Journal of Clinical Epidemiology 129 (2021) 86–96

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Diagnostic accuracy of FeNO [fractional exhaled nitric oxide] and asthma symptoms increased when evaluated with a superior reference standard

Christina Kellerer[a], Stefan Wagenpfeil[b], Luke Daines[c], Rudolf A. Jörres[d], Alexander Hapfelmeier[a,e], Antonius Schneider[a,*]

[a]*Institute of General Practice and Health Services Research, School of Medicine, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany*
[b]*Institute for Medical Biometry, Epidemiology and Medical Informatics (IMBEI), Saarland University, Homburg, Germany*
[c]*Asthma UK Centre for Applied Research, Usher Institute, The University of Edinburgh, Edinburgh, UK*
[d]*Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität München, Munich, Germany*
[e]*Institute for Medical Statistics and Epidemiology, School of Medicine, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany*

Accepted 15 September 2020; Published online 8 October 2020

Abstract

Objectives: The objective of the study is to determine the impact of changing reference standards (RS), namely spirometry vs. whole-body plethysmography (WBP), on estimation of the diagnostic accuracy of fractional exhaled nitric oxide (FeNO) and clinical signs and symptoms (CSS) as index tests regarding asthma diagnosis.

Study Design and Setting: This was a diagnostic study conducted in 393 patients attending a private practice of pneumologists with complaints suspicious of asthma. First, the index tests were compared with the diagnostic results of spirometry in terms of forced expiratory volume in the first second (FEV_1) responsiveness. Second, the index tests were compared with the results of WBP in terms of specific airway resistance and FEV_1 responsiveness. Areas under the curve (AUC) were compared with a generalized estimating equation approach based on binary logistic regression.

Results: FeNO values and CSS 'wheezing' and 'allergic rhinitis' showed higher specificities ($P < 0.001$) and sensitivities (not significant) when evaluated with WBP; also, Youden indices increased in these CSS ($P < 0.05$). AUC of FeNO in combination with 'wheezing' and 'allergic rhinitis' when WBP was used as RS (AUC = 0.724; 95% confidence interval 0.672 to 0.776) was higher compared with spirometry as RS (AUC = 0.654; 95% confidence interval 0.585 to 0.722) ($P < 0.001$).

Conclusion: In case of asthma, superior RS led to more favorable assessment of index tests. FeNO measurement might have been underestimated in some previous studies.

Keywords: Asthma; Diagnostic study; Sensitivity; Specificity; Area under the curve; Fractional exhaled nitric oxide

1. Introduction

The principle of a diagnostic study is to evaluate the diagnostic accuracy of an index test against a reference standard [1]. Often, the index test is a new diagnostic device. New devices could allow the replacement of some existing tests, may be used for triage or as an add-on test [2]. The diagnostic accuracy of clinical signs and symptoms (CSSs) could also be evaluated within a diagnostic study, which might help to determine the diagnostic accuracy of clinical patterns for distinct diseases [3,4]. It is often assumed that sensitivities and specificities are inherent diagnostic test properties. However, statistical modeling to investigate the effects of an imperfect reference standard on estimation of index tests found that the difference of the specificities of an index test increases with increasing disease prevalence, whereas the differences of sensitivities were higher in the area of low disease prevalence [5]. Beyond that, statistical modeling studies found that inaccurate reference standards will lead to underestimation of index test accuracy [5–7]. These effects were mostly evaluated with a comparison of diagnostic studies which

Conflicts of interest statement: All authors declare that there are no conflicts of interest.

* Corresponding author. TUM School of Medicine, Institute of General Practice and Health Services Research, Technical University of Munich, Orleansstraße 47, 81667 Munich, Germany. Tel.: +49 89 6146589-11; fax: +49 89 6146589-15.

E-mail address: antonius.schneider@tum.de (A. Schneider).

https://doi.org/10.1016/j.jclinepi.2020.09.021

What is new?

Key findings

- The choice of the reference standard has an influence on the assessment of index tests and on the pretest probability of a disease.
- In the case of asthma, superior reference standard favored the assessment of fractional exhaled nitric oxide (FeNO) and changed the predictive values of index tests considerably.

What this adds to what is known?

- Imperfect reference standards will distort the specificities and pretest probabilities of index tests.
- The diagnostic accuracy of FeNO measurement might have been underestimated in some previous studies.

What is the implication and what should change now?

- Index tests should be evaluated against an optimal reference standard.

were performed in different settings [5,8], but were rarely investigated within a coherent diagnostic study.

A perfect reference standard rarely exists [1,9], and disagreements about the ideal reference standard are particularly apparent for diagnostic decision-making in asthma. Guidelines suggest establishing the diagnosis based on medical history and verification of reversible airway obstruction [10]. Spirometry is considered to be a reference standard for diagnosing airway obstruction [11], and its accuracy in diagnosing severe asthma has been demonstrated [12]. Airway obstruction is often not persistent in mild asthma, thus leading to diagnostic uncertainty [13]. In the case of inconclusive spirometric results, bronchial provocation (BP) deserves as a reference standard for determining bronchial hyper-responsiveness (BHR) [14]. Thus, asthma could be defined on the basis of a positive bronchodilation (BD) test or positive results during BP [10]. In Germany, the results of BP are interpreted using whole-body plethysmography (WBP) as a reference standard, also in ambulatory care [15,16]. Patient investigation with WBP allows the determination of spirometric indices like forced expiratory volume in the first second (FEV_1) and vital capacity (VC) and specific airway resistance (sRaw) within a single diagnostic procedure without additional burden for a patient. However, the added value of WBP over spirometry for ruling in and ruling out asthma in the real world setting has been questioned [17] and is therefore not used regularly in other countries like the United Kingdom, where spirometry is used as a reference standard for interpreting BP results. On the other hand, there is increasing evidence supporting the diagnostic added value of WBP [15,16,18]. Our earlier work has shown that the sensitivity for the detection of BHR increased from 44.6% (when solely the spirometric parameter FEV_1 is used) to 95.2%, when sRaw (which can only be determined with WBP) is included in interpretation of BP, accompanied by a slight decline of specificity from 91.3% (FEV_1) to 81.7% (sRaw) [16].

