
Article:
Harnan, S.E., Essat, M. orcid.org/0000-0003-2397-402X, Gomersall, T. et al. (4 more authors) (2017) Exhaled nitric oxide in the diagnosis of asthma in adults: a systematic review. Clinical and Experimental Allergy, 47 (3). pp. 410-429. ISSN 0954-7894

https://doi.org/10.1111/cea.12867

Received Date : 20-Apr-2016

Revised Date : 24-Oct-2016

Accepted Date : 01-Nov-2016

Article type : Systematic reviews and Meta-analysis

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

Running head: $_{FE}$NO for asthma diagnosis

Ms Sue E Harnan, Health Economics and Decision Science, ScHARR, University of Sheffield
Dr Munira Essat, Health Economics and Decision Science, ScHARR, University of Sheffield
Dr Tim Gomersall, Health Economics and Decision Science, ScHARR, University of Sheffield
Dr Paul Tappenden, Health Economics and Decision Science, ScHARR, University of Sheffield
Professor Ian Pavord, Respiratory Medicine Unit, Nuffield Department of Medicine, University of Oxford
Professor Mark Everard, University of Western Australia, School of Paediatrics & Child Health, Princess Margaret Hospital, Australia
Dr Rod Lawson, Department of Respiratory Medicine, Royal Hallamshire Hospital

Corresponding author:

Ms Sue Harnan,
Senior Research Fellow
Health Economics and Decision Science
ScHARR
Regent Court
30 Regent Street
Sheffield S1 4DA
s.harnan@sheffield.ac.uk

Keywords: Asthma; diagnosis; systematic review; exhaled nitric oxide; meta analysis

Abstract

OBJECTIVES: To identify and synthesise 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 to 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: 4861 records were identified originally and 1312 in an update. 27 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 12ppb to 55ppb, but did not produce high accuracy. 100% sensitivity or 100% specificity were reported by some studies indicating potential use as a rule-in or rule-out strategy.

CONCLUSIONS: $_{FE}NO_{50}$ 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.

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;7-10] do not aim to diagnose steroid-responsive disease as a classification, but aim to diagnose asthma according to its classical definition. As such, the focus of this work was 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. 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 matches 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,[11-17] we are only aware of three systematic reviews. Two were only briefly reported,[9;18] and one had only a limited scope ($_{FE}$NO to diagnose exercise induced bronchoconstriction).[19] 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 (DAP). 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[20] and one in children[21]. 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[22]. In that report, we narrowed the evidence base to studies with relevance to UK practice. Here, we report both 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,[23] and an update in 2016,[24] 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)[25], 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 hyper-responsiveness (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;9] to identify patients who should respond to ICS therapy (position B) and avoid "false negatives" when using a trial of treatment e.g. 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 a lot of 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 down-stream 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 Online Supplement A, 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 and $_{FE}$NO, 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. RCT and diagnostic study filters were applied.

Study selection

Studies were considered for inclusion in the review by one reviewer against the inclusion criteria summarised here, and provided in detail in the online supplement B. Studies were included if they recruited participants presenting with the symptoms of asthma or reported a subgroup of such

patients. $_{FE}NO_{50}$ could be measured by any device (chemiluminescent or electrochemical), used in accordance with the ATS 2005 criteria (flow rate of 50mL/s, exhalation time ≥10 seconds (adults)/≥6 seconds (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 versus ICS-responsiveness were included, allowing an exploration of $_{FE}NO_{50}$ for identification of steroid-responsive disease. End-to-End studies (studies which 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[26] 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 (ie 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[27] and the Cochrane handbook[28] and piloted on two studies. A full list of extracted fields are provided in Online Supplement C, 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 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),[29] adapted to the specifics of the review (Online Supplement D). 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 sub-grouped according to the patient spectrum (position in Figure 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 (ppb), sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), and each of which characterised 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 Supplement E 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 (35citations) were included in the review (Figure 2).

