

Exhaled nitric oxide in the diagnosis of asthma in adults: a systematic review

S. E. Harnan¹ , M. Essat¹, T. Gomersall¹, P. Tappenden¹, I. Pavord², M. Everard³ and R. Lawson⁴

¹Health Economics and Decision Science, SchARR, University of Sheffield, Sheffield, ²Respiratory Medicine Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK, ³School of Paediatrics & Child Health, Princess Margaret Hospital, University of Western Australia, Crawley, WA, Australia and ⁴Department of Respiratory Medicine, Royal Hallamshire Hospital, Sheffield, UK

Clinical & Experimental Allergy

Summary

Objectives To identify and synthesize evidence on the diagnostic accuracy of FE_NO for asthma in adults.

Materials and Methods Systematic searches (nine key biomedical databases and trial registers) were carried out on November 2014. Records were included if they recruited patients with the symptoms of asthma; used a single set of inclusion criteria; measured FE_NO_{50} in accordance with American Thoracic Society guidelines, 2005 (off-line excluded); reported/allowed calculation of true-positive, true-negative, false-positive and false-negative patients as classified against any reference standard. Study quality was assessed using QUADAS II. Meta-analysis was planned where clinical study heterogeneity allowed. Rule-in and rule-out uses of FE_NO were considered.

Results A total of 4861 records were identified originally and 1312 in an update. Twenty-seven studies were included. Heterogeneity precluded meta-analysis. Results varied even within subgroups of studies. Cut-off values for the best sum of sensitivity and specificity varied from 12 to 55 p.p.b., but did not produce high accuracy. 100% sensitivity or 100% specificity was reported by some studies indicating potential use as a rule-in or rule-out strategy. **Conclusions and Clinical Relevance** FE_NO_{50} had variable diagnostic accuracy even within subgroups of studies with similar characteristics. Diagnostic accuracy, optimal cut-off values and best position for FE_NO_{50} within a pathway remain poorly evidenced.

Keywords asthma, diagnosis, exhaled nitric oxide, meta-analysis, systematic review

Submitted 20 April 2016; revised 24 October 2016; accepted 1 November 2016

Correspondence:

Sue E. Harnan, Health Economics and Decision Science, SchARR, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK.

E-mail: s.harnan@sheffield.ac.uk

Cite this as: S. E. Harnan, M. Essat, T. Gomersall, P. Tappenden, I. Pavord, M. Everard and R. Lawson, *Clinical & Experimental Allergy*, 2017 (47) 410–429.

Introduction

The fraction of nitric oxide in exhaled breath (FE_NO) is elevated in some patients with asthma as a result of interleukin (IL)-13-induced induction of nitric oxide synthase in airway epithelium. Elevated values are independent of allergy [1], but are associated with eosinophilic airway inflammation and with responsiveness to treatment with inhaled corticosteroids (ICS) [2, 3], and may therefore be a useful test in the diagnosis and management of asthma. The availability of hand-held monitors improves the practicality and affordability of including a FE_NO test in diagnostic pathways within primary care. However, the role for FE_NO within such pathways is currently unclear. As FE_NO is largely a marker of type-2 cytokine-mediated inflammation, it is unlikely to be useful as an absolute test for asthma as not all asthmatics have this feature.

Some current guidelines and commentaries suggest its best use may be to identify those patients who will respond to ICS therapy [4–6]. Indeed, there is increasing interest in the concept of diagnosing steroid-responsive disease as its own classification across what are currently defined as distinct airway disorders [i.e. asthma, chronic obstructive pulmonary disease (COPD)]. However, to date, national and international guidelines and diagnostic pathways [4, 5, 7–9] do not aim to diagnose steroid-responsive disease according to a classification, but aim to diagnose asthma according to its classical definition. As such, the focus of this work was on the use of FE_NO for the diagnosis of asthma, with data relating to its use to identify steroid-responsive disease included *de facto*.

Currently, there is no clear consensus relating to the cut-off between normal and abnormal FE_NO_{50} levels, or

whether FE_{NO}₅₀ should be used as a rule-in test, a rule-out test or both. In clinical practice, choosing different cut-off levels will alter sensitivity, specificity and positive and negative predictive values (NPVs). As we reviewed the evidence, it became clear that study designs varied greatly in terms of populations recruited and reference standards used; as sensitivity and specificity can theoretically vary depending on the characteristics of the population recruited, and as there is no definitive reference standard for asthma (and thus different reference standards will also affect sensitivity and specificity), the evidence base needed careful consideration. We have attempted to order the evidence sensibly to account for these variations and to enable readers to focus on the data that most closely match their own clinical scenario. However, the complexity of the evidence base makes both the presentation of results and the making of overall recommendations challenging, and highlights the need for critical thought in clinical applications.

Whilst there are many expert reviews relating to the use of FE_{NO} in the diagnosis of asthma [10–16], we are only aware of three systematic reviews. Two were only briefly reported [5, 17], and one had only a limited scope (FE_{NO} to diagnose exercise-induced bronchoconstriction; EIB) [18]. In comparison, our publication is the first to systematically search for and codify the evidence base, and as such provides a clear and comprehensive framework on which future research can be based. Further, through our synthesis, we aimed to qualitatively consider the question of where in the diagnostic pathway FE_{NO} may best be placed, and what cut-off points may be best for its use as a rule-in and rule-out test.

Materials and methods

Context

We were commissioned to conduct a systematic review of the evidence relating to the use of FE_{NO} in both asthma management and diagnosis as part of the National Institute of Health and Care Excellence (NICE) Diagnostic Appraisal Programme. The appraisal focussed specifically on the hand-held monitors NIOX MINO, NIOX VERO and NObreath, but we reviewed evidence produced using any on-line FE_{NO} monitoring device, including chemiluminescent and electrochemical devices. We have published two reviews from this project elsewhere, one on FE_{NO} for management in adults [19] and one in children [20]. We have also published an Health Technology Assessment (HTA) monograph, which included a review of the comparability of the devices, the two reviews on management, two reviews on diagnosis (in children and adults) and a cost-

effectiveness analysis [21]. In that report, we narrowed the evidence base to studies with relevance to UK practice. Here, we report an update on that work, and we do not restrict our inclusion criteria by relevance to UK practice, leading to a much wider evidence base. Notably, NICE produced draft guidelines in 2015 [22], and an update in 2016 [23], based on a separate piece of work with a different search strategy and different inclusion criteria.

Figure 1 presents some possible locations for FE_{NO} within a simplified diagnostic pathway (based on the UK BTS/SIGN guidelines) [24], denoted by letters A to E. In each position, FE_{NO} could be used as a rule-in test (high specificity) or a rule-out test (high sensitivity). We have used this schema to order our findings and to consider the clinical implications of different positions. One potentially useful position is to prevent referral of at least a proportion of patients to airway hyperresponsiveness (AHR), an unpleasant and costly test (positions C&D) with significant, and in rare cases severe, side-effects. Equally, it could be used as described elsewhere [4, 5] to identify patients who should respond to ICS therapy (position B) and avoid 'false negatives' when using a trial of treatment, for example by identifying patients who may need higher doses of ICS to achieve a response, or who may not have complied with the trial of treatment. Of course, a test may have high specificity, but very poor sensitivity, meaning many patients with disease will be missed by a rule-in scenario, unless further tests are performed; similarly, highly sensitive tests may have poor specificity and in a rule-out scenario may fail to eliminate some subjects. For FE_{NO}, the selection of the cut-off point dictates the balance between true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The downstream costs and consequences of each of these can in turn also affect the clinical utility and where best the balance between sensitivity and specificity lies.

Identifying the literature

Systematic searches were carried out between March 2013 and April 2013. Update searches were conducted in September 2013 and again in November 2014. The full strategy and an example from one bibliographic database are provided in Appendix S1, but in brief comprised searches of MEDLINE and MEDLINE In Process; EMBASE; Cochrane Library; Science Citation Index Expanded (SCIE); and Conference Proceedings Citation Index – Science (CPCI-S). Trial registers [ClinicalTrials.gov; metaRegister of Controlled Trials; FDA Manufacturer and User Facility Device (MAUDE); and EuroScan International Network] were also searched in March 2013 and for both updates. The search comprised subject heading and free-text terms for NIOX MINO, NObreath

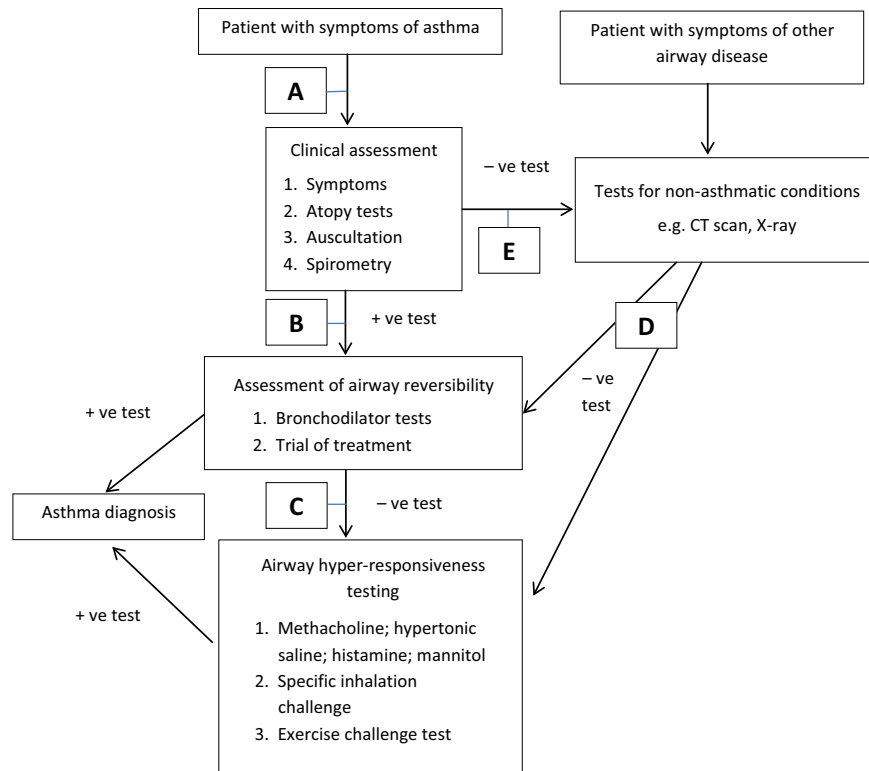


Fig. 1. Generic asthma diagnostic pathway. Based on the BTS/SIGN guidelines (2012) [24]. Additional tests used internationally include whole-body plethysmography, sputum eosinophils. A, B, C, D: potential positions for the use of FeNO in a diagnostic pathway. +ve, test suggestive of an asthma diagnosis; -ve, test not suggestive of an asthma diagnosis. 'Test' refers to the tests listed at that point in the pathway.

and FeNO , combined with terms for asthma and lower respiratory tract symptoms. Additional searches for studies relating to NIOX VERO were conducted in August 2013, and terms relating to NIOX VERO included in the update searches. randomised controlled trial (RCT) and diagnostic study filters were applied.

Study selection

Studies were considered for inclusion in the review by one reviewer against the inclusion criteria summarized here, and provided in detail in the Appendix S2. Studies were included if they recruited participants presenting with the symptoms of asthma or reported a subgroup of such patients. FeNO_{50} could be measured by any device (chemiluminescent or electrochemical), used in accordance with the ATS 2005 criteria [flow rate of 50 mL/s, exhalation time ≥ 10 s (adults)/ ≥ 6 s (children/adolescents)]. Off-line measurement techniques were excluded. Studies using a reference standard of any established diagnostic test or set of tests were eligible for inclusion and as such studies vs. ICS responsiveness were included, allowing an exploration of FeNO_{50} for the identification of steroid-responsive disease. End-to-end studies (studies that recruit patients before diagnosis and follow them through to long-term clinical

outcomes such as overall survival and quality of life, which are considered the highest quality primary research study design) and observational cohort studies as defined by the NICE methods guide [25] were eligible for inclusion. In the event, no end-to-end studies were found, and the review included only observational clinical validity (diagnostic accuracy) studies. These could be prospective or retrospective, cross-sectional or cohort designs, with one set of inclusion criteria (i.e. not case-control style studies), and could be either derivation (identification of cut-off value that best fits the data) or validation (cut-off value pre-specified) studies. Studies were included if they reported data that allowed the extraction or calculation of the numbers of patients who were TP, TN, FP and FN against the reference standard. Any clinical setting was acceptable. Conference abstracts were included where they provided sufficient data relating to population, reference standard, cut-off value used and diagnostic accuracy.

Data extraction

A data extraction form was developed following the guidelines given in the Centre for Reviews and Dissemination (CRD) Handbook [26] and the Cochrane Handbook [27] and piloted on two studies. A full list of

extracted fields are provided in Appendix S3 and included data relating to study characteristics, patient characteristics, reference standard, index tests and numerical data relating to diagnostic accuracy. Data were extracted from the studies by one of three reviewers and checked by a second reviewer. Discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. Authors were contacted for key missing or unclear data. Data from multiple publications relating to the same group of patients were extracted and quality was assessed as a single study. Separate quality assessment was performed for subgroup analyses or use of a different reference standard within a study.

Quality assessment

Diagnostic accuracy studies were assessed using Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS 2) [28], adapted to the specifics of the review (Appendix S4). All four of the QUADAS 2 risk of bias domains were assessed, namely patient spectrum (consecutive or random sample; case-control design avoided; inappropriate exclusions avoided); the conduct of the index test (blinded interpretation; pre-specified cut-off); the conduct of the reference standard test (appropriateness of the reference standard; blinded interpretation); and flow and timing of the study (time between index test and reference standard (to assess whether the underlying condition might have changed between tests, such that a different diagnosis would be made); same reference standard applied to all patients; inclusion of all patients in analysis). Each domain is given a summary score of high risk, low risk or unclear risk of bias. Questions relating to applicability were omitted as these were addressed through subgrouping of studies according to the patient spectrum and reference standard used. Quality assessment was conducted by one reviewer and checked by a second. A third reviewer was consulted in cases of disagreement.

Synthesis and analysis

A narrative synthesis was conducted and a hierarchical meta-analysis was planned where sufficient studies of acceptable clinical heterogeneity were available. Sources of clinical heterogeneity considered to be of importance included age, patient spectrum as indicated by inclusion criteria and baseline measurements [mean FE_{NO}₅₀, mean percentage of predicted forced expiratory volume in the first second (FEV1%), atopic status, smoking status], FE_{NO} measurement device used and reference standard used. As already described, studies were subgrouped according to the patient spectrum (position in Fig. 1) and reference standard used as both

factors can affect estimates of accuracy. For simplicity, bronchodilator reversibility testing (BDR) and trials of treatment with ICS were grouped together as airway reversibility tests (ARTs), with the test used specified in Table 2. We have focussed on three sets of data, each set comprising cut-off in parts per billion (p.p.b.), sensitivity, specificity, positive predictive value (PPV) and NPV, and each of which characterized a different potential use of the test. These were (i) the cut-off with the highest sum of sensitivity and specificity, which is the most commonly reported accuracy statistic, but is only useful if the test is to be interpreted absolutely (i.e. a positive test diagnoses asthma; a negative test indicates no asthma); (ii) the cut-off with the highest sensitivity reported, which indicates its potential value as a rule-out test where a negative test rules out asthma and a positive test leads to further tests for asthma; and (iii) the cut-off with the highest specificity reported, which indicates its potential value as a rule-in test, where a positive test diagnoses asthma and a negative test leads to further tests for asthma (see Appendix S5 for details). Studies that did not report a full range of cut-offs were not ideal for assessing the use of FE_{NO}₅₀ as a rule-in or rule-out test, as the cut-offs with the highest sensitivity or specificity were not always reported, but were included in the analysis.