However, BP is time consuming, costly, not widely available, and carries a small risk of inducing severe bronchospasm [19]. Therefore, new technologies like the measurement of fractional exhaled nitric oxide (FeNO), a noninvasive, easily available marker, are investigated with the aim of replacing BP; and increased FeNO level has been consistently demonstrated in asthma including milder forms of the disease [20,21]. FeNO has been evaluated against different reference standards, indicating a promising diagnostic value [22]. However, it remains unclear if the statistical assumptions prove true when index tests like FeNO measurement or CSSs are evaluated against different reference standards, namely in situations where WBP might be superior to spirometry for interpreting BP, thus leading to different definitions of asthma. Therefore, we sought to determine the diagnostic accuracy of individual patient-reported symptoms and FeNO in making an asthma diagnosis when compared with WBP or spirometry.

2. Methods

2.1. Study design and sample

We performed a secondary analysis of data from a diagnostic accuracy study conducted in a large private practice led by five pneumologists in Augsburg, Germany, between June 2010 and October 2011 [23]. A total of 400 patients attending the practice for the first time with a clinical history suggestive of asthma and giving written consent were consecutively included. Inclusion criteria were the presence of symptoms including dyspnea, cough, or phlegm for more than 2 months, leading to a clinical suspicion of obstructive airway disease (‘‘indicated population’’). Patients were advised not to smoke on the day of assessment. If patients were already using inhaler medication (prescribed by a general practitioner before referral), they were advised not to use it for 12 hours before the assessment. Patients were excluded if there was any contraindication to BP testing (pregnancy, heart disease) or had experienced a chest infection within the preceding 6 weeks. Seven patients were excluded from the analysis as they did not complete all necessary diagnostic tests [23]. The study was approved by the Ethical Committee of the Technical University of Munich.

2.2. Index tests: CSS and FeNO measurement

CSSs as index tests were drawn from the anamnestic data derived from the questionnaires (Table 1). Each patient

Table 1. Characteristics of patients

	Diagnosis of asthma based on WBP (FEV_1 and sRAW)			Diagnosis of asthma based on spirometry (FEV_1)		
Diagnosis	**Asthma, *n* = 154**	**No asthma, *n* = 239**	***P*-value**	**Asthma, *n* = 83**	**No asthma, *n* = 310**	***P*-value**
Age (mean in years [sd])	40.5 [15.4]	45.1 [16.1]	0.009	39.2 [14.5]	44.4 [16.8]	0.017
Female *n* (%)	91 (59.0)	144 (60.3)	0.819	57 (68.7)	178 (57.4)	0.063
Symptoms	*N* (%)	*N* (%)		*N* (%)	*N* (%)	
1. Wheeze in the past 12 months? (Yes)	97 (63.0)	84 (35.1)	<0.001	48 (57.8)	133 (42.9)	0.020
1.1. Short of breath when wheezing? (Yes)	70 (45.0)	39 (16.3)	<0.001	38 (45.8)	71 (22.9)	<0.001
1.2. Wheeze even when no cold* (Yes)	57 (37.0)	43 (18.0)	<0.001	27 (32.5)	73 (23.5)	<0.001
2. Suffer from shortness of breath? (Yes, any)	98 (63.0)	128 (53.6)	0.071	57 (68.7)	169 (54.5)	0.042
3. Woken up with shortness of breath at night (Yes)	35 (22.7)	29 (12.1)	0.004	21 (25.3)	43 (13.9)	0.012
4. Woken up at night with chest tightness? (Yes)	54 (35.1)	60 (25.1)	0.032	30 (36.1)	84 (27.1)	0.107
5. Woken up at night with coughing? (Yes)	92 (59.7)	136 (56.9)	0.520	52 (62.7)	176 (56.8)	0.435
6. Suffer from frequent cough? (Yes)	65 (42.2)	115 (48.1)	0.309	33 (39.8)	147 (47.4)	0.170
7. Do you often suffer from expectoration? (Yes)	44 (28.6)	61 (25.5)	0.575	19 (22.9)	86 (27.7)	0.346
8. Allergic rhinitis (yes)	76 (49.4)	47 (19.7)	<0.001	39 (47.0)	84 (27.1)	<0.001
9. Do you smoke? (Yes)	19 (12.3)	20 (8.4)	0.198	11 (13.3)	28 (9.0)	0.274
10. Have you smoked in the past? (Yes)	56 (36.0)	83 (34.7)	0.425	32 (38.6)	107 (34.5)	0.296
FeNO						
>12 ppb	131 (85.1)	171 (71.5)	0.002	70 (84.3)	232 (74.8)	0.068
>16 ppb	107 (69.5)	126 (52.7)	0.001	55 (66.3)	178 (57.4)	0.145
>20 ppb	92 (59.7)	88 (36.8)	<0.001	48 (57.8)	132 (42.6)	0.013
>35 ppb	50 (32.5)	29 (12.1)	<0.001	27 (32.5)	52 (16.8)	0.001
>46 ppb	36 (23.4)	19 (7.9)	<0.001	19 (22.9)	36 (11.6)	0.009
>50 ppb	35 (22.7)	15 (6.3)	<0.001	19 (22.9)	31 (10.0)	0.002
>71 ppb	27 (17.5)	7 (2.9)	<0.001	17 (20.5)	17 (5.5)	<0.001

Abbreviations: BP, bronchial provocation; FeNO, fractional exhaled nitric oxide; ppb, parts per billion; WBP, whole-body plethysmography; sRAW, specific airway resistance; FEV1, forced expiratory volume in the first second.

underwent FeNO testing using a NIOX MINO® device (Aerocrine, Solna, Sweden) following a standard protocol at a flow rate of 50 ml/s [24], which was indicated by the machine display. Measurements were recorded on a continuous scale in parts per billion (ppb). The FeNO measurement was performed before WBP and BP, as the breathing maneuvers involved could distort FeNO results. The responsible pneumologist was blinded to the FeNO and questionnaire results and made the diagnostic decision based solely on medical history, physical examination, spirometry, WBP, and BP results.

