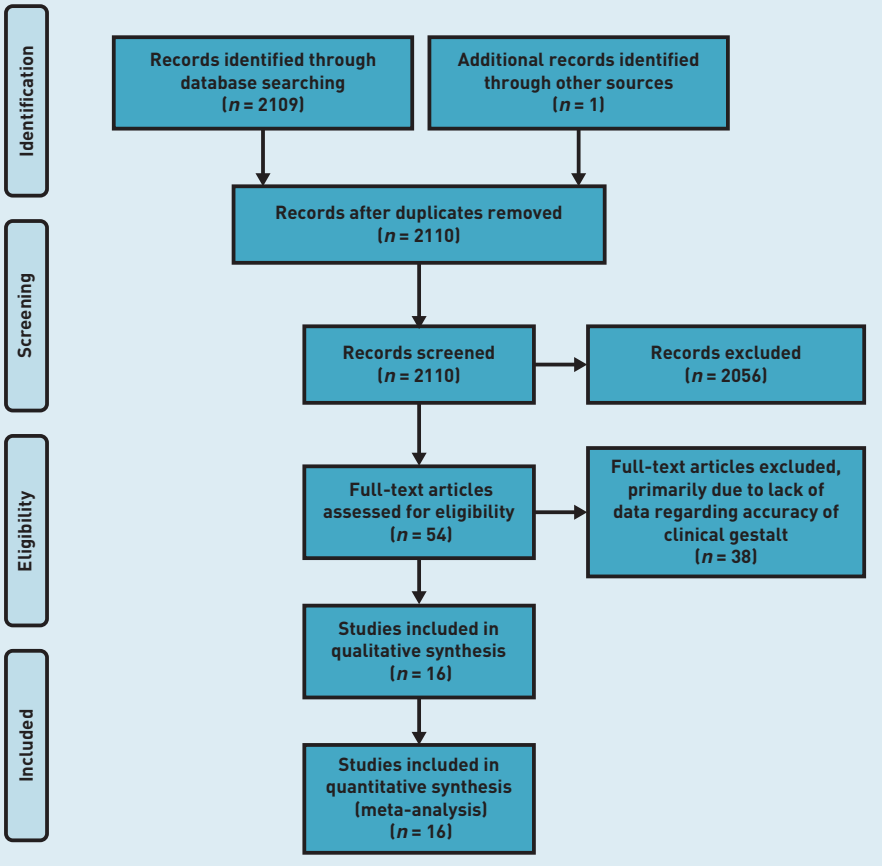

Figure 1. PRISMA flow diagram of study search.

Data abstraction

Each title and abstract was reviewed by two investigators to identify potential studies for inclusion. Any study identified for full-text analysis by one of the reviewers was reviewed independently by two investigators, and any discrepancies were resolved by a third reviewer (lead investigator). For studies that met the inclusion and exclusion criteria, two reviewers abstracted study characteristics, data regarding the accuracy of clinical gestalt, and study design characteristics for the quality assessment, with discrepancies resolved via consensus discussion or, if necessary, by the lead investigator. All of the included studies were reviewed a final time by the lead investigator to confirm the accuracy of data abstraction.

Where a study reported the accuracy of clinical gestalt using more than two categories (for example, 'sure', 'quite sure', and 'unsure'), the results were collapsed into two dichotomous categories, that is, 'sure' versus 'quite sure' or 'unsure'. The selection of category combinations was based on the combination that provided the highest diagnostic odds ratio (DOR; ratio of positive to negative likelihood ratio [LR]), a measure of discrimination. Where studies reported physician estimates of probability, >50% versus ≤50% was used. One study reported data in the form of a figure.[17] The figure was enlarged, digital vertical lines drawn to determine the intercept, and a ruler was used to calculate the number of patients in each category. Data were reported separately for the three study sites in this study (Illinois, Nebraska, and Virginia), as each site enrolled a distinct population and found somewhat different sensitivity and specificity.[17]

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) framework was adapted to evaluate the quality of the included studies. Studies at low risk of bias for all four domains (patient selection; index test; reference standard; and patient flow and timing) were judged to be at low risk of bias overall.[18] Those with a single domain at high risk of bias were judged to be at moderate risk of bias overall, and all others were judged to be at high risk of bias.

Statistical analysis

The authors performed the meta-analysis using the Reitsma function in the mada package in R (version 3.4.3), which uses a bivariate model equivalent to the hierarchical summary receiver operating characteristic (HSROC) model of Rutter and Gatsonis.[19] The authors used a summary receiver operating characteristic (ROC) curve to plot 95% confidence intervals for the summary estimates and calculated the area under the ROC curve (AUROCC), also called the C-statistic. Heterogeneity was evaluated using inspection of the summary ROC plots and confidence intervals, as I^2 is not recommended for use in diagnostic meta-analysis[20] or when there is a small number of primary studies.[21] To facilitate comparison with a dichotomous overall clinical impression for each diagnosis, clinical decision rules were dichotomised into low or moderate versus high risk, or low risk versus moderate or high risk depending on which approach provided the highest diagnostic odds ratio (DOR).

