

A novel statistical model for analyzing data of a systematic review generates optimal cutoff values for fractional exhaled nitric oxide for asthma diagnosis

Antonius Schneider^{a,*}, Klaus Linde^a, Johannes B. Reitsma^b, Susanne Steinhauser^c,
Gerta Rücker^d

^aInstitute of General Practice, University Hospital Klinikum rechts der Isar, Technische Universität München, Orleansstraße 47, 81667 München, Germany

^bJulius Center for Health Sciences and Primary Care, University Medical Center Utrecht, PO Box 85500, 3508 GA, Utrecht, The Netherlands

^cInstitute of Medical Statistics, Informatics and Epidemiology, University of Cologne, Bachemer Str. 86, 50931 Köln, Germany

^dInstitute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Stefan-Meier-Strasse 26, 79104 Freiburg, Germany

Accepted 1 September 2017; Published online 12 September 2017

Abstract

Objectives: Measurement of fractional exhaled nitric oxide (FE_{NO}) might substitute bronchial provocation for diagnosing asthma. However, optimal FE_{NO} thresholds for diagnosing asthma remain unclear. We reanalyzed data collected for a systematic review investigating the diagnostic accuracy of FE_{NO} measurement to exploit all available thresholds under consideration of pretest probabilities using a newly developed statistical model.

Study Design and Setting: One hundred and fifty data sets for a total of 53 different cutoffs extracted from 26 studies with 4,518 participants were analyzed with the multiple thresholds model. This model allows identifying thresholds at which the test is likely to perform best.

Results: Diagnosing asthma might only be possible in a meaningful manner when the pretest probability of asthma is at least 30%. In that case, FE_{NO} > 50 ppb leads to a positive predictive value of 0.76 [95% confidence interval (CI): 0.29–0.96]. Excluding asthma might only be possible, when the pretest probability of asthma is 30% at maximum. Then, FE_{NO} < 20 ppb leads to a negative predictive value of 0.86 (95% CI 0.66–0.95).

Conclusion: The multiple thresholds model generates a more comprehensive and more clinically useful picture of the effects of different thresholds, which facilitates the determination of optimal thresholds for diagnosing or excluding asthma with FE_{NO} measurement. © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Asthma; Fractional exhaled nitric oxide; Diagnostic accuracy; Sensitivity; Specificity; Receiver operating characteristic analysis

Conflict of interest: The authors declare that they have no conflict of interest.

Funding: A.S. received funding for the systematic review of the diagnostic accuracy of FE_{NO} by the German Federal Ministry of Education and Research (BMBF FKZ 01KG1211). G.R. was funded by the German Research Foundation (DFG) (RU 1747/1-1/2) for the development of the multiple thresholds model. The funding organizations had no role in the design or conduct of this research.

Availability of data and materials: The study data are provided as supplementary excel file. R code for analysis and plotting can also be found as supplemental material.

* Corresponding author. Tel.: +49-89-614658916; fax: +49-89-614658915.

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

1. Introduction

Asthma is a complex chronic inflammatory disorder of the bronchial tree, characterized by bronchial hyper-responsiveness and usually reversible airway obstruction. Many efforts continue to be undertaken to improve the diagnostic process to allow an early diagnosis, as early treatment is important for the management of the disease [1]. However, ruling-in and ruling-out the diagnosis of asthma in patients suffering from chronic symptoms like dyspnea, wheezing, and coughing is still challenging. Spirometry is considered a reference standard for diagnosing airway obstruction, but it is not possible to rule out milder forms of asthma, as obstruction is not present in these cases [2]. For these cases, bronchial provocation (BP) is a reference standard, particularly in cases with

What is new?

Key findings

- The multiple thresholds model allows using all available information on thresholds of studies investigating biomarkers, as demonstrated with a systematic review of studies investigating the diagnostic accuracy of FE_{NO} measurement.

What this study adds to what was known?

- The multiple thresholds model enables the calculation of summarized sensitivities and specificities of different cutoff points, and applying Bayes' theorem allows calculating the predictive values, given the prevalence of the disease.
- This allows deriving optimal thresholds related to the clinical setting where the index test is used, as defined by the pretest probability of a disease.

What is the implication and what should change now?

- A prerequisite for summarizing diagnostic test accuracy studies with the multiple thresholds model is that authors of primary studies report the pairs of sensitivities and specificities of all thresholds they examined of the index test or, even more preferably, the biomarker values of all individual patients along with their true disease status according to the reference standard.
- If reviewers encounter primary studies that report findings for multiple thresholds in sufficient detail, they should consider the multiple thresholds model or related methods.

inconclusive spirometric results, but BP is time-consuming, costly, and carries a small risk of inducing severe bronchospasm [3]. New biomarkers like fractional exhaled nitric oxide (FE_{NO}) might help to close this diagnostic gap. Nitric oxide (NO) in orally exhaled air mainly originates from the respiratory epithelium, where it is produced by the inducible NO synthase [4]. FE_{NO} was initially estimated by the guideline of the American Thoracic Society (ATS) as a specific marker for eosinophilic airway inflammation [5]. Results of recent studies suggested a more comprehensive concept which reclassified FE_{NO} more accurately as a broader marker of T-helper cell type 2-mediated allergic inflammation [6,7]. Fitting to the concept of increased FE_{NO} values correlating with allergic inflammation pattern, several studies have demonstrated a diagnostic value of FE_{NO} measurement [8–11]. We recently performed a systematic review of all available studies investigating the diagnostic accuracy of FE_{NO} in patients suspected to suffer

from asthma. Using standard meta-analytic methods, we found a summarized area under the curve of 0.80 (95% confidence interval: 0.77–0.85) [12].