Results

A total of 4861 citations were identified by the initial search and 1312 by the updates. After exclusion of irrelevant titles and abstracts, the full texts of 430 citations were obtained, and of these, 27 studies (35 citations) were included in the review (Fig. 2).

Study characteristics

Of the 27 studies included in the review, 15 studies (19 citations) were conducted in adults [29–47], four studies (six citations) in adults plus adolescents [3, 48–52], three studies (three citations) in all age groups [53–55] and five studies (seven citations) in an unspecified age range [56–62].

Key study and patient characteristics are presented in Table 1. Mean ages ranged from 37 to 49 years, except four studies recruited younger cohorts [29, 30, 50, 52] and two recruited older cohorts [33, 38]. Gender ranged from 33% male to 85%. One study excluded atopic patients [30], one recruited only atopic patients [40], and all others ranged widely in between or did not report this characteristic. FEV1% was often not reported, but ranged from 89% to 113% where it was. Mean FE_{NO}₅₀ at baseline was also often not reported, with the lowest mean at 15.4 p.p.b. [34] and the highest at 74.5 p.p.b. [31]. Most studies used chemiluminescent analysers ($n = 14$) [3, 29–31, 33–36, 40, 42, 43, 48, 49,

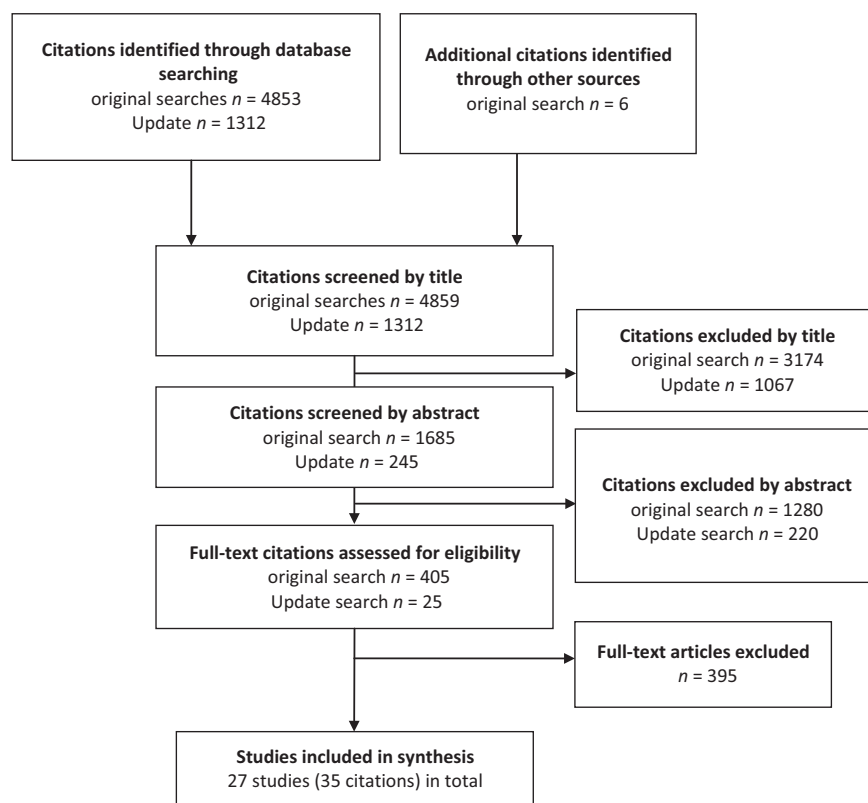


Fig. 2. PRISMA flow diagram (adapted from <http://www.prisma-statement.org/>) for the review of diagnostic accuracy studies.

54, 61], whilst nine studies used NIOX MINO (a handheld electrochemical device) [37, 41, 44–47, 50, 52, 55, 56, 59] and four studies did not report which device was used [38, 39, 53, 60, 62]. Twelve studies did not specify whether corticosteroids were used by patients prior to or during testing [34, 36, 38, 39, 48, 50–52, 55–62], whilst corticosteroid use was either stopped (three studies) [33, 41, 44–47] or patients who used them were excluded (11 studies) [3, 29–32, 37, 40, 42, 43, 53, 54]. One study had a mix of users and non-users [35]. Due to these differences across all study and patient characteristics, clinical heterogeneity was judged to be high and a meta-analysis was not performed. The cut-offs used to diagnose asthma varied, but were most commonly within the 20–40 p.p.b. range. Some studies reported sensitivity and specificity for multiple cut-off points, whilst others either pre-specified the value or selected the value with the best sum of sensitivity and specificity (best sum cut-off).

Diagnostic accuracy results are presented in Table 2 in summary and in full in Appendix S7. Twelve subgroups were identified (grouped by position in the pathway and reference standard used), and for each, methodological quality (Fig. 3) according to QUADAS 2 [28] and estimates of cut-offs and accuracy are described below.

Position A: Asthma symptoms, no previous tests

(i) *Reference standard of diagnostic pathway that included AHR and ART:* There were seven studies in this category [3, 31–33, 44–46, 53, 54]. Methodological quality was mixed (Fig. 3), with no study scoring particularly well. The index test domain scored high risk most often, as many studies did not use a pre-defined cut-off. Blinding of the index test scored well as it was usually performed before the reference standard. The reference standard domain and the flow and timing domain usually scored poorly due to poor reporting.

Cut-off values were pre-specified in two studies [31–33] (≥ 20 and 40 p.p.b.). Derived values for the best sum similarly ranged from 20 to 47 p.p.b. There was no pattern as to whether sensitivity was higher than specificity amongst any of these studies. A partial range of cut-off values was reported by three studies [3, 44–46], but a full range by only two (see Appendix S7) [44–46]. In these, rule-out scenario cut-off values were 9 and 12 p.p.b., respectively, with sensitivities of 96% (specificity of 13%) and 85% (specificity of 24%). Rule-in scenario cut-off values were 71 and 76 p.p.b., with specificities of 97% (sensitivity of 18%) and 100% (sensitivity of 13%), respectively, although specificity of

Table 1. Study and population characteristics

Study author (year), country	Population (adults unless otherwise stated)	Study design, N (analysed/ recruited)	Mean age in years (SD or range)	FEV1% (SD)	Mean FE _{NO} ₅₀ p.p.b. (SD)	Corticosteroid use during testing	Smokers	Atopy
Position A: Asthma symptoms, no previous tests								
(i) Reference standard of pathway including AHR and ART								
Cordeiro (2011) [54], the Netherlands	Asthma symptoms All ages	R 114/114 (100%)	Median (range) A: 39 (7–83); NA: 38 (7–87)	NR	Median (range) A: 44 (6–290); NA: 17 (5–45)	ICS or OCS in previous 6 weeks excluded	10%	71%
Fortuna (2007) [31, 32], Spain	Asthma symptoms	P, C 50/57 (88%)	A: 37 (18–68) NA: 38 (18–64)	A: 94 (19) NA: 99 (10)	A: 40 (31) NA: (23)	ICS or OCS in previous 4 weeks excluded	Current or Ex: 28%	NR
Fukuhara (2011) [33], Japan	Asthma symptoms	P 61/97 (63%)	55.6 (17–81)	96% (95% CI, 90.1–102.0)	74.5 (95% CI, 56.2–92.8)	ICS or OCS excluded	Current or Ex: 31%	23%
Schneider (2013) [46], Germany	Asthma symptoms	P, C 393/400 (98%)	43.2 (16.3)*	104.8 (17.1)*	29.9 (33.2)*	ICS use stopped 12 h before	Current: 10%; Ex: 35%	NR
Schneider (2009) [44, 45], Germany	Asthma symptoms	P, C 160/160 (100%)	43.9 (16.3)*	95.8 (20.0)*	31.5 (36.0)*	ICS use stopped 12 h before	Current and Ex: 54%	NR
Smith (2005) [3], New Zealand	Asthma symptoms Adults and adolescents	P, C 52/60 (87%)	40.5 (14–71)	97.8 (14.2)	Range 6.3–242.0 [48]	ICS or OCS in previous 4 weeks excluded	Current 6% Ex 19%	77% [48]
Smith (2004) [53], New Zealand	Asthma symptoms All ages	P, C 44/51 (86%)	A: 41.6 (9–72) NA: 31.8 (9–64)	102.9 (18.0)*	28.8 (28.6)*	ICS or OCS in previous 4 weeks excluded	Ex 11%	NR
(ii) Reference standard of AHR								
Backer (2014) [52], Denmark	Asthma symptoms 10–45 years old	R 141/348 (40.5%)	28.3 (9.8)	97%	NR	NR	Current: 4% Ex: 19%	90%
Giovannini (2014) [34], Italy	Asthma symptoms	NR 42/42 (100%)	38.17 (15.2)	NR	15.4 (15.8)	NR	Current: 19% Ex: 26.2%	50%
(iii) Reference standard of ART								
De La Barra (2011) [49], [†] New Zealand	Asthma symptoms Adults and adolescents	P, C 52/60 (87%)	40.5 (14–71)	97.8 (14.2)	Range 6.3–242.0 [48]	ICS or OCS in previous 4 weeks excluded	Current 6% Ex 19%	77%
(iv) Reference standard of pathway plus long-term follow-up								
Schneider (2014) [47], Germany	Asthma symptoms	P, C 302/400 (75.5%)	43.2 (16.3)*	104.8 (17.1)*	29.9 (33.2)*	ICS use stopped 12 h before	Current: 10%; Ex: 35%	NR
Position A: Asthma symptoms plus one additional symptom								
(v) Reference standard of pathway including AHR and ART								
Heffler (2006) [48], Italy	Rhinitis and asthma symptoms Adults and adolescents	P, C 48/48 (100%)	40.08 (NR)	89.2 (95% CI 80.1–98.4)	59.7 (95% CI 50.2–89)	NR	0	92%

(continued)

Table 1 (continued)

Study author (year), country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1% (SD)	Mean FEV_{50} p.p.b. (SD)	Corticosteroid use during testing	Smokers	Atopy
Position C: Negative art								
(vi) Reference standard of AHR Katsoulis (2013) [37], Greece	Negative ART	P 112/112 (100%)	Median 25 years (IQR 22–37)	Median 89% (IQR 83–99%)	Median 20.5 (IQR 12–34)	ICS users excluded	Recent ex-smokers excluded	45.5%
Position C or D: Referred for AHR								
(vii) Reference standard of AHR Brannan (2013) [61], Australia	Unclear, referred for mannitol challenge	R 401/401 (100%)	NR	NR	NR	NR	NR	NR
Position C or D: Referred for AHR								
Chancafe-Morgan (2013) [62], Spain	Unclear, referred for AHR testing in Spain	P, C 30/30 (100%)	44.2 (16.7)	NR	33.6 (18.7)	NR	NR	NR
Nickels (2014) [38, 39]	Adults at tertiary referral centre	R 1322/1322 (100%)	54.1 (15.5)	NR	NR	NR	NR	NR
Position E: Negative for clinical assessment (e.g. normal spirometry)								
(viii) Reference standard of AHR Schneider (2009) [44, 45], Germany	Normal spirometry	P, C 101/101 (100%)	NR	NR	NR	ICS use stopped 12 h before	NR	NR
Chronic cough								
Position D: Negative for other conditions (e.g. GORD medications ineffective, X-ray or CT scan clear)								
(ix) Reference standard of ART Hahn (2007) [35], USA	Normal chest radiograph	R 64/64 (100%)	46.8 (14.7)*	95.6 (9.0)*	41.0 (22.4)*	ICS not excluded	Current 0%; Ex 16%	NR
Hsu (2013) [36], Taiwan	No response to GORD or UACS	R 81/114 (71%)	49 (14)	91.8 (15.3)	Mean rank 47	NR	0%	NR
Prieto (2009) [40], Spain	Normal chest radiograph, CT scan, spirometry	P 43/43 (100%)	48 (95% CI 43–52)	113.2 (95% CI 108.0–118.3)	GM mean (95% CI): ICS responders 23.2 (17.5–30.7); non-responders 18.6 (14.7–24.0)	ICS or OCS excluded	0%	100%
(x) Reference standard including ART and AHR								
Sato (2008) [42], Japan	Normal chest radiograph or CT scan	P, C 71/71 (100%)	NR	NR	NR	ICS or OCS excluded	75%	NR
Zhang (2011) [56–58], China	Normal chest radiograph	U 106/106 (100%)	NR	NR	NR	NR	NR	NR

(continued)

Table 1 (continued)

Study author (year), country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1% (SD)	Mean FE _{NO} ₅₀ p.p.b. (SD)	Corticosteroid use during testing	Smokers	Atopy
(xi) Other studies								
Arora (2006) [29], USA	Mix of undiagnosed and diagnosed*	P 172/172 (100%)	20 (2.7)*	99.7 (13.6)*	27.8 (28.5)*	ICS or other anti-inflammatories excluded	0%	NR
Bobolea (2012) [55], Spain	Failed MCT All ages	P, C 30/30 (100%)	37.3 (13–69)	NR	NR	NR	NR	NR
El Halawani (2003) [30], USA	Suspected EIB	P, C 49/50 (98%)	27.9 (SD: NR)	NR	EIB +ve, 41; EIB –ve: 25.6	ICS or OCS excluded	0%	0%
Mathew (2011) [60], UK	Normal spirometry	P 84/84 (100%)	NR	NR	NR	NR	NR	NR
Pedrosa (2010) [50, 51], Spain	and negative ART Normal spirometry or negative ART	P, C 114/115 (99%)	34 (13)	104.29 (14.95)	34	NR	Current: 15% Ex: 10%	87%
Pizzimenti (2009) [59]	Adults and adolescents Chronic cough	P, C 156/156 (100%)	NR	NR	34.1 (95% CI: 28.5–39.5)	NR	9%	47%
Sastre (2013) [41], Spain	Unclear age group Suspected occupational asthma (various)	NR 68/68 (100%)	39.5 (10.3)	NR	Median 23 (IQR 15.0–32.9)	ICS stopped 1 week before	Current: 21% Ex: 11%	66.2%
Schleich (2012) [43], Belgium	Normal spirometry or negative ART	P, C 174/237 (73%)	41 (16)	97 (13)	Median 17 range (4–271)	ICS excluded	34%	48%

N, number; SD, standard deviation; FEV1%, forced expiratory volume in one second percent predicted; FE_{NO}₅₀, fraction of exhaled nitric oxide measured at a flow rate of 50 mL/s; p.p.b., parts per billion; R, retrospective data collection; P, prospective data collection; C, consecutive sample; A, asthmatics; NA, non-asthmatics; ART, airway obstruction reversibility testing; AHR, airway hyperresponsiveness; MCT, methacholine challenge test; CI, confidence interval; NR, not reported; GP, general practitioner; IQR, interquartile range; ICS, inhaled corticosteroid; EIB +ve, patients who tested positive for exercise-induced bronchoconstriction; EIB –ve, patients who tested negative for exercise-induced bronchoconstriction; GORD, gastro-oesophageal reflux disease.

*Reviewer calculated.

†Reanalysis of Smith et al. [3].

‡Army recruits, some of whom were thought to have lied about existing asthma diagnosis.