2.3. Reference tests: methacholine responsiveness was determined via WBP and via spirometry to diagnose asthma

Lung function tests including spirometry were performed in accordance with standard protocols, and reference values were adjusted for sex, age, and height [25]. Patients with $FEV_1 < 80\%$ predicted underwent a BD test using salbutamol with an additional WBP investigation 20 min later. Obstructive airway disease was diagnosed in patients with a pathological Tiffeneau index ($FEV_1/VC \leq 0.70$). An asthma diagnosis was made if clinical symptoms and history

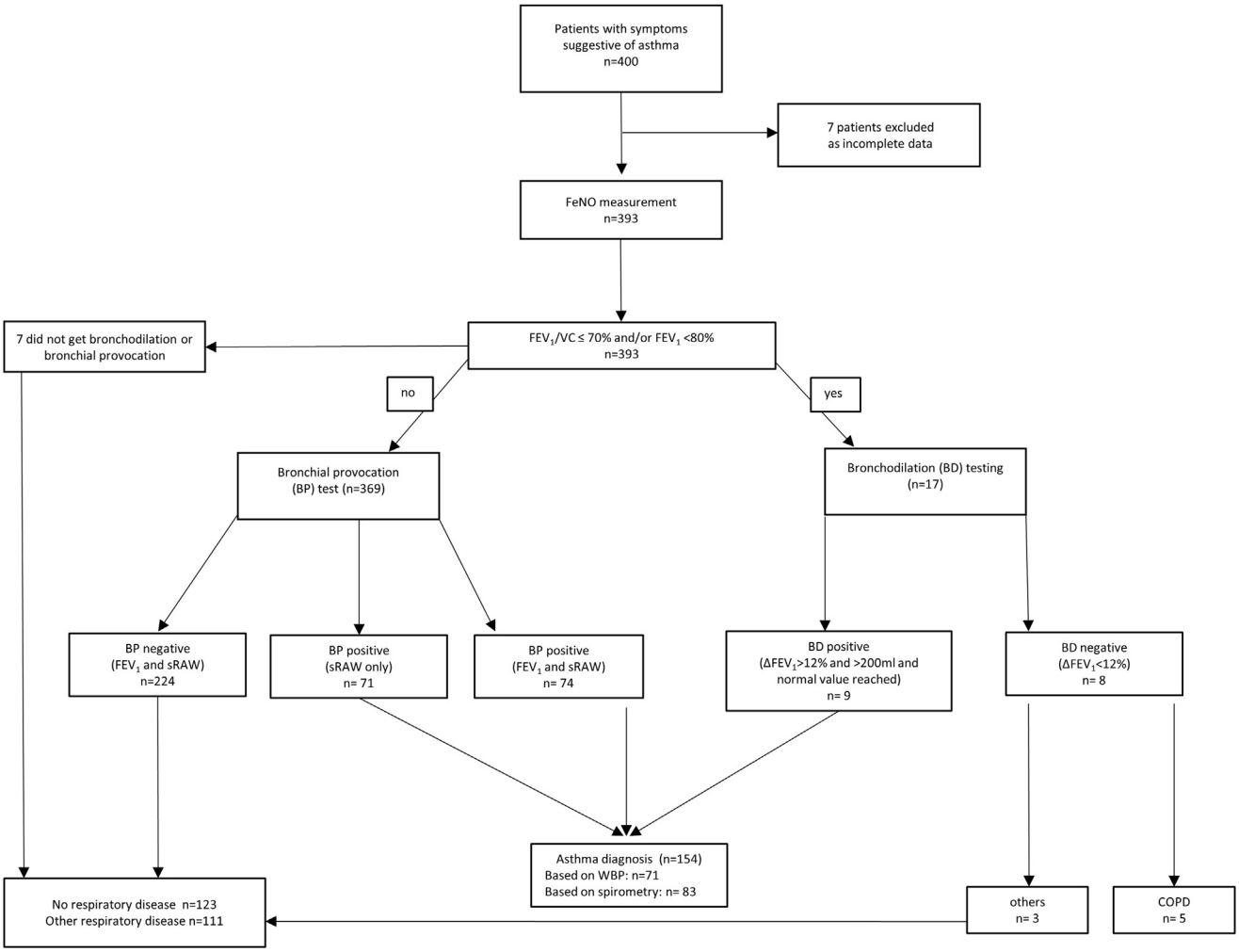

Fig. 1. Flowchart of diagnostic investigation. BP, bronchial provocation; BD, bronchodilation; FeNO, fractional exhaled nitric oxide; WBP, whole-body plethysmography; FEV1, forced expiratory volume in 1 s; sRAW, specific airway resistance; VC, vital capacity.

fitted, the change during BD test in FEV_1 was $\geq$12% compared with baseline and $\geq$200 ml, and lung function returned to the predicted normal range after salbutamol inhalation. An incomplete bronchodilator response was recorded if FEV_1 was $<$12% compared with baseline and $<$200 mL and lung volumes remained below predicted. A diagnosis of chronic obstructive pulmonary disease was given, if clinical symptoms and history fitted, and the FEV_1 bronchodilator response after salbutamol was $<$12% compared with baseline and $<$200 ml. If there was no bronchial obstruction, BP in accordance with the 1-concentration-4-step dosimeter protocol [26] was performed to determine BHR to methacholine. This yields similar results to the multiconcentration protocol of the American Thoracic Society [19] but offers advantages in clinical practice as it can be conducted more rapidly and simply. An "asthma" diagnosis required a 20% fall in FEV_1 from baseline after inhaling methacholine stepwise until the maximum concentration (16 mg/mL), alternatively a simultaneous increase in sRaw by at least 100% and to at least 2.0 kPa*s [15]. Spirometric and WBP test indices are determined at the same time within the diagnostic procedure.

2.4. Data analysis

Baseline data are presented descriptively. Hypothesis testing on differences between the patients with and without asthma, either diagnosed by spirometry (FEV_1) or WBP (sRaw) were assessed via the chi-square test or by the Mann-Whitney U-test. Each CSS was considered as an 'index test'. FeNO was measured as a scalar quantity and initially assessed as a continuous variable. To create several binary variables, FeNO measurements were dichotomized at six cutoff points in accordance with the literature (12 ppb [27], 16 ppb [27], 20 ppb [28], 35 ppb [29], 46 ppb [27], 50 ppb [30], 70 ppb [23]).