Systematic reviews of DTA studies have become a valuable tool to summarize the available evidence and generate recommendations about the use of diagnostic tests in practice [13]. If a test produces a binary result, the methods for meta-analyzing such data are now well established. In these situations, there is a single two-by-two table that can be analyzed using the bivariate or hierarchical summary approach [13]. The analysis becomes more complex if a test produces a continuous result as in the case of FE_{NO} measurement. The actual choice of threshold now becomes an additional factor to evaluate as changing the threshold will alter the corresponding sensitivity and specificity. Clinical decision-making about how to use a test is driven by the absolute number of correct and incorrect decisions based on the results of the test and the relative weights attached to consequences due to a false-positive or false-negative test result. Of course, these weights will depend on the clinical scenario at hand, but it is clear that clinical decision-making of tests producing a continuous result has to take into account the prevalence, the choice of threshold, and the accuracy corresponding to each threshold. Therefore, primary studies often present pairs of sensitivity and specificity for multiple thresholds or a receiver operating characteristic (ROC) curve. Unfortunately, current diagnostic reviews typically select one pair of sensitivity and specificity to be analyzed using the standard bivariate meta-analysis model [14,15]. For this purpose, often the most commonly used threshold is chosen, thus ignoring all information coming from other thresholds and wasting relevant information about test performance. Therefore, it was not possible to identify optimal thresholds for medical decision-making with the traditional way of a systematic review [12].

In a recent paper [16], two authors of the present article (G.R. and S.S.) proposed a statistical approach for meta-analysis of diagnostic test accuracy (DTA) studies, henceforth called the multiple thresholds model, that allows using all available information on thresholds and the corresponding two-by-two tables from all studies. The aim was to apply the new model to reanalyze the diagnostic accuracy of FE_{NO} measurement and to derive optimal decision rules under consideration of pretest probabilities for diagnosing asthma.

2. Methods

The overall methods used for our systematic review have been described previously [12]. In brief, we searched MEDLINE, Embase, and Scopus up to November 29, 2015, for potentially relevant studies. To be eligible, studies had to allow the generation of two-by-two tables for asthma diagnosis by FE_{NO} as compared to a reference standard. The reference standard could be BP, measurement of FEV₁ with

bronchodilation, peak flow variability, expert’s opinion, or a combination of these. Study participants were patients with suspected asthma and at least 75% had to be steroid naïve. Measurement of FE_{NO} had to be done using a mean exhalation flow rate of 50 mL/s and instantaneous flows within the range of 45–55 mL/s according to the international guidelines. Study quality was assessed with the QUADAS-2 tool. For meta-analysis, we used data for the cutoff point with the highest sum of sensitivity and specificity (i.e., the Youden Index) whenever possible.

For the analyses presented in this paper, we extracted the numbers of true test positives, true negatives, false positives, and false negatives based on each cutoff point for which these figures were available or could be calculated. The data were collected in an Excel sheet (data file in the online Appendix at www.jclinepi.com).

The multiple thresholds model creates a link between the range of thresholds and the respective pairs of sensitivity and specificity and thus allows identifying thresholds at which the test is likely to perform best [16]. Different studies may contribute a varying number of thresholds as well as different sets of thresholds. The multiple thresholds model is a multilevel random effects model. At the study level, for the group of patients without the target condition (in short disease free), the specificities at all available thresholds together provide an estimate of the cumulative distribution function (cdf) of the test results within the disease-free individuals. Likewise, for patients with the target condition (in short diseased), via the observed sensitivities at all observed thresholds, we obtain an estimate of the cdf of the test results within the diseased patients. At the meta-analytic level, the model fits the data for both groups and all available thresholds over all studies. Based on a log-logistic model, it provides estimates of the two cdfs for the two groups across all studies, accounting for the between-study heterogeneity and correlation between groups.

To be more precise, in order to fit a linear model to the data, we first applied the natural logarithm (ln) to the FE_{NO} values. We then transformed all specificities (for disease-free patients) and all one minus sensitivities (for diseased patients) using the logit transformation, $\text{logit}(x) = \ln(x) - \ln(1 - x)$ and fitted a linear mixed-effects model to all data points. Each data point was weighted with the inverse variance of the respective logit-transformed proportion. The model (called DIDS* in [16], which stands for different random intercepts and different random slopes) is given by

$$\text{logit}(Sp_{ki}) = \alpha_0 + a_{0k} + (\beta_0 + b_{0k})x_{ki} + \varepsilon_{ki}$$

$$\text{logit}(1 - Se_{ki}) = \alpha_1 + a_{1k} + (\beta_1 + b_{1k})x_{ki} + \delta_{ki}$$

where Sp_{ki} and Se_{ki} are the specificity and the sensitivity of the i ’th threshold x_{ki} (on the log scale) in study k , α_0 and α_1 are the fixed intercepts, β_0 and β_1 the fixed slopes for the disease-free and the diseased individuals, respectively. The terms a_{0k} , a_{1k} , b_{0k} , and b_{1k} denote the random intercepts and slopes which are assumed to follow a common

four-dimensional normal distribution, that is, to correlate across studies. With ε_{ki} and δ_{ki} , we denote the within-study random errors. After using the inverse of the logit function for back-transforming the fixed effects part of the model equations, we obtain model-based distribution functions for disease-free and diseased individuals from which a model-based summary ROC curve is obtained in the usual way. The model was implemented in the free software environment R [17]. For more details of the modeling, we refer to Steinhäuser et al. [16] and provide R code in the supplemental material.

We then can read off the “pooled” sensitivity and specificity values at every threshold; and a multiple thresholds summary ROC (mtsROC) curve naturally follows while preserving threshold information. Additionally, we obtain positive and negative predictive values, depending on the threshold and different possibilities with respect to the prevalence of the disease observed in the included studies. The 95% CIs of sensitivities, specificities, and predictive values are estimated using the delta method [16].

For sensitivity analysis, we investigated the influence of age by comparison of children and adults as subgroup analysis. A further subgroup analysis was performed regarding the subgroup of patients where solely BP was used as a reference standard.

3. Results

We included a total of 26 studies [11,18–42] with 29 patient groups (two studies presented results for separate subgroups) and 4,518 participants (see Table 1 for characteristics).

In our original meta-analysis using the bivariate model, we found a fair diagnostic accuracy of FE_{NO} for diagnosing asthma [12]. The overall sensitivity was 0.65 (95% CI: 0.58–0.72), and the overall specificity was 0.82 (0.76–0.86). The standard sROC curve according to Rutter and Gatsonis [43] is displayed in Fig. 1A. However, valuable information on thresholds was lost because the standard bivariate approach focuses on a single pair of sensitivity and specificity from each study, thereby ignoring the impact and value of the different thresholds applied.