Study characteristics

Of the 27 studies included in the review, 15 studies (19 citations) were conducted in adults,[30-48] four studies (six citations) in adults plus adolescents,[49-54] three studies (three citations) in all age groups[55-57] and five studies (seven citations) in an unspecified age range.[58-64]

Key study and patient characteristics are presented in Table 1. Mean ages ranged from 37 to 49 years, except four studies recruited younger cohorts[30;31;52;54] and two recruited older cohorts.[34;39] Gender ranged from 33% male to 85%. One study excluded atopic patients,[31] one recruited only atopic patients[41] 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 ppb[35] and the highest at 74.5 ppb.[32] Most studies used chemiluminescent analysers (n=14), [30-32;34-37;41;43;44;49-51;56;63] whilst nine studies used NIOX MINO (a hand-held electrochemical device), [38;42;45-48;52;54;57;58;61] and four studies did not report which device was used. [39;40;55;62;64] Twelve studies did not specify whether corticosteroids were used by patients prior to or during testing[35;37;39;40;49;52-54;57-64] whilst corticosteroid use was either stopped (3 studies) [34;42;45-48] or patients who used them were excluded (11 studies).[30-33;38;41;43;44;50;55;56] One study had a mix of users and non-users. [36] 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 to 40 ppb 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 Online Supplement E. Twelve subgroups were identified (grouped by position in the pathway and reference standard used), and for each, methodological quality (Figure 3) according to QUADAS 2[29] 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 & ART: There were seven studies in this category.[32-34;45-47;50;55;56] Methodological quality was mixed (Figure 3), with no study scoring particularly well. The index test domain scored high risk most often, as many studies did not use a predefined 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[32;34] (≥20 ppb and 40 ppb). Derived values for the best sum similarly ranged from 20 ppb to 47 ppb. 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,[45-47;50] but a full range by only two (see Online Supplement F).[45-47] In these, rule-out scenario cut-off values were 9 ppb and 12 ppb respectively, with sensitivities of 96% (specificity 13%) and 85% (specificity 24%). Rule-in scenario cut-off values were 71 ppb and 76 ppb, with specificities of 97% (sensitivity 18%) and 100% (sensitivity 13%) respectively, though specificity of >90% was achieved at the lower cut-offs of >41 and 46 ppb respectively.

ii) Reference standard of airway hyper-responsiveness only: [35;54] Methodological quality was poor or unclear. Both studies have limited generalisability (low mean age [54]; very low mean $_{FE}NO_{50}$[35]). Both studies used a pre-specified cut-off value (25 ppb and 30 ppb). Both reported high specificity (86.8% and 100%). No other cut-off values were reported.

iii) Reference standard of ART only: This study[51] was a re-analysis of Smith 2005 et al.[50] using a different reference standard (ART instead of a sequence of tests). Methodological quality was poor or unknown. A low cut-off of 25 ppb (rule-out) had a sensitivity of 83.3% (specificity 57.5%) and a high cut-off of 90 ppb (rule-in) had a specificity of 92.5% (sensitivity 41.7%).

iv) Reference standard of pathway plus long-term follow-up: This study[48] was a long-term follow-up of Schneider et al, 2013.[47] 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 26 ppb) were very similar to the original article,[47] which used a reference standard of the pathway alone (Table 2).

Position A: Asthma symptoms plus one additional symptom

v) Reference standard of airway hyper-responsiveness: In this study[49] patients had rhinitis with asthma symptoms. Only patient selection was at low risk of bias. The derived best sum cut-off (36 ppb) was within the same range as other studies at position A, though accuracy was unimpressive at 77.8% sensitivity and 60% specificity. However, 100% sensitivity was achieved at 25 ppb and 100% specificity at 100 ppb, and 90% specificity at 75 ppb.

Position C: negative ART

vi) Reference standard of airway hyper-responsiveness: One study recruited patients who had a negative ART,[38] but all were army recruits (introducing spectrum bias as excludes certain ages, comorbidities etc). Cut-offs from 10 ppb to 30 ppb were reported, which is unlikely to include the cut-off with the best specificity. The derived best sum cut-off was 32 ppb (47% and 85% respectively). A cut-off of 10 ppb yielded 81% sensitivity (specificity 39%), whilst the 30 ppb cut-off yielded 82% specificity (sensitivity 49%).

