

ORIGINAL ARTICLE

Accuracy of FE_{NO} for diagnosing asthma: a systematic review

Stefan Karrasch,^{1,2,3} Klaus Linde,¹ Gerta Rücker,^{4,5} Harriet Sommer,^{4,5} Marlies Karsch-Völk,¹ Jos Kleijnen,^{6,7} Rudolf A Jörres,³ Antonius Schneider¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2016-208704>).

For numbered affiliations see end of article.

Correspondence to

Professor Antonius Schneider, Institute of General Practice, Klinikum rechts der Isar, Technische Universität München, Orleansstraße 47, München 81667, Germany; antonius.schneider@tum.de

Received 29 March 2016

Accepted 2 June 2016

Published Online First

7 July 2016

ABSTRACT

Background Measurement of FE_{NO} might substitute bronchial provocation for diagnosing asthma. We aimed to investigate the diagnostic accuracy of FE_{NO} measurement compared with established reference standard.

Methods Systematic review and diagnostic meta-analysis. Data sources were Medline, Embase and Scopus up to 29 November 2015. Sensitivity and specificity were estimated using a bivariate model. Additionally, summary receiver-operating characteristic curves were estimated.

Results 26 studies with 4518 participants (median 113) were included. Risk of bias was considered low for six of seven items in five studies and for five items in seven studies. The overall sensitivity in the meta-analysis was 0.65 (95% CI 0.58 to 0.72), the overall specificity 0.82 (0.76 to 0.86), the diagnostic OR 9.23 (6.55 to 13.01) and the area under the curve 0.80 (0.77 to 0.85). In meta-regression analyses, higher cut-off values were associated with increasing specificity (OR 1.46 per 10 ppb increase in cut-off) while there was no association with sensitivity. Sensitivities varied significantly within the different FE_{NO} devices, but not specificities. Neither prevalence, age, use of bronchoprovocation in >90% of participants or as exclusive reference standard test, nor risk of bias were significantly associated with diagnostic accuracy.

Conclusions There appears to be a fair accuracy of FE_{NO} for making the diagnosis of asthma. The overall specificity was higher than sensitivity, which indicates a higher diagnostic potential for ruling in than for ruling out the diagnosis of asthma.

INTRODUCTION

Bronchial asthma is one of the most frequent chronic diseases. It is characterised by a chronic inflammatory process in the respiratory tract mostly concomitant with unspecific bronchial hyper-reactivity as well as reversible airflow obstruction.¹ In mild asthma in particular, airway obstruction is often not present during investigation by spirometry, thus leading to diagnostic uncertainty.² Serial peak-flow measurement or bronchial provocation are recommended in international guidelines for these cases.¹ However, the low diagnostic value of peak-flow variability has already been demonstrated.^{3–4} Thus, bronchial provocation for determining bronchial hyper-responsiveness remains a reference standard in case of inconclusive spirometric results.^{5–6} On the other hand, bronchial

Key messages**What is the key question?**

- Is the measurement of FE_{NO} sufficiently accurate compared with established reference standard to diagnose or exclude bronchial asthma?

What is the bottom line?

- The overall specificity was higher than sensitivity, which indicates a higher diagnostic potential for ruling in than for ruling out the diagnosis of asthma; thus, FE_{NO} measurement might render bronchial provocation partially superfluous.

Why read on?

- This systematic review and meta-analysis summarises the methods, risk of bias and findings of all currently available diagnostic studies relevant to the question.

provocations are time-consuming, cost-intensive, often only available in specific lung function laboratories and also bear the risk of heavy bronchospasm.⁵

When compared with that, measurement of FE_{NO} concentration in exhaled air is a non-invasive method, which is used to estimate inflammatory processes in the lung. Patients with asthma, even in milder stages of the disease, have been shown to exhale FE_{NO} in higher concentrations, in correlation with high expression levels of the inducible NO synthase in airway epithelium.⁷ The major pathophysiological basis is that NO modulates airway hyper-responsiveness⁸ and is associated with eosinophilic inflammation.⁹ Thus, the concentration of FE_{NO} is regarded as an indirect marker for the extent of airway inflammation.

Corresponding to this, FE_{NO} measurement is discussed as an alternative procedure to diagnose or exclude bronchial asthma. This would facilitate asthma diagnosis without referral for bronchial provocation, and, consequently, FE_{NO} measurement would be well suited for application especially in primary care. A health technology assessment (HTA) found that the inclusion of FE_{NO} measurement into the diagnostic pathway might increase the diagnostic cost-effectiveness.¹⁰ However, the



To cite: Karrasch S, Linde K, Rücker G, et al. *Thorax* 2017;**72**:109–116.

overall diagnostic accuracy of FE_{NO} for asthma is still unclear. The aim of our systematic review and meta-analysis was to estimate the diagnostic accuracy of FE_{NO} measurement in patients suspected to suffer from this disease.

METHODS

The review was registered in the PROSPERO international prospective register of systematic reviews (CRD42014010810; full protocol provided in online supplementary material). The ideal studies for our review would be prospective studies recruiting consecutive, undiagnosed, mainly (>90%) steroid-naïve patients with symptoms suggestive of asthma in a natural setting (eg, patients referred by general practitioners for diagnostic testing of asthma), not restricted to highly specific patient groups (eg, bakers), testing FE_{NO} measurements according to current guidelines with well-defined cut-off values against adequate reference standards. However, as such studies were likely to be rare, we used more liberal inclusion criteria for our overall review. To be eligible, studies had to allow the generation of 2×2 tables for asthma diagnosis by FE_{NO} compared with a reference standard. The reference standard could be bronchial provocation, 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. Studies on patients with prediagnosed asthma or in populations in which >25% already had undergone a trial of inhaled corticosteroids were excluded as we expected too many a priori diagnosed patients in these studies (selection bias). 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 international guidelines.¹¹ Studies not explicitly stating the exhalation flow rate were included if the technical properties of the used instruments assured that a result could only be gained with a flow rate within the stated range.

