REVIEW

Narrative Review: Should Teaching of the Respiratory Physical Examination Be Restricted Only to Signs with Proven Reliability and Validity?

Jochanan Benbassat, MD¹ and Reuben Baumal, MD²

¹Myers-JDC-Brookdale Institute, Smokler Center for Health Policy Research, Jerusalem, Israel; ²Department of Laboratory Medicine and Pathobiology, Toronto Hospital for Sick Children and University of Toronto, Toronto, ON, Canada.

OBJECTIVE: To review the reported reliability (reproducibility, inter-examiner agreement) and validity (sensitivity, specificity and likelihood ratios) of respiratory physical examination (PE) signs, and suggest an approach to teaching these signs to medical students.

METHODS: Review of the literature. We searched Paper Chase between 1966 and June 2009 to identify and evaluate published studies on the diagnostic accuracy of respiratory PE signs.

RESULTS: Most studies have reported low to fair reliability and sensitivity values. However, some studies have found high specificities for selected PE signs. None of the studies that we reviewed adhered to all of the STARD criteria for reporting diagnostic accuracy.

CONCLUSIONS: Possible flaws in study designs may have led to underestimates of the observed diagnostic accuracy of respiratory PE signs. The reported poor reliabilities may have been due to differences in the PE skills of the participating examiners, while the sensitivities may have been confounded by variations in the severity of the diseases of the participating patients.

IMPLICATION FOR PRACTICE AND MEDICAL EDUCATION: Pending the results of properly controlled studies, the reported poor reliability and sensitivity of most respiratory PE signs do not necessarily detract from their clinical utility. Therefore, we believe that a meticulously performed respiratory PE, which aims to *explore a diagnostic hypothesis*, as opposed to a PE that aims to detect a disease in an asymptomatic person, remains a cornerstone of clinical practice. We propose teaching the respiratory PE signs according to their importance, beginning with signs of life-threatening conditions and those that have been reported to have a high specificity, and ending with signs that are "nice to know," but are no longer employed because of the availability of more easily performed tests.

KEY WORDS: respiratory physical examination; diagnostic accuracy; medical students.

J Gen Intern Med 25(8):865-72

DOI: 10.1007/s11606-010-1327-8

© Society of General Internal Medicine 2010

INTRODUCTION

Despite the evidence that supports the role of the physical examination (PE) in the assessment of a patient's disease^{1,2}, there is considerable controversy as to whether the PE has outlived its usefulness. Some suggest that imaging may provide a more direct view into the body and prevent errors³. Others argue that the value of ancillary testing is overrated⁴; that it cannot replace a physician's ability to recognize familiar patterns of disease^{5,6}; and that the failure to perceive the importance of the PE is due to poor teaching and learning of basic clinical skills⁵⁻⁸.

This later criticism has led to major changes in the way the skills of patient interviewing^{9,10} and diagnostic reasoning^{11,12} are taught. However, teaching of the PE has remained unchanged. From the 1950s¹³ to the present time¹², textbooks of the PE continue to offer a comprehensive and unselective compilation of signs, which include those no longer considered to be useful¹⁴, while often excluding important PE signs¹⁵. Some medical schools do encourage students to form diagnostic hypotheses early on while listening to the patient's narrative and conduct a targeted PE aimed at confirming or refuting these hypotheses⁸. However, we suspect that even in these schools, this educational approach is not reinforced during the subsequent clinical undergraduate training. During their clerkship and internship, students are expected to perform a complete PE along a predetermined sequence, which begins with the patient's appearance and vital signs, and moves on to an examination from head to toes. Consequently, students often perform hasty PEs with frequent shortcuts that are unlikely to detect physical findings. This may explain the observed decline of students' breast examination skills over the course of their training¹⁶ and students' reliance on "hard" laboratory and imaging data rather than on the clinical assessment of their patients⁴.

We believe that, in order to enhance students' appreciation of the PE, its teaching should focus on *selected* and *important* PE skills⁸ rather than overwhelm the learners with an all-inclusive list of signs. One possible way to discriminate between important and less important components of the PE would be to select for signs with proven diagnostic accuracy. Indeed, the call by Sackett and Rennie¹⁷ for a rational approach to the clinical examination triggered a series of publications (e.g.,¹⁸), reviews (e.g.,¹⁴) and a textbook¹⁹ dealing with the evidence base of the PE. Most of them emphasized the need for additional studies of the reliability and validity of PE signs. However, we know of no suggestions for reconciling the teaching of the PE with the paucity of evidence on the diagnostic accuracy of most PE signs.

Received October 3, 2009

Revised December 2, 2009

Accepted March 4, 2010

Published online March 27, 2010

The objective of this paper is to update the presently available evidence, or lack thereof, for the reliability and validity of PE signs, and suggest a modified approach to their teaching to undergraduate medical students. We chose the PE of the respiratory system because its importance has been debated ever since the advent of chest radiography²⁰.

METHODS

We used Paper Chase²¹ to search Medline and Old Medline between 1966 and June 2009. An effort was made to identify all *original* published studies of the diagnostic accuracy of *specific* respiratory PE signs. We excluded *reviews* and studies of the diagnostic value of *combinations* of PE signs. First, we used the terms ['physical examination'] and ['respiratory' or 'pulmonary diseases' or 'lung diseases']. Of the 392 hits, only 9 were original studies of the reliability and validity of respiratory PE signs. Second, we used the terms ['physical examination' or 'auscultation' or 'percussion'] and ['pneumonia' or 'pleural effusion' or 'airway obstruction' or 'pneumothorax' or 'asthma' or 'pulmonary embolism'] and obtained 701 hits, which included 13 additional studies. Finally, we searched the reference sections of all relevant studies and identified another 38, some of which were published before 1966. After excluding 17 studies of the validity and 3 studies of the reliability of respiratory signs in children and infants, we were left with a total of 40 studies: 13 of the reliability²²⁻³⁴, 20 of the validity³⁵⁻⁵⁴ and 7 of both⁵⁵⁻⁶¹.