Table 2. Diagnostic accuracy data for the highest sum of sensitivity and specificity, the pre-specified cut-off, the highest sensitivity (rule-out) and the highest specificity (rule-in) reported in each study

Study author (year)	Device	Reference standard	Derived or pre- specified	Highest sum of sens and spec or pre-specified cut-off						Rule-out						Rule-in																			
				FE _{NO50} Cut-off (p.p.b.)						Cut- off						Cut- off																			
				Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV																
Position A: Asthma symptoms, no previous tests																																			
(i) Reference standard of pathway including AHR and ART																																			
Cordeiro (2011) [54]	Niox-Flex	ART (BDR), AHR	D	27	78	92	85	88	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR						
	Niox-Flex; airway reversibility	(histamine)	D	27	87	90	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR							
Fortuna (2007) [31, 32]	SIR N-6008, Madrid, Spain	GINA guidelines [spirometry, ART (BDR), AHR (MCT)]	P	≥ 20	77	64	63.0	78.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR							
Fukuhara (2011) [33]	NA623 (Chest MI, Tokyo, Japan)	At least two of: induced sputum eosinophilia, or ART plus exclusion of other lung diseases	P	40	78.6	89.5	94.3	65.4	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR							
Schneider (2013) [46]	NIOX MINO	Spirometry, ART (BDR), AHR (MCT)	D	25	49	75	56.0	69.5	9	96	13	41.6	83.8	71	18	97	79.5	64.6																	
Schneider (2009) [44, 45]	NIOX MINO	Spirometry, ART (BDR), AHR (MCT)	D	46	32	93	77.3	59.5	12	85	24	49.6	64.5	76	13	100	100	56.7																	
Smith (2005) [3]	Niox	ATS 1987 guidelines plus one of: ART (ICS), AHR (MCT).	D	47	55.6	92	88.2	65.7	15	81.5	48	29.4	37.1	47	55.6	92	88.2	65.7																	
Smith (2004) [53]	NR	ATS 1987 guidelines plus: AHR (saline) and/or ART (BDR)	D	20	88	79	70	91.7	NR	NR	NR	NR	NR	NR	NR	NR	NR																		
(ii) Reference standard of AHR				P	25	36.2	86.8	65.6	66.1	NR	NR	NR	NR	NR	NR	NR	NR																		
	NIOX MINO	AHR (mannitol)																																	

(continued)

Table 2 (continued)

Study author (year)	Device	Reference standard	Derived or pre- specified	Highest sum of sens and spec or pre-specified cut-off					Rule-out					Rule-in				
				FE _{NO} Cut-off (p.p.b.)					Cut- off					Cut- off				
				Sens	Spec	PPV	NPV		Sens	Spec	PPV	NPV		Sens	Spec	PPV	NPV	
Backer (2014) [52]	HypAir FeNO	AHR (MCT)	P	30	14.3	100	100	53.9	NR	NR	NR	NR	NR	NR	NR	NR	NR	
Giovannini (2014) [34]																		
(iii) Reference standard of ART																		
De La Barra (2011) [49]	Niox	ART (ICS)	D	41.7	NR	NR	NR	NR	25	83.3	57.5	37.0	92	110	25	95	60	80.9
(iv) Reference standard of pathway plus long-term follow-up																		
Schneider (2014) [47]	NIOX MINO	Spirometry, ART (BDR), AHR (MCT) and long-term follow-up of patients and GPs	D	26	47	73.1	39.8	78.4	NR	NR	NR	NR	NR	NR	NR	NR	NR	
Position A: Asthma symptoms plus one additional symptom.																		
(v) Reference standard of pathway including AHR and ART																		
Heffler (2006) [48]	Niox	AHR (MCT) or ART (BDR)	D	36	77.8	60	53.8	81.8	25	100	46.7	52.9	100	100	27.8	100	100	69.8
Position C: Negative art																		
(vi) Reference standard of AHR																		
Katsoulis (2013) [37]	NIOX MINO	AHR (MCT)	D* D [†]	32 26	47 55	85 85	70.1 NR	68.1 NR	10 10	81 90	39 10	49.9 NR	73.2 NR	30 30	49 48	82 85	67.1 NR	68.2 NR
Position C or D: Referred for AHR																		
(vii) Reference standard of AHR																		
Brannan (2013) [61]	HypAir	AHR (mannitol)	NR	47	30.2	96.3	65.7	85.5	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Chancafe- Morgan (2013) [62]	NR	AHR (MCT)	P	35	75	83.3	75	83.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Nickels (2014) [38, 39]	NR	AHR (MCT)	NR	NR	33.3	84.8	35.7	83.4	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Position E: Negative for clinical assessment (e.g. normal spirometry)																		
(viii) Reference standard of AHR																		
Schneider (2009) [44, 45]	NIOX MINO	AHR (MCT)	D	46	35	90	NR	NR	16	78	45	NR	NR	NR	NR	NR	NR	NR
(continued)																		

(continued)

Table 2 (continued)

Study author (year)	Device	Reference standard	Derived or pre- specified	Highest sum of sens and spec or pre-specified cut-off								Rule-out				Rule-in						
				FE _{NO} ₅₀ Cut-off (p.p.b.)								Cut- off				Cut- off						
				Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV			
Chronic cough																						
Position D: Negative for other conditions (e.g. GERD medications ineffective, X-ray or CT scan clear)																						
(ix) Reference standard of ART																						
Hahn (2007) [35]	280i Sievers	ART (ICS)	D (P also available)	38	90	85	89.5	84.6	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Hsu (2013) [36]	280i Sievers	ART (ICS)	D (P also available)	33.9	94.7	76.3	80	94	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Prieto (2009) [40]	Niox	ART (ICS)	D	20	53	63	52.6	62.5	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
(x) Reference standard including ART and AHR																						
Sato (2008) [42]	Chemiluminescence analyzer (Kimoto, Osaka, Japan)	Sputum eosinophilia, ART (BDR), AHR (MCT)	D	38.8	79.2	91.3	95.0	67.7	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Zhang (2011) [56–58]	NIOX MINO	Sputum cell counts, spirometry, AHR, 24-h oesophageal pH monitoring, SPT and serum IgE	D	40	75	86	76.3	85.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
(xi) Other studies																						
Aroa (2006) [29]	Niox	AHR (histamine)	D	> 17	63	58.8	86.1	28.2	> 6	96.4	0	79.6	0.00	> 46	16.7	100	100	22.8				
Bobolea (2012) [55]	NIOX MINO	AHR (adenosine)	P	30 [‡]	100	29.2	26	100	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
El Halawani (2003) [30]	Sievers 280A	Exercise challenge	D	12 [‡]	100	31	19.4	100	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Mathew (2011) [60]	NR	AHR (MCT)	P	NR	10	67.2	8.7	70.5	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Pedrosa (2010) [50, 51]	NIOX MINO	AHR (MCT)	D	40	74.3	72.5	54.1	86.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Pizzimenti (2009) [59]	NIOX MINO	AHR (methacholine)	D	55	10	67.2	39.3	97.7	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				
Sastre (2013) [41]	NIOX MINO		D	25	60	80	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR				

(continued)

Table 2 (continued)

Study author (year)	Device	Reference standard	Derived or pre-specified	Highest sum of sens and spec or pre-specified cut-off						Rule-out			Rule-in		
				FE _{NO} ₅₀ Cut-off (p.p.b.)	Sens	Spec	PPV	NPV	Cut-off	Sens	Spec	PPV	NPV	Cut-off	Sens
Schleich (2012) [43]	Niox	Specific inhalation challenge AHR (MCT)	D	34	35	95	87.8	62.4	NR	NR	NR	NR	NR	NR	NR
	Niox		D	34	24.4	98.9	NR	NR	NR	NR	NR	NR	NR	NR	NR
	FEV1 ≤ 101%														

AHR, airway hyperresponsiveness; ART, airway reversibility test; N, number analysed; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; FE_{NO}₅₀, fractional exhaled nitric oxide measured at a flow rate of 50 mL/s; FEV1, forced expiratory volume in first second; FVC, forced vital capacity; BDR, bronchodilator reversibility test; MCT, methacholine challenge test; ICS, inhaled corticosteroids; NR, not reported; EIB, exercise-induced bronchoconstriction; SPT, skin prick test for atopy.

*This row of data relates to all recruited patients, a mix of atopic and non-atopic ($n = 112$).

*This row of data relates to atopic patients only ($n = 51$).

*Assumed to be the best sum of sensitivity and specificity.

> 90% was achieved at the lower cut-offs of > 41 and 46 p.p.b., respectively.

(ii) *Reference standard of AHR only* [34, 52]: Methodological quality was poor or unclear. Both studies have limited generalizability (low mean age [52]; very low mean FE_{NO}₅₀ [34]). Both studies used a pre-specified cut-off value (25 and 30 p.p.b.). Both reported high specificity (86.8% and 100%). No other cut-off values were reported.

(iii) *Reference standard of ART only*: This study [49] was a re-analysis of Smith et al. [3] using a different reference standard (ART instead of a sequence of tests). Methodological quality was poor or unknown. A low cut-off of 25 p.p.b. (rule-out) had a sensitivity of 83.3% (specificity of 57.5%), and a high cut-off of 90 p.p.b. (rule-in) had a specificity of 92.5% (sensitivity of 41.7%).

(iv) *Reference standard of pathway plus long-term follow-up*: This study [47] was a long-term follow-up of Schneider et al. [46]. Methodological quality was poor overall, mainly due to the loss of patients during follow-up, and the use of different follow-up methods for some patients. The results (cut-off of 26 p.p.b.) were very similar to the original article [46], which used a reference standard of the pathway alone (Table 2).

Position A: Asthma symptoms plus one additional symptom

(v) *Reference standard of AHR*: In this study [48], patients had rhinitis with asthma symptoms. Only patient selection was at low risk of bias. The derived best sum cut-off (36 p.p.b.) was within the same range as other studies at position A, although accuracy was unimpressive at 77.8% sensitivity and 60% specificity. However, 100% sensitivity was achieved at 25 p.p.b. and 100% specificity at 100 p.p.b., and 90% specificity at 75 p.p.b.

Position C: Negative airway reversibility test

(vi) *Reference standard of AHR*: One study recruited patients who had a negative ART [37], but all were army recruits (introducing spectrum bias as excludes certain ages, comorbidities, etc). Cut-offs from 10 to 30 p.p.b. were reported, which is unlikely to include the cut-off with the best specificity. The derived best sum cut-off was 32 p.p.b. (47% and 85%, respectively). A cut-off of 10 p.p.b. yielded 81% sensitivity (specificity of 39%), whilst the 30 p.p.b. cut-off yielded 82% specificity (sensitivity of 49%).

Position C or D: Referred for airway hyperresponsiveness

(vii) *Reference standard of AHR*: Two studies stated that patients were referred for AHR testing, but did not state

Study author, year	Patient selection	Index test	Reference standard	Flow and timing
Position A: Asthma symptoms, no previous tests				
i) Reference standard of pathway including AHR & ART				
Cordeiro 2011 [54]	unclear	high	High	high
Fortuna, 2007 [31, 32]	low	low	Unclear	unclear
Fukuhara 2011 [33]	unclear	low	Unclear	unclear
Schneider 2013 [46]	low	high	low	high
Schneider 2009 [44, 45]	low	high	low	unclear
Smith 2005 [3]	low	high	unclear	unclear
Smith 2004 [53]	low	high	unclear	unclear
ii) Reference standard of AHR				
Backer 2014 [52]	High	unclear	Unclear	unclear
Giovannini 2014 [34]	high	low	Unclear	unclear
iii) Reference standard of ART				
De La Barra 2011 [49]	low	high	unclear	unclear
iv) Reference standard of pathway plus long-term follow-up				
Schneider 2014 [47]	low	high	unclear	high
Position A: Asthma symptoms plus one additional symptom. No previous tests				
v) Reference standard of pathway including AHR & ART				
Heffler 2006 [48]	low	high	Unclear	unclear
Position C: negative ART				
vi) Reference standard of AHR				
Katsoulis 2013 [37]	high	high	unclear	low
Position C or D: referred for AHR				
vii) Reference standard of AHR				
Brannan 2013 [61]	high	unclear	unclear	unclear
Chancafe-Morgan 2013 [62]	high	unclear	unclear	unclear
Nickels 2014 [38, 39]	high	high	high	low
Position E: negative for clinical assessment (e.g. normal spirometry)				
viii) Reference standard of AHR				
Schneider 2009 [44, 45]	low	high	low	unclear
Chronic cough.				
Position D: negative for other conditions (e.g. GERD, X-ray or CT scan clear)				
ix) Reference standard of ART				
Hahn 2007 [35]	high	unclear	unclear	unclear
Hsu 2013 [36]	high	unclear	unclear	unclear
Prieto 2009 [40]	unclear	high	low	unclear
x) Reference standard including ART and AHR				
Sato 2008 [42]	low	high	unclear	unclear
Zhang 2011 [56-58]	unclear	high	unclear	unclear
xi) Other studies				
Arora 2006 [29]	high	high	low	unclear
Bobolea 2012 [55]	low	low	unclear	unclear
El Halawani 2003 [30]	high	high	unclear	unclear
Mathew 2011 [60]	unclear	low	unclear	unclear
Pedrosa 2010 [50, 51]	low	high	unclear	low
Pizzimenti 2009 [59]	high	high	unclear	unclear
Sastre 2013 [41]	low	high	unclear	low
Schleich, 2012 [43]	low	high	unclear	unclear

Fig. 3. Author's judgement of risk of bias of included studies, according to adapted version of QUADAS II [28].

for what reason [61, 62], and one stated that patients were at a tertiary referral centre [38, 39]. Patients are assumed to be at position C or D. Studies were at unclear risk of bias due to poor reporting and selection of patients with data for both FE_{NO}₅₀ and AHR tests [38, 62]. One study used mannitol [61], whilst the other two used methacholine [38, 62]. Although none reported multiple cut-offs, specificity did reach 96.3% (sensitivity of 30.2%) at a cut-off of 47 p.p.b. in one study [61]. One other study reported similar sensitivity and specificity at an unknown cut-off (sensitivity of 33.3%, specificity of 84.8%) [38]. The third study reported a cut-off of 35 p.p.b. with sensitivity of 75% and specificity of 83.3% [62].

Position E: Patients with normal spirometry

(viii) *Reference standard of AHR*: one study [44, 45] analysed a subgroup of patients with normal spirometry and the symptoms of asthma. The study was of reasonable quality and only scored poorly due to not pre-specifying a cut-off point and using different reference standards. Results at two cut-offs only were reported, with a specificity of 90% (sensitivity of 35%) at 46 p.p.b. and a sensitivity of 78% (specificity of 45%) at 16 p.p.b.

Position D: Patients with chronic cough negative for other conditions

Five studies recruited patients with chronic cough who had undergone tests for other conditions [e.g. gastro-oesophageal reflux disease (GORD) medication response, X-ray, computed tomography (CT) scan], but without reaching a diagnosis [35, 36, 40, 42, 56–58].

(ix) *Reference standard of ART to ICS*: Patients had normal chest radiographs or CT scans [35, 40], or had not responded to GORD or upper airway cough syndrome treatments [36]. These studies assess the use of FE_{NO}₅₀ at position D in the pathway, to replace or be used as well as one possible subsequent test. Key methodological weaknesses included the exclusion of smokers (spectrum bias) [35, 36] and a lack of clarity about the timing, sequence and blinding of tests. No study reported a range of cut-offs. Estimates were very similar in the two studies that excluded smokers [35, 36], with similar optimal cut-offs (38 and 33.9 p.p.b.), high sensitivity values (90% and 94.7%) and reasonable specificity (85% and 76.3%, respectively). The third study (which also had 0% smokers but did not state they were specifically excluded) reported sensitivity and specificity in the opposite direction than expected for the low cut-off value of 20 p.p.b. (Table 2).