A systematic literature search was executed over the databases Medline, Embase and Scopus (last update search 29 November 2015). Furthermore, reference lists of identified papers and systematic reviews were searched for eligible papers. For searches in Medline and Embase, the interface supplied by Ovid (Wolters Kluwer) was applied. The main Medline search from 30 September 2014 is shown in the online supplementary table S1. Search terms for asthma or asthma indication were combined with terms for the index test. For Scopus, the offered functionality generating secondary documents based on the primary list of sources was additionally applied. Methodological search filters to more specifically identify diagnostic accuracy studies were not used.¹²

After de-duplication, two reviewers independently screened titles and abstracts of search hits and excluded all studies that clearly did not fit the study question. The remaining references were retrieved in full text and examined for eligibility according to the selection criteria described above. The following relevant information was extracted from included studies by two reviewers independently using a pre-tested form: setting, asthma prevalence, study type and recruitment strategy, number of study centres, type, brand and model of NO analyses, adherence to measurement guidelines; number of participants, age, symptoms raising the suspicion of asthma, inclusion of smokers or patients with recent respiratory tract infection; cut-off levels; prevalence of asthma; and reported findings. If the numbers of true positive, true negative, false positive and true negative subjects were reported, these values were used. Whenever possible, we used data for the cut-off point with the highest sum of

sensitivity and specificity (ie, the Youden index); for two studies, we had to use data for a cut-off chosen by the authors referring to previous studies, and for three the reasons choosing the cut-off presented were not clear. Otherwise, 2×2 tables were generated from the reported diagnostic accuracy parameter, that is, sensitivity and specificity, prevalence or predictive values. All studies included were assessed for quality using the Quadas-2 tool¹³ (seven items).

Based on the extracted 2×2 tables (a=true positive, b=false positive, c=false negative, d=true negative), sensitivity ($a/(a+c)$), specificity ($d/(b+d)$) and diagnostic odds ratio (DOR) ($a/b/c/d$) with corresponding 95% CIs were calculated for the reported cut-off of each study. For further statistical analysis, the bivariate random-effects model by Reitsma *et al*¹⁴ for meta-analysis of diagnostic accuracy studies was applied using the original form of normal-normal approximation.^{15–16} Sensitivity and specificity were pooled using the 'reitsma' function of the R package 'mada' (R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. <http://www.R-project.org/>; accessed 29 March 2016). Additionally, summary receiver-operator characteristic (sROC) curves were estimated. To this end, both the Rutter and Gatsonis approach¹⁶ and the Rucker and Schumacher approach¹⁷ were applied. The latter is based on the assumption that the investigators of the primary studies have chosen the optimal cut-off for each study based on maximum Youden index (sensitivity+specificity –1). The curve adjusts for the resulting cut-off optimisation. Calculation of pooled DORs was based on the random-effects model with variance estimation according to DerSimonian-Laird. The heterogeneity of the DOR was measured by Cochran's Q and Higgins's I².¹⁸ To examine small study effects as an indicator of publication bias, a funnel plot according to Deeks *et al*¹⁹ was generated. We performed an additional best evidence analysis limited to 'key studies' meeting the criteria listed above for ideal studies. In meta-regression analyses, we investigated the influence of asthma prevalence, type (chemoluminescence or electrochemical) and brand (Niox Mino, Niox Flex or other) of the FE_{NO} measurement device, cut-off values, age group, whether or not >90% of all study participants underwent bronchial provocation with a biochemical agent, whether or not a positive bronchial provocation test was the only mean to diagnose asthma and risk of bias. Variables for meta-regression analyses were partly chosen after qualitative review of the available studies.

RESULTS

The literature search identified a total of 6070 references of which 287 were retrieved in full text (figure 1). In total, 234 of these full texts were clearly irrelevant while further 53 were excluded after closer scrutiny (see online supplementary table S2). Also, 26 studies with 4518 participants (median 113, range 30–923) were included in the review (see characteristics of studies in table 1 and additional information on participants in online supplementary table S3).^{20–45} Two studies only presented results for participants divided into subgroups (never smokers, ex-smoker and current smoker in one study and depending on the specific reference standard used in another); therefore, the 26 studies provided 29 contingency tables for data analysis. Studies were published between 2003 and 2015 and originated from 16 different countries. Apart from one case-control study (nested in a cohort), all were cohort studies. Three studies exclusively included children, all other studies included adults or predominantly adults (one study without information on age). Only a few studies were performed in patient groups with distinct symptoms, like cough

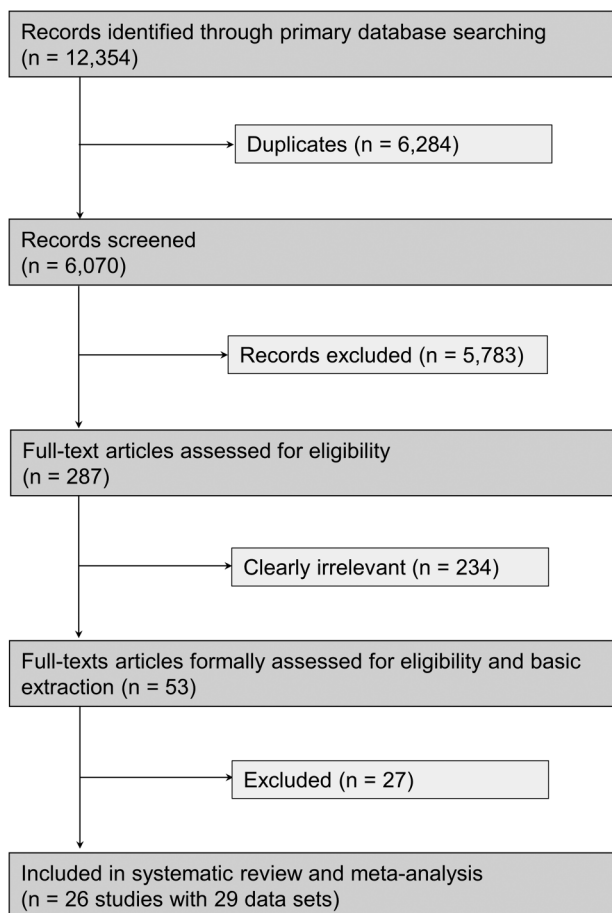


Figure 1 Flow chart.