Other search strategies, such as those using the terms ['physical examination'] and ['diagnostic accuracy'], or ['physical examination'] and ['sensitivity' or 'specificity' or 'reliability'] and ['respiratory'], did not identify additional studies. Our failure to capture most studies of the diagnostic accuracy of PE signs using conventional literature searches may have been due to less than optimal indexing.

Studies of diagnostic accuracy are subject to various sources of bias that may result from their design, selection of patients, performance of the test and analysis of data. The recognition that the quality of reporting of these studies is often deficient led to the development of a 25-item list of Standards for Reporting of Diagnostic Accuracy (STARD)⁶². Similar to the experts who developed this list, we realize that the methodology for designing and conducting studies of diagnostic accuracy requires further development. However, at present, these criteria are the best available measure of the value of such studies, including those of the diagnostic accuracy of the PE⁶³. Therefore, we chose to evaluate the studies included in the present review by the degree of their adherence to the STARD check list.

We reviewed all 40 publications and tabulated (1) the degree of their adherence to the STARD criteria⁶², (2) the reliability of the various respiratory PE signs and (3) their validity for detecting defined disorders. We use the term "reliability" interchangeably with the reproducibility of the findings that are obtained when the same PE is repeated on the same patient by the same or different examiners. Most commonly, reliability was reported as kappa statistics on a -1 (complete disagreement) to 0 (chance agreement) to +1 (perfect agreement) scale. Values between 0 and 0.4 are commonly accepted as indicating low agreement, those between 0.4 and 0.6, fair agreement; between 0.6 and 0.8, good agreement; between 0.8 and 1, perfect agreement⁶⁴. Less frequently, reliability was presented as agreement rates between examiners (e.g.,²²), coefficients of

correlation (e.g.,²⁹) or the "standard deviation agreement index" (e.g.,²⁵), which is, similar to the kappa statistics, a measure of inter-examiner agreement beyond the one expected by chance. "Validity" refers to the ability of a PE sign to discriminate between patients with and without the disease under consideration, and it is expressed as sensitivity and specificity relative to an agreed upon gold standard. The phrase "diagnostic accuracy," as used here, refers to the contribution of a given PE sign to establishing a diagnosis and to its usefulness in clinical practice.

RESULTS

(1) Adherence to the STARD criteria for diagnostic accuracy

Most of the reviewed studies were published before the development of the STARD criteria in 2003, and none of them adhered to all of these criteria. Most studies complied with the following criteria: they were prospective (33/40), provided an explicit or implicit definition of the study objectives (37/40) and data on the study populations (39/40), methods of patient recruitment (27/40) and methods of presentation of results (38/40). All validity studies presented their findings relative to a gold standard of diagnosis. In most studies, the test results were dichotomous (PE signs present or absent); in some studies, such as those of respiratory rates, the results were presented as above or below cutoff values. The examiners in 16 of the 20 reliability studies were blinded to the findings of other examiners, and those in 20 of the 27 validity studies were blinded to the results of the gold standard (data not shown in table format).

However, only few of the reviewed studies adhered to the following criteria: only 2 of the 20 reliability studies, and 4 of the 27 validity studies reported data on the severity of disease of the participating patients; only 10 reliability studies and 14 validity studies reported on attempts to use standardized PE procedures in order to enhance the consistency of the examiners' PE techniques; only 1 reliability study and 4 validity studies presented data on the number of qualifying patients who did not participate in the study; only 7 validity studies reported the reliability of the various PE signs (data not shown in table format).

(2) Reliability (reproducibility) of respiratory PE signs (Tables 1, 2)

The only study of the *intra*-examiner reproducibility that we know of found that the examiners disagreed with themselves in 11–26% of the cases and that pulmonary specialists were significantly less self-consistent than medical students³². Of the 20 reliability studies, only 4 studies^{32,34,60,61} reported inter-examiner agreement rates above $k=0.6$ for one or more PE signs. Another five studies^{24,25,29,56,57} reported more than 90% or "almost total" inter-examiner agreement for at least one PE sign.

The following PE signs were reported by some, *but not most* studies to have reliabilities of $kappa=0.6$ –1.0 or disagreement rates of 10% or less: chest movements, clubbing, vocal fremitus, dullness on percussion and reduced auscultatory percussion (Table 1), breath sound intensity, crepitations, vocal resonance (diminished) and wheezes (Table 2). Low to fair

Table 1. Reliability of Respiratory Physical Examination Signs Elicited by Inspection, Palpation and Percussion