(x) *Reference standard of ART and AHR*: One study recruited patients with normal chest radiograph and CT scan [42], whilst another recruited patients with normal

chest radiograph [56–58]. Key methodological weaknesses included a lack of information relating to the timing, sequence and blinding of tests. A range of cut-offs was not reported. The derived best cut-offs were similar to the previous studies at 38.8 and 40 p.p.b. However, specificities (91.3% and 86%, respectively) were higher than sensitivities (79.2% and 75%, respectively).

Other studies

(xi) Eight studies did not fall into the above categories. Arora et al. [29] have limited generalizability due to the cohort recruited (Table 1). Mathew et al. recruited patients with normal spirometry and no evidence of reversibility, a difficult group to place on our generic pathway as patients with normal spirometry would not normally undergo reversibility testing (Fig. 1). Pedrosa et al. [50, 51] and Schleich et al. [43] recruited patients with normal spirometry or negative ART (implying fixed airflow obstruction in some), which is likewise a difficult group to place as it mixes patients from two positions in the pathway. The diagnostic accuracy of FE_{NO}₅₀ in these groups is presented in Table 2 and in Appendix S7.

The three remaining studies reported useful results. Bobolea et al. [55] recruited patients who had a negative methacholine challenge test (MCT), to see whether there might be evidence of asthma despite this. The reference standard was adenosine challenge and the study quality was good except for information about blinding and timing of tests. FE_{NO}₅₀ demonstrated a sensitivity of 100% (specificity of 29.2%) at a cut-off of 30 p.p.b. El Halawani et al. [30] recruited patients suspected of EIB and used a reference standard of exercise challenge testing. Study quality was unclear/low as patients over 40 years were excluded to avoid patients with emphysema (spectrum bias), and missing details about timing, blinding and sequence of tests. FE_{NO}₅₀ demonstrated 100% sensitivity (specificity of 31%) at a cut-off of 12 p.p.b. Sastre et al. [41] recruited patients with suspected occupational asthma and used a reference standard of a specific inhalation challenge. In this case, specificity was reasonable at 80% (sensitivity of 60%) at a cut-off of 25 p.p.b.

(xii) *FE_{NO}₅₀ used in conjunction with another test*: Three studies reported results for FE_{NO}₅₀ in conjunction with another test [31, 43, 54]. In Cordeiro et al. [54], a positive test was FE_{NO}₅₀ > 27 p.p.b. and/or ART. The reference standard, ART and/or AHR, introduced incorporation bias as ART formed part of both the index test and the reference standard. Diagnostic accuracy improved compared to FE_{NO}₅₀ alone. In Schleich et al. [43], a positive index test was FEV₁ ≤ 101% with FE_{NO}₅₀ > 34 p.p.b. The reference standard was AHR by

MCT. The patients selected were a mix of positions B and C. Addition of FEV1 $\leq$ 101% improved specificity, but decreased sensitivity (Table 2). In Fortuna *et al.* [31, 32], using sputum eosinophilia in conjunction with FENO_{50} increased specificity from 64% to 76%; specificity was not reported. They also reported that the addition of bronchodilator test and lung function tests did not increase accuracy, but actual data were not reported.

Discussion

This systematic review of the diagnostic accuracy of FENO_{50} for asthma included 27 studies in adults/all ages. Across this heterogeneous literature, it is extremely difficult to draw any robust conclusions about optimal cut-off points, whether accuracy varies according to patient populations and reference standards, where FENO_{50} should be placed within a diagnostic pathway, or whether FENO_{50} would be best used as a rule-in test or a rule-out test. Several categories of studies were identified (i–xii), each of which relates to a specific clinical scenario. Across these categories, no studies reported 100% accuracy, although rule-in and rule-out scenarios appeared more promising with high and perfect sensitivities or specificities often being reported.

Discussion of results

Estimates of accuracy in category i–iv studies (FENO_{50} to replace the whole pathway in newly presenting patients, position A) varied. Even at its best, FENO_{50} could not be considered an adequate replacement for the whole diagnostic pathway as many patients would be incorrectly diagnosed. As a rule-in/rule-out test, high cut-off values (41 p.p.b.) for FENO_{50} achieved high specificity ($> 90\%$), indicating it could be used as a rule-in test, but the best sensitivity reported was less good (81.5–96%, cut-off values of 9–25 p.p.b.), indicating only moderate potential as a rule-out test. Despite the fact that not all asthmatics have eosinophilic inflammation, three studies reported 100% sensitivity [30, 47, 55]. However, all were from relatively small samples ($n < 50$) and highly selected populations, which may account for the unexpected results.

The category v study recruited patients with rhinitis and asthma symptoms and reported excellent sensitivity (100% at 25 p.p.b.) and specificity (100% at 100 p.p.b.). There is not enough evidence to conclude there is better accuracy in patients with rhinitis, due to the small number of studies ($n = 1$).

Category vi and vii studies assessed the use of FENO_{50} before or in place of AHR testing. This position was identified *a priori* as potentially the most useful

position for FENO_{50} for use in primary care to avoid the need for referral to a secondary care facility for unpleasant and expensive AHR testing in at least a proportion of patients, by either ruling asthma in or out. None of the studies reported a full range of cut-off values. High specificity (96.3%, sensitivity 30%) was reported in one study [60], using a cut-off of 47 p.p.b., indicating the potential for FENO_{50} to be a useful rule-in test at this point, although the small evidence base precludes a firm conclusion.

Categories ix and x recruited patients with chronic cough who tested negative for imaging tests such as X-ray or CT scan (position D). Some tested FENO_{50} against a reference standard of ICS responsiveness (trial of treatment). A trial of ICS is an unlikely test at this point, according to some guidelines [24], but could be a useful test for steroid-responsive disease in these patients. A trial of ICS treatment can lead to misdiagnosis of asthma through spontaneous remission of symptoms, through subjective interpretations of response or through poor adherence to ICS treatment or need for higher ICS dose. As such, these studies are difficult to interpret if, for example, FENO_{50} is better at identifying patients with truly steroid-responsive disease than a trial of treatment. Other studies tested against AHR, which is a likely next test at least in some jurisdictions; avoiding this test has advantages as already discussed. As no study reported a range of cut-offs, the full potential of FENO_{50} as a rule-in/rule-out test at this point is unclear, although good sensitivity vs. ICS responsiveness (90%, 94.7%) [35, 36] and good specificity vs. AHR (91.3%) [46] were reported.

Category viii assessed the use of FENO_{50} before or in place of referring patients to tests for other conditions (position E). We had not explicitly identified this position *a priori*, but believe it may be a useful position as it could, if used in primary care, hasten the correct diagnosis, through accelerated and more appropriate patient pathways. Only one study conducted this analysis [44, 45] and a full range of cut-off values was not reported, although a specificity of 90% at 46 p.p.b. and a sensitivity of 78% at 16 p.p.b. are promising.

In category xi, very specific uses of FENO_{50} were tested. The small evidence base suggests FENO_{50} could be used successfully (100% sensitivity) in patients who had a negative MCT to identify patients who would score negatively by adenosine challenge testing [55] and for those under 40 years of age referred for exercise challenge testing [30]. The other study in this category showed 80% sensitivity at a derived cut-off of 25 p.p.b.; promising better results may be achieved at different cut-offs. As such, FENO_{50} appears to have the potential to be a useful rule-out test before these very specific bronchial challenge tests. Category xii studies were, in theory, useful to construct steps mid-way in

the diagnostic pathway. However, the reported scenarios were limited and not useful for this purpose.

Strengths and limitations of this study

This study is the first fully reported systematic review of the diagnostic accuracy of FE_{NO}₅₀ in asthma. It benefits from a high-quality, extensive search strategy, quality assessment using the QUADAS-2 checklist and double data checking by the authors. Whilst a large number of studies were identified, the results remain inconclusive. This is in part due to the methodological weaknesses of the primary studies themselves, poor reporting of key methodological points and a lack of reporting of a full range of cut-off values. It is also due to the variability in results between studies. Ideally, the impact on estimates of diagnostic accuracy of factors such as age, severity at inclusion, atopy, smoking status, ICS use and FE_{NO} device used would have been investigated, but the heterogeneity between these factors as well as in the major characteristics of study population and reference standard made any sensitivity or subgroup comparisons between studies with a given characteristic problematic. As such, no sensitivity analyses were conducted and the causes of heterogeneity in results remain largely uninvestigated.

Of particular note was the low level of reporting of atopic status and corticosteroid use prior to testing. Both factors have been shown to affect FE_{NO} levels, so it is surprising that they were not more carefully considered by study authors. For atopy, one study reported results both for the whole recruited cohort and for a subgroup of patients who had atopy [37]. Whilst the overall accuracy was similar in both cohorts (area under the curve 0.69 (95% CI 0.6–0.775) and 0.68 (95% CI 0.53–0.8), respectively), the cut-off with the highest sum of sensitivity and specificity was higher in non-atopic patients than in atopic patients (32 vs. 26 p.p.b., respectively), which is not as expected. Equally, an exploratory analysis comprising a simple ordering of studies according to % atopic patients did not reveal a trend towards lower cut-offs (see Appendix S7) for studies that recruited a higher proportion of atopic patients. Reasons for this remain unclear, although the heterogeneity in other characteristics may play a part. For corticosteroid use, the studies that did not mention whether this was stopped prior to testing may have recruited patients before any treatments were prescribed, but this was largely unclear. The study that included patients who were on ICS as well as patients who were not [35] aimed to identify patients who would respond to an increase in or new treatment with ICS, so the inclusion of those already treated with ICS was appropriate to the aim of that study. The mean FE_{NO}₅₀ in patients who already

took ICS and responded to a dose increase was 44.8 p.p.b. (SD 3.1), compared to 51.25 p.p.b. (SD 20.1) in patients of any ICS status who responded to an increase in or new treatment with ICS, indicating that FE_{NO}₅₀ could be useful in identifying patients who require an increase in ICS dose, even in those pre-treated with ICS, and potentially with a similar cut-off.

The use of FE_{NO} in the diagnosis of asthma or steroid-responsive disease?

The variability of results seen in our review does not detract from the potential that FE_{NO} has to make a useful contribution in the diagnosis of asthma evidenced by the very good specificities and sensitivities reported in some studies. Some current thinking about asthma is moving towards the identification of steroid-responsive disease rather than classical asthma diagnosis. For the identification of patients with asthma who will respond to ICS, a recent international review [4] recommended that patients are placed into one of three categories based on their FE_{NO}₅₀ levels: < 20–25 p.p.b. (normal), patients should be considered for other diagnoses (rule-out asthma); 20–25 to 50 p.p.b. (elevated), patients should be considered asthmatic and a low dose of ICS prescribed (rule-in asthma); and > 50 p.p.b. (high), patients should be considered asthmatic and prescribed a moderate dose of ICS therapy (rule-in asthma). The ATS guidelines [5] reported that levels below 25 p.p.b. indicate a low likelihood of eosinophilic inflammation and ICS responsiveness and levels above 50 p.p.b. indicate a high likelihood. It also reports similar cut-offs for ICS responsiveness. Our review found only four studies that used ICS responsiveness as the reference standard [35, 36, 40, 49]. None reported 100% sensitivity or specificity, despite one study reporting a wide range of cut-offs (25–150 p.p.b.). Two studies reported highest sum cut-offs of 20–25 p.p.b., with one reporting poor sensitivity (53%) and specificity (63%) [40] and one reporting somewhat better values at 83% and 57% [49], respectively. The latter study also reported a cut-off of 50 p.p.b. and showed almost inverse values of 58% and 80%, respectively [48]. Better values were reported by the other studies, indicating good overall accuracy, with sensitivities and specificities of 90% and 85%, and 95% and 76%, respectively, at higher cut-offs of 38 and 33.9 p.p.b. [35, 36]. All four studies are limited in terms of generalizability (three recruited patients with chronic cough who were negative for tests for other conditions [35, 36, 40], and in one case some patients were already on ICS treatment [35] and one recruited adults and adolescents with the symptoms of asthma [49]). The imperfect reference standard, as previously discussed, may account for some of the

differences between index test (FENO_{50}) and reference standard (ICS responsiveness). Whilst the two guidelines recommended specific cut-offs for identifying steroid-responsive disease, the results of our review do not lend themselves to making such a recommendation. This is in part due to the more restrictive inclusion criteria we have applied. In particular, this means we do not include studies cited in the guidelines which are on paediatric [63–65] patients, COPD [66–70] and diagnosed asthmatics [71]. Given this smaller evidence base, we conclude that the use of FENO_{50} as a marker of steroid-responsive disease in adult patients with the symptoms of asthma remains very much understudied.

When considering FENO_{50} to diagnose asthma, it was interesting to see that results did not obviously differ between categories, with very similar ranges of cut-offs and accuracies across them all. This was surprising and may indicate that the theoretical assumption that patient spectrum and reference standard used would affect accuracy estimates does not hold. Equally, any differences may be obscured by the large amount of ‘noise’ from confounders between studies and relatively small number of studies in any one category. However, because of this generality of results, and because FENO_{50} may be used at any point in a diagnostic pathway to indicate asthma, an exploration of which cut-offs may provide the best sensitivity and specificity in any situation was conducted using all available studies (Appendix S7). When study results were ordered according to cut-offs, studies using 20–25 p.p.b. did not consistently deliver good sensitivity (range 36–100%). Sensitivities above 80% were only consistently delivered by studies with cut-offs ≤ 12 p.p.b., and sensitivities above 70% were only consistently provided by studies with cut-offs ≤ 15 p.p.b. This perhaps suggests that the cut-off point for ‘normal’ should be set somewhat lower if FENO_{50} is used in a pathway to diagnose asthma as opposed to steroid-responsive airway disease. Similarly, results suggested that a higher cut-off than 50 p.p.b. may be more likely to consistently deliver high levels of specificity, with specificities $\geq 90\%$ only consistently being reported in studies with cut-offs > 70 p.p.b.

Research recommendations

The benefits and harms to patients of being TP, TN, FP or FN remain largely unquantified and could affect which cut-off is preferred; the consequences of being wrongly diagnosed (either FP or FN) may mean it is preferable to sacrifice some sensitivity or specificity to minimize these consequences. As part of the wider project for this review, a cost-utility analysis was planned with one aim being to ascertain the best cut-

off to use by modelling the consequences of each category. However, the evidence base was inadequate to populate such a model, mainly due to the lack of a full range of cut-offs being available and to the lack of data relating to each individual step in the pathway. To address this, future study authors should aim to report a full range of cut-off values. The best use of FENO in a diagnostic pathway could be investigated using a study design where all diagnostic tests are given to all patients (as index tests) and different algorithms are modelled in a cost-utility analysis. This may allow the best position for FENO to be assessed taking into account costs, benefits and harms.

Problems with the reference standard currently hamper all studies in this review. One study [47] attempted to surmount this issue through long-term follow-up of patients, but lost a large proportion of patients to follow up which may have introduced bias. Future study designs could include intensive long-term follow-up. Simultaneously, researchers could consider the capture of long-term clinical outcomes in end-to-end studies (considered the highest standard in the NICE evidence hierarchy) [25], which would be directly useful in cost-effectiveness models.

Another key research question is whether diagnosis of steroid-responsive disease is a better diagnostic target than asthma. Any new consensus regarding this may necessitate a different evidence base including wider populations (e.g. patients with symptoms of airway obstruction) and different reference standard methods (e.g. that aim to address adherence and dose issues when using a trial of ICS treatment). Publication checklists such as STARD [72] would facilitate better reporting and better assessment of methodological quality. Until high-quality evidence is available, the balance of costs and clinical benefits of the use of FENO can only be assumed through clinical interpretation.