variant asthma or wheezing children (see online supplementary table S3). In most studies, participants had been referred to the specialised centres performing the study, but details of the referral process were not always described in detail. The reference standard used varied considerably; bronchial provocation and combinations of different procedures were used frequently. The asthma prevalence determined varied between 9% and 80% (median 39%; 25th and 75th percentile 32% and 51%). In 17 studies, all or >90% of participants underwent a bronchial provocation test. In seven of these studies, the diagnosis of asthma was exclusively based on this test while all other studies used a combination of tests or stepped approaches. Cut-off levels used for FE_{NO} for determining the diagnosis of asthma varied between 10.5 and 64 ppb (median 30 ppb; 25th and 75th percentile 20 and 40 ppb). The most frequently used devices for measuring FE_{NO} were Niox Mino (12 studies) and Niox Flex (5 studies). Nine studies met the criteria of 'key studies'; the most frequent reasons for excluding other studies from this group were doubts that studies were naturalistic and recruitment strictly consecutive avoiding exclusions of relevant patient groups.

Study quality was variable but no study was considered to have low risk of bias in all seven items assessed (see online supplementary table S4 for detailed assessments of all studies). This was mainly due to the fact that the cut-off level for FE_{NO} measurement was defined post hoc in almost all studies (which can lead to overestimation of diagnostic accuracy). Many studies were rated as having unclear risk of bias for items V (Could the reference standard, its conduct or its interpretation have introduced bias?) and VII (Could the patient flow have introduced bias?) due to

insufficient reporting. Problems with other items were minor. Overall risk of bias was considered low for six of seven items in five studies, for five items in seven studies and for four items in eight studies. For the remaining six studies, the risk of bias was considered low in three or fewer items.

In individual studies, calculated sensitivity varied between 0.16 and 0.94, and specificity between 0.31 and 0.98 (see figure 2A, B). Pooled sensitivity was 0.65 (95% CI 0.58 to 0.72), pooled specificity was 0.82 (0.76 to 0.86) and the pooled DOR was 9.23 (6.55 to 13.01). Heterogeneity of the DOR was considerable ($Q=112$, $df=28$ ($p<0.0001$), $I^2=75%$ (64%, 83%)). The area under the curve according to Rutter–Gatsonis was 0.80 (0.77 to 0.85). Results are shown in an ROC scatterplot in figure 3. All included studies are plotted individually as well as the pooled estimate from the bivariate model including the corresponding 95% confidence area and the 95% prediction area, where 95% of future studies are expected. Furthermore, sROC curves according to Rutter and Gatsonis as well as Rucker and Schumacher are shown. The funnel plot over the included studies (see online supplementary figure S1) did not suggest relevant publication bias ($p=0.71$) for the Egger test modified according to Deeks *et al.*¹⁹ In the nine key studies, the pooled sensitivity was 0.68 (0.53 to 0.79), specificity 0.83 (0.74 to 0.89), the pooled DOR 10.21 (4.78 to 21.82) and the area under the curve 0.83 (0.75 to 0.92) (see online supplementary figure S2); heterogeneity remained considerable ($Q=51$, $df=8$ ($p<0.0001$), $I^2=84%$ (72%, 91%)).

In meta-regression analyses, higher cut-off values were associated with increasing specificity (OR 1.46 per 10 ppb increase in cut-off) while there was no association with sensitivity (see online supplementary table S5). Neither prevalence, age, use of bronchial provocation in >90% of participants or as an exclusive reference standard test, nor risk of bias were significantly associated with diagnostic accuracy. However, sensitivities varied within the different FE_{NO} devices, with Niox devices showing lower sensitivities compared with the other chemoluminescence devices ($p<0.01$). The resulting sensitivities and specificities of the different FE_{NO} devices are given in table 2. Multivariable regression analysis with prevalence, choice of cut-off points and FE_{NO} devices as independent variables provided similar results.

DISCUSSION

In the present systematic review, we found a fair diagnostic accuracy of FE_{NO} in general for discrimination of asthma in patients suspected to suffer from asthma. We received a good summary area under the curve. Specificities were mostly higher than sensitivities, and the observed cut-off points were predominantly in the range reported in guidelines regarding FE_{NO} interpretation.⁴⁶

Because the bivariate model is based on pairs of sensitivity and specificity, not directly on the cut-offs, it does not allow to identify ideal cut-off points for medical decision-making. However, the results of our meta-analysis allow a better estimation of the diagnostic usefulness of FE_{NO} measurement. Values of specificity were observed to be often superior to those of sensitivity. This suggests that FE_{NO} measurement is more suitable for ruling in than for ruling out the disease. Fitting to this, if the summary sensitivity and specificity values found in the present meta-analysis are applied to the median asthma prevalence over all included studies (0.39), the resulting positive predictive value (PPV) is 0.70. Such a PPV is well comparable to those of established bronchial provocation testing procedures.^{5 47} The mediocre specificity of bronchial provocation might result because of postinfectious bronchial hyper-responsiveness, gastro-oesophageal reflux disease, allergic rhinitis and many