Sign	Inter-examiner reliability ^a	Reference
Respiratory distress (use of accessory muscle/intercostal retractions)	"31% disagreement" "26% disagreement" ^b	22 26 25,28
Cyanosis	"18% disagreement" 0.36 ^b	26 33 25
Deformities of thorax/ barrel chest and spinal column/ kyphosis	"29% disagreement" "16% disagreement" "24% disagreement" ^b	22 24 26 25
Respiratory rate	0.25 ^b	33 25
Chest movements/ expansion (decreased/ asymetric)	"22% disagreement" "19% disagreement" 0.14 0.38 0.85 ^b	24 26 31 33 61 25
Clubbing	"Considerable disagreement" "8% disagreement" "Almost total agreement" 0.45	23 24 25 33
Crico-sternal distance	0.28	33
Position of trachea	0.01 ^b	33 25
Vocal fremitus (reduced)	0.24 0.86 ^b	33 61 25
Dullness on percussion	0.34 0.16 to 0.46 ^c "3% disagreement" 0.52 0.84 ^b	31 32 56 33 61 25
Hyperresonance on percussion	"33% disagreement" "23% disagreement" 0.26 0.25 to 0.43 ^c 0.50 ^b	22 26 31 32 33 25,28
Diaphragmatic expansion	"20% disagreement" "60% agreement" 0.05	26 30 59
Cardiac dullness (impaired/absent)	"32% disagreement" "24% disagreement" 0.47	22 26 59
Auscultatory percussion (hyperresonance)	0.37	59
Auscultatory percussion (reduced)	"10% disagreement" -0.04 to 0.45 ^c 0.76	56 60 61

^aUnless stated otherwise, the inter-examiner reliability is expressed as kappa

^bInter-examiner reliability was expressed as "standard deviation agreement index," and it was about midway between perfect agreement rates and those expected by chance alone

^cRange of reproducibility between multiple pairs of examiners

inter-examiner agreement rates, i.e., kappa=0.0–0.6, were consistently reported in one or more studies for the following PE signs: cyanosis, respiratory rate, crico-sternal distance, deformities of the thorax, respiratory distress, position of trachea, hyperresonance on percussion, diaphragmatic expansion and

cardiac dullness (Table 1); bronchial breathing, pectoriloquy, bronchophony, egophony, rhonchi, vocal resonance (increased), prolonged expiratory phase and pleural friction rub (Table 2).

- (3) Sensitivity, specificity and likelihood ratios of respiratory PE signs for defined diseases (Tables 3, 4, 5)

Table 2. Reliability of Respiratory Physical Examination Signs Elicited by Auscultation

Sign	Inter-examiner reliability ^a	Reference
Bronchial breathing	0.19 0.32 -0.14 to 0.14 ^c	31 35 60
Whispering pectoriloquy	0.11	33
Whispering pectoriloquy	0.16 to 0.25 ^c	60
Bronchophony	-0.10 to 0.18 ^c	60
Egophony	^b	25
Breath sound intensity	"Differences among examiners—not statistically significant" "29% disagreement" Coefficient of correlation: 0.58 Coefficient of correlation: 0.97 0.16 0.23 to 0.43 ^c 0.43 0.38 0.42 0.89 ^b	55 22,26 29 29 31 32 33 59 58 61 25,28
Rales/crackles/crepitations	0.36 0.32 0.30 to 0.63 ^c 0.41 0.24 to 0.65 ^c 0.67 ^b	31 58 32 33 60 61 25,28
Rhonchi	0.55 -0.06 to 0.18 ^c ^b	31 60 25
Vocal resonance (diminished)	0.78 ^b	61 25
Vocal resonance (increased)	^b	25
Wheezes	0.54 to 0.70 ^c 0.51 0.50 0.43 0.85 -0.05 to 0.65 ^c ^b	32 33 59 58 34 60 28
Wheezing at maximal forced exhalation	"90% agreement"	57
Expiratory phase (prolonged)	0.53 0.35	31 33
Pleural friction rub	-0.02 0.51 -0.02	31 33 61

^aUnless stated otherwise, the inter-examiner reliability is expressed as kappa

^bInter-examiner reliability expressed as "standard deviation agreement index," and it was about midway between perfect agreement rates and those expected by chance alone

^cRange of reproducibility between multiple pairs of examiners

The vast majority of studies found sensitivity values of 0.5 or less (Tables 3, 4). Sensitivity values with likelihood ratios-negative (LR-) of 0.2 or less were reported only for dullness on percussion in diagnosing pleural effusion among inpatients with respiratory symptoms (Tables 3, 5), forced expiratory time of more than 6 s duration for obstructive airway disease (OAD) in asymptomatic plumbers screened for lung diseases and for diminished breath sounds in diagnosing hemo-pneumothorax in the context of trauma (Tables 4, 5).

On the other hand, high specificity values with LR+ of 4.0 or more have been reported for PE signs, such as those of pulmonary consolidation, in as many as 17 of the 27 validity studies (Tables 3, 4). Table 5 lists the settings and clinical contexts in which respiratory PE signs may be useful for increasing or reducing the post-test odds of a specific diagnosis.

DISCUSSION

Two main findings emerge from the present review. First, most studies have found low to fair reliability values for respiratory PE signs. This finding detracts from the credibility of the validity studies and is consistent with the view that "clinical skills textbooks fail evidence-based examination"⁶⁵. Second, since none of the reviewed studies complied with all of the STARD criteria⁶², their design may have been flawed. Consequently, the reported reliability (Tables 1, 2) and sensitivity (Tables 3, 4) values should be interpreted with caution.