RESULTS

The initial search identified 2109 articles, of which 54 were evaluated as full text and 15 met the inclusion criteria. A review of references of included studies identified no additional studies for full-text review. The Google search identified no additional studies, whereas the review of previous systematic reviews identified one additional study of pharyngitis[22] for a final total of 16 included studies (three acute pharyngitis, nine CAP, and four ARS or ABRS). The search is summarised in Figure 1 using the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) framework.

Characteristics of included studies

The characteristics of the included studies are summarised in Table 1. A total of six studies took place in the US, four in Sweden, and one each in Ireland, Israel, Lesotho, Norway, Spain, and a consortium of 12 European countries. Most gathered data in either a primary care clinic or the ED or a combination of those sites. Regarding age group, 11 studies enrolled only adults, four only children, and one both adults and children. All studies of pneumonia diagnosis used chest radiography as the reference standard, all studies of pharyngitis used throat culture, and studies of rhinosinusitis used either antral puncture revealing purulent fluid[23–25] or sinus radiography.[26] The rhinosinusitis and pneumonia studies generally included patients where there was already some clinical suspicion for these diagnoses; an exception was the study by van Vugt and colleagues that included any patient with acute cough.[11] The prevalence of pneumonia varied from 5% in the van Vugt study to 44%; the median prevalence was 15%. The pharyngitis studies had broad

Table 1. Characteristics of included studies

Study author, year	Patient sample size, *n*	Setting (number of physicians)	Presenting symptoms	Operationalisation of clinical gestalt definition	Age group	Mean or median age (range)	Prevalence of disease, %	Country, date of data collection
Pneumonia								
Grossman, 1988[27]	155	ED (NR)	Diagnosis of pneumonia was considered	Clinical impression	Children	(0 to 18 years, 62% <2 years of age)	33	US, NR
Mahabee-Gittens, 2005[28]	510	ED (5)	Presenting with one or more of laboured, rapid, or noisy breathing; chest or abdominal pain; fever	Physician probability ≤50% versus >50%	Children	Mean 15.3 months	9	US, 2000–2002
Redd, 1994[29]	226	Outpatient department of a hospital (NR)	Children at high risk for pneumonia	Clinical diagnosis of pneumonia	Children	Mean 15.1 months (3 months to 5 years)	17	Lesotho, NR
Gonzalez Ortiz, 1995[30]	141	ED (NR)	Fever for 48 hours plus lower respiratory symptoms	Diagnostic judgement	Adult	(≥15 years)	38	Spain, NR
Lieberman, 2003[31]	250	Primary care clinic and ED (15 primary care, ED NR)	Acute febrile illness plus at least one of cough, sore throat, coryza, or hoarseness	Physician's judgement	Adult	Mean 39 years	8	Israel, 1999
Melbye, 1988[32]	71	Primary care clinic (25)	Patients with clinically suspected pneumonia by their GP	Physician probability >75% versus ≤75%	Adult	Mean 48 years (15 to 79)	15	Norway, 1986
Moberg, 2016[33]	100	Primary care clinic (NR)	Clinically suspected pneumonia	Suspicion of pneumonia sure versus quite sure or unsure	Adult	Mean 56 years	44	Sweden, 2011–2014
Tape, 1991[17]	1364	ED (NR)	Physician thought pneumonia was a possibility, or at least two of fever, cough, sputum, pleuritic pain, dyspnoea, wheeze, or altered mentation	Probability of pneumonia estimated >50%	Adult	Mean 45.4 years in Illinois, 47.6 years in Nebraska, and 41 years in Virginia	14	US, 1987–1988
van Vugt, 2013[11]	2810	Primary care clinics (294)	Acute or worsened cough (<28 days) or judged to have lower respiratory infection	Clinical judgement of GP	Adult	Mean 50 years	5	12 European countries, 2007–2010
GABHS pharyngitis								
Attia, 2001[34]	58	ED and two outpatient clinics (NR)	Signs or symptoms of acute pharyngitis	Physician probability estimate 0 to 5 versus 6 to 10	Children	Mean 6.8 years (1 to 18)	37	US, 1999–2000
Centor, 1981[8]	234	ED (NR)	Complaint of sore throat	Resident physician's 'guess'	Adult	NR	17	US, 1980
Dobbs, 1996[22]	206	Primary care clinic (1)	Main symptom of sore throat	Physician clinical diagnosis	Both	(≥4 years)	35	Ireland, 1988–1991
Sinusitis								
Williams, 1992[26]	247	General medicine clinic (86% by one of three authors)	Self-described sinusitis, nasal discharge, or non-traumatic facial pain	Clinical impression high versus intermediate or low	Adult	Median 50 years	38	US, NR
Berg,1981[23]	50	Otolaryngology clinic (NR)	Suspected acute sinusitis	Probably or definitely purulent sinusitis versus other	Adult	Mean 46 years	50	Sweden, NR
Berg, 1985[24]	90	Otolaryngology clinic (NR)	Suspected acute sinusitis	Evidence of suppuration clinically	Adult	NR	48	Sweden, NR
Berg, 1988[25]	155	ED (3)	Suspected acute maxillary sinusitis	Sinusitis versus no sinusitis	Adult	Mean 38 years	51	Sweden, NR