For the present analysis, we extracted true test positives, true negatives, false positives, and false negatives for further 121 data points in addition to the 29 used in the standard analysis. Data were now available for a total of 150 data points, the median number of cutoffs per study was 4 with interquartile range 1–7, and a maximum of 19 cutoffs. There were 53 different cutoffs used over all studies (Table 2).

The application of the multiple thresholds model [16] leads to the mtsROC curve which allows estimating the sensitivities and specificities (Fig. 1B). For example, at the threshold where the Youden index was maximized, that was FE_{NO} = 26 parts per billion (ppb), we found a sensitivity

Table 1. Characteristics of included studies

First author	Year	Country	Age group (range years or mean and SD)	n	Prevalence	FE _{NO} measurement device	Number cutoffs available	Reference standard
Arora et al. [18]	2006	USA	Adults (17–38)	172	0.80	NIOX—no further specification	19	Bronchial provocation
Cordeiro et al. [19]	2011	The Netherlands	Mainly adults (7–87)	114	0.37	NIOX Flex (c)	1	FEV1/ VC + reversibility and bronchial provocation
El Halawani et al. [20]	2003	USA	Adults (18–40)	49	0.14	Sievers 280 (c)	4	Exercise challenge
Florentin et al. [21]	2014	France	Young adults (mean 25 SD 3)	178	0.11	NIOX Mino (e)	1	FEV1/ VC + reversibility and PEF variability
Fortuna et al. [22]	2007	Spain	Adults (18–64)	50	0.44	SIR N-6008 (c)	1	Bronchial provocation
Fukuhara et al. [23]	2011	Japan	Adults (48–66)	61	0.69	NA623N (c)	1	FEV1/VC (+reversibility), bronchial provocation, sputum eosinophilia
Giovannini et al. [24]	2014	Italy	Mainly adults (mean 38 SD 15)	42	0.50	HypAir FE _{NO} (e)	1	Bronchial provocation
Heffler et al. [25]	2006	Italy	Mainly adults (11–75)	48	0.38	NIOX Flex (c)	17	FEV1/ VC + reversibility and bronchial provocation
Katsoulis et al. [26]	2013	Greece	Adults (22–37)	112	0.43	NIOX Mino (e)	6	Bronchial provocation
Kostikas et al. [27]	2008	Greece	Young adults (mean 21 SD 2)	149	0.42	NIOX Mino (e)	6	FEV1/ VC + reversibility, bronchial provocation
Kowal et al. [28]	2009	Poland	Adults (18–45)	540	0.33	Sievers 280i (c)	10	FEV1/ VC + reversibility, bronchial provocation, PEF variability, expert's opinion
Linkosalo et al. [29]	2012	Finland	Children (6–19)	30	0.60	Sievers 280 (c)	5	Free running test
Malinowski et al. [30]	2012	Denmark	Mainly adults (14–44)	108	0.42	NIOX Mino (e)	7	Expert's assessment based on FEV1/ VC + reversibility, bronchial provocation and medication use
				62	0.31		8	
				112	0.29		8	
Pedrosa et al. [31]	2010	Spain	Adults (mean 34 SD 13)	114	0.31	NIOX Mino (e)	1	Bronchial provocation
Pizzimenti et al. [32]	2009	Italy	Not stated (probably adults)	156	0.09	NIOX Mino (e)	1	Bronchial challenge, spirometry
Sato et al. [33]	2008	Japan	Adults (20–78)	71	0.68	Kimoto, Osaka, Japan (c)	5	FEV1/VC (+reversibility) and/or bronchial provocation, symptoms
Schleich et al. [34]	2012	Belgium	Adults (mean 41 SD 16)	174	0.47	NIOX Flex (c)	1	Bronchial provocation
Schneider et al. [11]	2013	Germany	Mainly adults (mean 43)	393	0.39	NIOX Mino (e)	13	FEV1/ VC + reversibility and bronchial provocation

(Continued)

Table 1. Continued

First author	Year	Country	Age group (range years or mean and SD)	n	Prevalence	FE _{NO} measurement device	Number cutoffs available	Reference standard
Sivan et al. [35]	2009	Israel	Children (5–18)	113	0.61	Eco Physics CLD88 (c)	3	Expert assessment based on FEV1/VC (+reversibility) and/or bronchial provocation, medication
Smith et al. [36]	2004	New Zealand	Mainly adults (9–72)	47	0.36	Logan LR2000 (c)	4	FEV1/VC + reversibility, bronchial provocation
Smith et al. [37]	2005	New Zealand	Mainly adults (14–71)	52	0.52	NIOX Flex (c)	2	FEV1/FVC reversibility, meta-choline challenge, ICS response
Tilemann et al. [38]	2011	Germany	Adults (38–58)	210	0.41	NIOX Mino (e)	6	Bronchial provocation or reversibility
Voutilainen et al. [39]	2013	Finland	Adolescents and adults (14–31)	87	0.34	NIOX Flex (c)	1	Expert's assessment based on FEV1/VC + reversibility, bronchial provocation, PEF variability
Wang et al. [40]	2015	China	Mainly adults (13–89)	515	0.35	Nano Coulomb NO analyzer, Shangwo (u)	1	Depending on FEV1 predicted positive bronchodilatation or bronchial provocation
Bronchodilatation Bronchial provocat.				408	0.31		1	
Woo et al. [41]	2012	Korea	Children (8–16)	245	0.68	NIOX Mino (e)	15	FEV1/VC + reversibility and bronchial provocation
Zhang et al. [42]	2011	China	Adults (15–70)	106	0.37	NIOX Mino (e)	1	FEV1 with bronchodilator, bronchial challenge test

Abbreviations: SD, standard deviation; ppb, parts per billion; FE_{NO}, fractional exhaled nitric oxide; FEV1, forced expiratory volume in 1 second; VC, vital capacity; PEF, peak expiratory flow rate; reversibility, positive reaction to a bronchodilator; ICS, inhaled corticosteroids; e, electrochemical; c, chemiluminescence; u, unclear.

of 0.66 (95% CI: 0.56–0.74) and a specificity of 0.79 (95% CI: 0.73–0.84). For a clinician, presenting these test indices might be still not satisfying, as he or she wants to know the probability that the patient has or has not asthma when a resulting FE_{NO} value exceeds a certain threshold. Applying Bayes' theorem allows calculating the predictive values, given the prevalence of the disease and the sensitivities and specificities for a threshold [44]. The multiple thresholds model now includes all this information, thus allowing to calculate the positive and negative predictive values related to the full range of threshold values and for different prevalences. Fig. 2A illustrates that making the diagnosis of asthma might only be possible in a meaningful manner when the prevalence of asthma (pretest probability) is at least 30%. In that case, a FE_{NO} > 50 ppb leads to a positive predictive value (PPV) of 0.76 (95% CI: 0.29–0.96).