To conclude, this review reports a large and heterogeneous evidence base of high to moderate, and often unclear, risk of bias. Study designs do not allow a full assessment of the clinical impact of FENO_{50} when used in a pathway. Estimates of diagnostic accuracy and cut-off values for the diagnosis of asthma by its current definition varied greatly, even within groups of similar studies, probably due to heterogeneity in multiple study and patient characteristics, and study quality. Whilst optimal cut-off values often failed to produce impressive accuracy, very high sensitivities and specificities were reported at low and high cut-offs in several studies, indicating FENO_{50} could be a useful rule-in and/or rule-out test. However, the diagnostic accuracy, cut-off values that should be used and optimal position for FENO_{50} within a pathway remain poorly evidenced.

Acknowledgements

We would like to thank John W Stevens, Reader in Decision Science, SchARR, for providing statistical support.

Funding

This project was funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) Programme (Project Number: 12/60/01) as part of a review on 'Measurement of exhaled nitric oxide concentration in asthma; NIOX MINO, NIOX VERO and NObreath', and published as part of a full report in Health Technology Assessment [16, 17] (PROSPERO registration number: CRD42013004149) (<http://www.crd.york.ac.uk/>

prospero/). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NHS, the NIHR, MRC, CCF, NETSCC, the NIHR HTA programme or the UK Department of Health.

Conflict of interest

Professor Ian Pavord received personal fees from GSK, Astra Zeneca, Almirall, Novartis, Aerocrine, Genentech, Teva, Regeneron and Boehringer Ingelheim, outside this work. Dr Rod Lawson received research support in the form of a grant from GSK and Novartis for diagnostic imaging, personal fees from GSK and Novartis for advisory board meetings and educational meetings and personal fees from Astra Zeneca, Almirall and Boehringer Ingelheim for educational meetings, all outside this work.

References

- Berry MA, Shaw DE, Green RH *et al.* The use of exhaled nitric oxide concentration to identify eosinophilic airway inflammation: an observational study in adults with asthma. *Clin Exp Allergy* 2005; 35:1175–9.
- Cowan DC, Taylor DR, Peterson LE *et al.* Biomarker-based asthma phenotypes of corticosteroid response. *J Allergy Clin Immunol* 2015; 135:877–83.
- 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.
- Bjerner 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.
- 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.
- Pavord ID, Hilvering B. Biomarkers and inhaled corticosteroid responsiveness in asthmatic patients. *J Allergy Clin Immunol* 2015; 135:884–5.
- British Thoracic Society and Scottish Intercollegiate Guidelines Network. British guideline on the management of asthma, 2007 update. <http://www.sign.ac.uk/guidelines/fulltext/63/index.html> (accessed April 2008).
- Bateman ED, Hurd SS, Barnes PJ *et al.* Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J* 2008; 31:143–78.
- Moral VP, Gutierrez F, Alvarez Gutierrez P *et al.* Guía española para el manejo del asma. *Arch Bronconeumol* 2003; 39(Suppl. 5):3–42.
- Mahr TA, Malka J, Spahn JD. Inflammometry in pediatric asthma: a review of fractional exhaled nitric oxide in clinical practice. *Allergy Asthma Proc* 2013; 34:210–9.
- Taylor DR, Taylor DR. Advances in the clinical applications of exhaled nitric oxide measurements. *J Breath Res* 2012; 6:047102.
- Rodway GW, Choi J, Hoffman LA *et al.* Exhaled nitric oxide in the diagnosis and management of asthma: clinical implications. *Chron Respir Dis* 2009; 6:19–29.
- Kercsmar C, Kercsmar C. Exhaled nitric oxide in the diagnosis and management of childhood asthma [review] [57 refs]. *Ther Adv Respir Dis* 2010; 4:71–82.
- Barnes PJ, Dweik RA, Gelb AF *et al.* Exhaled nitric oxide in pulmonary diseases: a comprehensive review. *Chest* 2010; 138:682–92.
- Dodig S, Richter D, Zrinski-Topic R, Dodig S, Richter D, Zrinski-Topic R. Inflammatory markers in childhood asthma [review]. *Clin Chem Lab Med* 2011; 49:587–99.
- Spitale N, Popat N, McIvor A. Update on exhaled nitric oxide in pulmonary disease. *Expert Rev Respir Med* 2012; 6:105–15.
- Szeffler SJ, Wenzel S, Brown R *et al.* Asthma outcomes: biomarkers [review]. *J Allergy Clin Immunol* 2012; 129(3 Suppl.):S9–23.
- Feitosa LA, Dornelas de AA, Reinaux CM, Britto MC. Diagnostic accuracy of exhaled nitric oxide in exercise-induced bronchospasm: systematic review. *Rev Port Pneumol* 2012; 18:198–204.
- Essat M, Harnan S, Gomersall T *et al.* Fractional exhaled nitric oxide for the management of asthma in adults: a systematic review. *Eur Respir J* 2016; 47:751–68.
- Gomersall T, Harnan S, Essat M *et al.* A systematic review of fractional exhaled nitric oxide in the routine management of childhood asthma. *Pediatr Pulmonol* 2016; 51:316–28.
- 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:(82).
- National Clinical Guideline Centre. Asthma: diagnosis and monitoring of asthma in adults, children and young people. Clinical guideline. Methods, evidence and recommendations. Draft for Consultation. 2015.
- Asthma: diagnosis and monitoring of asthma in adults, children and young people. Clinical guideline. Methods, evidence and recommendations. Draft for consultation. 2016.
- British Thoracic Society SIGN. British guideline on the management of asthma: A national clinical guideline. 2013. <http://www.brit-thoracic.org.uk/>

- guidelines/asthma-guidelines.aspx (accessed 2 February 2013).
- 25 National Institute for Health and Care Excellence. *Diagnostics assessment programme manual*. Manchester: National Institute for Health and Care Excellence, 2011.
 - 26 Centre for Reviews and Dissemination. Systematic reviews. CRD's guidance for undertaking reviews in healthcare. 2009. <https://www.york.ac.uk/crd/guidance/>.
 - 27 Higgins J, Green S, eds. *Cochrane handbook for systematic reviews of interventions Version 5.1.0*. [Updated March 2011]. www.cochrane-handbook.org (accessed 21 July 2015).
 - 28 Whiting P, Rutjes A, Westwood M *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155:529–36.
 - 29 Arora R, Thornblade CE, Dauby PAL, Flanagan JW, Bush AC, Hagan LL. Exhaled nitric oxide levels in military recruits with new onset asthma. *Allergy Asthma Proc* 2006; 27:493–8.
 - 30 El Halawani SM, Ly NT, Mahon RT, Amundson DE. Exhaled nitric oxide as a predictor of exercise-induced bronchoconstriction. *Chest* 2003; 124: 639–43.
 - 31 Fortuna AM, Feixas T, Gonzalez M *et al*. Diagnostic utility of inflammatory biomarkers in asthma: exhaled nitric oxide and induced sputum eosinophil count [Erratum appears in *Respir Med* 2012; 106(4):599]. *Respir Med* 2007; 101:2416–21.
 - 32 Fortuna AM, Feixas T, Gonzalez M, Casan P. Diagnostic utility of inflammatory biomarkers in asthma: exhaled nitric oxide and induced sputum eosinophil count (vol. 101, pg 2416, 2007). *Respir Med* 2012; 106:599.
 - 33 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.
 - 34 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.
 - 35 Hahn PY, Morgenthaler TI, Lim KG. Use of exhaled nitric oxide in predicting response to inhaled corticosteroids for chronic cough. *Mayo Clin Proc* 2007; 82:1350–5.
 - 36 Hsu JY, Wang CY, Cheng YW, Chou MC. Optimal value of fractional exhaled nitric oxide in inhaled corticosteroid treatment for patients with chronic cough of unknown cause. *J Chin Med Assoc* 2013; 76:15–9.
 - 37 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.
 - 38 Nickels A. Exhaled nitric oxide performance compared to methacholine challenge in asthma. *J Allergy Clin Immunol* 2014; 133:AB148.
 - 39 Nickels A, Parker K, Scanlon P, Lim K. The role of a positive exhaled nitric oxide in evaluating the pulmonary patient: exhaled nitric oxide versus methacholine challenge? *CHEST J* 2014; 145(3 Meeting Abstracts):467A.
 - 40 Prieto L, Ferrer A, Ponce S, Palop J, Marin J. Exhaled nitric oxide measurement is not useful for predicting the response to inhaled corticosteroids in subjects with chronic cough. *Chest* 2009; 136:816–22.
 - 41 Sastre J, Costa C, del Garcia Potro M, Aguado E, Mahillo I, Fernández-Nieto M. Changes in exhaled nitric oxide after inhalation challenge with occupational agents. *J Investig Allergol Clin Immunol* 2013; 23:421–7.
 - 42 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.
 - 43 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.
 - 44 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 ≤ 16 ppb better than FENO ≤ 12 ppb to rule out mild and moderate to severe asthma. *Respir Res* 2009; 10:64.
 - 45 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.
 - 46 Schneider A, Schwarzbach J, Faderl B, Welker L, Karsch-Volk M, Jorres RA. FENO measurement and sputum analysis for diagnosing asthma in clinical practice. *Respir Med* 2013; 107: 209–16.
 - 47 Schneider A, Faderl B, Schwarzbach J, Welker L, Karsch-Völkl M, Jörres RA. Prognostic value of bronchial provocation and FENO measurement for asthma diagnosis – results of a delayed type of diagnostic study. *Respir Med* 2014; 108:34–40.
 - 48 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.
 - 49 de la Barra SL, Smith AD, Cowan JO, Herbison GP, Taylor DR. Predicted versus absolute values in the application of exhaled nitric oxide measurements. *Respir Med* 2011; 105:1629–34.
 - 50 Pedrosa M, Cancelliere N, Barranco P, Lopez-Carrasco V, Quirce S. Usefulness of exhaled nitric oxide for diagnosing asthma. *J Asthma* 2010; 47:817–21.
 - 51 Pedrosa M, Cancelliere N, Barranco P, San JB, Quirce S. Usefulness of exhaled nitric oxide in asthma diagnosis. *J Allergy Clin Immunol* 2010; 2 (Suppl. 1):AB7.
 - 52 Backer V, Sverrild A, Porsbjerg C. FENO and AHR mannitol in patients referred to an out-of-hospital asthma clinic: a real-life study. *J Asthma* 2014; 51:411–6.
 - 53 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.
 - 54 Cordeiro D, Rudolphus A, Snoey E, Braunstahl G-J. Utility of nitric oxide for the diagnosis of asthma in an allergy clinic population. *Allergy Asthma Proc* 2011; 32:119–26.
 - 55 Bobolea ID, Barranco P, Lopez-Carrasco V, Calderon O, Guillen D, Quirce S. Is methacholine challenge sufficient to rule out bronchial hyperresponsiveness in patients with suspected asthma? *J Allergy Clin Immunol* 2012; 2(Suppl. 1):AB3.
 - 56 Zhang YM, Lin JT, Su N *et al*. [Values of fractional exhaled nitric oxide in the diagnosis of chronic cough] [Chinese]. *Chung-Hua i Hsueh Tsa Chih [Chin Med J]* 2011; 91:1254–8.
 - 57 Zhang Y. The values of fractional exhaled nitric oxide in the diagnosis and treatment of chronic cough. *Respirology* 2011; 16:213.

- 58 Zhang Y-M, Lin J-T. The values of fractional exhaled nitric oxide for the diagnosis and treatment response evaluation of chronic cough. *Respirology* 2011; 16:219.
- 59 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.
- 60 Mathew S, Cliff I, Agarwal S, Lim A, Allen M, Mustfa N. Relationship between exhaled nitric oxide and methacholine challenge test in suspected asthma. *Am J Respir Crit Care Med* 2011; 183(1 Meeting Abstracts):183;2011:A5554.
- 61 Brannan JD, Adoni H, Daw L, Huang HC, Hurwitz M, Figurski D. Fraction exhaled no in patients referred to pulmonary function laboratory (PFLAB) for mannitol challenge. *Respirology* 2013; 18:43.
- 62 Chancafe-Morgan J, Ramos-Quispe Y, Gomez-Garcia R, Vargas-Espinal J, Puente-Maestu L. Validity of the fractional exhaled nitric oxide (FeNO) for identification of bronchial hyperresponsiveness in a pulmonary function laboratory. *European Respiratory Society Annual Congress, Barcelona, Spain* 2013. 2013.
- 63 Szeffler SJ, Martin RJ, King TS *et al.* Significant variability in response to inhaled corticosteroids for persistent asthma. *J Allergy Clin Immunol* 2002; 109:410–8.
- 64 Knuffman JE, Sorkness CA, Lemanske RF *et al.* Phenotypic predictors of long-term response to inhaled corticosteroid and leukotriene modifier therapies in pediatric asthma. *J Allergy Clin Immunol* 2009; 123:411–6.
- 65 Klaassen EM, van de Kant KD, Jobsis Q *et al.* Symptoms, but not a biomarker response to inhaled corticosteroids, predict asthma in preschool children with recurrent wheeze. *Mediators Inflamm* 2012; 2012:162571.
- 66 Zietkowski Z, Kucharewicz I, Bodzenta-Lukaszyk A, Zietkowski Z, Kucharewicz I, Bodzenta-Lukaszyk A. The influence of inhaled corticosteroids on exhaled nitric oxide in stable chronic obstructive pulmonary disease. *Respir Med* 2005; 99:816–24.
- 67 Antus B. [Role of exhaled nitric oxide in predicting steroid response in chronic obstructive pulmonary disease]. *Orv Hetil* 2010; 151:2083–8.
- 68 Gorska K, Krenke R, Korczynski P, Kosciuch J, Domagala-Kulawik J, Chazan R. Eosinophilic airway inflammation in chronic obstructive pulmonary disease and asthma. *J Physiol Pharmacol* 2008; 59(Suppl. 6):261–70.
- 69 Karen A, Scott I, Andrew J. Wardlaw. *Eosinophilic airway disorders*. New York, NY: Thieme Medical Publishers, Inc., 2006.
- 70 Dummer JF, Epton MJ, Cowan JO *et al.* Predicting corticosteroid response in chronic obstructive pulmonary disease using exhaled nitric oxide. *Am J Respir Crit Care Med* 2009; 180:846–52.
- 71 Malinovschi A, Van Muylen A, Michils A. Both intermediate and high exhaled nitric oxide levels predict improvement in asthma control after new-onset of inhaled corticosteroids. *Allergy* 2013; 68(Suppl. s97):164.
- 72 Bossuyt PM, Reitsma JB, Bruns DE *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138:W1–12.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Example search strategy.

Appendix S2. Inclusion and exclusion criteria.

Appendix S3. Data extraction form.

Appendix S4. Quality Assessment form.

Appendix S5. Rule in and rule out scenarios.

Appendix S6. Table of studies ordered according to % atopic patients.

Appendix S7. Table of all data reported.

Supplemental Information for “Exhaled nitric oxide in the diagnosis of asthma in adults: a systematic review” by Harnan et al.

Contents	Page
Appendix S1 –Example search strategy	2
Appendix S2 – Inclusion and exclusion criteria	13
Appendix S3 – Data extraction form	18
Appendix S4 – Quality Assessment form	19
Appendix S5 – Rule in and Rule out scenarios	24
Appendix S7 - Table of all data reported	
i) ordered by author	24
ii) ordered by cut-off value	28
iii) studies which used ICS as the reference standard	31

Appendix S1

Example search strategy: Medline.