ED = emergency department. NR = not reported. US = United States.

Table 2. Assessment of study quality using the QUADAS-2 framework

	Domains																	
	Patient selection						Index test			Reference standard				Flow and timing				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Overall
Study, year	Consecutive	Not case control	Exclusion criteria	Risk of bias	Applicability	Index blinded	Threshold prespecified	Risk of bias	Applicability	Antral puncture used	Reference blinded	Risk of bias	Applicability	All had reference standard	All had same reference standard	All accounted for	Risk of bias	
Attia, 2001[34]	Y	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	L
Centor, 1981[8]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	L
Dobbs, 1996[22]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	L
Gonzalez Ortiz, 1995[30]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	L
Moberg, 2016[33]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	Y	**L**	L
Williams, 1992[26]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	L
Tape[a] (Nebraska, Illinois), 1991[17]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	Y	**L**	L
van Vugt, 2013[11]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	L
Berg, 1985[24]	U	Y	Y	**H**	H	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	Y	**L**	M
Berg, 1988[25]	Y	Y	Y	**L**	H	Y	Y	**L**	H	Y	N	**H**	H	Y	Y	Y	**L**	M
Grossman, 1988[27]	N	Y	Y	**H**	H	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	Y	**L**	M
Lieberman, 2003[31]	U	Y	N	**H**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	M
Melbye, 1988[32]	N	Y	Y	**H**	H	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	Y	**L**	M
Redd, 1994[29]	N	Y	N	**H**	H	Y	Y	**L**	H	Y	Y	**L**	H	Y	Y	Y	**L**	M
Mahabee-Gittens, 2005[28]	N	Y	Y	**H**	H	Y	Y	**L**	H	Y	U	**L**	H	Y	Y	Y	**L**	H
Tape[a] (Virginia), 1991[17]	N	Y	Y	**H**	H	Y	Y	**L**	H	Y	N	**H**	H	Y	Y	Y	**L**	H
Berg, 1981[23]	N	Y	Y	**H**	H	U	Y	**H**	H	Y	Y	**L**	H	Y	Y	Y	**L**	H

[a]The quality Tape, 1991 was evaluated separately for data gathered in two sites (Nebraska and Illinois) versus data gathered in Virginia because of different methods. L = 0, M = 1, and H = 2+ domains with high likelihood of bias. Y = Yes. N = No. L = low risk of bias. H = high risk of bias. M = moderate risk of bias. QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies-2.[18] U = unclear risk of bias.

inclusion criteria of any patient with a sore throat, with prevalence of GABHS pharyngitis ranging from 17% to 31%.

Quality assessment

The assessment of study quality using the QUADAS-2 framework is summarised in Table 2. The authors judged nine studies to be at low risk of bias, six to be at moderate risk of bias, and three to be at high risk of bias. One study reported data from three sites, two of which were judged low risk of bias and one high risk of bias.[17]

Accuracy of the overall clinical impression ('clinical gestalt')

The accuracy of clinical gestalt as a diagnostic test for GABHS pharyngitis, ARS, and CAP is summarised in Table 3. Due to differences in the clinical presentation of pneumonia in children and adults, as well as observed heterogeneity in the summary ROC curve, results for the accuracy of CAP in adults and children with suspected pneumonia are reported separately. The summary estimates for the positive (LR+) and negative (LR-) likelihood ratios were LR+ 7.7, 95% confidence interval (CI) = 4.8 to 11.5 and LR– 0.54, 95% CI = 0.42 to 0.65 for the diagnosis of CAP in adults; LR+ 2.7, 95% CI = 1.1 to 4.3, and LR– 0.63, 95% CI = 0.20 to 0.98 for the diagnosis of CAP in children; LR+ 3.0, 95% CI = 2.1 to 4.4 and LR– 0.37, 95% CI = 0.29 to 0.46 for ARS in adults; LR+ 3.9, 95% CI = 2.4 to 5.9 and LR– 0.33, 95% CI = 0.20 to 0.50 for ABRS in adults; and LR+ 2.1, 95% CI = 1.6 to 2.8 and LR– 0.47, 95% CI = 0.36 to 0.60 for GABHS pharyngitis