First, the index tests were compared with the diagnostic results of spirometry in terms of FEV_1 responsiveness (determined with BD or BP). Second, the index tests were compared with the results of WBP in terms of sRAW and

Table 2. Sensitivity and specificity of each clinical symptom based on WBP as a reference standard compared with the respective sensitivity and specificity based on spirometry as a reference standard

	Asthma assessed by WBP	Asthma assessed by spirometry	
Symptoms	**Sensitivity (95% CI)**	**Sensitivity (95% CI)**	***P*-value**
Wheezing in the past 12 months	63.8% (55.6%, 71.4%)	58.5% (47.1%, 69.3%)	0.171
Shortness of breath when wheezing	48.6% (40.2%, 57.1%)	47.5% (36.2%, 59.0%)	0.781
Wheezing even when no cold	40.1% (32.0%, 48.7%)	33.8% (23.6%, 45.2%)	0.105
Ever suffer from shortness of breath	66.2% (58.0%, 73.8%)	70.4% (59.2%, 80.0%)	0.293
Woken up with shortness of breath at night	24.0% (17.30%, 31.73%)	26.3% (17.0%, 37.3%)	0.491
Woken up at night with chest tightness	36.0% (28.3%, 44.2%)	37.0% (26.6%, 48.5%)	0.788
Woken up at night with coughing	60.9% (52.7%, 68.8%)	62.7% (51.3%, 73.0%)	0.649
Suffer from frequent cough	44.2% (36.0%, 52.6%)	40.7% (30.0%, 52.2%)	0.390
Expectoration	31.4% (23.9%, 39.8%)	25.3% (16.0%, 36.7%)	0.127
Allergic rhinitis	53.5% (45.0%, 61.9%)	50.0% (38.5%, 61.5%)	0.382
Smoker	12.5% (7.7%, 18.8%)	13.3% (6.8%, 22.5%)	0.753
Ex-smoker	40.6% (32.3%, 49.3%)	43.2% (31.8%, 55.3%)	0.515

Abbreviations: CI, confidence interval; WBP, whole-body plethysmography.

FEV_1 responsiveness (determined with BD or BP). Two-by-two contingency tables related to spirometric asthma diagnosis vs. body plethysmographic asthma diagnosis were prepared, which allowed the calculation of sensitivities, specificities, positive predictive values (PPVs), and negative predictive values (NPVs) for each table. Ninety-five percent confidence intervals (95% CIs) were calculated using Wilson's method [31].

Receiver operating characteristic (ROC) curves of FeNO for the diagnosis of asthma assessed by two different reference standards, WBP and spirometry, were constructed and quantified by the area under the curve (AUC) and corresponding 95% CI. In addition, ROC analysis was performed with a multiple logistic regression model used to combine FeNO, wheezing, and allergic rhinitis to obtain probabilities of asthma, again using the alternative reference standards WBP and spirometry. These clinical symptoms turned out to be significant predictors of asthma in a previously performed multiple logistic regression analysis [32].

For statistical hypothesis testing, a dependence structure has to be taken into account because each of the study participants has a WBP as well as spirometric examination. In this way, comparisons of positive and negative predictive values of each of clinical symptoms between WBP and spirometry as reference standard are due to McNemar's test. For comparisons of sensitivities as well as specificities between WBP and spirometry, a generalized estimating equation (GEE) approach was used based on binary logistic regression with respective clinical symptom as dependent and diagnostic measurement (either WBP or spirometry) as factor variable. Comparison of AUCs is due to a GEE approach based on linear regression analyses with either quantitative FeNO measurements or asthma probabilities as dependent and way of diagnostic measurement as well as diagnosis as factor variables. Hypothesis testing on differences in the Youden index was performed by the nonparametric bootstrap using 5000 bootstrap replicates [33]. All analyses were performed using the software package SPSS (Version 25, IBM, Armonk, NY, USA) and R 3.6.1. (The R Foundation for Statistical Computing, Vienna, Austria), and the two-sided level of statistical significance was prespecified at $\alpha = 0.05$.

3. Results

3.1. Study population

A total of 393 patients were included in the analysis (Fig. 1), of whom 235 (59.8%) were women, with a mean age of 43.3 (standard deviation 16.4) years. A total of 154 patients received an asthma diagnosis which corresponded to a prevalence of 39.2%. Nine (2.3%) patients showed positive results during BD testing. A total of 145 (36.9%) patients showed positive BP test results. Of these, 71 (18.1%) had a pathological sRAW reaction, but no pathological FEV_1 reaction, so could only be diagnosed with WBP. Seventy-four (18.8%) had a pathological FEV_1 reaction, which could be diagnosed based on spirometry (these patients also show increased sRaw). Thus, the prevalence of asthma diagnosable with spirometry was 21.1%. Five (1.3%) patients received the diagnosis of chronic obstructive pulmonary disease, and 234 (59.5%) had no obstructive airway disease [23]. Patients' characteristics including CSS and FeNO results are presented in Table 1.

3.2. CSS and FeNO results

'Wheezing', 'shortness of breath when wheezing', 'wheezing even when not suffering from a cold', and

Asthma assessed by WBP	Asthma assessed by spirometry		Youden index	Youden index	
Specificity (95% CI)	Specificity (95% CI)	*P*-value	WBP as a reference standard	Spirometry as a reference standard	*P*-value
63.8% (57.3%, 70.0%)	56.0% (50.2%, 61.6%)	<0.001	0,28	0,15	0.011
81.2% (75.2%, 86.3%)	73.8% (68.1%, 78.9%)	<0.001	0,30	0,21	0.120
78.8% (72.6%, 84.2%)	72.5% (66.7%, 77.7%)	0.001	0,19	0,06	0.016
43.1% (36.6%, 49.9%)	42.1% (36.4%, 48.0%)	0.555	0,09	0,12	0.544
87.6% (82.6%, 91.5%)	85.6% (81.1%, 89.4%)	0.158	0,12	0,12	0.933
74.3% (68.1%, 79.7%)	72.2% (66.8%, 77.2%)	0.205	0,10	0,09	0.837
42.4% (36.0%, 49.0%)	42.1% (36.5%, 47.9%)	0.867	0,03	0,05	0.766
50.4% (43.8%, 57.0%)	50.7% (44.8%, 56.5%)	0.885	−0,05	−0,09	0.549
71.4% (64.8%, 77.3%)	69.0% (63.3%, 74.5%)	0.170	0,03	−0,06	0.100
79.8% (74.1%, 84.8%)	71.7% (66.2%, 76.8%)	<0.001	0,33	0,22	0.029
91.5% (87.2%, 94.8%)	90.8% (87.0%, 93.8%)	0.481	0,04	0,04	0.989
63.6% (57.0%, 69.9%)	63.4% (57.5%, 68.9%)	0.880	0,04	0,07	0.669

'allergic rhinitis' showed significantly higher specificities when WBP was used as the reference test, compared with spirometry (Table 2). All further CSSs, with the exception of 'frequent cough', showed higher specificities when evaluated with WBP; however, the differences were not significant (Table 2). The sensitivities of 'wheezing' and 'allergic rhinitis' also increased when evaluated with WBP; however, these differences were not significant. There were remarkable differences in the Youden indices, in particular for wheezing ($P = 0.011$) and allergic rhinitis ($P = 0.029$). The PPVs of CSSs for an asthma diagnosis based on WBP as the reference standard were higher ($P < 0.001$) and the