Fig. 2B is probably even more useful for clinical practice. Here, the posttest probability is not estimated for whether the observed value is below or above a given threshold but simply based on the specific values actually measured. Again, Bayes' formula is used for calculating the conditional probabilities of asthma, given a patient's individual FE_{NO} values and based on the results of the meta-analysis incorporating all relevant and reported data of the primary studies. For example, the probability of having asthma is 0.62 (95% CI: 0.20–0.91), when FE_{NO} = 50 ppb in case for a pretest probability of 30%; 0.69 (95% CI: 0.24–0.94) when FE_{NO} = 60 ppb; and 0.73 (95% CI: 0.26–0.96) when FE_{NO} = 70 ppb. Though there is some uncertainty in these numbers, as illustrated by the 95% CI, this has an interesting interpretation: instead of only using the (dichotomous) information that the

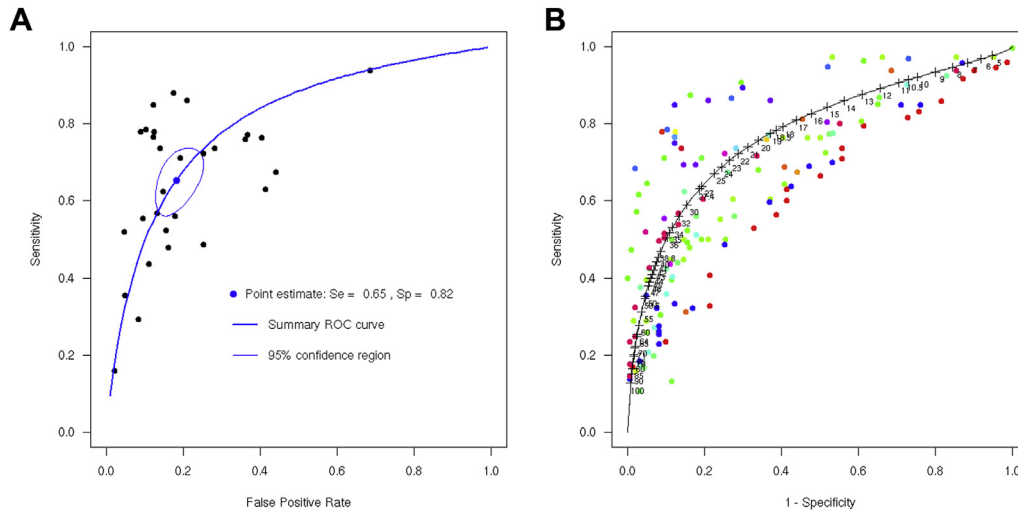


Fig. 1. (A) Summary receiver operating characteristic (sROC-) curve based on the bivariate model (using one pair of sensitivity and specificity) in 29 studies on the accuracy of fractional exhaled nitric oxide (FE_{NO}) for diagnosing asthma. Circles represent information on sensitivity and specificity in the individual studies. The blue circle represents the pooled point estimate, the ellipsoid its 95% confidence region. (B) Multiple thresholds sROC (mtsROC) curve based on the multiple thresholds model using all available information. Circles represent information on sensitivity and specificity in the individual studies (one color per study). Numbers are cutoff values of fractional exhaled nitric oxide (FE_{NO}) [ppb].

patient's value exceeds a certain threshold, we use the patient's actual value, based on the consideration that the individual's probability of having asthma continuously increases with the actual value. The potential for excluding asthma is illustrated in Figs 2C and D. Fig. 2C shows that excluding asthma might only be possible in a useful manner, when the pretest probability of asthma is 30% at maximum. In that case, a $FE_{NO} < 20$ ppb leads to a negative predictive value of 0.86 (95% CI: 0.66–0.95). Fig. 2D illustrates that in case of a pretest probability of 30%, the probability of not having the disease is 0.79 (95% CI: 0.61–0.90) when $FE_{NO} = 20$ ppb.

Moreover, we asked whether the diagnostic performance of FE_{NO} depended on covariates such as the diagnostic reference standard (diagnosis only with provocation) and the age group included in the study (adults vs. children). We found no significant impact for the diagnostic reference standard, which is in line with our earlier analysis using the bivariate model [12]. The specificities of FE_{NO} values were significantly higher in the children group when $FE_{NO} < 25$ ppb.

4. Discussion

The multiple thresholds model generates a more complete and more clinically useful picture of the effects of different thresholds in relation to their sensitivities, specificities, and predictive values. This enhances decision making about whether and how to use diagnostic tests devices or biomarkers in clinical practice. The new model shows that an optimal threshold for making the diagnosis of asthma might be around 60 ppb, and exclusion of asthma might be possible with 20 ppb, when the pretest probability of asthma is 30%.

For calculating posttest probabilities, which are crucial for thinking through downstream consequence, we need pretest probabilities. These can be derived from the included diagnostic studies or tailored toward the specific setting in which the test will be used. For example, previous studies have shown that the pretest probability of asthma is at least 30% in primary care setting, when patients with complaints suggestive of asthma are coming for the first diagnostic workup [45]. Related to the scenario of our clinical example based on the multiple thresholds model, a primary care physician could start the treatment with inhaled corticosteroids without the necessity to confirm the diagnosis with BP when $FE_{NO} > 50$ ppb is given, corresponding to a PPV of 0.76 (95% CI: 0.29–0.96). Such a PPV is similar to established BP testing procedures, which was shown to be around 70% when the pretest probability of asthma is around 30% [3,46]. Therefore, FE_{NO} measurement might be equivalent to BP. However, the probability of asthma increases remarkably with increasing FE_{NO} values as illustrated by the calculation of the exact conditional probabilities. Further studies have to evaluate if FE_{NO} measurement might be even superior to BP in some circumstances. For example, a recent study found that patients with negative BP response might benefit from inhaled corticoid steroids (ICSs) when $FE_{NO} > 33$ ppb because FE_{NO} measurement turned out to be more useful for predicting ICS responsiveness than diagnosing asthma [47]. However, the authors state that the chosen criteria were likely to be “oversensitive” in detecting a response.