Table 3. Summary estimates of diagnostic accuracy of clinical gestalt for the diagnosis of common respiratory infections

Author, year	TP	FP	FN	TN	Sensitivity	Specificity	LR+	LR–	DOR	AUC
CAP (adults)										
Gonzalez-Ortiz, 1995[30]	24	6	29	82	0.45	0.93	6.6	0.59		
Lieberman, 2003[31]	14	37	5	194	0.74	0.84	4.6	0.31		
Melbye, 1988[32]	3	3	8	57	0.27	0.95	5.5	0.77		
Moberg, 2016[33]	14	2	30	53	0.32	0.96	8.6	0.71		
Tape (Illinois), 1991[17]	58	108	63	876	0.48	0.89	4.4	0.58		
Tape (Nebraska), 1991[17]	24	9	14	72	0.63	0.89	5.7	0.41		
Tape (Virginia), 1991[17]	24	14	6	96	0.80	0.87	6.3	0.23		
van Vugt, 2013[11]	41	31	99	2639	0.29	0.99	25.2	0.72		
Summary estimate (95% CI)					0.50 (0.37 to 0.62)	0.93 (0.87 to 0.97)	7.7 (4.8 to 11.5)	0.54 (0.42 to 0.65)	14.2 (9.0 to 21.0)	0.80
CAP (children)										
Grossman, 1988[27]	41	33	10	77	0.80	0.70	2.7	0.28		
Mahabee-Gittens, 2005[28]	6	56	38	410	0.14	0.88	1.1	0.98		
Redd, 1994[29]	19	20	21	166	0.48	0.89	4.4	0.59		
Summary estimate (95% CI)					0.46 (0.12 to 0.84)	0.84 (0.70 to 0.92)	2.7 (1.1 to 4.3)	0.63 (0.20 to 0.98)	5.5 (1.1 to 16)	0.80
ARS (adults)[a]										
Berg, 1981[23]	21	2	4	23	0.84	0.92	10.5	0.17		
Berg, 1985[24]	33	11	10	36	0.77	0.77	3.3	0.30		
Berg, 1988[25]	52	16	27	60	0.66	0.78	3.0	0.44		
Williams, 1992[26]	72	55	23	97	0.76	0.64	2.1	0.38		
Summary estimate (95% CI)					0.73 (0.66 to 0.79)	0.75 (0.64 to 0.84)	3.0 (2.1 to 4.4)	0.37 (0.29 to 0.46)	8.3 (4.9 to 13.1)	0.77
ABRS (adults)[a]										
Berg, 1981[23]	21	2	4	23	0.84	0.92	10.5	0.17		
Berg, 1985[24]	33	11	10	36	0.77	0.77	3.3	0.30		
Berg, 1988[25]	52	16	27	60	0.66	0.78	3.0	0.44		
Summary estimate (95% CI)					0.74 (0.61 to 0.84)	0.80 (0.72 to 0.87)	3.9 (2.4 to 5.9)	0.33 (0.20 to 0.50)	13.0 (5.0 to 27)	0.84
GABHS pharyngitis (both)										
Attia, 2001[34]	157	148	61	221	0.72	0.60	1.8	0.47		
Centor, 1981[8]	29	47	11	147	0.73	0.76	3.0	0.36		
Dobbs, 1996[22]	44	47	28	87	0.61	0.65	1.7	0.60		
Summary estimate (95% CI)					0.69 (0.61 to 0.76)	0.67 (0.57 to 0.76)	2.1 (1.6 to 2.8)	0.47 (0.36 to 0.60)	4.6 (2.6 to 7.8)	0.73

[a]ARS was defined as sinusitis diagnosed using any reference standard, whereas ABRS was defined as sinusitis diagnosed using antral puncture fluid inspection as the reference standard. ABRS = acute bacterial rhinosinusitis. ARS = acute rhinosinusitis. AUC = area under the receiver operating characteristic curve. CAP = community-acquired pneumonia. DOR = diagnostic odds ratio (LR+/LR–). FN = false negative. FP = false positive. GABHS = group A beta-haemolytic streptococcal. LR– = negative likelihood ratio. LR+ = positive likelihood ratio. TN = true negative. TP = true positive.

in both adults and children. Based on the diagnostic odds ratio, clinical gestalt was most accurate for diagnosis of CAP in adults (DOR 14.2, 95% CI = 9.0 to 21.0), ABRS in adults (DOR 13.0, 95% CI = 5.0 to 27.0), and ARS in adults (DOR 8.3, 95% CI = 4.9 to 13.1). It was less accurate for the diagnosis of CAP in children (DOR 5.5) and GABHS pharyngitis (DOR 4.6).