Table 3. Positive and negative predictive values (PPVs and NPVs) of clinical signs and symptoms for the diagnosis of asthma when using whole-body plethysmography (WBP) compared with spirometry as a reference standard

	Asthma assessed by WBP	Asthma assessed by spirometry		Asthma assessed by WBP	Asthma assessed by spirometry	
Symptoms	PPV (95% CI)	PPV (95% CI)	*P*-value	NPV (95% CI)	NPV (95% CI)	*P*-value
Wheezing in the past 12 months	53.6% (48.4%, 58.7%)	26.5% (22.4%, 31.1%)	<0.001	72.9% (68.1%, 77.2%)	83.3% (79.0%, 86.8%)	<0.001
Shortness of breath when wheezing	64.2% (56.4%, 71.4%)	34.9% (28.3%, 42.1%)	<0.001	69.4% (65.7%, 72.9%)	82.6% (79.3%, 85.6%)	<0.001
Wheezing even when no cold	57.0% (48.7%, 64.9%)	27.0% (20.5%, 34.7%)	<0.001	65.3% (61.8%, 68.7%)	78.4% (75.3%, 81.2%)	<0.001
Ever suffer from shortness of breath	43.4% (39.4%, 47.4%)	25.2% (22.1%, 28.6%)	<0.001	66.0% (59.7%, 71.8%)	83.7% (78.1%, 88.0%)	<0.001
Woken up with shortness of breath at night	54.7% (43.6%, 65.4%)	32.8% (23.6%, 43.6%)	0.001	64.8% (62.4%, 67.1%)	81.3% (79.1%, 83.3%)	<0.001
Woken up at night with chest tightness	47.4% (39.9%, 55.0%)	26.3% (20.3%, 33.4%)	<0.001	64.3% (61.0%, 67.5%)	81.0% (78.1%, 83.7%)	<0.001
Woken up at night with coughing	40.4% (36.4%, 44.5%)	22.8% (19.6%, 26.4%)	<0.001	62.9% (56.9%, 68.5%)	80.5% (75.2%, 84.9%)	<0.001
Suffer from frequent cough	36.1% (31.1%, 41.4%)	18.3% (14.4%, 23.0%)	<0.001	58.8% (54.1%, 63.4%)	75.9% (71.8%, 79.6%)	<0.001
Cough up sputum regularly	41.9% (34.3%, 49.9%)	18.1% (12.6%, 25.3%)	<0.001	61.3% (57.9%, 64.6%)	77.4% (74.6%, 80.0%)	<0.001
Allergic rhinitis	61.8% (54.6%, 68.5%)	31.7% (25.9%, 38.2%)	<0.001	73.8% (70.0%, 77.3%)	84.5% (81.2%, 87.3%)	<0.001
Smoker	48.7% (34.4%, 63.2%)	28.2% (17.0%, 43.0%)	0.008	61.9% (60.2%, 63.6%)	79.4% (77.8%, 80.8%)	<0.001
Ex-smoker	40.3% (34.1%, 46.8%)	23.0% (18.1%, 28.8%)	<0.001	40.3% (34.1%, 46.8%)	81.5% (78.0%, 84.6%)	<0.001

Abbreviation: CI, confidence interval.

Table 4. Sensitivity and specificity of different FeNO cutoffs for the diagnosis of asthma using whole-body plethysmography (WBP) compared with spirometry as a reference standard

	Asthma assessed by WBP	Asthma assessed by spirometry	
FeNO (ppb)	Sensitivity (95% CI)	Sensitivity (95% CI)	*P*-value
>12 ppb	85.1% (78.4%, 90.3%)	84.3% (74.7%, 91.4%)	0.797
>16 ppb	69.5% (61.6%, 76.6%)	66.3% (55.1%, 76.3%)	0.372
>20 ppb	59.7% (51.5%, 67.6%)	57.8% (46.5%, 68.6%)	0.626
>35 ppb	32.5% (25.2%, 40.5%)	32.5% (22.7%, 43.7%)	0.986
>46 ppb	23.4% (16.9%, 30.9%)	22.9% (14.4%, 33.4%)	0.881
>50 ppb	22.7% (16.4%, 30.2%)	22.9% (14.4%, 33.4%)	0.959
>71 ppb	17.5% (11.9%, 24.5%)	20.5% (12.4%, 30.8%)	0.270

Abbreviations: CI, confidence interval; ppb, parts per billion; FeNO, fractional exhaled nitric oxide.

NPVs were lower ($P < 0.001$) than those for an asthma diagnosis based on spirometry (Table 3). The differences of the PPVs were highest in wheezing and allergic rhinitis. The differences of the NPVs were generally less pronounced.

The FeNO cutoff values showed significantly higher specificities ($P < 0.01$) in the WBP group, whereas the sensitivities were only slightly increased in case of FeNO < 50 ppb (Table 4). The Youden indices were insignificantly higher when WBP instead of spirometry was used as a reference standard in case of FENO <71 ppb. The PPVs of the different FeNO cutoffs assessed by WBP were higher ($P < 0.001$), and the NPVs significantly lower ($P < 0.001$), than those assessed by spirometry (Table 5).

3.3. ROC analysis

The AUC for FeNO based on WBP as a reference standard (0.66; 95% CI 0.60 to 0.71) was slightly but not significantly ($P = 0.608$) greater than the AUC when spirometry was used as the reference standard (0.62; 95% CI 0.55 to 0.69) (Fig. 2). The diagnostic accuracy increased for both reference standards when FeNO was combined with wheezing and allergic rhinitis (Fig. 3). The AUC when WBP was used as a reference standard (0.724; 95% CI 0.672 to 0.776) was significantly ($P < 0.001$) higher than the AUC when spirometry was used (0.654; 95% CI 0.585 to 0.722).

4. Discussion

The diagnostic accuracy of FeNO measurement and several CSSs increased when they were evaluated with WBP compared with spirometry indices. Beyond this, the prevalence of asthma increased when diagnostic decision-making was based on WBP, which is accompanied by increased PPVs and decreased NPVs of FeNO and CSSs.