A finding of a poor reliability (i.e., high examiner variability) of a PE sign may indicate either that it has a poor diagnostic accuracy or that some of the examiners had deficient PE skills. This latter possibility is consistent with the reported lack of

Table 3. Sensitivity, Specificity and Likelihood Ratios of Respiratory PE Signs Elicited by Inspection, Palpation and Percussion

Sign	Diagnosis	Gold standard	Sensitivity	Specificity	LR positive ^a	LR negative ^a	Reference
Laryngeal height <4 cm	OAD ^b	Spirometry	0.24	0.76	1.0	1.0	47
Clubbing	Lung cancer	Histology	0.10	-	-	-	45
Hoover's sign (inward motion of the lower lateral rib cage with inspiration)	OAD ^b - Mild	FEV1 >0.8 normal	0.00	-	-	-	51
	- Moderate	0.5-0.8	0.36				
	- Severe	0.3-0.5	0.43				
	- Very severe	<0.3	0.76				
Cyanosis	Pneumonia	Chest x-ray	0.15	0.97	5.0	0.9	54
Heart rate >100/min	Pneumonia	Chest x-ray	0.17-0.50	0.62-0.92	1.0-2.5	0.6-0.9	38,39,41,48,52,54
>120/min			0.18	0.96	4.5	0.8	53
Heart rate >100/min	Pulmonary emboli	Pulmonary angio	0.58	-	-	-	37
Respiratory rate >20/min	Pneumonia	Chest x-ray	0.45,0.90	0.31,0.79	1.3, 2.1	0.3, 0.7	52,54
>25			0.07,0.28	0.92, 0.99	3.5, 7.0	0.8,0.9	38,53
>30			0.29-0.41	0.72-0.88	1.5-2.6	0.8-0.9	39,41,48,54
>40			0.05	0.97	1.7	1.0	39
Respiratory rate >20/min	Hemo-pneumothorax, after trauma- blunt	Chest x-ray	0.43	1.00	∞	0.6	49
	- Penetrating		0.32	0.99	32.0	0.7	
Respiratory rate >20/min	Pulmonary emboli	Pulmonary angio	0.85	-	-	-	37
Pulsus paradoxus	Bronchial asthma	Spirometry	0.31	1.00	∞	0.7	36d
Asymetric expansion of the chest	Pneumonia	Chest x-ray	0.04	1.00	∞	1.0	38
Asymetric expansion of the chest	Pleural effusion	Chest x-ray	0.74	0.91	8.2	0.3	61
Vocal fremitus (increased)	Pneumonia	Chest x-ray	0.08	0.99	8.0	0.9	38
Vocal fremitus (reduced)	Pleural effusion	Chest x-ray	0.76	0.88	6.3	0.3	61
Dullness on percussion	Pneumonia	Chest x-ray	0.04	0.99	4.0	1.0	38
			0.14	0.96	3.5	0.9	43
			0.26	0.93	3.7	0.8	41
			0.13	0.94	2.2	0.9	39
Dullness on percussion	Any chest x-ray lesion	Chest x-ray	0.15	0.97	5.0	0.9	40
			0.08-0.09	0.98-0.99	4.0-9.0	0.9-0.9	56c
Dullness on percussion	Pleural effusion	Chest x-ray	0.89	0.81	4.7	0.1	61
Diaphragm expansion	OAD ^b	Spirometry	0.37	0.74	1.4	0.9	59
Cardiac dullness (absent)	OAD ^b	Spirometry	0.26	0.96	6.5	0.8	59
Auscultatory percussion (reduced)	Any chest x-ray lesion	Chest x-ray	0.10-0.15	0.88-0.88	0.8-1.3	1.0-1.0	56c
			0.19	0.85	1.3	1.0	40
Auscultatory percussion (reduced)	Pneumonia	Chest x-ray	0.00-0.50	0.88-0.97	2.0-12.5	0.5-1.0	60c
Auscultatory percussion (reduced)	Pleural effusion	Chest x-ray	0.96	1.0	∞	0.04	44
			0.58	0.85	3.9	0.5	61
Auscultatory percussion (hyperesonant)	OAD ^b	Spirometry	0.42	0.86	3.0	0.7	59

^aLikelihood ratio positive = sensitivity/(1-specificity); likelihood ratio negative = (1-sensitivity)/specificity

^bOAD = obstructive airway disease

^cRange of data obtained by more than one examiner

^dDerived from data on a total of 25 patients presented in figure form

Table 4. Sensitivity, Specificity and Likelihood Ratios of Respiratory PE Signs Elicited by Auscultation