The summary ROC curves are shown in Figure 2. The summary AUROCC of the overall clinical impression as a test for CAP was 0.80 in both children and adults, 0.77 for ARS in adults, 0.84 for ABRS in adults, and 0.73 for GABHS pharyngitis in adults and children. Note that the C-statistic for CAP in children was unreliable in the authors' judgement based on the small number of studies and high heterogeneity. Inspection of the summary ROC curves in Figure 2 reveals different patterns of heterogeneity for each disease. There was good homogeneity for the diagnosis of acute pharyngitis, despite the fact that the three studies enrolled children in one, adults in another, and both in a third. For sinusitis, there was good homogeneity with regards to sensitivity (range 0.71 to 0.84) but

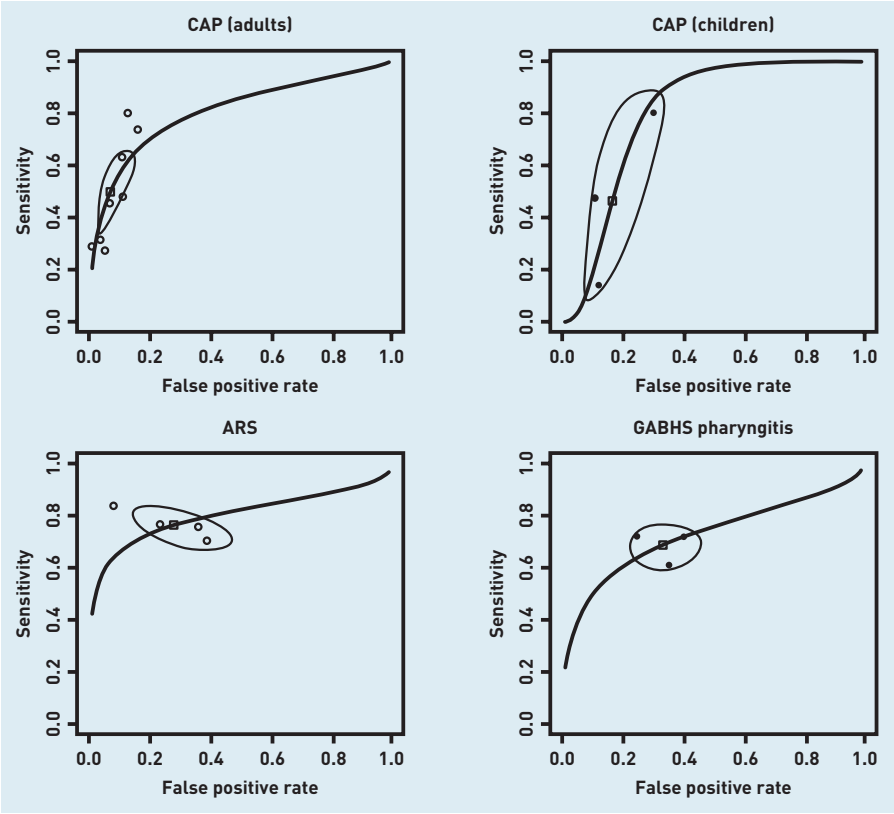

Figure 2. Summary receiver operating characteristic curves (ROC) are shown for the accuracy of clinical gestalt in the diagnosis of community-acquired pneumonia (CAP) in adults, CAP in children, group A beta-haemolytic streptococcal (GABHS) pharyngitis, and acute rhinosinusitis (ARS).

less with regards to specificity (range 0.61 to 0.92).

For the diagnosis of CAP in adults, the ROC curve showed a pattern that was consistent with a threshold effect. That is, as sensitivity increases, specificity decreases, with the points arrayed along the ROC curve. There was also better homogeneity for studies of CAP in adults compared with studies in children, which are presented separately in the ROC curves. As noted before, most studies in this group were limited to patients with clinically suspected disease. The one study with very broad inclusion criteria of any patient with cough had the highest specificity (0.99) but among the lowest sensitivities (0.29), perhaps a consequence of the low prevalence of CAP.[11]

Accuracy of clinical decision rules

For comparison with the overall clinical impression, the authors determined the accuracy of CDRs for GABHS pharyngitis in children and adults,[8,35] CAP,[36] and acute bacterial rhinosinusitis (ABRS).[10] The accuracy of the Strep Score for GABHS pharyngitis in adults and children was obtained from recent systematic reviews.[13,35] The accuracy of the CDR for CAP was obtained from a large European study of

Table 4. Accuracy of selected clinical decision rules for pneumonia, pharyngitis, and acute rhinosinusitis