The impact of different reference standards on the estimation of the diagnostic accuracy of FeNO measurement as an index test might partly explain the variation of optimal cutoff points described in different studies [22]. The AUC of FeNO evaluated against WBP was higher compared with spirometry, but this difference was not significant. However, we found a significant difference in the combined score of FeNO, wheezing, and allergic rhinitis. Beyond this, the specificities of all FeNO cutoff values were significantly higher. The impact of patient selection and clinical setting on diagnostic test results is well known [8,34], but up to now, the impact of different reference standards on test evaluation is rather vague. Statistic modeling pointed toward increased differences of specificities within increasing disease prevalence and increased index test accuracy when the reference standard is optimized [5–7], but this has not yet been investigated within a coherent study design. Our results fit with these theoretical considerations and therefore demonstrate that in the case of asthma, index tests perform better when they are evaluated against a superior reference standard. This aspect is important as evaluation with spirometry might underestimate the diagnostic accuracy of new diagnostic devices like FeNO measurement.

Another important point is that the PPVs of FENO and CSSs increased remarkably. This increase can be derived by the Bayes' Theorem, as the pretest probability of asthma is much higher in the WBP group, which in turn is explained by the higher sensitivity of the WBP compared with spirometry [16]. This effect became apparent by the specific design as a coherent diagnostic study where different reference standards are used in the same population. The salient aspect is that the clinical patterns of patients suspected to suffer from asthma vary considerably depending on the clinical setting. For example, a pneumologist in the German primary care setting has to be convinced of the PPV of classical CSSs such as 'wheezing' and 'allergic rhinitis' and high FeNO values when the diagnosis of asthma is established by BP in WBP. In contrast, a GP with a special interest in pneumology in the United Kingdom

Asthma assessed by WBP	Asthma assessed by spirometry		Youden index	Youden index	
Specificity (95% CI)	Specificity (95% CI)	*P*-value	WBP as a reference standard	Spirometry as a reference standard	*P*-value
28.5% (22.8%, 34.6%)	25.2% (20.4%, 30.4%)	0.008	0,14	0,10	0.297
47.3% (40.8%, 53.8%)	42.6% (37.0%, 48.3%)	0.003	0,17	0,09	0.108
63.2% (56.7%, 69.3%)	57.4% (51.7%, 63.0%)	0.001	0,23	0,15	0.140
87.9% (83.0%, 91.7%)	83.2% (78.6%, 87.2%)	0.003	0,20	0,15	0.318
92.1% (87.9%, 95.2%)	88.4% (84.3%, 91.7%)	0.009	0,15	0,11	0.328
93.7% (89.9%, 96.5%)	90.0% (86.1%, 93.1%)	0.007	0,16	0,13	0.389
97.1% (94.1%, 98.8%)	94.5% (91.4%, 96.8%)	0.025	0,15	0,15	0.911

might be less trusting of the diagnostic accuracy of CSSs and FeNO when the diagnosis of asthma is established by BP in spirometry. On the other hand, ruling out asthma when CSSs are not present (or FeNO values are low) seems to be more reliable when the diagnostic decision is made by spirometry as compared with WBP. Further studies are needed to evaluate the impact of pattern recognition and clinical decision-making under varying diagnostic circumstances.

The extent that different diagnostic techniques generate different definitions of asthma remains a subject for debate. A significant airway obstruction with positive bronchial dilation testing points clearly toward asthma. Positive BP response indicates BHR, which is a core symptom of asthma, but the PPV for asthma diagnosis is only 70% [35]. Previous studies have shown that FeNO has a high diagnostic accuracy for detecting airway hyper-reactivity [36]. Beyond that, FeNO might be superior to BP for detecting allergic inflammatory alterations of the respiratory tract and responsiveness to inhaled corticosteroids (ICS) [37]. Therefore, a delayed type of diagnostic study [38] would be necessary to compare the different diagnostic strategies which are related to the various definitions of asthma, ideally including the responsiveness to therapy with ICS. For such an evaluation, newly developed reference equations using factors such as age, height, and gender should be included as they might improve the diagnostic accuracy of FeNO measurement [39].

A limitation is that the results were derived by a secondary analysis. Therefore, the findings should be validated within other diagnostic studies. Second, the data were gathered in 2011. However, this would not distort the results as the diagnostic rules regarding interpreting BP did not change. Beyond that, it might be speculated that diagnostic decision-making in a clinical setting using WBP leads to more people being considered ill due to higher sensitivity. On the other hand, this might lead to earlier optimization of therapy. However, false classification by WBP investigation seems unlikely because the overarching test indices like AUC and the Youden index of the index tests increased with WBP as a reference standard.

5. Conclusion

The reference standard, WBP or spirometry, had a meaningful influence on the estimation of the diagnostic accuracy of the index tests 'FeNO measurement' and CSSs. A superior reference standard leads to more favorable assessment of index tests when patients suspected to suffer from asthma were evaluated. Therefore, the diagnostic accuracy of FeNO measurement might have been underestimated in

Table 5. Positive and negative predictive values (PPVs and NPVs) of different FeNO cutoffs for the diagnosis of asthma using whole-body plethysmography (WBP) compared with spirometry as a reference standard are shown

	Asthma assessed by WBP	Asthma assessed by spirometry		Asthma assessed by WBP	Asthma assessed by spirometry	
FeNO (ppb)	PPV (95% CI)	PPV (95% CI)	*P*-value	NPV (95% CI)	NPV (95% CI)	*P*-value
>12 ppb	43.4% (40.9%, 45.9%)	23.2% (21.2%, 25.3%)	<0.001	74.7% (65.9%, 81.9%)	85.7% (77.9%, 91.1%)	0.006
>16 ppb	45.9% (42.0%, 49.9%)	23.6% (20.5%, 27.0%)	<0.001	70.6% (64.7%, 76.0%)	82.5% (77.3%, 86.8%)	<0.001
>20 ppb	51.1% (45.9%, 56.3%)	26.7% (22.5%, 31.3%)	<0.001	70.9% (66.3%, 75.1%)	83.6% (79.5%, 86.9%)	<0.001
>35 ppb	63.3% (53.4%, 72.2%)	34.2% (25.9%, 43.6%)	<0.001	66.9% (64.2%, 69.5%)	82.2% (79.7%, 84.4%)	<0.001
>46 ppb	65.5% (53.0%, 76.1%)	34.6% (24.2%, 46.5%)	<0.001	65.1% (62.9%, 67.2%)	81.1% (79.1%, 82.9%)	<0.001
>50 ppb	70.0% (56.9%, 80.5%)	38.0% (26.8%, 50.7%)	<0.001	65.3% (63.2%, 67.4%)	81.3% (79.4%, 83.1%)	<0.001
>71 ppb	79.4% (63.3%, 89.6%)	50.0% (34.8%, 65.2%)	0.002	64.6% (62.9%, 66.3%)	81.6% (79.9%, 83.2%)	<0.001

Abbreviations: CI, confidence interval; ppb, parts per billion; FeNO, fractional exhaled nitric oxide.