Alternating grey and white shading: groups of related terms

1. Review string #1

- 1 NIOX MINO.mp.
- 2 aerocrine.mp.
- 3 (niox adj5 (monitor\$ or chemiluminescence or analyser\$ or sensor or device\$ or desktop)).mp.
- 4 NObreath.mp.
- 5 bedfont.mp.
- 6 or/1-5
- 7 exp cough/
- 8 cough\$.mp.
- 9 phlegm.mp.
- 10 sputum.mp.
- 11 mucus.mp.
- 12 wheez\$.mp.
- 13 chest pain/
- 14 chest pain\$.mp.
- 15 (chest adj5 tight\$.tw.
- 16 ((lower respiratory or lrt) adj5 symptom\$.tw.
- 17 (lower airway adj5 symptom\$.tw.
- 18 ((trache\$ or wind pipe or lung\$ or bronch\$) adj3 symptom\$.tw.
- 19 exp lung/ or trachea/
- 20 symptom\$.tw.
- 21 19 and 20
- 22 or/7-18,21

23 exp asthma/
 24 asthma\$.mp.
 25 exp respiratory hypersensitivity/
 26 exp bronchial hyperreactivity/
 27 bronchial spasm/
 28 bronchospas\$.mp.
 29 exp Bronchoconstriction/
 30 bronchoconstric\$.mp.
 31 (bronch\$ adj3 constrict\$).mp.
 32 (bronch\$ adj5 spas\$).mp.
 33 (airway\$ adj5 (obstruct\$ or inflammation\$)).mp.
 34 ((bronchial\$ or respiratory or airway\$ or lung\$) adj3 (hypersensitiv\$ or hyperreactiv\$ or allerg\$ or insufficiency)).mp.

35 or/23-34

36 Nitric Oxide/
 37 nitric oxide.mp.
 38 36 or 37
 39 (exhal\$ or expir\$ or alveolar or fractional).mp.

40 38 and 39 (5228)

41 exhaled NO.mp.
 42 eno.mp.
 43 fe?no\$.mp.
 44 (fractional adj2 NO).mp.

45 or/40-44

46 22 and 45

47 35 and 45

48 6 or 46 or 47

49 limit 48 to yr="2009 -Current"

2. Review string #2

- 1 NIOX MINO.mp.
- 2 aerocrine.mp.
- 3 (niox adj5 (monitor\$ or chemiluminescence or analyser\$ or sensor or device\$ or desktop)).mp.
- 4 NObreath.mp.
- 5 bedfont.mp.

6 or/1-5

- 7 exp cough/
- 8 cough\$.mp.
- 9 phlegm.mp.
- 10 sputum.mp.
- 11 mucus.mp.
- 12 wheez\$.mp.
- 13 chest pain/
- 14 chest pain\$.mp.
- 15 (chest adj5 tight\$).tw.
- 16 ((lower respiratory or lrt) adj5 symptom\$).tw.
- 17 (lower airway adj5 symptom\$).tw.
- 18 ((trache\$ or wind pipe or lung\$ or bronch\$) adj3 symptom\$).tw.

19 exp lung/ or trachea/

20 symptom\$.tw.

21 19 and 20

22 or/7-18,21

23 exp asthma/
24 asthma\$.mp.
25 exp respiratory hypersensitivity/
26 exp bronchial hyperreactivity/
27 bronchial spasm/
28 bronchospas\$.mp.
29 exp Bronchoconstriction/
30 bronchoconstric\$.mp.
31 (bronch\$ adj3 constrict\$).mp.
32 (bronch\$ adj5 spas\$).mp.
33 (airway\$ adj5 (obstruct\$ or inflammation\$)).mp.
34 ((bronchial\$ or respiratory or airway\$ or lung\$) adj3 (hypersensitiv\$ or hyperreactiv\$ or allerg\$ or insufficiency)).mp.

35 or/23-34

36 Nitric Oxide/
37 nitric oxide.mp.
38 36 or 37
39 (exhal\$ or expir\$ or alveolar or fractional).mp.

40 38 and 39 (5228)

41 exhaled NO.mp.
42 eno.mp.
43 fe?no\$.mp.
44 (fractional adj2 NO).mp.

45 or/40-44

46 22 and 45

47 35 and 45

48 6 or 46 or 47

49 meta-analysis as topic/
 50 (meta analy\$ or metaanaly\$).tw.
 51 Meta-Analysis/
 52 (systematic adj (review\$1 or overview\$1)).tw.
 53 "Review Literature as Topic"/
 54 or/49-53 (96944)
 55 (cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or cinhal or science
 citation index or bids or cancerlit).ab.
 56 ((reference adj list\$) or bibliograph\$ or hand-search\$ or (relevant adj journals) or (manual adj
 search\$)).ab.
 57 ((selection adj criteria) or (data adj extraction)).ab.
 58 "review"/
 59 57 and 58
 60 comment/ or editorial/ or letter/
 61 Animals/
 62 Humans/
 63 61 not (61 and 62)
 64 60 or 63
 65 54 or 55 or 56 or 59
 66 65 not 64
 67 48 and 66

3. Review string #3

1 NIOX MINO.mp.
 2 aerocrine.mp.
 3 (niox adj5 (monitor\$ or chemiluminescence or analyser\$ or sensor or device\$ or desktop)).mp.
 4 NObreath.mp.
 5 bedfont.mp.
 6 or/1-5

7 exp cough/
 8 cough\$.mp.
 9 phlegm.mp.
 10 sputum.mp.
 11 mucus.mp.
 12 wheez\$.mp.
 13 chest pain/
 14 chest pain\$.mp.
 15 (chest adj5 tight\$.tw.
 16 ((lower respiratory or lrt) adj5 symptom\$.tw.
 17 (lower airway adj5 symptom\$.tw.
 18 ((trache\$ or wind pipe or lung\$ or bronch\$) adj3 symptom\$.tw.

19 exp lung/ or trachea/
 20 symptom\$.tw.

21 19 and 20

22 or/7-18,21

23 exp asthma/
 24 asthma\$.mp.
 25 exp respiratory hypersensitivity/
 26 exp bronchial hyperreactivity/
 27 bronchial spasm/
 28 bronchospas\$.mp.
 29 exp Bronchoconstriction/
 30 bronchoconstric\$.mp.
 31 (bronch\$ adj3 constrict\$.mp.
 32 (bronch\$ adj5 spas\$.mp.
 33 (airway\$ adj5 (obstruct\$ or inflammation\$)).mp.
 34 ((bronchial\$ or respiratory or airway\$ or lung\$) adj3 (hypersensitiv\$ or hyperreactiv\$ or allerg\$

or insufficiency)).mp.

35 or/23-34

36 Nitric Oxide/

37 nitric oxide.mp.

38 36 or 37

39 (exhal\$ or expir\$ or alveolar or fractional).mp.

40 38 and 39 (5228)

41 exhaled NO.mp.

42 eno.mp.

43 fe?no\$.mp.

44 (fractional adj2 NO).mp.

45 or/40-44

46 22 and 45

47 35 and 45

48 6 or 46 or 47

1 NIOX MINO.mp.

2 aerocrine.mp.

3 (niox adj5 (monitor\$ or chemiluminescence or analyser\$ or sensor or device\$ or desktop)).mp.

4 NObreath.mp.

5 bedfont.mp.

6 or/1-5

49 Randomized controlled trials as Topic/
 50 Randomized controlled trial/
 51 Random allocation/
 52 randomized controlled trial.pt.
 53 Double blind method/
 54 Single blind method/
 55 Clinical trial/
 56 exp Clinical Trials as Topic/
 57 controlled clinical trial.pt.
 58 or/49-57
 59 (clinic\$ adj25 trial\$).ti,ab.
 60 ((singl\$ or doubl\$ or treb\$ or tripl\$) adj (blind\$ or mask\$)).tw.
 61 Placebos/
 62 Placebo\$.tw.
 63 (allocated adj2 random).tw.

64 or/59-63

65 58 or 64

66 Case report.tw.

67 Letter/

68 Historical article/

69 66 or 67 or 68

70 exp Animals/

71 Humans/

72 70 not (70 and 71)

73 69 or 72

74 65 not 73

75 48 and 74

4. Review string #4

- 1 NIOX MINO.mp.
- 2 aerocrine.mp.
- 3 (niox adj5 (monitor\$ or chemiluminescence or analyser\$ or sensor or device\$ or desktop)).mp.
- 4 NObreath.mp.
- 5 bedfont.mp.

6 or/1-5

- 7 exp cough/
- 8 cough\$.mp.
- 9 phlegm.mp.
- 10 sputum.mp.
- 11 mucus.mp.
- 12 wheez\$.mp.
- 13 chest pain/
- 14 chest pain\$.mp.
- 15 (chest adj5 tight\$.tw.
- 16 ((lower respiratory or lrt) adj5 symptom\$.tw.
- 17 (lower airway adj5 symptom\$.tw.
- 18 ((trache\$ or wind pipe or lung\$ or bronch\$) adj3 symptom\$.tw.

19 exp lung/ or trachea/

20 symptom\$.tw.

21 19 and 20

22 or/7-18,21

23 exp asthma/
24 asthma\$.mp.
25 exp respiratory hypersensitivity/
26 exp bronchial hyperreactivity/
27 bronchial spasm/
28 bronchospas\$.mp.
29 exp Bronchoconstriction/
30 bronchoconstric\$.mp.
31 (bronch\$ adj3 constrict\$).mp.
32 (bronch\$ adj5 spas\$).mp.
33 (airway\$ adj5 (obstruct\$ or inflammation\$)).mp.
34 ((bronchial\$ or respiratory or airway\$ or lung\$) adj3 (hypersensitiv\$ or hyperreactiv\$ or allerg\$ or insufficiency)).mp.

35 or/23-34

36 Nitric Oxide/
37 nitric oxide.mp.
38 36 or 37
39 (exhal\$ or expir\$ or alveolar or fractional).mp.

40 38 and 39 (5228)

41 exhaled NO.mp.
42 eno.mp.
43 fe?no\$.mp.
44 (fractional adj2 NO).mp.

45 or/40-44

46 22 and 45

47 35 and 45

48 6 or 46 or 47

49 exp "Sensitivity and Specificity"/
 50 sensitivity.tw.
 51 specificity.tw.
 52 ((pre-test or pretest) adj probability).tw.
 53 post-test probability.tw.
 54 predictive value\$.tw.
 55 likelihood ratio\$.tw.
 56 or/49-55
 57 48 and 56

5. Review string #5

1 NIOX MINO.mp.
 2 aerocrine.mp.
 3 (niox adj5 (monitor\$ or chemiluminescence or analyser\$ or sensor or device\$ or desktop)).mp.
 4 NObreath.mp.
 5 bedfont.mp.
 6 or/1-5

Appendix S2

Inclusion and exclusion criteria

	Inclusion	Exclusion	Change from protocol
Population	Primary population are patients presenting with clinical characteristics suggestive of asthma. The main relevant subgroups within this population are: • Those presenting with clinical characteristics suggestive of asthma and who are difficult to diagnose. • Women during pregnancy • Older people • Smokers Studies were included if they recruited a wider population but report <i>a priori</i> subgroup analyses for the populations of relevance to this review.	Children <5 years old Studies that recruited a wider population and did not report <i>a priori</i> subgroup analyses for the populations of relevance to this review Animal models Unselected specific population (e.g fire fighters, obese, athletes)	None

	Inclusion	Exclusion	Change from protocol
Intervention	Use of NIOX MINO or NObreath in the diagnosis of asthma, either with or without another test. NIOX MINO and NObreath devices are set to record according to ATS 2005 criteria: • Expiratory flow rate of 50mL per second (0.05L/sec) • An exhalation time of $\geq$ 10 seconds for adults, $\geq$ 6 seconds for children If data were not available for the above interventions, studies were included if they reported clinical validity of $_{FE}NO_{50}$ measured by any chemiluminescent device with appropriate measurement methods: • Expiratory flow rate of 50mL per second (0.05L/sec) • An exhalation time of $\geq$ 10 seconds for adults, $\geq$ 6 seconds for children • Online measurement Studies that did not report any of	Expiratory flow rate not 50mL per second (0.05L/sec) An exhalation time of < 10 seconds in adults, < 6 seconds in children Offline measurements Alveolar nitrogen oxide or nasal nitrogen oxide measurements	The protocol stated that studies using the following cut-off would be included: • $_{FE}NO_{50}$ less than 25 ppb (< 20 ppb in children) indicates that eosinophilic inflammation and responsiveness to corticosteroids are less likely • $_{FE}NO_{50}$ greater than 50 ppb (> 35 ppb in children) indicates that eosinophilic inflammation and, in symptomatic patients, responsiveness to corticosteroids are likely • $_{FE}NO_{50}$ values between 25 ppb and 50 ppb (20–35 ppb in children) should be interpreted cautiously and with reference to

	Inclusion	Exclusion	Change from protocol
	these details were included and discussed in the narrative review.		the clinical context
	Studies using any cut-off value or combination of cut-off values were included		However, no studies using these exact cut-off values and protocol were found, so all cut-off values were included.
Comparator	Comprises any combination or selection of the tests and clinical characteristics described in the UK guideline for the diagnosis of asthma.	Uses tests to diagnose asthma that are not included in the BTS/SIGN guidelines or if the comparator includes the use of FE_{NO} measurements.	Studies using tests not in routine use in the UK but mentioned in the BTS/SIGN guidelines were included in the review
Outcome	End-to-end studies – include studies with relevant clinical outcomes (see management study outcomes) Clinical validity studies - Include studies that report data that allow the extraction of the numbers of patients who are true-positive, true-negative, false-positive and false-negative against the reference standard. Studies which report test failure rates were also included.	Does not report useable diagnostic validity data (i.e. extraction of the numbers of patients who are true-positive, true-negative, false-positive and false-negative against the reference standard).	None

	Inclusion	Exclusion	Change from protocol
Study design	End-to-end studies – (which follow patients from diagnostic test to clinical outcomes) If no evidence was found at this level, clinical validity studies (which compare the diagnosis of patients by the intervention with a reference standard) were included. These should be prospective cohort studies, cross sectional studies or retrospective cohort studies. If studies of these designs were not located, other study designs were considered (e.g. case control studies). Both studies deriving cut-off values for diagnosis and studies validating existing cut-off values for diagnosis will be included. Abstracts with comparable data that do not exist in full published studies	Preclinical and biological studies Editorials and opinion pieces Studies only published in languages other than English	Studies published as abstracts not reporting sufficient methodological details to allow critical appraisal of study quality were NOT excluded.

	Inclusion	Exclusion	Change from protocol
Setting	Primary care, secondary care, out-patient clinic or specialist clinic	Emergency care diagnosis of exacerbation	None

_{FE}NO, fractional exhaled nitric oxide; _{FE}NO₅₀, fraction of exhaled nitric oxide measured at a flow rate of 50 mL/s; pb, parts per billion; ATS, American Thoracic Society; BTS, British Thoracic Society; SIGN, Scottish Intercollegiate Guidelines Network.

Appendix S3

Data extraction forms

Review type	Data extracted for:			
	Study background	Methods and devices	Study sample characteristics	Results
Diagnostic	Author/ year Source of funding Study design (prospective / retrospective) Validation or derivation? Dates Description of study timeline Setting Inclusion/ exclusion criteria Inclusion age range Code for age (children/ adolescents/ adults) Code for population (e.g. symptoms of asthma; difficult to diagnose) Medication permitted during study	Name of device Code for device(Niox mino; Nobreath; Chemiluminescence) Code for $FENO_{50}$ measurement method (e.g. ATS/ERS; Bespoke) Measurement method details if bespoke $FENO_{50}$ cut-off points Description of reference standard Code for reference standard (e.g. ATS/ERS 2005) Definition of asthma	Total N patients recruited Number not followed up (with reasons) Number of patients analysed Mean age ($\pm$ SD) Gender n/N male (%) FEV1%predicated Mean $FENO_{50}$ ppb ($\pm$ SD)] Symptoms Smokers Atopic/ allergic status Other symptoms Statistical methods	Prevalence of asthma Prevalence of positive result by reference standard Cut-off value True positive False positive False negative True negative Total Prevalence of ICS responsiveness (95% CI) True positive False positive False negative True positive Sensitivity asthma (95% CI) Specificity asthma (95% CI) Sensitivity ICS responsiveness (95% CI) Specificity ICS responsiveness (95% CI)

Abbreviations: AUC, area under the curve; CI, confidence interval; FEV1%, forced expiratory volume

Appendix S4

Quality assessment scoring criteria for diagnostic accuracy studies.