Sign	Diagnosis	Gold standard	Sensitivity	Specificity	LR positive ^a	LR negative ^a	Reference
Forced expiratory time >6 s	OAD ^b	Spirometry	0.92	0.43	1.6	0.2	42
Bronchial breathing	Pneumonia	Chest x-ray	0.11–0.63	0.64–0.94	1.1–3.3	0.5–1.0	60c
			0.14	0.96	3.5	0.9	41
		Chest CT	0.08	1.0	∞	0.9	50
Bronchophony	Pneumonia	Chest x-ray	0.11–0.69	0.61–0.96	1.2–12.5	0.5–1.0	60c
			0.04	0.99	4.0	1.0	54
Egophony	Pneumonia	Chest x-ray	0.13–0.54	0.61–0.98	1.0–12.5	0.8–1.0	60c
			0.16	0.97	5.3	0.9	41
			0.08	0.96	2.0	1.0	39
			0.04	1.0	∞	1.0	38
Breath sounds (diminished)	Pleural effusion	Chest CT	0.42	0.90	4.2	0.6	50
Breath sounds (diminished)	OAD ^b	Chest x-ray	0.88	0.83	5.2	0.1	61
		Spirometry	0.57	0.88	4.8	0.5	59
			0.29 ^d	0.85	1.9	0.8	58
			-	-	-	-	55
Breath sounds (diminished)	Pneumonia	Chest x-ray	0.15	0.95	3.0	0.9	43
			0.49	0.80	2.5	0.6	41
			0.34	0.85	2.3	0.8	39
Breath sounds (diminished)	Hemo-pneumothorax, after trauma-blunt - penetrating	Chest x-ray	1.00	1.00	∞	0.0	49
Rales/crackles/crepitations	Pneumonia	Chest x-ray	0.50	1.00	∞	0.5	60c
			0.54–1.0	0.39–0.79	1.2–4.8	0.5–0.8	54
			0.59	0.70	2.0	0.6	41
			0.50	0.81	2.6	0.6	39
			0.36	0.77	1.6	0.8	38
			0.19	0.93	2.7	0.9	48
			0.67	0.44	1.2	0.8	35
			0.26	-	-	-	58
Rales/crackles/crepitations	OAD ^b	Spirometry	0.01	0.99	1.0	1.0	46
Rales/crackles/crepitations	Ventilator-associated pneumonia	Histological evidence	0.73	0.43	1.3	0.6	37
Rales/crackles/crepitations	Pulmonary emboli	Pulmonary angio	0.56	-	-	-	41
Rhonchi	Pneumonia	Chest x-ray	0.26	0.82	1.4	0.9	39
			0.34	0.78	1.5	0.8	38
			0.15	0.90	1.5	0.9	61
Reduced vocal resonance	Pleural effusion	Chest x-ray	0.76	0.88	6.3	0.3	39
Wheezes	Pneumonia	Chest x-ray	0.22	0.75	0.9	1.0	48
			0.22	0.71	0.8	1.1	47
Wheezes	OAD ^b	Spirometry	0.26	0.74	1.0	1.0	59
			0.06	0.99	6.0	0.9	58
			0.14	0.99	14.0	0.9	57
Wheezes at maximal forced exhalation	Asthma	Metacholine challenge	0.57	0.37	0.9	1.2	61
Pleural friction rub	Pleural effusion	Chest x-ray	0.05	0.99	5.0	0.7	39
Pleural friction rub	Pneumonia	Chest x-ray	0.17	0.96	4.3	0.9	43
			0.10	0.98	5.0	0.9	38
			0.04	0.99	4.0	1.0	37
Accentuated P-2 heart sound	Pulmonary emboli	Pulmonary angio	0.57	-	-	-	

^aLikelihood ratio positive = sensitivity/(1-specificity); likelihood ratio negative = (1-sensitivity)/specificity

^bOAD = Obstructive airway disease

^cRange of data obtained by more than one examiners

^d"Significant correlation between breath sound intensity scores and spirometry"

improvement, or even *deterioration*, of respiratory^{6,32}, breast¹⁶ and cardiac⁶⁶ PE skills with seniority and experience. The possibility that examiners differed in their PE skills is also suggested by the reported variability in the approach to the respiratory PE of 403 members of the British Thoracic Society⁷. A lack of adherence to appropriate PE technique also explains the low reliability found in studies of tachypnea^{25,33}; examiners appear to rely on their subjective impression rather than on counting⁶⁷. On the other hand, an adherence to

technique may explain the high reliabilities of respiratory PE signs that have been reported by some authors (e.g.,⁶¹).

The reported low to moderate sensitivity of respiratory PE signs may have been similarly due to deficient examination skills: an examiner with poor skills is likely to miss a PE finding. Alternatively, the reported sensitivities may have been confounded by differences in the severity of the diseases of the examined patients. Indeed, it has been reported that the sensitivity of reduced breath sounds⁵⁵, pulsus paradoxus³⁶

Table 5. Clinical Contexts and Respiratory Physical Examination Signs with Probable Diagnostic Value

Setting and context	Physical examination sign	LR positive ^a	LR negative ^a	Diagnosis	Reference
^c OPD: Patients with acute cough	Asymetric expansion of the chest	∞	1.0	Pneumonia	38
	Vocal fremitus (increased)	8.0	0.9		
	Dullness on percussion	4.0	1.0		
	Egophony	∞	1.0		
	Pleural friction rub	4.0	1.0		
^b ER: Fever and acute respiratory symptoms	Cyanosis	5.0	0.9	Pneumonia	54
	Heart rate >120/min	4.5	0.8		53
	Dullness on percussion	2.2–3.7	0.8–0.9		39,41,43
	Bronchial breathing	1.1–3.5	0.5–1.0		41,60
	Bronchophony	1.2–12.5	0.5–1.0		54,60
	Egophony	1.0–12.5	0.8–1.0		41,60
	Pleural friction rub	4.3,5.0	0.9		39,43
^d ICU: Acute respiratory distress syndrome	Bronchial breathing	∞	0.9	Pneumonia	50
^c OPD: Patients with acute cough	Asymetric expansion of the chest	8.2	0.3	Pleural effusion	38
Inpatients with respiratory symptoms	Vocal fremitus (diminished)	6.3	0.3	Pleural effusion	61
	Dullness on percussion	4.7	0.1		
	Breath sounds (diminished)	5.2	0.1		
	Pleural friction rub	5.0	0.7		
	Vocal resonance (diminished)	6.3	0.3		
^d ICU: Acute respiratory distress syndrome	Breath sounds (diminished)	4.2	0.6	Pleural effusion	50
^c OPD: screening of plumbers	Forced expiratory time >6 s	1.6	0.2	OAD ^e	42
^c OPD: self-reported history of smoking, asthma, chronic bronchitis	Wheezes	6.0	0.9	OAD ^e	59
	Cardiac dullness (absent)	6.5	0.8		
^c OPD for preoperative evaluation	Wheezes	14.0	0.9	OAD ^e	58
^b ER: Chest trauma – blunt	Respiratory rate >20/min	∞	0.6	Hemo-pneumothorax	49
	Breath sounds (diminished)	∞	0.0		
^b ER: Chest trauma – penetrating	Respiratory rate >20/min	32.0	0.7	Hemo-pneumothorax	49
	Breath sounds (diminished)	∞	0.5		

^aLikelihood ratio positive = sensitivity/(1-specificity); likelihood ratio negative = (1-sensitivity)/specificity

^bER = Emergency room

^cOPD = outpatient department

^dICU = Intensive care unit

^eOAD= Obstructive airway disease

and Hoover's sign (an inward motion of the lower lateral rib cage with inspiration)⁵¹ increases with the severity of airway obstruction.