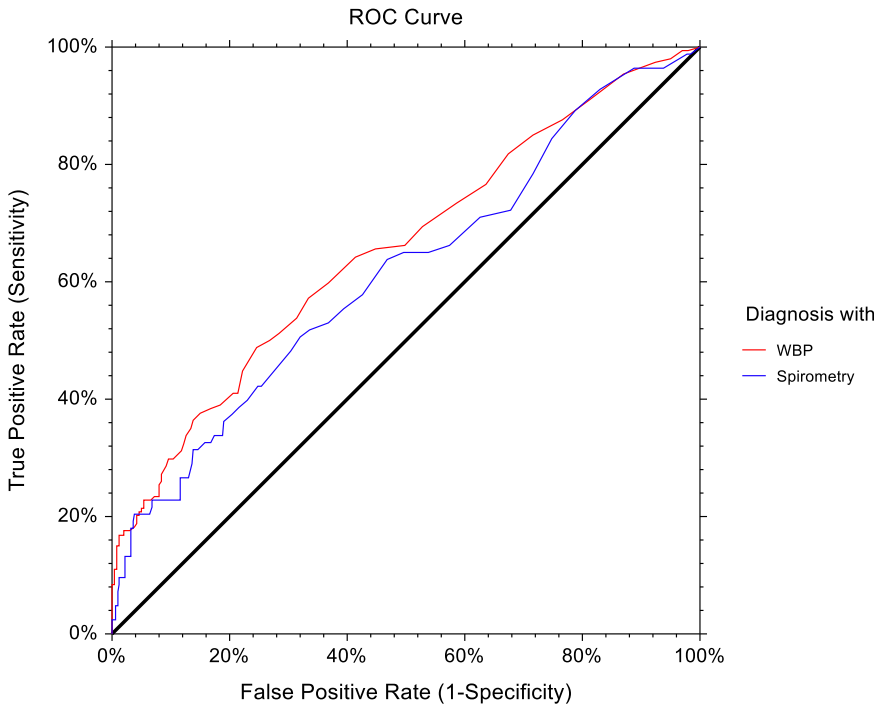

Fig. 2. ROC curves for FeNO for the diagnosis of asthma compared against alternative reference standards. The area under the curve (AUC) was 0.66 (95% CI 0.60 to 0.71) when whole-body plethysmography (WBP) was used as a reference standard. When using spirometry as a reference standard, the AUC was 0.62 (95% CI 0.55 to 0.69). There was no statistical difference when comparing both AUCs ($P = 0.608$). ROC, receiver operating characteristic; CI, confidence interval.

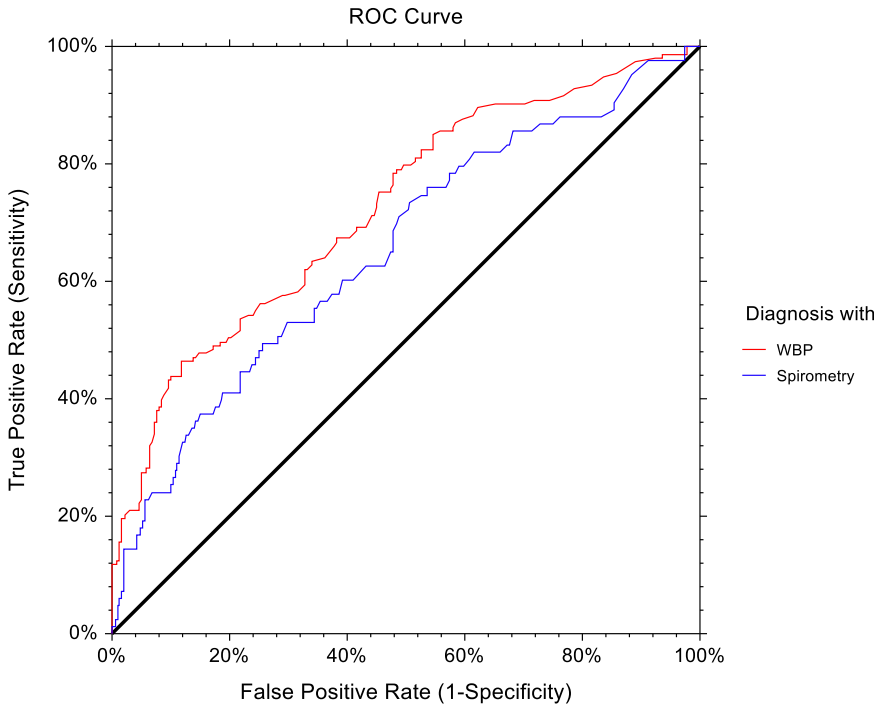

Fig. 3. ROC curve of a combined score comprising FeNO, wheezing, and allergic rhinitis for the diagnosis of asthma when using alternative reference standards. The area under the curve (AUC) was significantly higher ($P < 0.00.1$) when whole-body plethysmography (WBP) was used as a reference standard (AUC = 0.724 (95% CI 0.672 to 0.776)) than the AUC when spirometry was used as a reference standard (AUC = 0.654 (95% CI 0.585 to 0.722)). ROC, receiver operating characteristic; CI, confidence interval.

some previous studies. Consequently, where possible, index tests should be evaluated against an optimal reference standard.

Acknowledgments

We thank Klaus Linde for proof reading of the manuscript.

Authors' contributions: A.S. had the study idea and conceptualized the study. C.K. and A.S. wrote the first draft of the manuscript. C.K., S.W., and A.H. performed the statistical analysis. L.D. and R.J. helped with interpretation of data and writing.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.09.021.