Risk of bias in diagnostic studies (for both children and adult populations) was assessed and described using the Bristol University QUADAS-2 tool. QUADAS-2 is structured around four domains of potential sources of bias in primary diagnostic studies: patient selection; index test; reference standard; and flow and timing. There are signalling questions within each of these domains which allow researchers to overview the potential sources of bias therein, and a summary domain score can be generated to provide an indication of overall potential for bias in each aspect of a study's methodology. These signalling questions, and our approach to scoring them, are detailed below. It should be noted that, in our risk of bias tables, we report only the domain summary scores, although we also narratively summarise our responses to signalling questions to support these judgments in the review text.

Domain 1: Patient selection

Signalling question 1: Was a consecutive or random sample of patients enrolled?

- Score 'yes' if the report states enrolment was consecutive or random
- Score 'no' if the report states another method of sampling
- Score 'unclear' if insufficient information was provided to make a judgment

Signalling question 2: Was a case-control design avoided?

- Score 'yes' if the study was not case-control
- Score 'no' if the study was a case-control
- Score 'unclear' if insufficient information was provided to make a judgment

Signalling question 3: did the study avoid inappropriate exclusions?

With respect to the current review, the population of interest was patients presenting with clinical characteristics of asthma, or those who are "difficult to diagnose", that is, patients who have already undergone some of the tests for asthma in the UK pathway and have not yet been confirmed to have asthma. The review scope also sought data on subgroups, in particular, women during pregnancy, older people, and smokers/passive smokers. Hence, this question was answered with respect to these groups. Where there were ambiguities, two reviewers would discuss whether the population was appropriate.

During the course of the review, it became clear that some studies made exclusions relating to previous respiratory infections, comorbidities and use of asthma treatments. We considered whether

this constituted an inappropriate exclusion, but concluded that it was reasonable to assume these patients would be treated differently in clinical practice. Consequently, such exclusions did not lead to a negative score. Studies which imposed inappropriate limits on recruitment such as only army recruits, non-smokers or only up to a certain age were scored negatively as these changes were likely to alter the patient spectrum with respect to difficult to diagnose patients.

- Score 'yes' if the study had appropriately selected patients conforming to the groups outlined above
- Score 'no' if the study made inappropriate exclusions from the group it set out to select, e.g. because of missing test data
- Score 'unclear' if insufficient information was provided to make a judgment

Summary domain score: Could the selection of patients have introduced bias?

- Score as 'low risk' if the study scored 'yes' on all signalling questions above
- Score as 'high risk' if the study scored 'no' or 'unclear' on ≥ 2 of the individual items, or on either of case-control design and inappropriate exclusions
- Score as 'unclear risk' if the study scored 'no' or 'unclear' on signalling question 1.

Domain 2: Index tests

Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?

- Score 'yes' if index test was interpreted blind to the reference standard or the index test was clearly interpreted before the reference standard was known. Measurement before the reference standard alone was sufficient to score well as the test is objective.
- Score 'no' if results of reference standard (UK guidelines pathway) were already known, or if parts of the reference standard downstream of the position of the test in the UK pathway were already known. This will need to be scored with reference to the patient population.
- Score 'unclear' if unclear

Signalling question 2: If a threshold was used, was it pre-specified?

- Score 'yes' if pre-specified cut of values were used (validation study)
- Score 'no' if cut-off values were fitted to the data (derivation study)
- Score 'unclear' if unclear

Summary domain score: Could the conduct or interpretation of the index tests have introduced bias?

- Score as 'low risk' if the study scored 'yes' on both signalling questions
- Score as 'high risk' if cut-off values were fitted to the data (since derivation studies are likely to over-estimate the true diagnostic accuracy of a technology relative to clinical practice), or if the study scored 'no' or 'unclear' on both signalling questions
- Score as 'unclear' if the study scored 'no' or 'unclear' on signalling question 1 only

During the course of the study, it was noticed that there is no contingency in this description for studies that scored unclear in question 2 only. These studies scored unclear overall.

Domain 3: Reference standard

Signalling question 1: Is the reference standard likely to correctly classify the target condition?

No reference standard for Asthma is 100% sensitive or specific. Possibly the only way this could be achieved is with long term follow-up, but even this might be confounded by the fact that asthma can remiss and develop (eg as a comorbidity) over time. Hence, in the original NICE project, this item was scored with respect to UK guidelines:

- Score ‘yes’ if the reference standard is correct according to UK guidelines for that patient group/position in the diagnostic pathway
- Score ‘no’ if the reference standard does not conform with UK guidelines i.e. uses a test that is not within the UK guidelines
- Score ‘unclear’ if unclear

However, to facilitate a wider international perspective, this was changed for this publication to score as follows:

- Score “yes” if the reference standard is either the next step in the generic pathway (Figure 1), or is a highly accurate reference standard such as airway hyper-responsiveness testing, or a sequence of tests including airway hyper-responsiveness testing and airway reversibility testing.
- Score “no” if the reference standard is not the next step in the diagnostic pathway, and is a test with low accuracy such as ICS responsiveness.
- Score “unclear” if the reference standard or population was not clear

Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?

- Score yes if the reference standard was interpreted blind to the index test or the reference standard was clearly interpreted before the index test was known.
- Score no if the results of the index test were known.
- Score unclear if unclear

Summary domain score: Could the conduct or interpretation of the index tests have introduced bias?

- Score as ‘low risk’ if the study scored ‘yes’ on both signalling questions
- Score as ‘high risk’ if the study scored ‘no’ or ‘unclear’ on both signalling questions

- Score as 'unclear' if the study scored 'no' or 'unclear' on either of the two signalling questions

Domain 4: Flow and timing

Signalling question 1: Was there an appropriate interval between index test(s) and reference standard?

- Score 'yes' if tests were conducted consecutively
- Score 'no' if index and reference tests were conducted >1 week apart
- Score unclear if unclear

Signalling question 2: Did all patients receive a reference standard?

- Score yes if yes
- Score no if no
- Score unclear if unclear

Signalling question 3: Did patients receive the same reference standard (this could be a series of tests applied according to an algorithm)?

- Score yes if yes
- Score no if no
- Score unclear if unclear

Signalling question 4: Were all patients included in the analysis?

- Score yes if yes (note: studies where patients were excluded from the analysis at selection e.g. where test data was missing are scored negatively for patient spectrum and should not be scored negatively again here)
- Score no if no
- Score unclear if unclear

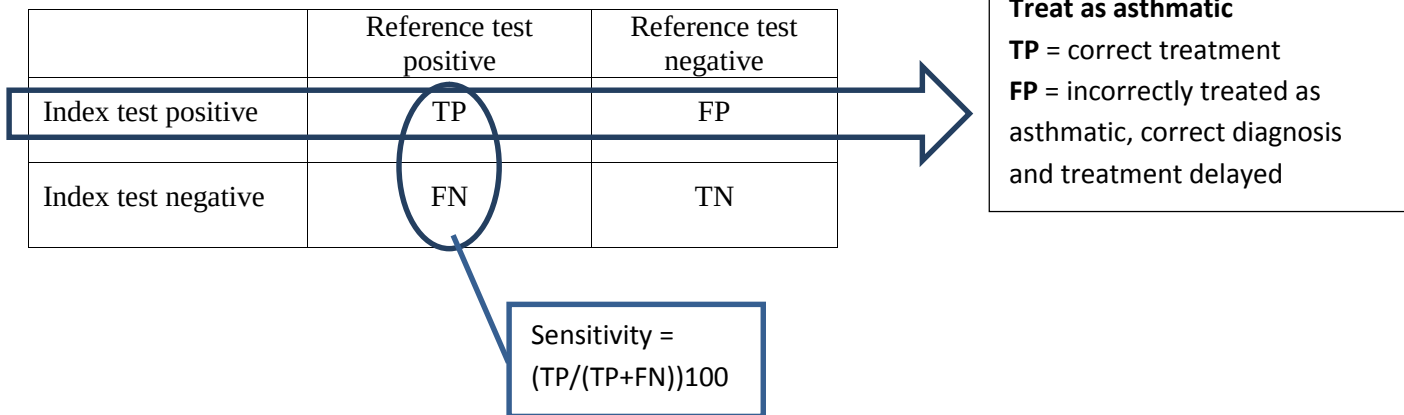
Summary domain score: Could the patient flow have introduced bias?

- Score 'low risk' if the study scored 'yes' on all signalling questions
- Score 'high risk' if the study scored 'no' or 'unclear' on ≥ 3 items
- Score 'unclear risk' if the study scored 'no' or 'unclear' on up to two items.

Appendix S5

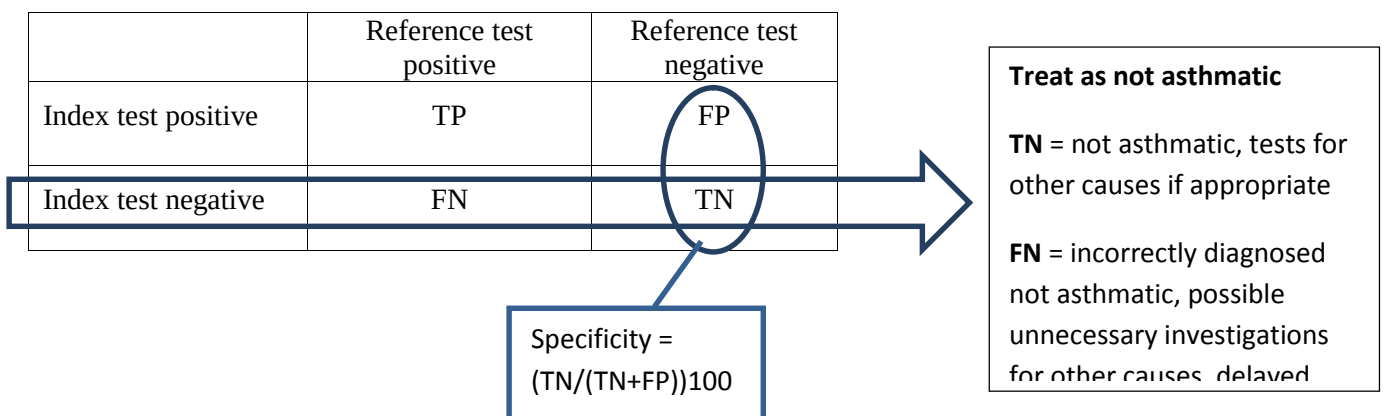
Rule-in and rule-out scenarios

a. Rule-in scenario



TP, true positive; FP, false positive; TN, true negative; FN, false negative

b. Rule-out scenario



TP, true positive; FP, false positive; TN, true negative; FN, false negative

The arrows indicate how patients in those categories will be treated. Note that high sensitivity alone will not prevent false positive results and high specificity will not prevent false negative results. As such, specificity also has a part to play in determining the best rule-out scenario and vice versa for sensitivity and rule-in scenarios. However, as the relative clinical consequences of TP, FP, TN, FN are not known, it is difficult to know where the balance between benefits lies by considering only the sensitivity and specificity.

Appendix S6

Table of studies ordered according to % atopic patients. Data are shown for the highest sum of sensitivity and specificity.

<u>Study author, year</u>	<u>Device</u>	<u>Reference standard</u>	<u>Cut-off</u>	<u>Sens</u>	<u>Spec</u>	<u>% Atopic</u>
El Halawani 2003	Sievers 280A	Exercise challenge	12	100	31	0
Fukuhara, 2011	NA623 (Chest MI, Tokyo, Japan)	At least 2 of: induced sputum eosinophilia, AHR, or ART plus exclusion of other lung diseases.	40	78.6	89.5	23
Katsoulis 2013	NIOX MINO	AHR (MCT)	32	47	85	45.5
Pizzimenti 2009	NIOX MINO	AHR (Methacholine)	55	10	67.2	47
Schleich, 2012	Niox	AHR (MCT)	34	35	95	48
Giovannini 2014	HypAir FeNO;	AHR (MCT)	30	14.3	100	50
Sastre 2013	NIOX MINO	Specific inhalation challenge	25	60	80	66.2
Cordeiro 2011	Niox-Flex	ART, AHR (histamine)	27	78	92	71
	Niox-Flex; Airway reversibility		27	87	90	71
De La Barra 2011	Niox	ART	41.7	NR	NR	77
Smith 2005	Niox	ATS 1987 guidelines plus one of: ART (ICS), AHR (MCT).	47	55.6	92	77
Pedrosa 2010	NIOX MINO	AHR (MCT)	40	74.3	72.5	87
Backer 2014	NIOX MINO	AHR (mannitol)	25	36.2	86.8	90
Heffler 2006	Niox	AHR (MCT) or ART	36	77.8	60	92
Katsoulis 2013	NIOX MINO	AHR (MCT)	26	55	85	100
Prieto 2009	Niox	ART (ICS)	20	53	63	100

Appendix S7

- i) Table of all data reported, ordered by study author.