Table 1 Characteristics of included studies

First author	Year	Country	Age group (range years or mean and SD)	n	Prevalence	FE _{NO} measurement device	Cut-off level (ppb)	Reference standard	Key study	True positive*	False negative*	False positive*	True negative*
Arora ²⁰	2006	USA	Adults (17–38)	172	0.80	NIOX—no further specification	17	Bronchial provocation	Yes	87	51	14	20
Cordeiro ²¹	2011	Netherlands	Mainly adults (7–87)	114	0.37	NIOX Flex (c)	27	FEV ₁ /VC+reversibility and bronchial provocation	Yes	33	9	6	66
ElHalawani ²²	2003	USA	Adults (18–40)	49	0.14	Sievers 280 (c)	12	Exercise challenge	No	7	0†	29	13
Florentin ²³	2014	France	Young adults (mean 25 SD 3)	178	0.11	NIOX Mino (e)	10.5	FEV ₁ /VC+reversibility and PEF variability	No	13	6	70	89
Fortuna ²⁴	2007	Spain	Adults (18–64)	50	0.44	SIR N-6008 (c)	20	Bronchial provocation	Yes	17	5	10	18
Fukuhara ²⁵	2011	Japan	Adults (48–66)	61	0.69	NA623N (c)	40	FEV ₁ /VC (+reversibility), bronchial provocation, sputum eosinophilia	No	33	9	2	17
Giovannini ²⁶	2014	Italy	Mainly adults (mean 38 SD 15)	42	0.50	HypAir FENO (e)	30	Bronchial provocation	No	3	18	0†	21
Heffler ²⁷	2006	Italy	Mainly adults (11–75)	48	0.38	NIOX Flex (c)	36	FEV ₁ /VC+reversibility and bronchial provocation	No	14	4	12	18
Katsoulis ²⁸	2013	Greece	Adults (22–37)	112	0.43	NIOX Mino (e)	32	Bronchial provocation	No	23	25	10	54
Kostikas ²⁹	2008	Greece	Young adults (mean 21 SD 2)	149	0.42	NIOX Mino (e)	19	FEV ₁ /VC+reversibility, bronchial provocation	No	33	30	13	73
Kowal ³⁰	2009	Poland	Adults (18–45)	540	0.33	Sievers 280i (c)	40	FEV ₁ /VC+reversibility, bronchial provocation, PEF variability, expert's opinion	No	157	21	63	299
Linkosalo ³¹	2012	Finland	Children (6–19)	30	0.60	Sievers 280 (c)	20	Free running test	No	13	5	2	10
Malinowski ³²													
Never smokers	2012	Denmark	Mainly adults (14–44)	108	0.42	NIOX Mino (e)	15	Expert's assessment based on FEV ₁ /VC +reversibility, bronchial provocation and medication use	No	18	14	14	66
Ex-smokers				62	0.31		22		No	12	7	6	37
Current smokers				112	0.29		17		No	35	10	23	40
Pedrosa ³³	2010	Spain	Adults (mean 34 SD 13)	114	0.31	NIOX Mino (e)	40	Bronchial provocation	No	26	9	22	57
Pizzimenti ³⁴	2009	Italy	Not stated (probably adults)	156	0.09	NIOX Mino (e)	55	Bronchial challenge, spirometry	No	11	3	17	125
Sato ³⁵	2008	Japan	Adults (20–78)	71	0.68	Kimoto, Osaka, Japan (c)	38.8	FEV ₁ /VC (+reversibility) and/or bronchial provocation, symptoms	Yes	38	10	2	21
Schleich ³⁶	2012	Belgium	Adults (mean 41 SD 16)	174	0.47	NIOX Flex (c)	34	Bronchial provocation	No	29	53	4	88
Schneider ³⁷	2013	Germany	Mainly adults (mean 43)	393	0.39	NIOX Mino (e)	25	FEV ₁ /VC+reversibility and bronchial provocation	Yes	75	79	60	179
Sivan ³⁸	2009	Israel	Children (5–18)	113	0.61	Eco Physics CLD88 (c)	19	Expert assessment based on FEV ₁ /VC (+reversibility) and/or bronchial provocation, medication	Yes	59	10	5	39
Smith ³⁹	2004	New Zealand	Mainly adults (9–72)	47	0.36	Logan LR2000 (c)	20	FEV ₁ /VC+reversibility, bronchial provocation	Yes	15	2	6	24
Smith ⁴⁰	2005	New Zealand	Mainly adults (14–71)	52	0.52	NIOX Flex (c)	47	FEV ₁ /FVC reversibility, meta-choline challenge, ICS response	No	15	12	2	23
Tilemann ⁴¹	2011	Germany	Adults (38–58)	210	0.41	NIOX Mino (e)	46	Bronchial provocation or reversibility	Yes	25	61	10	114
Voutilainen ⁴²	2013	Finland	Adolescents and adults (14–31)	87	0.34	NIOX Flex (c)	30	Expert's assessment based on FEV ₁ /VC + reversibility, bronchial provocation, PEF variability	No	13	17	6	51

Continued

Table 1 Continued

First author	Year	Country	Age group (range years or mean and SD)	n	Prevalence	FE _{NO} measurement device	Cut-off level (ppb)	Reference standard	Key study	True positive*	False negative*	False positive*	True negative*
Wang ⁴³ –Bronchodilatation –Bronchial provocation	2015	China	Mainly adults (13–89)	515 408	0.35 0.31	Nano Coulomb NO analyser, Shangwo (u)	41 64	Depending on FEV ₁ , predicted positive bronchodilatation or bronchial provocation	No	134 65	51 60	83 16	247 327
Woo ⁴⁴	2012	Korea	Children (8–16)	245	0.68	NIOX Mino (e)	21	FEV ₁ /VC ₊ reversibility and bronchial provocation	No	95	72	10	68
Zhang ⁴⁵	2011	China	Adults (15–70)	106	0.37	NIOX Mino (e)	40	FEV ₁ with bronchodilator, bronchial challenge test	Yes	29	10	9	58

*Values are absolute frequencies.

†In studies (data sets) in which a cell in the contingency table was empty, a continuity correction of 0.5 was used for data analyses. c, chemoluminescence; e, electrochemical; ICS, inhaled corticoid steroids; PEF, peak expiratory flow rate; reversibility, positive reaction to a bronchodilator; u, unclear; VC, vital capacity.

more, thus leading to false positive diagnoses.⁵ It can therefore be assumed that FE_{NO} measurement might render bronchial provocation partially superfluous. In this context, for the guidance of therapy the prediction of steroid responsiveness might be more relevant than solely making the diagnosis of asthma. It has been shown that FE_{NO} measurement has some prognostic value for assessing steroid responsiveness.^{40–48} Fitting to this, the American Thoracic Society (ATS) guideline recommends that FE_{NO} values >50 ppb can be used to indicate eosinophilic inflammation and that, in symptomatic patients, responsiveness to corticosteroids is likely.⁴⁶ Originally, we also wanted to evaluate whether FE_{NO} has in particular an added value in subgroups of patients, for instance, with wheezing of chronic cough. However, we identified only a few studies that did not allow pooling of data for further subgroup analyses.