Study	Sensitivity, %	Specificity, %	LR+	LR–	DOR
CAP (adults)[36]	45	90	4.4	0.61	7.2
Sinus Score for ARS[a] (adults)[10]	89	53	1.9	0.53	3.6
Sinus Score for ABRS[b] (adults)[10]	78	63	2.1	0.35	5.9
Strep (Centor) Score (adults)[13]	82	49	1.6	0.37	4.2
Strep (Centor) Score (children)[35]	41	85	2.0	0.8	2.5

[a]Reference standard for ARS was abnormal CT. [b]Reference standard for ABRS was antral fluid culture positive for a pathogen. ABRS = acute bacterial rhinosinusitis. ARS = acute rhinosinusitis. CAP = community-acquired pneumonia. DOR = diagnostic odds ratio (LR+/LR–). LR– = negative likelihood ratio. LR+ = positive likelihood ratio.

outpatients with acute cough where all received a chest radiograph.[36] The CDRs for ARS and ABRS were developed by the author based on a study of 175 primary care patients who all underwent CT, and antral puncture for fluid and culture if fluid was seen on CT.[10] ARS was defined as abnormal CT, and ABRS as abnormal culture of antral puncture fluid, as in the clinical gestalt studies. The accuracy of the CDRs are summarised in Table 4.

DISCUSSION

Summary

This is the first systematic review of the accuracy of clinical gestalt or the overall clinical impression as a diagnostic test. The authors found that the overall clinical impression is an accurate diagnostic test for CAP, ARS, and ABRS in adults (DOR 14.2, 8.3, and 13.0, respectively), and is moderately accurate for the diagnosis of GABHS pharyngitis in adults and children (DOR 4.6) and for the diagnosis of CAP in children (DOR 5.5).

Clinical gestalt is more accurate than individual signs and symptoms for all three conditions, and compares well with clinical decision rules. For example, using a cut-off of three or more out of four symptoms as a positive test, the Strep Score had diagnostic odds ratios of 4.2 in adults and 2.5 in children, compared with a DOR of 4.6 for the overall clinical impression in mixed populations of adults and children. The CDR for CAP in adults had a DOR of 7.2, compared with a DOR of 14.2 for the overall clinical impression in adults. For ARS, the CDR had a DOR of 3.6 compared with 8.3 for clinical gestalt. For ABRS the CDR had a DOR 5.9, compared with 13.0 for clinical gestalt. In all cases, the overall clinical impression performed as well or better than the clinical decision rule.

Patterns of heterogeneity differed between conditions. There was good homogeneity around estimates of the accuracy of gestalt for pharyngitis, for ABRS using antral puncture as the reference standard, and for CAP in adults. A threshold effect can be observed for the diagnosis of CAP. A threshold effect is the result of a trade-off between sensitivity and specificity, and may occur when different definitions of the outcome of interest are used, such as different thresholds for diagnosis of CAP. Some physicians may prioritise sensitivity at the price of specificity, and others specificity at the price of sensitivity.

Strengths and limitations

A strength of this study is the fact that the results for the accuracy of clinical gestalt were fairly consistent for adults with CAP, ABRS, and pharyngitis based on inspection of the summary ROC curves. Other strengths of the present study include the use of modern methods for diagnostic meta-analysis, a comprehensive search, and that only three of 18 studies were judged to be at high risk of bias. This study had several limitations as well: the clinical decision rules discussed above for ARS and CAP have not been prospectively validated. However, accuracy usually suffers during prospective validation, so the fact that gestalt was as accurate as these proposed CDRs is notable. There were a fairly small number of studies, several were quite old, some were at high risk of bias, and three of the four for ARS were by the same author. There was also considerable heterogeneity with regards to inclusion criteria, the age of participants, and the reference standards used. Finally, the studies of pneumonia generally only included studies where there was already some clinical suspicion of CAP. However, only a minority in each of the nine studies had CAP diagnosed by radiography.

Comparison with existing literature

The authors conclude that clinical gestalt is either similarly accurate to or more accurate than CDRs based on usual metrics of diagnostic accuracy. Since clinical gestalt requires no calculations, no algorithm, and no computer, it is not surprising that it is far more widely used than CDRs for clinical decision making. That said, the ability to use clinical gestalt as an accurate test for pneumonia or acute rhinosinusitis is not innate. It must be developed and cultivated, as any skill, and likely requires exposure to a great many cases with a known outcome ('patterns') before it is fully developed and accurate. Artificial neural networks can be 'trained' to create a complex algorithm by exposing the network to a large number of patterns with known outcomes, eventually developing the ability to accurately make predictions for new cases.