These two main flaws in the design of most studies, namely their failure to control for disease severity and examiners' skills, have probably led to underestimates of the reported reliability, sensitivity and specificity values. Future studies may avoid these biases by adhering to the STARD requirements. However, pending the publication of properly controlled studies, the reliability and validity of most respiratory PE signs remain uncertain. In 1986, Mulrow et al.³² concluded that "despite their routine use, most physical examination techniques, including pulmonary auscultation and percussion, are poorly standardized and of uncertain [diagnostic] value." This conclusion is also pertinent today.

The uncertain sensitivity of the respiratory PE argues against its utility for screening of *asymptomatic persons*. Screening for disease requires that the test used be highly sensitive, well above 0.7, which is not the case for most respiratory PE signs (Tables 3, 4). However, their low or uncertain reliability and validity do not preclude their usefulness in *patients with suspected respiratory diseases* for two reasons. First, the observed reliabilities and sensitivities may have been confounded by flaws in the study design. Second, the possible bias produced by these flaws would be toward *underestimating* the diagnostic

accuracy of the various respiratory PE signs. Therefore, the high specificities and high LR_s⁺ reported for some of these signs appear to be credible and to indicate that they may be useful in specific clinical contexts. For example, assuming that the pretest probability of pneumonia in outpatients with acute cough is 10% (odds 1:9)³⁸, a finding of asymetric expansion of the chest would increase the odds to 8:9, i.e., increase the post-test probability to 47%. Assuming a pretest probability of pneumonia of 12–30% (odds 1:9–3:7) among emergency room patients with fever and acute respiratory symptoms⁴¹, a finding of pleural friction rub would increase the odds of pneumonia to 5:9–15:7 or to a post-test probability of 36–68%. Future studies should explore whether the various respiratory PE signs provide independent information. In other words, it is at present uncertain whether a *combination* of PE signs (e.g., of pulmonary consolidation such as dullness on percussion, bronchial breathing and egophony) allows the multiplication of each of their LR_s in order to assess the post-test probability of pneumonia.

Therefore, we believe that a meticulously performed respiratory PE, which aims to *explore a diagnostic hypothesis*, as opposed to a PE that aims to detect a disease in an asymptomatic person, remains a cornerstone of clinical practice. We propose that teaching of the PE should not discriminate among respiratory signs according to their presently uncertain reliabilities and sensitivities, but rather according to the importance of

the disease under consideration and to their specificity.

The most important PE signs are, first, those of life-threatening conditions *in any clinical context*. For example, a patient, who presents with any degree of *respiratory abnormality* (tachypnea, bradypnea, apnea, labored breathing, stridor, accessory muscle recruitment or paradoxical breathing) is in respiratory distress. Its detection mandates immediate treatment with oxygen and a sustained effort to establish the cause by looking for *stridor* (croup, epiglottitis), *wheezes* (bronchial asthma, bronchitis), *reduced breath sounds* and *changes in percussion note* (pneumothorax or pleural effusion) and for signs suggesting pulmonary emboli. Second, we believe that teaching should emphasize the respiratory signs that have been reported to have high positive or low negative LR (Table 5).

At the other end of the spectrum, the least important respiratory PE signs are those that are no longer employed in clinical practice because of the availability of more easily performed ancillary tests. For example, hand-held spirometry provides an easier and more precise assessment of obstructive airway disease than Hoover's sign and pulsus paradoxus. Spirometry may also alert physicians to the possibility of mild pulmonary disorders, and it may be used for monitoring patients with conditions such as asthma and cystic fibrosis. Similarly, pulse oxymetry may detect reduced blood oxygenation at earlier stages than central cyanosis⁶⁸. Therefore, we join the calls to incorporate pulse oximetry and spirometry into the PE, and add hand-held oximeters and spirometers to the stethoscope, sphygmomanometer and reflex hammer that a doctor already uses during patient examination⁶⁹.

Conflict of Interest: None disclosed.

Corresponding Author: Jochanan Benbassat, MD; Myers-JDC-Brookdale Institute, PO Box 3886, Jerusalem 91037, Israel (e-mail: benbasat@jdc.org.il).