The reported FE_{NO} cut-off values were generally at the lower end of the range defined as ‘intermediate’ by the ATS guideline⁴⁶ for adults (25–50 ppb) or children (20–35 ppb). Even considering the typically lognormal distribution of FE_{NO} values, the lower threshold of this intermediate range might be too high for a reliable exclusion of asthma. This view is supported by findings that suggested lower cut-off values ranging between 9 and 16 ppb for exclusion of asthma.^{37–49} This lower performance for ruling out asthma might be explained by a weakness of FE_{NO} regarding non-eosinophilic inflammation.^{37–50} However, recent studies suggest that FE_{NO} is more representative of a Th2-driven local inflammation, specifically of the bronchial mucosa, rather than general eosinophilic inflammation.^{51–53} This might explain why exclusion of asthma could be still possible, but only at low FE_{NO} values, which might be around 9–16 ppb. However, the difficulty to rule out non-eosinophilic asthma should be kept in mind when the presence of asthma is suggested by symptoms and clinical history but FE_{NO} is low. Very low FE_{NO} values might indicate the absence of airway inflammatory processes, with a negative predictive value of around 80%, provided this fits into the patient’s medical history.⁵⁴

In clinical practice consequently, the diagnostic pathway could start with FE_{NO} measurement at first. Bronchial provocation would be superfluous when FE_{NO} exceeds a distinct cut-off value, which needs to provide a meaningful PPV. Referral for bronchial provocation would be necessary when FE_{NO} values are lower. A positive response during bronchial provocation would help to rule in the diagnosis of asthma in patients with low FE_{NO} values, and a negative bronchial provocation response would rule out asthma.^{5–47} In the light of previous research, it remains unclear to which extent FE_{NO} measurement is more promising than bronchial provocation to guide therapy with inhaled corticosteroids. A recent study suggests that also some patients with negative bronchial provocation response might benefit from inhaled corticoid steroids (ICS) when FE_{NO}>33 ppb because FE_{NO} measurement turned out to be more useful for predicting ICS responsiveness than diagnosing asthma.⁵⁵ However, there might be some overestimation of the predictive value as the authors state that the chosen response criteria were likely to be ‘oversensitive’. Therefore, further research has to compare the diagnostic usefulness of bronchial provocation and FE_{NO} measurement, and to evaluate the informative value of their combination, ideally in long-term studies.

Regression analyses showed that the diagnostic accuracy in terms of sensitivities differed within the different FE_{NO} devices. It might be speculated if the lower sensitivities of Niox instruments compared with the other chemoluminescence devices are

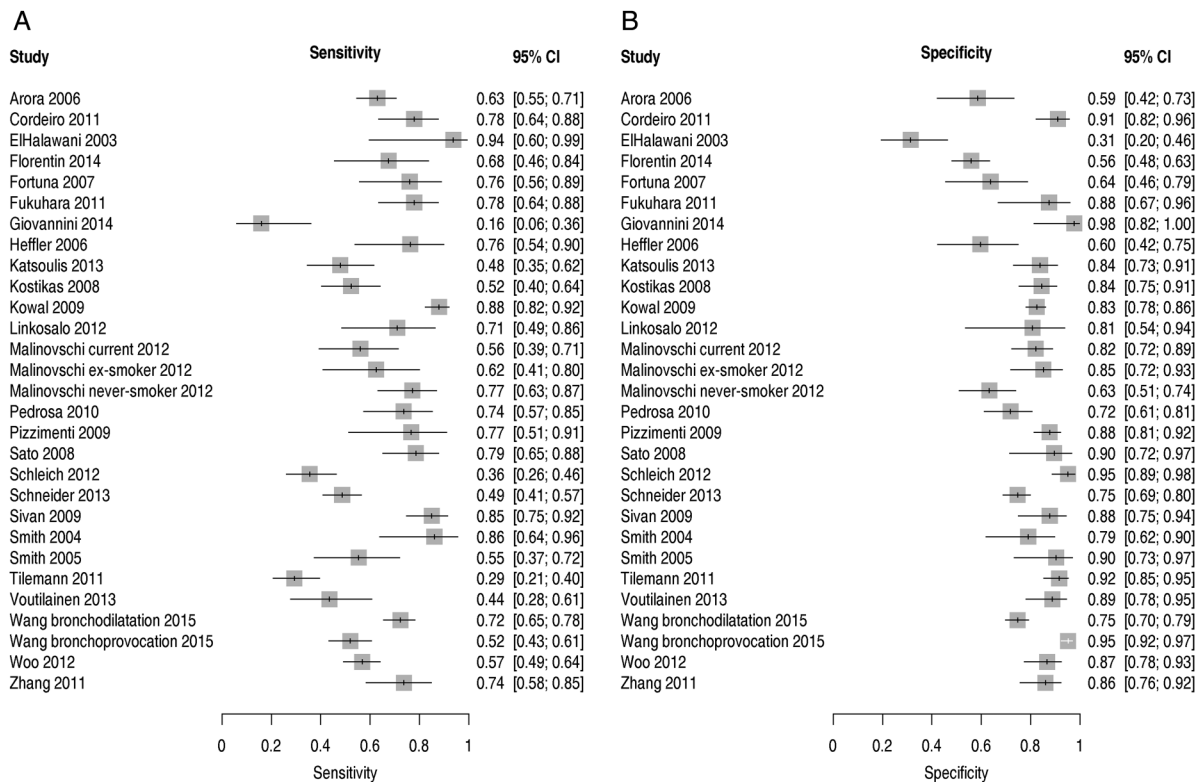


Figure 2 Sensitivity (A) and specificity (B) in included studies.