Multivariate models and neural networks typically require several hundred or more patterns to create a predictive model. How many of these known cases or 'patterns' are required before the human brain is trained remains unclear. Bierema proposes a model for professional knowledge development that identifies stages of novice, beginner, competent, proficient, expert, and generative leader.[37] For novice and beginner learners, CDRs can be used to hone diagnostic skills and teach them the best independent predictors of disease,

providing focus and a framework for their diagnostic training. For the proficient and expert physician, the CDR moves to the background, while a physician who is a generative leader may further develop and improve CDRs.

Implications for research and practice

The authors propose that use of formal CDRs is potentially most useful for early-stage clinicians, who have not yet been exposed to a large number of patterns. As they develop their own clinical gestalt, informed by repeated use of validated CDRs, they may eventually rely less and less on the CDR. But even for experienced clinicians CDRs can serve as a back-up to their clinical gestalt. For example, if a physician judges that a patient with CAP can be treated as an outpatient, it is still worthwhile to double-check that judgement by calculating the CRB-65 prognostic score for pneumonia.[38] In fact, both the clinical decision rule and clinical gestalt only identified about half of the patients with pneumonia, missing the other half. Thus, use of a CDR and clinical gestalt may be complementary and supportive of each other rather than an either/or proposition.

In conclusion, clinical gestalt is accurate for the diagnosis of CAP, ARS, and ABRS in adults, and the overall accuracy is similar to or better than that of clinical decision rules. Experienced clinicians should be confident in their use of the overall clinical impression and use clinical decision rules as a backstop to that judgement. Trainees, on the other hand, may benefit more from explicit use of CDRs until they develop their clinical skills. Further work is needed to understand how to best teach clinical gestalt to trainees.

Future studies of clinical diagnosis should primarily include an 'overall clinical impression' question to gather further data on the accuracy of clinical gestalt for a range of conditions, including of course non-infectious conditions such as chest pain, deep vein thrombosis, and pulmonary embolism. If found to be accurate and reliable for the diagnosis of a disease, the overall clinical impression could be built into guidelines regarding the evaluation of a range of conditions such as suspected sepsis, myocardial infarction, depression, and early diagnosis of cancer. It will also be important to consider how an overall judgement about the likelihood of disease fits with the threshold framework for decision making, such that a judgement of 'disease is unlikely' also falls below the test threshold for that disease.[39]

Funding

No funding was received for this study.

Ethical approval

This study was exempt from ethical approval as it was limited to secondary data analysis.

Provenance

Freely submitted; externally peer reviewed.

Competing interests

The authors have declared no competing interests.