The ATS guideline recommends to diagnose eosinophilic inflammation when $FE_{NO} > 50$ ppb [5]. However, the multiple thresholds model under consideration of pretest probabilities clearly illustrates the impact of the clinical setting where diagnoses are made. Making the diagnosis of

Table 2. Sensitivities and specificities of the identified thresholds

Threshold	Number of identified studies	Sensitivity (95% CI)	Specificity (95% CI)
5	1	0.98 (0.97–0.99)	0.05 (0.03–0.08)
6	1	0.97 (0.95–0.98)	0.08 (0.06–0.12)
7	1	0.96 (0.94–0.97)	0.12 (0.08–0.16)
8	1	0.95 (0.92–0.96)	0.16 (0.11–0.21)
9	2	0.94 (0.91–0.96)	0.20 (0.15–0.27)
10	10	0.92 (0.89–0.95)	0.25 (0.19–0.32)
10.5	1	0.91 (0.88–0.94)	0.27 (0.21–0.35)
11	1	0.91 (0.87–0.94)	0.30 (0.23–0.37)
12	4	0.89 (0.85–0.93)	0.34 (0.27–0.42)
13	1	0.88 (0.83–0.91)	0.39 (0.32–0.47)
14	2	0.86 (0.81–0.90)	0.44 (0.36–0.52)
15	11	0.84 (0.79–0.89)	0.48 (0.40–0.56)
16	3	0.83 (0.76–0.88)	0.52 (0.44–0.60)
17	2	0.81 (0.74–0.86)	0.56 (0.48–0.64)
18	2	0.79 (0.72–0.85)	0.60 (0.52–0.67)
18.5	1	0.78 (0.71–0.84)	0.61 (0.54–0.69)
19	3	0.78 (0.70–0.84)	0.63 (0.55–0.70)
20	14	0.76 (0.68–0.82)	0.66 (0.59–0.73)
21	1	0.74 (0.66–0.81)	0.69 (0.62–0.75)
22	3	0.72 (0.64–0.79)	0.71 (0.64–0.77)
23	1	0.71 (0.62–0.78)	0.74 (0.67–0.79)
24	1	0.69 (0.60–0.77)	0.76 (0.69–0.81)
25	11	0.67 (0.58–0.75)	0.77 (0.71–0.83)
27	1	0.64 (0.54–0.72)	0.81 (0.75–0.85)
27.4	1	0.63 (0.54–0.72)	0.81 (0.76–0.86)
30	13	0.59 (0.49–0.68)	0.85 (0.80–0.89)
32	1	0.56 (0.46–0.66)	0.87 (0.82–0.90)
34	2	0.53 (0.43–0.63)	0.88 (0.84–0.91)
35	6	0.52 (0.42–0.62)	0.89 (0.85–0.92)
36	1	0.50 (0.40–0.61)	0.90 (0.86–0.93)
38.8	1	0.47 (0.37–0.57)	0.91 (0.88–0.94)
40	8	0.45 (0.35–0.56)	0.92 (0.89–0.94)
41	3	0.44 (0.34–0.55)	0.92 (0.89–0.95)
42	1	0.43 (0.33–0.54)	0.93 (0.90–0.95)
43	1	0.42 (0.32–0.53)	0.93 (0.91–0.95)
44	1	0.41 (0.31–0.52)	0.94 (0.91–0.96)
45	3	0.40 (0.30–0.50)	0.94 (0.91–0.96)
46	3	0.39 (0.29–0.49)	0.94 (0.92–0.96)
47	2	0.38 (0.29–0.48)	0.95 (0.92–0.96)
50	7	0.35 (0.26–0.46)	0.95 (0.93–0.97)
50.6	1	0.35 (0.26–0.45)	0.95 (0.93–0.97)
55	2	0.31 (0.23–0.41)	0.96 (0.95–0.98)
60	2	0.28 (0.20–0.38)	0.97 (0.96–0.98)
64	1	0.25 (0.18–0.35)	0.97 (0.96–0.98)
65	1	0.25 (0.17–0.34)	0.98 (0.96–0.98)
70	1	0.22 (0.15–0.31)	0.98 (0.97–0.99)
71	1	0.22 (0.15–0.31)	0.98 (0.97–0.99)
75	1	0.20 (0.14–0.29)	0.98 (0.97–0.99)
76	1	0.20 (0.13–0.28)	0.98 (0.97–0.99)
80	2	0.18 (0.12–0.26)	0.99 (0.98–0.99)
85	1	0.17 (0.11–0.24)	0.99 (0.98–0.99)
90	1	0.15 (0.10–0.23)	0.99 (0.98–0.99)
100	2	0.13 (0.08–0.19)	0.99 (0.99–1.00)

Abbreviation: CI, confidence interval.

asthma with FE_{NO} values lower than 60 ppb might only be possible when enhanced pretest probabilities are given, like in selected patient in the hospital setting, and even FE_{NO} = 60 ppb still gives a potential 31% chance of the diagnosis being incorrect, when the pretest probability is

30%. A valuable strategy in primary care to enhance the diagnostic added value of point-of-care-testing instruments is to combine testing results with clinical signs and symptoms. It was shown recently that the diagnostic information of C-reactive protein measurement increases remarkably when interpreted in the clinical context of medical history and clinical examination [48,49]. This might also be true for FE_{NO} measurement, when information about wheezing and allergic rhinitis is used for clinical interpretation [45]. In this prediction model, the probability of asthma was 78% when FE_{NO} was at least 30 ppb and both clinical signs were present. However, this prediction model needs to be validated in a confirmatory study.