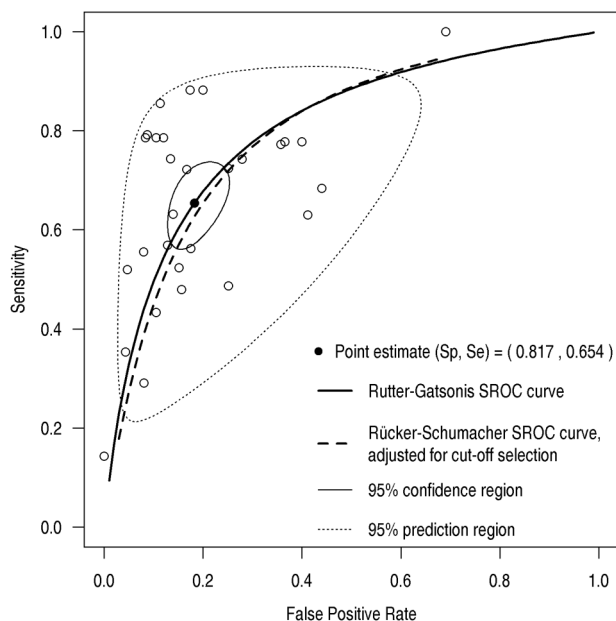


Figure 3 Pooled sensitivity (Se) and specificity (Sp) as well as estimated summary receiver-operator characteristic (sROC) curves based on calculation according to Rutter and Gatsonis¹⁶ (continuous line) and Rücker and Schumacher¹⁷ (dashed line). The 95% confidence region (black solid line) as well as the 95% prediction region (black dotted line) are given.

explainable by the different clinical settings of the diagnostic studies. However, we found no relationship between prevalence, cut-off points and FE_{NO} devices in the multivariable regression analysis, which might be, on the other hand, explainable by the low number of studies. However, this has rather low relevance

as specificities were shown to be of higher importance. Altogether, the different devices showed comparatively high specificities, which increased in meta-regression analysis with increasing cut-off points. This once again indicates that FE_{NO} measurement is more suitable for ruling in than ruling out the disease.

Our systematic review and meta-analysis provides a comprehensive summary of the currently available studies investigating the diagnostic accuracy of FE_{NO} in patients with symptoms suggestive of asthma. It is based on a protocol predefining methods and was performed following the recommendations of the Cochrane Diagnostic Test Accuracy Working Group.⁵⁶ Our study selection differs to some extent from two similar recent reviews.^{10 57} The careful systematic review performed as a part of a Diagnostic Assessment Report commissioned by the National Institute for Health Research (NIHR) HTA Programme on behalf of the National Institute for Health and Care Excellence (NICE) by Harnan *et al*¹⁰ included 24 studies, of which we excluded 6, mainly because of highly selected patient populations (eg, only patients with negative metacholine challenge), investigating prediction of steroid responsiveness rather than diagnosis of asthma, and insufficient data reported for meta-analysis (see online supplementary table S1 for details). Seven of eight additional studies included by us were either published only recently or not identified; one study³² was excluded. It should be noted that the HTA report focuses less on estimating diagnostic accuracy per se but focused on processes compatible with usual clinical pathways in the UK. The authors of the report decided to not conduct meta-analysis because of the strong clinical heterogeneity of the studies. While we think that the pooled estimates in our meta-analysis must be interpreted with considerable caution due to clinical and statistical heterogeneity, we consider meta-analysis desirable and justifiable for

Table 2 Meta-regression for FE_{NO} devices

FE _{NO} device (measurement technique)	n	Sensitivity (95% CI)	Specificity (95% CI)
Niox Flex (chemoluminescence)	12	0.58 (0.27 to 0.83)	0.88 (0.59 to 0.97)
Other chemoluminescence	8	0.84 (0.76 to 0.90)	0.77 (0.65 to 0.86)
Niox Mino (electrochemical)	5	0.59 (0.31 to 0.82)	0.81 (0.51 to 0.94)
Other electrochemical	1	0.16 (0.02 to 0.63)	0.98 (0.48 to 1.00)

n, number of data sets in analysis (two studies/three data sets without sufficient information).

several reasons. First, some of the heterogeneity of the studies included by Harnan *et al* is due to their in some aspects wider inclusion criteria. Second, the narrative review leaves the reader without an intuitive idea about the overall findings available. Third, our approach did allow us to investigate potential reasons for heterogeneity between study findings empirically by meta-regression. Fourth, the very restrictive approach using solely key studies showed similar results as the main analysis. However, heterogeneity remained high also in the key study set.

The recent systematic review by Li *et al*⁵⁷ actually also includes meta-analysis and meta-regression analyses. However, this review might have important shortcomings. In total, 6 of the 19 studies analysed were excluded by us due to highly selected populations, unusual techniques for FE_{NO} measurement or because we were unable to construct a plausible and consistent contingency table from the data presented. No less than 12 studies included by us were not mentioned. While one study published in 2015 might not have been available to the authors, the reason for this large discrepancy remains unclear as the selection process is not transparent. Pooled sensitivity in this review was higher than in our review (0.78 vs 0.65), specificity slightly lower (0.74 vs 0.82) and the DOR slightly higher (11.37 vs 9.23).

The significant association between increasing cut-off points and increasing specificities in meta-regression analysis is plausible and in line with the clinical decision rule, which allows ruling in asthma with higher certainty when higher FE_{NO} values are given.^{46 54} We are uncertain whether the differences according to type and brand of FE_{NO} measurement device are a valid or a chance finding. On the other hand, given the limited number of studies and the multiplicity of potential influencing variables, existing differences according to other factors might have been missed. Furthermore, factors like infection, smoking habits and allergy that might modify results of FE_{NO} testing⁵⁸ are difficult to investigate without access to individual patient data.