REFERENCES

- Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *Br Med J*. 1975;2:486-9.
- Reilly BM. Physical examination in the care of medical inpatients: an observational study. *Lancet*. 2003;362:1100-05.
- Todd IK. A thorough pulmonary exam and other myths. *Acad Med*. 2000;75:50-1.
- Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: getting diagnostic testing into perspective. *QJM*. 1998;91:247-58.
- Feddock CA. The lost art of clinical skills. *Am J Med*. 2007;120:374-378.
- Bradding P, Cookson JB. The dos and don'ts of examining the respiratory system: a survey of British Thoracic Society members. *J R Soc Med*. 1999;92:632-4.
- Mangione S, Nieman LZ. Pulmonary auscultatory skills during training in internal medicine and family practice. *Am J Respir Crit Care Med*. 1999;159:1119-24.
- Benbassat J, Bauml R, Heyman SN, Brezis M. Viewpoint: suggestions for a shift in teaching clinical skills to medical students: the reflective clinical examination. *Acad Med*. 2005;80:1121-6.
- Morgan WL, Engel GL, eds. *The Clinical Approach to the Patient*. Philadelphia: WB Saunders Co; 1969: 197-204.
- Hargie O, Dickson D, Boohan M, Hughes K. A survey of interviewing skills training in UK schools of medicine: present practices and prospective proposals. *Med Educ*. 1998;32:25-34.
- DeGowin EL. *Bedside diagnostic examination*. 1st ed. New York: Macmillan Co; 1965.
- LeBlond RF, Brown DD, DeGowin RL. *DeGowin's diagnostic examination*. 9th ed. New York: McGraw-Hill; 2009.
- Kampmeier RH. *Physical examination in health and disease*. 2nd ed. Philadelphia: FA Davis Co; 1957.
- Joshua AM, Celermajor DS, Stockler MR. Beauty is in the eye of the examiner: reaching agreement about physical signs and their value. *Intern Med J*. 2005;35:178-87.
- Cook CJ, Smith GB. Do textbooks of clinical examination contain information regarding the assessment of critically ill patients? *Resuscitation*. 2004;60:129-36.
- Lee KC, Dunlop D, Dolan NC. Do clinical breast examination skills improve during medical school? *Acad Med*. 1998;73:1013-9.
- Sackett DL, Rennie D. The science of the art of the clinical examination. *JAMA*. 1992;267:2650-2.
- Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. *JAMA*. 1997;278:1440-5.
- McGee S. *Evidence based physical diagnosis*. 2nd ed. Philadelphia: WB Saunders Co; 2007.
- Auld AG. The roentgen rays in the diagnosis of pulmonary disease. *Lancet*. 1903;162:341-2.
- Horowitz GL, Bleich HL. PaperChase: a computer program to search the medical literature. *New Engl J Med*. 1981;305:924-930.
- Fletcher CM. The clinical diagnosis of pulmonary emphysema. *Proc R Soc Med*. 1952;45:577-84.
- Pyke DA. Finger clubbing. *Lancet*. 1954;2:352-354.
- Schilling RS, Hughes JP, Dingwall-Fordyce I. Disagreement between observers in an epidemiological study of respiratory disease. *Br Med J*. 1955;1:65-8.
- Smylie HC, Blendis IM, Armitage P. Observer disagreement in physical signs of the respiratory system. *Lancet*. 1965;2:412-3.
- Schneider IC, Anderson AE. Correlation of clinical signs with ventilatory function in obstructive lung disease. *Ann Intern Med*. 1965;62:477-85.
- Nairn JR, Turner-Warwick M. Breath sounds in Emphysema. *Brit J Dis Chest*. 1969;63:29-37.
- Godfrey S, Edwards RHT, Campbell EJM, Armitage P, Oppenheimer EA. Repeatability of physical signs in airway obstruction. *Thorax*. 1969;24:4-9.
- Bohadana AB, Peslin R, Uffholtz H. Breath sounds in the clinical assessment of airflow obstruction. *Thorax*. 1978;33:345-51.
- Williams TJ, Ahmad D, Morgan WK. A clinical and roentgenographic correlation of diaphragmatic movement. *Arch Intern Med*. 1981;141:878-80.
- Gjorup T, Bugge PM, Jensen AM. Interobserver variation in assessment of respiratory signs. Physicians' guesses as to interobserver variation. *Acta Med Scand*. 1984;216:61-6.
- Mulrow CD, Dolmatch BL, Delong ER, Feussner JR, Benyunes MC, Dietz JL, Lucas SK, Pisano ED, Svetkey LP, Volpp BD, et al. Observer variability in the pulmonary examination. *J Gen Intern Med*. 1986;1:364-7.
- Spiteri MA, Cook DG, Clarke SW. Reliability of eliciting physical signs in examination of the chest. *Lancet*. 1988;1:873-5.
- Purohit A, Bohadana A, Kopferschmitt-Kubler MC, Mahr L, Linder J, Pauli G. Lung auscultation in airway challenge testing. *Respir Med*. 1997;91:151-7.
- Osmer JC, Cole BK. The stethoscope and roentgenogram in acute pneumonia. *South Med J*. 1966;59:75-7.
- Bilgi C, Jones RL, Sproule BJ. Relation between pulsus paradoxus and pulmonary function in patients with chronic airways obstruction. *Can Med Assoc J*. 1977;117:1389-92.
- Stein PD, Willis PW 3rd, DeMets DL. History and physical examination in acute pulmonary embolism in patients without preexisting cardiac or pulmonary disease. *Am J Cardiol*. 1981;47:218-23.
- Diehr P, Wood RW, Bushyhead J, Krueger L, Wolcott B, Tompkins RK. Prediction of pneumonia in outpatients with acute cough—a statistical approach. *J Chronic Dis*. 1984;37:215-25.
- Gennis P, Gallagher J, Falvo C, Baker S, Than W. Clinical criteria for the detection of pneumonia in adults: guidelines for ordering chest roentgenograms in the emergency department. *J Emerg Med*. 1989;7:263-8.
- Bourke S, Nunes D, Stafford F, Hurley G, Graham I. Percussion of the chest re-visited: a comparison of the diagnostic value of auscultatory and conventional chest percussion. *Ir J Med Sci*. 1989;158:82-4.
- Heckerling PS, Tape TG, Wigton RS, Hissong KK, Leikin JB, Ornato JP, Cameron JL, Racht EM. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med*. 1990;113:664-70.
- Kern DG, Patel SR. Auscultated forced expiratory time as a clinical and epidemiologic test of airway obstruction. *Chest*. 1991;100:636-9.