Excluding asthma in primary care might be possible with FE_{NO} < 20 ppb. It must be considered that the pretest probability of being healthy might be around 70% and is only enhanced up to 79% (95% CI: 0.61–0.90) when FE_{NO} = 20 ppb. On the other hand, the NPV of FE_{NO} < 20 ppb would be equal to the NPV of a 20% fall of FEV₁ during BP, which can be detected with spirometry maneuvers [50]. However, there is no consensus about strategies how to deal with uncertainty in terms of accepting probabilities for clinical decision-making. Beyond this, it needs being kept in mind that FE_{NO} had a diagnostic blind spot in noneosinophilic asthma [11,51]. This is of importance when the presence of asthma is suggested by symptoms and clinical history, but FE_{NO} is low.

While our findings represent substantial increases to the PPV and NPV at the thresholds, respectively, they illustrate also that the chance of the diagnosis being incorrect remains sufficiently large, so that asthma would have been neither “ruled in” nor “ruled out” as a possibility. This underlines that there is still some inherent uncertainty in making the diagnosis of asthma, also depending from the course of the disease [52,53]. Consequently, from a clinical point of view, patients should be reevaluated regularly to confirm, or eventually reject, the diagnosis of asthma.

4.1. Limitations

Some limitations of the new model have yet to be discussed. As well as the bivariate model, the multiple thresholds model is a two-stage approach, based on the estimated study-specific sensitivities, but ignoring the uncertainty of their variances at study level. In case of zeros in the two-by-two tables, continuity correction is needed; a practice that has been criticized [54]. The estimated CIs around the posttests probabilities are relatively wide, reflecting the uncertainty that was still present within this review about the exact shape of the summary ROC curve. A limitation of our model is that it sometimes fails to estimate a valid mtsROC curve if there is large between-study heterogeneity or if too many studies report only one threshold [16]. There are other approaches that have been suggested, such as the nonparametric gamma-frailty model presented by Putter et al. [55] and the multivariate model by Hamza

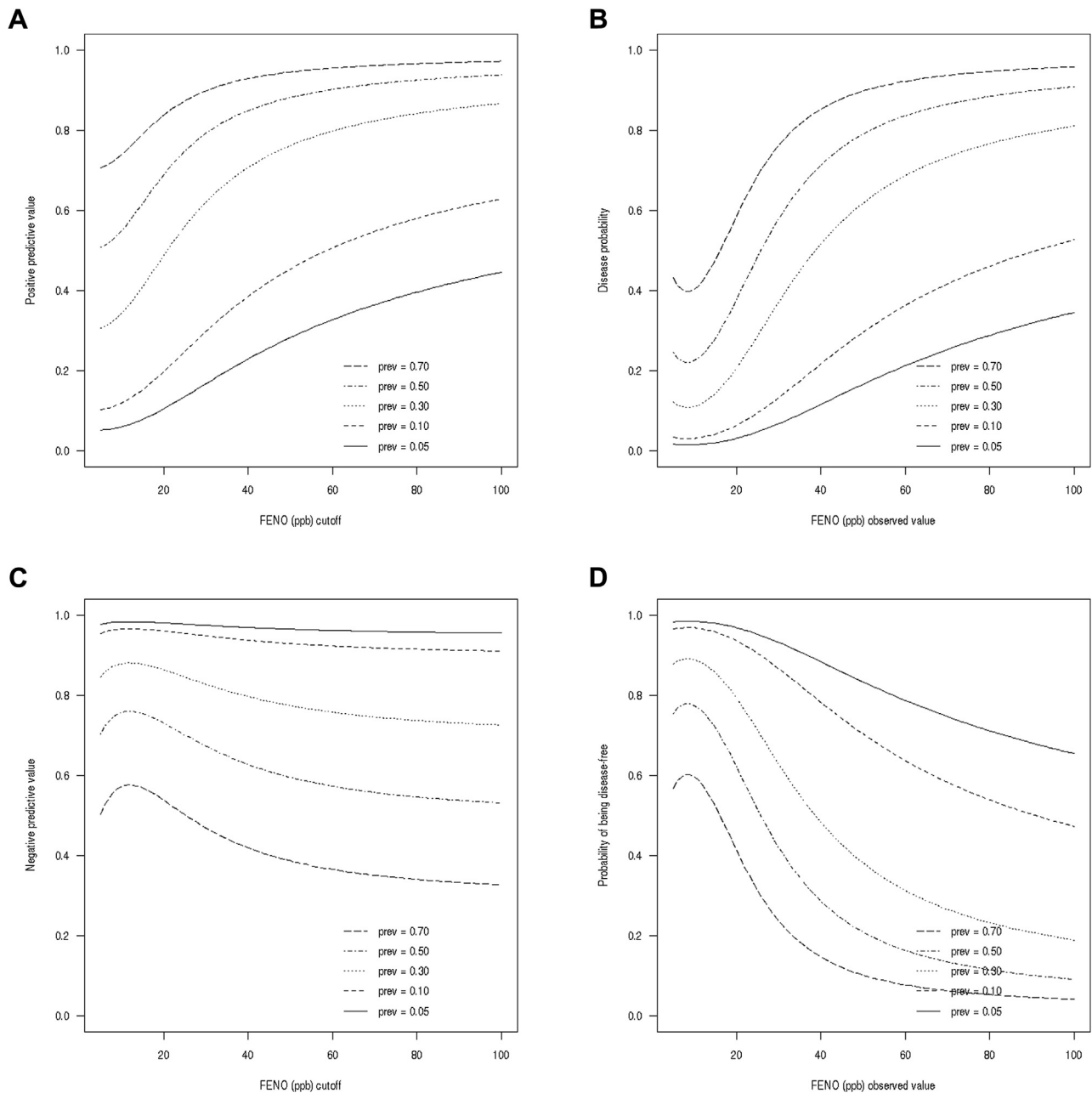


Fig. 2. (A) Positive predictive values at different FE_{NO} cutoff values for different prevalences (prev). (B) Disease probability in individual patients for observed FE_{NO} values for different prevalences (prev). (C) Negative predictive values at different FE_{NO} cutoff values for different prevalences (prev). (D) Probability of not having asthma in individual patients for observed FE_{NO} values for different prevalences (prev).

et al. [56]. Both models assume a fixed set of thresholds for all studies, which for most biomarkers is not realistic in practice. Martínez-Cambor [57] suggested a nonparametric averaging of ROC curves, not using any threshold information. A more refined approach, a multivariate regression model, has been proposed by Riley et al. [58,59]. Our model makes a parametric assumption about the distribution of the biomarker. We used a log-logistic assumption, but other parametric assumptions are possible as well.