It should also be taken into account that the included studies mostly defined the optimal cut-off levels based on the observations made in the particular patient populations. This probably led to an overestimation of the diagnostic accuracy per study and consequently in the meta-analysis. We used the alternative approach according to Rücker and Schumacher¹⁷ for calculating the sROC curve that implies a correction for this problem. The resulting values were well within the confidence area of the classic calculatory approach by Rutter and Gatsonis.¹⁶

In conclusion, the systematic review and meta-analysis showed promising test indices of FE_{NO} measurement with good specificity, DORs and ROC area under the curve. The high specificity indicates a diagnostic potential for ruling in asthma, notwithstanding sensitivity and prevalence of disease are also important. However, sensitivity is comparatively low, suggesting that ruling out might be rather difficult with FE_{NO}

measurement. Despite the impossibility to provide definite cut-points with optimal sensitivities and specificities, our results point towards the necessity of a confirmatory diagnostic study. With this respect, a diagnostic study in terms of a triage test in a fully paired study design with an a priori defined cut-off value would be necessary to evaluate the potential for partial replacement of bronchial provocation.⁵⁹ It should be kept in mind that higher cut-off values seem to be more suitable for ruling in asthma. This means that the ideal a priori cut-off point should not be determined solely on the basis of the highest sum of sensitivity and specificity, but rather under consideration of an optimal PPV, which might correspond with higher specificity but lower sensitivity. Studies have shown a PPV > 70% when a cut-off >45 ppb is chosen.² The ATS guideline⁴⁶ suggests FE_{NO}>50 ppb to detect ICS responsiveness by referring to the study of Smith *et al*.⁴⁰ Therefore, a cut-point around 50 ppb might guarantee for a sufficient PPV for ruling in asthma and to determine ICS responsiveness at the same time, with subgroup analysis to keep possibly lower values in mind. Future diagnostic accuracy studies should optimally report their full ROC curve numerically to allow a more in-depth use in meta-analyses.

Author affiliations

¹Institute of General Practice, University Hospital Klinikum rechts der Isar, Technische Universität München, Munich, Germany

²Institute of Epidemiology I, Helmholtz Zentrum München—German Research Center for Environmental Health, Munich, Germany

³Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany

⁴Institute for Medical Biometry and Statistics, Medical Center—University of Freiburg, Freiburg, Germany

⁵Faculty of Medicine, University of Freiburg, Freiburg, Germany

⁶Kleijnen Systematic Reviews Ltd, Escrick, York, UK

⁷School for Public Health and Primary Care (CAPRI), Maastricht University, Maastricht, The Netherlands

Contributors AS conceived the study and wrote the review protocol. GR, JK, RAJ, SK and KL contributed to the protocol development. SK, MK-V and KL performed the literature search, selection, data extraction quality assessment and data entry. HS and GR planned and performed statistical analyses. AS, KL and SK drafted the manuscript. All authors contributed to data interpretation and critical revision of the draft manuscript. AS is the guarantor.

Funding The project was supported by the German Federal Ministry of Education and Research (BMBF FKZ 01KG1211).

Competing interests SK declares that he uses a NIOX Vero testing device provided by Aerocrine without charge within a study outside the submitted work.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The complete data set for the analyses performed is available as an MS Excel file in the online supplemental material.

REFERENCES

- GINA. Global Initiative for Asthma. Global strategy for asthma management and prevention (updated 2015). <http://ginasthma.org/gina-report-global-strategy-for-asthma-management-and-prevention/> (accessed 29 Mar 2016).
- Schneider A, Gindner L, Tilemann L, *et al*. Diagnostic accuracy of spirometry in primary care. *BMC Pulm Med* 2009;9:31.
- den Otter JJ, Reijnen GM, van den Bosch WJ, *et al*. Testing bronchial hyper-responsiveness: provocation or peak expiratory flow variability? *Br J Gen Pract* 1997;47:487–92.
- Tilemann L, Gindner L, Meyer FJ, *et al*. [Diagnostic value of peak flow variability in patients with suspected diagnosis of bronchial asthma in general practice]. *Dtsch Med Wochenschr* 2009;134:2053–8.
- Crapo RO, Casaburi R, Coates AL, *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.
- Hunter CJ, Brightling CE, Woltmann G, *et al*. A comparison of the validity of different diagnostic tests in adults with asthma. *Chest* 2002;121:1051–7.
- Lane C, Knight D, Burgess S, *et al*. Epithelial inducible nitric oxide synthase activity is the major determinant of nitric oxide concentration in exhaled breath. *Thorax* 2004;59:757–60.