43. **Melbye H, Straume B, Aasebo U, Dale K.** Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. *Scand J Prim Health Care.* 1992;10:226-33.
44. **Guarino JR, Guarino JC.** Auscultatory percussion: a simple method to detect pleural effusion. *J Gen Intern Med.* 1994;9:71-4.
45. **Baughman RP, Gunther KL, Buchsbaum JA, Lower EE.** Prevalence of digital clubbing in bronchogenic carcinoma by a new digital index. *Clin Exp Rheumatol.* 1998;16:21-6.
46. **Petersen IS, Aru A, Skodt V, Behrendt N, Bols B, Kiss K, Simonsen K.** Evaluation of pneumonia diagnosis in intensive care patients. *Scand J Infect Dis.* 1999;31:299-303.
47. **Straus SE, McAlister FA, Sackett DL, Deeks JJ.** The accuracy of patient history, wheezing, and laryngeal measurements in diagnosing obstructive airway disease. CARE-COAD1 Group. Clinical Assessment of the Reliability of the Examination-Chronic Obstructive Airways Disease. *JAMA.* 2000;283:1853-7.
48. **Mehr DR, Binder EF, Kruse RL, Zweig SC, Madsen RW, D'Agostino RB.** Clinical findings associated with radiographic pneumonia in nursing home residents. *J Fam Pract.* 2001;50:931-7.
49. **Bokhari F, Brakenridge S, Nagy K, Roberts R, Smith R, Joseph K, An G, Wiley D, Barrett J.** Prospective evaluation of the sensitivity of physical examination in chest trauma. *J Trauma.* 2002;53:1135-8.
50. **Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ.** Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. *Anesthesiology.* 2004;100:9-15.
51. **Garcia-Pachon E, Padilla-Navas I.** Frequency of Hoover's sign in stable patients with chronic obstructive pulmonary disease. *Int J Clin Pract.* 2006;60:514-7.
52. **O'Brien WT, Rohweder DA, Lattin GE, Thornton JA, Dutton JP, Ebert-Long DL, Duncan MD.** Clinical indicators of radiographic findings in patients with suspected community-acquired pneumonia: who needs a chest x-ray? *J Am Coll Radiol.* 2006;3:703-6.
53. **Nolt BR, Gonzales R, Maselli J, Aagaard E, Camargo CA, Metlay JP.** Vital-sign abnormalities as predictors of pneumonia in adults with acute cough illness. *Am J Emerg Med.* 2007;25:631-6.
54. **Saldias PF, Cabrera TD, de Solminihaç LI, Hernandez AP, Gederlini GA, Diaz FA.** Predictive value of history and physical examination for the diagnosis of community-acquired pneumonia in adults. *Rev Med Chil.* 2007;135:143-150.
55. **Pardee EN, Martin CJ, Morgan EH.** A test of the practical value of estimating breath sound intensity. *Chest.* 1976;70:341-344.
56. **Bohadana AB, Coimbra FT, Santiago JR.** Detection of lung abnormalities by auscultatory percussion: a comparative study with conventional percussion. *Respiration.* 1986;50:218-25.
57. **King DK.** Wheezing on maximal forced exhalation in the diagnosis of atypical asthma. *Ann Intern Med.* 1989;110:451-55.
58. **Holleman DR, Simel DL, Goldberg JS.** Diagnosis of obstructive airways disease from the clinical examination. *J Gen Intern Med.* 1993;8:63-8.
59. **Badgett RG, Tanaka DJ, Hunt DK, Jelley MJ, Feinberg LE, Steiner JF, Petty TL.** The clinical evaluation for diagnosing obstructive airways disease in high-risk patients. *Chest.* 1994;106:1427-31.
60. **Wipf JE, Lipsky BA, Hirschmann JV, Boyko EJ, Takasugi J, Peugeot RL, Davis CL.** Diagnosing pneumonia by physical examination: relevant or relic? *Arch Intern Med.* 1999;159:1082-7.
61. **Kalantri S, Joshi R, Lokhande T, Singh A, Morgan M, Colford JM, Pai M.** Accuracy and reliability of physical signs in the diagnosis of pleural effusion. *Respir Med.* 2007;101:431-8.
62. **Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glaziou PP, Irwig LM, Moher D, Rennie D, De Vet HCV, Lijmer JG.** The STARD statement for reporting studies of diagnostic accuracy: explanations and elaboration. *Ann Intern Med.* 2003;138:w1-w12.
63. **Simmel DL, Rennie D, Bossuyt PM.** The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination. *J Gen Intern Med.* 2008;23:768-74.
64. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
65. **King DB, Dickinson JA, Boulton MR, Toumpas C.** Clinical skills textbooks fail evidence-based examination. *ACP J Club.* 2006;144:A10-1.
66. **Vukanovic-Criley JM, Criley S, Warde CM, Boker JR, Guevara-Matheus L, Churchill WH, Nelson WP, Criley JM.** Competency in cardiac examination skills in medical students, trainees, physicians and faculty. *Arch Intern Med.* 2006;166:610-616.
67. **Mukkamala SG, Gennings C, Wenzel RP.** R = 20: bias in the reporting of respiratory rates. *Am J Emerg Med.* 2008;26:237-9.
68. **Weber MW, Usen S, Palmer A, Jaffar S, Mulholland EK.** Predictors of hypoxaemia in hospital admissions with acute lower respiratory tract infection in a developing country. *Arch Dis Child.* 1997;76:310-4.
69. **Mower W, Sachs C, Nicklin E, et al.** Pulse oximetry as a fifth vital sign. *Pediatrics.* 1997;99:681-6.