Discuss this article

Contribute and read comments about this article: **bjgp.org/letters**

REFERENCES

1. Ridderikhoff J. Medical problem-solving: an exploration of strategies. *Med Educ* 1991; **25(3):** 196–207.
2. Ridderikhoff J. Problem-solving in general practice. *Theor Med* 1993; **14(4):** 343–363.
3. Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ* 2003; **37(8):** 695–703.
4. Stiell IG, Greenberg GH, McKnight RD, *et al.* A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Ann Emerg Med* 1992; **21(4):** 384–390.
5. Lucassen W, Geersing GJ, Erkens PM, *et al.* Clinical decision rules for excluding pulmonary embolism: a meta-analysis. *Ann Intern Med* 2011; **155(7):** 448–460.
6. McIsaac WJ, Goel V. Effect of an explicit decision-support tool on decisions to prescribe antibiotics for sore throat. *Med Decis Making* 1998; **18(2):** 220–228.
7. Ebell M. AHRQ White Paper: use of clinical decision rules for point-of-care decision support. *Med Decis Making* 2010; **30(6):** 712–721.
8. Centor RM, Witherspoon JM, Dalton HP, *et al.* The diagnosis of strep throat in adults in the emergency room. *Med Decis Making* 1981; **1(3):** 239–246.
9. McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. *CMAJ* 2000; **163(7):** 811–815.
10. Ebell MH, Hansen JG. Proposed clinical decision rules to diagnose acute rhinosinusitis among adults in primary care. *Ann Fam Med* 2017; **15(4):** 347–354.
11. van Vugt SF, Verheij TJ, de Jong PA, *et al.* Diagnosing pneumonia in patients with acute cough: clinical judgment compared to chest radiography. *Eur Respir J* 2013; **42(4):** 1076–1082.
12. Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. *JAMA* 1997; **278(17):** 1440–1445.
13. Aalbers J, O'Brien KK, Chan WS, *et al.* Predicting streptococcal pharyngitis in adults in primary care: a systematic review of the diagnostic accuracy of symptoms and signs and validation of the Centor score. *BMC Med* 2011; **9:** 67.
14. Shaikh N, Swaminathan N, Hooper EG. Accuracy and precision of the signs and symptoms of streptococcal pharyngitis in children: a systematic review. *J Pediatr* 2012; **160(3):** 487–493.
15. Shah SN, Bachur RG, Simel DL, Neuman MI. Does this child have pneumonia? The rational clinical examination systematic review. *JAMA* 2017; **318(5):** 462–471.
16. Ebell MH, Smith MA, Barry HC, *et al.* The rational clinical examination. Does this patient have strep throat? *JAMA* 2000; **284(22):** 2912–2918.
17. Tape TG, Heckerling PS, Ornato JP, Wigton RS. Use of clinical judgment analysis to explain regional variations in physicians' accuracies in diagnosing pneumonia. *Med Decis Making* 1991; **11(3):** 189–197.
18. 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(8):** 529–536.
19. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20(19):** 2865–2884.
20. Zhou Y, Dendukuri N. Statistics for quantifying heterogeneity in univariate and bivariate meta-analyses of binary data: the case of meta–analyses of diagnostic accuracy. *Stat Med* 2014; **33(16):** 2701–2717.
21. von Hippel PT. The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med Res Methodol* 2015; **15:** 35.
22. Dobbs F. A scoring system for predicting group A streptococcal throat infection. *Br J Gen Pract* 1996; **46(409):** 461–464.
23. Berg O, Bergstedt H, Carenfelt C, *et al.* Discrimination of purulent from nonpurulent maxillary sinusitis. Clinical and radiographic diagnosis. *Ann Otol Rhinol Laryngol* 1981; **90(3 Pt 1):** 272–275.
24. Berg O, Carenfelt C. Etiological diagnosis in sinusitis: ultrasonography as clinical complement. *Laryngoscope* 1985; **95(7 Pt 1):** 851–853.
25. Berg O, Carenfelt C. Analysis of symptoms and clinical signs in the maxillary sinus empyema. *Acta Otolaryngol* 1988; **105(3-4):** 343–349.
26. Williams JW, Jr., Simel DL, Roberts L, Samsa GP. Clinical evaluation for sinusitis. Making the diagnosis by history and physical examination. *Ann Intern Med* 1992; **117(9):** 705–710.
27. Grossman LK, Caplan SE. Clinical, laboratory, and radiological information in the diagnosis of pneumonia in children. *Ann Emerg Med* 1988; **17(1):** 43–46.
28. Mahabee-Gittens EM, Grupp-Phelan J, Brody AS, *et al.* Identifying children with pneumonia in the emergency department. *Clin Pediatr (Phila)* 2005; **44(5):** 427–435.
29. Redd SC, Patrick E, Vreuls R, *et al.* Comparison of the clinical and radiographic diagnosis of paediatric pneumonia. *Trans R Soc Trop Med Hyg* 1994; **88(3):** 307–310.
30. Gonzalez Ortiz MA, Carnicero Bujarrabal M, Varela Entrecanales M. Prediction of the presence of pneumonia in adults with fever. *Med Clin (Barc)* 1995; **105(14):** 521–524.
31. Lieberman D, Shvartzman P, Korsonsky I, Lieberman D. Diagnosis of ambulatory community-acquired pneumonia. Comparison of clinical assessment versus chest X-ray. *Scand J Prim Health Care* 2003; **21(1):** 57–60.
32. Melbye H, Straume B, Aasebo U, Brox J. The diagnosis of adult pneumonia in general practice. The diagnostic value of history, physical examination and some blood tests. *Scand J Prim Health Care* 1988; **6(2):** 111–117.
33. Moberg AB, Taleus U, Garvin P, *et al.* Community-acquired pneumonia in primary care: clinical assessment and the usability of chest radiography. *Scand J Prim Health Care* 2016; **34(1):** 21–27.
34. Attia MW, Zaoutis T, Klein JD, Meier FA. Performance of a predictive model for streptococcal pharyngitis in children. *Arch Pediatr Adolesc Med* 2001; **155(6):** 687–691.
35. Le Marechal F, Martinot A, Duhamel A, *et al.* Streptococcal pharyngitis in children: a meta-analysis of clinical decision rules and their clinical variables. *BMJ Open* 2013; **3(3):** pii: e001482.
36. van Vugt SF, Broekhuizen BD, Lammens C, *et al.* Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic study. *BMJ* 2013; **346:** f2450.
37. Bierema LL. Navigating professional white water: rethinking continuing professional education at work. *New Dir Adult Contin Educ* 2016; **2016(151):** 53–67.
38. McNally M, Curtain J, O'Brien KK, *et al.* Validity of British Thoracic Society guidance (the CRB-65 rule) for predicting the severity of pneumonia in general practice: systematic review and meta-analysis. *Br J Gen Pract* 2010; DOI: https://doi.org/10.3399/bjgp10X532422.
39. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; **302(20):** 1109–1117.