Regarding the sensitivity analysis, we found no impact of the reference standard on the results of the multiple

thresholds model, which is in line with our earlier results using the bivariate model [12]. However, there were significant differences when FE_{NO} < 25 ppb between the groups of children and adults. It must be kept in mind that the children's group was dominated by a single study that contributed a large number of cutoff values [41].

5. Conclusion

The multiple thresholds model allowed to generate a more complete and more clinical useful picture of the

effects of different thresholds, which alleviates the derivation of decision rules for diagnosing or excluding asthma. However, a confirmatory study using the identified FE_{NO} values as predefined thresholds would be necessary to validate these findings. A prerequisite for summarizing DTA studies with the multiple thresholds model is that authors of primary studies report the pairs of sensitivities and specificities of all thresholds they examined of the index test or, even more preferably, the biomarker values of all individual patients along with their true disease status according to the reference standard. If reviewers encounter primary studies that report findings for multiple thresholds in sufficient detail, they should consider the multiple thresholds model or related methods. This would allow a more comprehensive interpretation of the clinical value of diagnostic tests producing continuous results.

Acknowledgments

Authors' contributions: A.S., K.L., and G.R. had the idea to apply the multiple thresholds model of S.S. and G.R. to the data of the systematic review regarding the diagnostic accuracy of FE_{NO} . G.R. calculated and generated the figures. S.S. helped with calculating, interpreting, and writing. J.B.R. contributed to the concept and writing. A.S. and K.L. wrote the first draft. A.S. and G.R. revised the manuscript. All authors revised it critically for important intellectual content and approved the final version.

Consent for publication: Not applicable.

Ethics approval and consent to participate: No ethics committee approval was needed for this systematic review.

Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.jclinepi.2017.09.001>.

References

- [1] National Institute of Health. Global Initiative for asthma - Global strategy for asthma management and Prevention (GINA), 2017. Available at www.ginasthma.com. Accessed September 27, 2017.
- [2] 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.
- [3] 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.
- [4] Ricciardolo FL, Sterk PJ, Gaston B, Folkerts G. Nitric oxide in health and disease of the respiratory system. *Physiol Rev* 2004;84:731–65.
- [5] 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.
- [6] Corren J, Lemanske RF, Hanania NA, Korenblat PE, Parsey MV, Arron JR, et al. Lebrizumab treatment in adults with asthma. *N Engl J Med* 2011;365:1088–98.
- [7] Haldar P, Brightling CE, Hargadon B, Gupta S, Monteiro W, Sousa A, et al. Mepolizumab and exacerbations of refractory eosinophilic asthma. *N Engl J Med* 2009;360:973–84.
- [8] 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;108:830–41.
- [9] 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;17:97–103.
- [10] 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.
- [11] Schneider A, Schwarzbach J, Faderl B, Welker L, Karsch-Volk M, Jörres RA. FENO measurement and sputum analysis for diagnosing asthma in clinical practice. *Respir Med* 2013;107:209–16.
- [12] 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.
- [13] Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration; 2010. Available at <http://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/Chapter%2010%20-%20Version%201.0.pdf>. Accessed September 27, 2017.
- [14] Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwiderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
- [15] Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59:1331–2.
- [16] Steinhäuser S, Schumacher M, Rucker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol* 2016;16:97.
- [17] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2017. Available at <https://www.R-project.org/>. Accessed September 27, 2017.
- [18] Arora R, Thornblade CE, Dauby PA, Flanagan JW, Bush AC, Hagan LL. Exhaled nitric oxide levels in military recruits with new onset asthma. *Allergy Asthma Proc* 2006;27:493–8.
- [19] Cordeiro D, Rudolphus A, Snoey E, Braunstahl GJ. Utility of nitric oxide for the diagnosis of asthma in an allergy clinic population. *Allergy Asthma Proc* 2011;32:119–26.
- [20] ElHalawani SM, Ly NT, Mahon RT, Amundson DE. Exhaled nitric oxide as a predictor of exercise-induced bronchoconstriction. *Chest* 2003;124:639–43.
- [21] Florentin A, Acouetey DS, Remen T, Penven E, Thaon I, Zmirou-Navier D, et al. Exhaled nitric oxide and screening for occupational asthma in two at-risk sectors: bakery and hairdressing. *Int J Tuberc Lung Dis* 2014;18:744–50.
- [22] Fortuna AM, Feixas T, Gonzalez M, Casan P. Diagnostic utility of inflammatory biomarkers in asthma: exhaled nitric oxide and induced sputum eosinophil count. *Respir Med* 2007;101:2416–21.
- [23] Fukuhara A, Saito J, Sato S, Sato Y, Nikaïdo T, Saito K, et al. Validation study of asthma screening criteria based on subjective symptoms and fractional exhaled nitric oxide. *Ann Allergy Asthma Immunol* 2011;107:480–6.
- [24] Giovannini M, Valli M, Ribuffo V, Cappiello G, Businarolo E, Andreani A. Relationship between Methacholine Challenge Testing and exhaled nitric oxide in adult patients with suspected bronchial asthma. *Eur Ann Allergy Clin Immunol* 2014;46:109–13.