Author, year	Prevalence	FeNO_{50} cut-off ppb	TP	FP	FN	TN	Sensitivity	Specificity
Arora 2006	138/172 (80.2%)	6	133	34	5	0	96.4	0
Arora 2006	138/172 (80.2%)	7	131	33	7	1	94.6	2.9
Arora 2006	138/172 (80.2%)	8	130	31	8	3	94.2	8.8
Arora 2006	138/172 (80.2%)	9	127	30	11	4	92	11.8
Arora 2006	138/172 (80.2%)	10	119	28	19	6	86.2	17.6
Arora 2006	138/172 (80.2%)	11	115	26	23	8	83.3	23.5
Arora 2006	138/172 (80.2%)	12	113	25	25	9	81.9	26.5
Arora 2006	138/172 (80.2%)	13	110	21	28	13	79.7	38.2
Arora 2006	138/172 (80.2%)	14	102	19	36	15	73.9	44.1
Arora 2006	138/172 (80.2%)	15	98	19	40	15	71	44.1
Arora 2006	138/172 (80.2%)	17	87	14	51	20	63	58.8
Arora 2006	138/172 (80.2%)	18	83	14	55	20	60.1	58.8
Arora 2006	138/172 (80.2%)	19	78	13	60	21	56.5	61.8
Arora 2006	138/172 (80.2%)	20	73	11	65	23	52.9	67.6
Arora 2006	138/172 (80.2%)	25	56	7	82	27	40.6	79.4
Arora 2006	138/172 (80.2%)	30	45	7	93	27	32.6	79.4
Arora 2006	138/172 (80.2%)	40	32	3	106	31	23.2	91.2
Arora 2006	138/172 (80.2%)	46	23	0	115	34	16.7	100
Arora 2006	138/172 (80.2%)	16	92	17	46	17	66.7	50
Backer 2014	58/141 (41%)	25	21	11	37	72	36	87
Bobolea 2012	6/30 (20.0%)	30	6	17	0	7	100	29.2
Brannan 2013	76/401 (19%)	47	23	12	44	322	30.2	96.3
Chancafe-Morgan 2013	12/30 (40%)	35	9	3	3	15	75	83.3
Cordeiro 2011	42/114	27	33	6	9	66	78	92

	(36.8%)							
De La Barra 2011	NR	25	10	17	2	23	83.3	57.5
De La Barra 2011	NR	40	9	12	3	28	75	70
De La Barra 2011	NR	50	7	8	5	32	58.3	80
De La Barra 2011	NR	70	5	5	7	35	41.7	87.5
De La Barra 2011	NR	90	5	3	7	37	41.7	92.5
De La Barra 2011	NR	110	3	2	9	38	25	95
De La Barra 2011	NR	130	2	2	10	38	16.7	95
De La Barra 2011	NR	150	2	2	10	38	16.7	95
El Halawani 2003	7/49 (14.3%)	12	7	29	0	13	100	31
Fortuna, 2007		20	17	10	5	18	77	64
Fukuhara, 2011	42/61 (68.85%)	40	33	2	9	17	78.6	89.5
Giovannini 2014	21/42 (50%)	30	3	0	18	21	14	100
Hahn 2007	38/64 (59.4%)	35	36	5	2	21	95	80
Hahn 2007	38/64 (59.4%)	38	34	4	4	22	90	85
Heffler 2006	18/48	10	18	29	0	1	100	3.3
Heffler 2006	18/48	15	18	26	0	4	100	13.3
Heffler 2006	18/48	20	18	20	0	10	100	33.3
Heffler 2006	18/48	25	18	16	0	14	100	46.7
Heffler 2006	18/48	30	14	15	4	15	77.8	50
Heffler 2006	18/48	34	14	14	4	16	77.8	53.3
Heffler 2006	18/48	36	14	12	4	18	77.8	60
Heffler 2006	18/48	40	11	11	7	19	61.1	63.3
Heffler 2006	18/48	45	11	8	7	22	61.1	73.3
Heffler 2006	18/48	50	10	7	8	23	55.6	76.7
Heffler 2006	18/48	55	9	6	9	24	50	80
Heffler 2006	18/48	60	9	4	9	26	50	86.7
Heffler 2006	18/48	65	8	4	10	26	44.4	86.7
Heffler 2006	18/48	75	8	3	10	27	44.4	90
Heffler 2006	18/48	80	7	1	11	29	38.9	96.7
Heffler 2006	18/48	85	5	1	13	29	27.8	96.7
Heffler 2006	18/48	100	5	0	13	30	27.8	100
Hsu 2013	38/81 (46.9%)	30	37	14	1	29	97.4	65.8
Hsu 2013	38/81 (46.9%)	33.9	36	9	2	34	94.7	76.3
Katsoulis 2013	48/112 (42.9%)	10	39	39	9	25	81	39
Katsoulis 2013	48/112 (42.9%)	15	35	33	13	31	73	49
Katsoulis 2013	48/112 (42.9%)	20	31	26	17	38	64	60
Katsoulis 2013	48/112	25	24	16	24	48	51	75

	(42.9%)							
Katsoulis 2013	48/112 (42.9%)	30	24	12	24	52	49	82
Mathew 2011	20/84 (23.8%)	NR	2	21	18	43	10	67
Nickels 2014	267/1322 (20.2%)	NR	89	160	178	895	33	85
Pedrosa 2010	35/114 (30.7%)	40	26	22	9	57	74.3	72.5
Pizzimenti 2009	14/156 (9.0%)	55	11	17	3	125	78	88
Prieto 2009	19/43 (44%)	20	10	9	9	15	53	63
Sastre 2014	45/68 (66.2%)	25	27	5	18	18	60	78
Sato 2008	48/71 (67.6%)	38.8	38	2	10	21	79.2	91.3
Schleich, 2012	82/174 (47.1%)	34	29	4	53	88	35.4	95.4
Schneider 2009	75/160 (46.9%)	12	64	65	11	20	85	24
Schneider 2009	49/101 (48.5%)	15	38	29	11	23	78	45
Schneider 2009	75/160 (46.9%)	20	48	36	27	49	64	58
Schneider 2009	75/160 (46.9%)	35	24	14	51	71	32	84
Schneider 2009	49/101 (48.5%)	46	17	5	32	47	35	90
Schneider 2009	75/160 (46.9%)	46	24	6	51	79	32	93
Schneider 2009	75/160 (46.9%)	76	10	0	65	85	13	100
Schneider 2009	75/160 (46.9%)	16	52	40	23	45	69	53
Schneider 2013	154/393 (39.2%)	9	146	209	7	31	96	13
Schneider 2013	154/393 (39.2%)	12	135	167	23	68	85	29
Schneider 2013	154/393 (39.2%)	16	105	128	46	114	70	47
Schneider 2013	154/393 (39.2%)	20	91	89	62	151	60	63
Schneider 2013	154/393 (39.2%)	25	75	59	79	180	49	75
Schneider 2014	83/302 (27.5%)	26	39	59	44	160	47	73
Schneider 2013	154/393 (39.2%)	35	50	29	104	210	33	88
Schneider 2013	154/393 (39.2%)	41	42	20	112	219	27	92
Schneider 2013	154/393 (39.2%)	42	40	20	114	219	26	92
Schneider 2013	154/393	43	39	19	115	220	25	92

	(39.2%)							
Schneider 2013	154/393 (39.2%)	44	39	19	115	220	25	92
Schneider 2013	154/393 (39.2%)	45	38	19	116	220	23	92
Schneider 2013	154/393 (39.2%)	46	38	17	116	222	27	92
Schneider 2013	154/393 (39.2%)	71	27	7	127	232	18	97
Smith 2004	17/47 (36.2%)	20	14	6	2	22	88	79
Smith 2005	27/52 (51.92%)	15	5	12	22	13	18	52
Smith 2005	27/52 (51.92%)	15	22	13	5	12	81	48
Smith 2005	27/52 (51.92%)	47	15	2	12	23	55	92
Zhang 2011	39/106 (36.8%)	36	32	5	7	62	82	93
Zhang 2011	39/106 (36.8%)	40	29	9	10	58	75	86

ii) Table of all data reported, ordered by cut-off value

Alternating grey and white areas delineate categories detailed in column 1.

	Author, year	Prevalence	NO ₅₀ cut-off ppb	TP	FP	FN	TN	Sensitivity	Specificity
Sensitivity consistently ≥80%	Arora 2006	138/172 (80.2%)	6	133	34	5	0	96.4	0
	Arora 2006	138/172 (80.2%)	7	131	33	7	1	94.6	2.9
	Arora 2006	138/172 (80.2%)	8	130	31	8	3	94.2	8.8
	Arora 2006	138/172 (80.2%)	9	127	30	11	4	92	11.8
	Schneider 2013	154/393 (39.2%)	9	146	209	7	31	96	13
	Heffler 2006	18/48	10	18	29	0	1	100	3.3
	Arora 2006	138/172 (80.2%)	10	119	28	19	6	86.2	17.6
	Katsoulis 2013	48/112 (42.9%)	10	39	39	9	25	81	39
	Arora 2006	138/172 (80.2%)	11	115	26	23	8	83.3	23.5

	El Halawani 2003	7/49 (14.3%)	12	7	29	0	13	100	31
	Schneider 2009	75/160 (46.9%)	12	64	65	11	20	85	24
	Arora 2006	138/172 (80.2%)	12	11 3	25	25	9	81.9	26.5
	Schneider 2013	154/393 (39.2%)	12	13 5	16 7	23	68	85	29
	Arora 2006	138/172 (80.2%)	13	11 0	21	28	13	79.7	38.2
	Arora 2006	138/172 (80.2%)	14	10 2	19	36	15	73.9	44.1
	Smith 2005	27/52 (51.92%)	15	5	12	22	13	18	52
	Heffler 2006	18/48	15	18	26	0	4	100	13.3
	Arora 2006	138/172 (80.2%)	15	98	19	40	15	71	44.1
	Schneider 2009	49/101 (48.5%)	15	38	29	11	23	78	45
	Smith 2005	27/52 (51.92%)	15	22	13	5	12	81	48
	Katsoulis 2013	48/112 (42.9%)	15	35	33	13	31	73	49
	Schneider 2013	154/393 (39.2%)	16	10 5	12 8	46	114	70	47
	Arora 200690	138/172 (80.2%)	16	92	17	46	17	66.7	50
	Schneider 200982,83	75/160 (46.9%)	16	52	40	23	45	69	53
	Arora 2006	138/172 (80.2%)	17	87	14	51	20	63	58.8
	Arora 2006	138/172 (80.2%)	18	83	14	55	20	60.1	58.8
	Arora 2006	138/172 (80.2%)	19	78	13	60	21	56.5	61.8
20-25 ppb range	Prieto 2009	19/43 (44%)	20	10	9	9	15	53	63
	Heffler 2006	18/48	20	18	20	0	10	100	33.3
	Schneider 2009	75/160 (46.9%)	20	48	36	27	49	64	58
	Schneider 2013	154/393 (39.2%)	20	91	89	62	151	60	63
	Arora 2006	138/172 (80.2%)	20	73	11	65	23	52.9	67.6
	Smith 2004	17/47 (36.2%)	20	14	6	2	22	88	79
	Fortuna, 2007		20	17	10	5	18	77	64
	Katsoulis 2013	48/112 (42.9%)	20	31	26	17	38	64	60
	De La Barra 2011	NR	25	10	17	2	23	83.3	57.5
	Sastre 2014	45/68 (66.2%)	25	27	5	18	18	60	78
	Backer 2014	58/141 (41%)	25	21	11	37	72	36	87
	Heffler 2006	18/48	25	18	16	0	14	100	46.7
	Schneider 2013	154/393 (39.2%)	25	75	59	79	180	49	75
	Arora 2006	138/172 (80.2%)	25	56	7	82	27	40.6	79.4
	Katsoulis 2013	48/112 (42.9%)	25	24	16	24	48	51	75
	Schneider 2013	83/302 (27.5%)	26	39	- 16 0	44	160	47	73
	Cordeiro 2011	42/114 (36.8%)	27	33	6	9	66	78	92
	Hsu 2013	38/81 (46.9%)	30	37	14	1	29	97.4	65.8
	Giovannini 2014	21/42 (50%)	30	3	0	18	21	14	100
	Katsoulis 2013	48/112 (42.9%)	30	24	12	24	52	49	82
	Bobolea 2012	6/30 (20.0%)	30	6	17	0	7	100	29.2
	Heffler 2006	18/48	30	14	15	4	15	77.8	50

	Arora 2006	138/172 (80.2%)	30	45	7	93	27	32.6	79.4
	Hsu 2013	38/81 (46.9%)	33.9	36	9	2	34	94.7	76.3
	Schleich, 2012	82/174 (47.1%)	34	29	4	53	88	35.4	95.4
	Heffler 2006	18/48	34	14	14	4	16	77.8	53.3
	Hahn 2007	38/64 (59.4%)	35	36	5	2	21	95	80
	Chancafe-Morgan 2013	12/30 (40%)	35	9	3	3	15	75	83.3
	Schneider 2009	75/160 (46.9%)	35	24	14	51	71	32	84
	Schneider 2013	154/393 (39.2%)	35	50	29	10 4	210	33	88
	Zhang 2011	39/106 (36.8%)	36	32	5	7	62	82	93
	Heffler 2006	18/48	36	14	12	4	18	77.8	60
	Hahn 2007	38/64 (59.4%)	38	34	4	4	22	90	85
	Sato 2008	48/71 (67.6%)	38.8	38	2	10	21	79.2	91.3
	De La Barra 2011	NR	40	9	12	3	28	75	70
	Pedrosa 2010	35/114 (30.7%)	40	26	22	9	57	74.3	72.5
	Zhang 2011	39/106 (36.8%)	40	29	9	10	58	75	86
	Arora 2006	138/172 (80.2%)	40	32	3	10 6	31	23.2	91.2
	Fukuhara, 2011	42/61 (68.85%)	40	33	2	9	17	78.6	89.5
	Heffler 2006	18/48	40	11	11	7	19	61.1	63.3
	Schneider 2013	154/393 (39.2%)	41	42	20	11 2	219	27	92
	Schneider 2013	154/393 (39.2%)	42	40	20	11 4	219	26	92
	Schneider 2013	154/393 (39.2%)	43	39	19	11 5	220	25	92
	Schneider 2013	154/393 (39.2%)	44	39	19	11 5	220	25	92
	Schneider 2013	154/393 (39.2%)	45	38	19	11 6	220	23	92
	Heffler 2006	18/48	45	11	8	7	22	61.1	73.3
	Schneider 2009	49/101 (48.5%)	46	17	5	32	47	35	90
	Schneider 2013	154/393 (39.2%)	46	38	17	11 6	222	27	92
	Schneider 2009	75/160 (46.9%)	46	24	6	51	79	32	93
	Arora 2006	138/172 (80.2%)	46	23	0	11 5	34	16.7	100
	Smith 2005	27/52 (51.92%)	47	15	2	12	23	55	92
	Brannan 2013	76/401 (19%)	47	23	12	44	322	30.2	96.3
50p pb	De La Barra 2011	NR	50	7	8	5	32	58.3	80
	Heffler 2006	18/48	50	10	7	8	23	55.6	76.7
	Pizzimenti 2009	14/156 (9.0%)	55	11	17	3	125	78	88
	Heffler 2006	18/48	55	9	6	9	24	50	80
	Heffler 2006	18/48	60	9	4	9	26	50	86.7
	Heffler 2006	18/48	65	8	4	10	26	44.4	86.7
Spe cif	De La Barra 2011	NR	70	5	5	7	35	41.7	87.5
	Schneider 2013	154/393 (39.2%)	71	27	7	12	232	18	97

						7			
	Heffler 2006	18/48	75	8	3	10	27	44.4	90
	Schneider 2009	75/160 (46.9%)	76	10	0	65	85	13	100
	Heffler 2006	18/48	80	7	1	11	29	38.9	96.7
	Heffler 2006	18/48	85	5	1	13	29	27.8	96.7
	De La Barra 2011	NR	90	5	3	7	37	41.7	92.5
	Heffler 2006	18/48	100	5	0	13	30	27.8	100
	De La Barra 2011	NR	110	3	2	9	38	25	95
	De La Barra 2011	NR	130	2	2	10	38	16.7	95
	De La Barra 2011	NR	150	2	2	10	38	16.7	95
	Mathew 2011	20/84 (23.8%)	NR	2	21	18	43	10	67
	Nickels 2014	267/1322 (20.2%)	NR	89	16	17	895	33	85

iii) Table of studies which used ICS responsiveness as the reference standard.

Author, year	Prevalence	FeNO_{50} cut-off ppb	TP	FP	FN	TN	Sensitivity	Specificity
Prieto 2009	19/43 (44%)	20	10	9	9	15	53	63
De La Barra 2011	NR	25	10	17	2	23	83.3	57.5
Hsu 2013	38/81 (46.9%)	30	37	14	1	29	97.4	65.8
Hsu 2013	38/81 (46.9%)	33.9	36	9	2	34	94.7	76.3
Hahn 2007	38/64 (59.4%)	35	36	5	2	21	95	80
Hahn 2007	38/64 (59.4%)	38	34	4	4	22	90	85
De La Barra 2011	NR	40	9	12	3	28	75	70
De La Barra 2011	NR	50	7	8	5	32	58.3	80
De La Barra 2011	NR	70	5	5	7	35	41.7	87.5
De La Barra 2011	NR	90	5	3	7	37	41.7	92.5
De La Barra 2011	NR	110	3	2	9	38	25	95
De La Barra 2011	NR	130	2	2	10	38	16.7	95
De La Barra 2011	NR	150	2	2	10	38	16.7	95