- 8 Ricciardolo FL. Multiple roles of nitric oxide in the airways. *Thorax* 2003;58:175–82.
- 9 Jatakanon A, Lim S, Kharitonov SA, et al. Correlation between exhaled nitric oxide, sputum eosinophils, and methacholine responsiveness in patients with mild asthma. *Thorax* 1998;53:91–5.
- 10 Harnan SE, Tappenden P, Essat M, et al. Measurement of exhaled nitric oxide concentration in asthma: a systematic review and economic evaluation of NIOX MINO, NIOX VERO and NObreath. *Health Technol Assess* 2015;19:1–330.
- 11 American Thoracic S, European Respiratory S. 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.
- 12 Leeflang MM, Scholten RJ, Rutjes AW, et al. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *J Clin Epidemiol* 2006;59:234–40.
- 13 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- 14 Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
- 15 Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239–51.
- 16 Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865–84.
- 17 Rucker G, Schumacher M. Summary ROC curve based on a weighted Youden index for selecting an optimal cutpoint in meta-analysis of diagnostic accuracy. *Stat Med* 2010;29:3069–78.
- 18 Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539–58.
- 19 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005;58:882–93.
- 20 Arora R, Thornblade CE, Dauby PA, et al. Exhaled nitric oxide levels in military recruits with new onset asthma. *Allergy Asthma Proc* 2006;27:493–8.
- 21 Cordeiro D, Rudolphus A, Snoey E, et al. Utility of nitric oxide for the diagnosis of asthma in an allergy clinic population. *Allergy Asthma Proc* 2011;32:119–26.
- 22 ElHalawani SM, Ly NT, Mahon RT, et al. Exhaled nitric oxide as a predictor of exercise-induced bronchoconstriction. *Chest* 2003;124:639–43.
- 23 Florentin A, Acouetey DS, Remen T, 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.
- 24 Fortuna AM, Feixas T, González M, et al. Diagnostic utility of inflammatory biomarkers in asthma: exhaled nitric oxide and induced sputum eosinophil count. *Respir Med* 2007;101:2416–21.
- 25 Fukuhara A, Saito J, Sato S, 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.
- 26 Giovannini M, Valli M, Ribuffo V, et al. 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.
- 27 Heffler E, Guida G, Marsico P, et al. Exhaled nitric oxide as a diagnostic test for asthma in rhinitic patients with asthmatic symptoms. *Respir Med* 2006;100:1981–7.
- 28 Katsoulis K, Ganavias L, Michailopoulos P, et al. Exhaled nitric oxide as screening tool in subjects with suspected asthma without reversibility. *Int Arch Allergy Immunol* 2013;162:58–64.
- 29 Kostikas K, Papaioannou AI, Tanou K, et al. Portable exhaled nitric oxide as a screening tool for asthma in young adults during pollen season. *Chest* 2008;133:906–13.
- 30 Kowal K, Bodzenta-Lukaszyk A, Zukowski S. Exhaled nitric oxide in evaluation of young adults with chronic cough. *J Asthma* 2009;46:692–8.
- 31 Linkosalo L, Lehtimäki L, Holm K, et al. Relation of bronchial and alveolar nitric oxide to exercise-induced bronchoconstriction in atopic children and adolescents. *Pediatr Allergy Immunol* 2012;23:360–6.
- 32 Malinowski A, Backer V, Harving H, et al. The value of exhaled nitric oxide to identify asthma in smoking patients with asthma-like symptoms. *Respir Med* 2012;106:794–801.
- 33 Pedrosa M, Cancelliere N, Barranco PL, et al. Usefulness of exhaled nitric oxide for diagnosing asthma. *J Asthma* 2010;47:817–21.
- 34 Pizzimenti S, Heffler E, Piccioni P, 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;64:395–5.
- 35 Sato S, Saito J, Sato Y, et al. Clinical usefulness of fractional exhaled nitric oxide for diagnosing prolonged cough. *Respir Med* 2008;102:1452–9.
- 36 Schleich FN, Asandei R, Manise M, et al. Is FENO50 useful diagnostic tool in suspected asthma?. *Int J Clin Pract* 2012;66:158–65.
- 37 Schneider A, Schwarzbach J, Faderl B, et al. FENO measurement and sputum analysis for diagnosing asthma in clinical practice. *Respir Med* 2013;107:209–16.
- 38 Sivan Y, Gadish T, Fireman E, et al. The use of exhaled nitric oxide in the diagnosis of asthma in school children. *J Pediatr* 2009;155:211–16.
- 39 Smith AD, Cowan JO, Filsell S, et al. Diagnosing asthma: comparisons between exhaled nitric oxide measurements and conventional tests. *Am J Respir Crit Care Med* 2004;169:473–8.
- 40 Smith AD, Cowan JO, Brassett KP, et al. Exhaled nitric oxide: a predictor of steroid response. *Am J Respir Crit Care Med* 2005;172:453–9.
- 41 Tilemann L, Gindner L, Meyer F, et al. Differences in local and systemic inflammatory markers in patients with obstructive airways disease. *Prim Care Respir J* 2011;20:407–14.
- 42 Voutilainen M, Malmberg LP, Vasankari T, et al. Exhaled nitric oxide indicates poorly athlete's asthma. *Clin Respir J* 2013;7:347–53.
- 43 Wang Y, Li L, Han R, 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.
- 44 Woo SI, Lee JH, Kim H, et al. Utility of fractional exhaled nitric oxide (FENO) measurements in diagnosing asthma. *Respir Med* 2012;106:1103–9.
- 45 Zhang YM, Lin JT, Su N, et al. [Values of fractional exhaled nitric oxide in the diagnosis of chronic cough]. *Chung Hua I Hsueh Tsa Chih* 2011;91:1254–8.
- 46 Dweik RA, Boggs PB, Erzurum SC, 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.
- 47 Schneider A, Schwarzbach J, Faderl B, et al. Whole-Body Plethysmography in Suspected Asthma: A Prospective Study of Its Added Diagnostic Value in 302 Patients. *Dtsch Arztebl Int* 2015;112:405–11.
- 48 van der Schee MP, Palmay R, Cowan JO, et al. Predicting steroid responsiveness in patients with asthma using exhaled breath profiling. *Clin Exp Allergy* 2013;43:1217–25.
- 49 Schneider A, Tilemann L, Schermer T, et al. Diagnosing asthma in general practice with portable exhaled nitric oxide measurement—results of a prospective diagnostic study: FENO<or = 16 ppb better than FENO<or = 12 ppb to rule out mild and moderate to severe asthma [added]. *Respir Res* 2009;10:15.
- 50 Porsbjerg C, Lund TK, Pedersen L, et al. Inflammatory subtypes in asthma are related to airway hyperresponsiveness to mannitol and exhaled NO. *J Asthma* 2009;46:606–12.
- 51 Corren J, Lemanske RF, Hanania NA, et al. Lebrikizumab treatment in adults with asthma. *N Engl J Med* 2011;365:1088–98.
- 52 Haldar P, Brightling CE, Hargadon B, et al. Mepolizumab and exacerbations of refractory eosinophilic asthma. *N Engl J Med* 2009;360:973–84.
- 53 Bjermer L, Alving K, Diamant Z, et al. Current evidence and future research needs for FeNO measurement in respiratory diseases. *Respir Med* 2014;108:830–41.
- 54 Schneider A, Wagenpfeil G, Jörres RA, et al. 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.
- 55 Martin MJ, Wilson E, Gerrard-Tarpey W, et al. The utility of exhaled nitric oxide in patients with suspected asthma. *Thorax* 2016;71:562–4.
- 56 Reitsma JB, Rutjes AWS, Whiting P, et al. Chapter 9: assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 100*. The Cochrane Collaboration, 2009. <http://methods.cochrane.org/sdt/handbook-dta-reviews> (accessed 29 Mar 2016).
- 57 Li Z, Qin W, Li L, et al. Diagnostic accuracy of exhaled nitric oxide in asthma: a meta-analysis of 4,691 participants. *Int J Clin Exp Med* 2015;8:8516–24.
- 58 Karrasch S, Ernst K, Behr J, et al. Exhaled nitric oxide and influencing factors in a random population sample. *Respir Med* 2011;105:713–18.
- 59 Bossuyt PM, Irwig L, Craig J, et al. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006;332:1089–92.