- [25] Heffler E, Guida G, Marsico P, Bergia R, Bommarito L, Ferrero N. Exhaled nitric oxide as a diagnostic test for asthma in rhinitic patients with asthmatic symptoms. *Respir Med* 2006;100:1981–7.
- [26] Katsoulis K, Ganavias L, Michailopoulos P, Bikas C, Dinapogias E, Kontakiotis T, et al. Exhaled nitric oxide as screening tool in subjects with suspected asthma without reversibility. *Int Arch Allergy Immunol* 2013;162:58–64.
- [27] Kostikas K, Papaioannou AI, Tanou K, Bikas C, Dinapogias E, Kontakiotis T. Portable exhaled nitric oxide as a screening tool for asthma in young adults during pollen season. *Chest* 2008;133:906–13.
- [28] Kowal K, Bodzenta-Lukaszyk A, Zukowski S. Exhaled nitric oxide in evaluation of young adults with chronic cough. *J Asthma* 2009;46:692–8.
- [29] Linkosalo L, Lehtimäki L, Holm K, Kaila M, Moilanen E. Relation of bronchial and alveolar nitric oxide to exercise-induced bronchoconstriction in atopic children and adolescents. *Pediatr Allergy Immunol* 2012;23:360–6.
- [30] Malinowski A, Backer V, Harving H, Porsbjerg C. The value of exhaled nitric oxide to identify asthma in smoking patients with asthma-like symptoms. *Respir Med* 2012;106:794–801.
- [31] Pedrosa M, Cancelliere N, Barranco P, López-Carrasco V, Quirce S. Usefulness of exhaled nitric oxide for diagnosing asthma. *J Asthma* 2010;47:817–21.
- [32] Pizzimenti S, Heffler E, Piccioni P, Bugiani M, Migliore E, Guida G, et al. Usefulness of exhaled nitric oxide (FeNO) measured by a portable analyzer to diagnose cough variant asthma in a clinical setting of chronic cough. *Allergy* 2009;64:395.
- [33] Sato S, Saito J, Sato Y, Ishii T, Xintao W, Tanino Y, et al. Clinical usefulness of fractional exhaled nitric oxide for diagnosing prolonged cough. *Respir Med* 2008;102:1452–9.
- [34] Schleich FN, Asandei R, Manise M, Sele J, Seidel L, Louis R. Is FENO50 useful diagnostic tool in suspected asthma? *Int J Clin Pract* 2012;66:158–65.
- [35] Sivan Y, Gadish T, Fireman E, Soferman R. The use of exhaled nitric oxide in the diagnosis of asthma in school children. *J Pediatr* 2009;155:211–6.
- [36] Smith AD, Cowan JO, Filsell S, McLachlan C, Monti-Sheehan G, Jackson P, et al. Diagnosing asthma: comparisons between exhaled nitric oxide measurements and conventional tests. *Am J Respir Crit Care Med* 2004;169:473–8.
- [37] Smith AD, Cowan JO, Brassett KP, Filsell S, McLachlan C, Monti-Sheehan G, et al. Exhaled nitric oxide: a predictor of steroid response. *Am J Respir Crit Care Med* 2005;172:453–9.
- [38] Tilemann L, Gindner L, Meyer F, Szecsenyi J, Schneider A. Differences in local and systemic inflammatory markers in patients with obstructive airways disease. *Prim Care Respir J* 2011;20:407–14.
- [39] Voutilainen M, Malmberg LP, Vasankari T, Haahtela T. Exhaled nitric oxide indicates poorly athlete's asthma. *Clin Respir J* 2013;7:347–53.
- [40] Wang Y, Li L, Han R, Lei W, Li Z, Li K, et al. Diagnostic value and influencing factors of fractional exhaled nitric oxide in suspected asthma patients. *Int J Clin Exp Pathol* 2015;8:5570–6.
- [41] Woo SI, Lee JH, Kim H, Kang JW, Sun YH, Hahn YS. Utility of fractional exhaled nitric oxide (F(E)NO) measurements in diagnosing asthma. *Respir Med* 2012;106:1103–9.
- [42] Zhang YM, Lin JT, Su N, Liu GL, Yu HX, Nong Y, et al. [Values of fractional exhaled nitric oxide in the diagnosis of chronic cough]. *Chung Hua I Hsueh Tsa Chih* 2011;91:1254–8.
- [43] Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865–84.
- [44] Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding how physicians reason. *Science* 1959;130:9–21.
- [45] Schneider A, Wagenpfeil G, Jorres 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.
- [46] 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.
- [47] Martin MJ, Wilson E, Gerrard-Tarpey W, Meakin G, Hearson G, McKeever TM, et al. The utility of exhaled nitric oxide in patients with suspected asthma. *Thorax* 2016;71:562–4.
- [48] Minnaard MC, de Groot JA, Hopstaken RM, Schierenberg A, de Wit NJ, Reitsma JB, et al. The added value of C-reactive protein measurement in diagnosing pneumonia in primary care: a meta-analysis of individual patient data. *CMAJ* 2017;189:E56–63.
- [49] van Vugt SF, Broekhuizen BD, Lammens C, Zuithoff NP, de Jong PA, Coenen S, 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.
- [50] 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.
- [51] Porsbjerg C, Lund TK, Pedersen L, Backer V. Inflammatory subtypes in asthma are related to airway hyperresponsiveness to mannitol and exhaled NO. *J Asthma* 2009;46:606–12.
- [52] Aaron SD, Vandemheen KL, FitzGerald JM, Ainslie M, Gupta S, Lemiere C, et al. Reevaluation of diagnosis in adults with physician-diagnosed asthma. *JAMA* 2017;317:269–79.
- [53] Agustí A, Bel E, Thomas M, Vogelmeier C, Brusselle G, Holgate S, et al. Treatable traits: toward precision medicine of chronic airway diseases. *Eur Respir J* 2016;47:410–9.
- [54] Kuss O. Statistical methods for meta-analyses including information from studies without any events-add nothing to nothing and succeed nevertheless. *Stat Med* 2015;34:1097–116.
- [55] Putter H, Fiocco M, Stijnen T. Meta-analysis of diagnostic test accuracy studies with multiple thresholds using survival methods. *Biom J* 2010;52:95–110.
- [56] Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med Res Methodol* 2009;9:73.
- [57] Martínez-Cambor P. Fully non-parametric receiver operating characteristic curve estimation for random effects meta-analysis. *Stat Methods Med Res* 2017;26:5–20.
- [58] Riley RD, Takwoingi Y, Trikalinos T, Guha A, Biswas A, Ensor J, et al. Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *J Biomet Biostat* 2014;5:3.
- [59] Riley RD, Ahmed I, Ensor J, Takwoingi Y, Kirkham A, Morris RK, et al. Meta-analysis of test accuracy studies: an exploratory method for investigating the impact of missing thresholds. *Syst Rev* 2015;4:12.