References

[1] Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. BMJ 2002;324:477−80.
[2] Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089−92.
[3] Schneider A, Ay M, Faderl B, Linde K, Wagenpfeil S. Diagnostic accuracy of clinical symptoms in obstructive airway diseases varied within different health care sectors. J Clin Epidemiol 2012;65:846−54.
[4] Bösner S, Haasenritter J, Becker A, Karatolios K, Vaucher P, Gencer B, et al. Ruling out coronary artery disease in primary care: development and validation of a simple prediction rule. CMAJ 2010;182:1295−300.
[5] Boyko EJ, Alderman BW, Baron AE. Reference test errors bias the evaluation of diagnostic tests for ischemic heart disease. J Gen Intern Med 1988;3:476−81.
[6] Phelps CE, Hutson A. Estimating diagnostic test accuracy using a" fuzzy gold standard". Med Decis Making 1995;15:44−57.
[7] Thibodeau L. Evaluating diagnostic tests. Biometrics 1981801−4.
[8] Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004;140:189−202.
[9] Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J Clin Epidemiol 2009;62:797−806.
[10] NIH. Global Initiative for asthma. 2020. Available at www.ginasthma.org. Accessed July 20, 2020.
[11] Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J 2005;26:948−68.
[12] Osborne ML, Pedula KL, O'Hollaren M, Ettinger KM, Stibolt T, Buist AS, et al. Assessing future need for acute care in adult asthmatics: the Profile of Asthma Risk Study: a prospective health maintenance organization-based study. Chest 2007;132:1151−61.
[13] Schneider A, Gindner L, Tilemann L, Schermer T, Dinant GJ, Meyer FJ, et al. Diagnostic accuracy of spirometry in primary care. BMC Pulm Med 2009;9:31.
[14] Hunter CJ, Brightling CE, Woltmann G, Wardlaw AJ, Pavord ID. A comparison of the validity of different diagnostic tests in adults with asthma. Chest 2002;121:1051−7.
[15] Criee CP, Sorichter S, Smith HJ, Kardos P, Merget R, Heise D, et al. Body plethysmography–its principles and clinical use. Respir Med 2011;105:959−71.
[16] Schneider A, Schwarzbach J, Faderl B, Hautmann H, Jorres RA. Whole-body plethysmography in suspected asthma: a prospective study of its added diagnostic value in 302 patients. Dtsch Arztebl Int 2015;112:405−11.
[17] Enright P. Spirometer + body box = VW beetle + Mercedes? Respir Med 2011;105:957−8.
[18] Decramer M, Janssens W, Derom E, Joos G, Ninane V, Deman R, et al. Contribution of four common pulmonary function tests to diagnosis of patients with respiratory symptoms: a prospective cohort study. Lancet Respir Med 2013;1:705−13.
[19] Crapo RO, Casaburi R, Coates AL, Enright PL, Hankinson JL, Irvin CG, et al. Guidelines for methacholine and exercise challenge testing-1999. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999. Am J Respir Crit Care Med 2000;161:309−29.
[20] Alving K, Weitzberg E, Lundberg JM. Increased amount of nitric oxide in exhaled air of asthmatics. Eur Respir J 1993;6:1368−70.
[21] Kharitonov SA, Yates D, Robbins RA, Logan-Sinclair R, Shinebourne EA, Barnes PJ. Increased nitric oxide in exhaled air of asthmatic patients. Lancet 1994;343:133−5.
[22] Karrasch S, Linde K, Rucker G, Sommer H, Karsch-Volk M, Kleijnen J, et al. Accuracy of FENO for diagnosing asthma: a systematic review. Thorax 2017;72:109−16.
[23] Schneider A, Schwarzbach J, Faderl B, Welker L, Karsch-Volk M, Jorres RA. FENO measurement and sputum analysis for diagnosing asthma in clinical practice. Respir Med 2013;107:209−16.
[24] Exhaled N. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. Am J Respir Crit Care Med 2005;171:912−30.
[25] Quanjer P. Standardized lung function testing. Bull Eur Physiopathol Respir 1983;19(Suppl 5):1−95.
[26] Merget R, Jörres RA, Heinze E, Haufs MG, Taeger D, Brüning T. Development of a 1-concentration-4-step dosimeter protocol for methacholine testing. Respir Med 2009;103:607−13.
[27] Schneider A, Tilemann L, Schermer T, Gindner L, Laux G, Szecsenyi J, et al. Diagnosing asthma in general practice with Portable exhaled nitric oxide measurement − results of a prospective diagnostic study. Respir Res 2009;10:15.
[28] Bjermer L, Alving K, Diamant Z, Magnussen H, Pavord I, Piacentini G, et al. Current evidence and future research needs for FeNO measurement in respiratory diseases. Respir Med 2014;108(6):830−41.
[29] Hewitt RS, Modrich CM, Medlicott T, Cowan JO, Taylor DR. Supporting the diagnosis of non-specific respiratory symptoms in primary care: the role of exhaled nitric oxide measurement and spirometry. Prim Care Respir J 2008;17(2):97.
[30] Dweik RA, Boggs PB, Erzurum SC, Irvin CG, Leigh MW, Lundberg JO, et al. An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications. Am J Respir Crit Care Med 2011;184:602−15.
[31] Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 1927;22:209−12.
[32] Schneider A, Wagenpfeil G, Jörres RA, Wagenpfeil S. Influence of the practice setting on diagnostic prediction rules using FENO measurement in combination with clinical signs and symptoms of asthma. BMJ Open 2015;5:e009676.
[33] Chen F, Xue Y, Tan MT, Chen P. Efficient statistical tests to compare Youden index: accounting for contingency correlation. Stat Med 2015;34:1560−76.

[34] Knottnerus JA, Knipschild PG, Sturmans F. Symptoms and selection bias: the influence of selection towards specialist care on the relationship between symptoms and diagnoses. Theor Med 1989;10:67−81.

[35] Perpina M, Pellicer C, de Diego A, Compte L, Macian V. Diagnostic value of the bronchial provocation test with methacholine in asthma. A Bayesian analysis approach. Chest 1993;104:149−54.

[36] Sverrild A, Malinovschi A, Porsbjerg C, Backer V, Alving K. Predicting airway hyperreactivity to mannitol using exhaled nitric oxide in an unselected sample of adolescents and young adults. Respir Med 2013;107:150−2.

[37] Alving K. FeNO and suspected asthma: better to identify responsiveness to treatment than to label with a diagnosis. Lancet Respir Med 2018;6:3−5.

[38] Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol 2003; 56:1118−28.

[39] Jacinto T, Amaral R, Malinovschi A, Janson C, Fonseca J, Alving K. Exhaled NO reference limits in a large population-based sample using the Lambda-Mu-Sigma method. J Appl Physiol (1985) 2018;125: 1620−6.