Position C or D: referred for AHR

vii) Reference standard of airway hyper-responsiveness: Two studies stated patients were referred for AHR testing, but did not state for what reason,[63;64] and one stated patients were at a tertiary referral centre.[39;40] 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.[39;64] One study used mannitol[63] whilst the other two used methacholine.[39;64] Though none reported multiple cut-offs, specificity did reach 96.3% (sensitivity 30.2%) at a cut-off of 47 ppb in one study.[63] One other study reported similar sensitivity and specificity at an unknown cut-off (sensitivity 33.3%, specificity 84.8%)[39]. The third study reported a cut-off of 35ppb with sensitivity 75% and specificity 83.3%.[64]

Position E: patients with normal spirometry

viii) Reference standard of airway hyper-responsiveness: One study 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 35%) at 46 ppb and a sensitivity of 78% (specificity 45%) at 16 ppb.

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.[36;37;41;43;58-60]

*ix) Reference standard of ART to ICS: P*atients had normal chest radiographs or CT scans,[36;41] or had not responded to GORD or upper airway cough syndrome treatments.[37] 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)[36;37] 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,[36;37] with similar optimal cut-offs (38 ppb and 33.9 ppb), 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 ppb (Table 2).

x) Reference standard of ART & AHR: One study recruited patients with normal chest radiograph and CT scan,[43] whilst another recruited patients with normal chest radiograph.[58-60] 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 ppb and 40 ppb. 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[30] has limited generalisability 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 (Figure 1). Pedrosa et al[52] and Schleich et al[44] 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 Online Supplement F.

The three remaining studies reported useful results. Bobolea et al [57] 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 29.2%) at a cut-off of 30 ppb. El Halawani et al[31] recruited patients suspected of exercise-induced bronchoconstriction (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 31%) at a cut-off of 12 ppb. Sastre et al[42] 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 60%) at a cut off of 25 ppb.

xii) $_{FE}NO_{50}$ *used in conjunction with another test:* Three studies reported results for $_{FE}NO_{50}$ in conjunction with another test[32;44;56] In Cordeiro et al[56] a positive test was $_{FE}NO_{50}$ >27ppb 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[44] a positive index test was FEV1 ≤101% with $_{FE}NO_{50}$ >34 ppb. The reference standard was AHR by MCT. The patients selected were a mix of positions B and C. Addition of FEV1 ≤101% improved specificity, but decreased sensitivity (Table 2). In Fortuna et al,[32] using sputum eosinophilia in conjunction with $_{FE}NO_{50}$ 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 $_{FE}NO_{50}$ 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 $_{FE}NO_{50}$ should be placed within a diagnostic pathway, or whether $_{FE}NO_{50}$ would be best used as a rule-in test or a rule-out test. Several categories of studies were identified (i to xii), each of which relates to a specific clinical scenario. Across these categories, no studies reported 100% accuracy, though 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 to iv) studies ($_{FE}NO_{50}$ to replace the whole pathway in newly presenting patients, position A) varied. Even at its best, $_{FE}NO_{50}$ 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 ppb) for $_{FE}NO_{50}$ achieved high specificity (>90%), indicating it could be used as a rule-in test, but the best sensitivity reported was less good (81.5% to 96%, cut-off values 9 ppb to 25 ppb), indicating only moderate potential as a rule-out test. Despite the fact that not all asthmatics have eosinophilic inflammation, three studies reported 100% sensitivity.[31;49;57] 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 pbb) and specificity (100% at 100ppb). 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 $_{FE}NO_{50}$ before or in place of AHR testing. This position was identified *a priori* as potentially the most useful position for $_{FE}NO_{50}$ 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[63], using a cut-off of 47ppb, indicating the potential for $_{FE}NO_{50}$ to be a useful rule-in test at this point, though 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 $_{FE}NO_{50}$ 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,[25] 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, 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, $_{FE}NO_{50}$ 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 $_{FE}NO_{50}$

as a rule-in/rule-out test at this point is unclear, though good sensitivity versus ICS-responsiveness (90%, 94.7%)[36;37] and good specificity versus AHR (91.3%)[47] were reported.

Category viii) assessed the use of $_{FE}NO_{50}$ 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[46] and a full range of cut-off values was not reported, though a specificity of 90% at 46 ppb and a sensitivity of 78% at 16 ppb are promising.

In category xi) very specific uses of $_{FE}NO_{50}$ were tested. The small evidence base suggests $_{FE}NO_{50}$ could be used successfully (100% sensitivity) in patients who had a negative MCT to identify patients who would score negatively by adenosine challenge testing[57] and for those under 40 years of age referred for exercise challenge testing.[31] The other study in this category showed 80% sensitivity at a derived cut-off of 25 ppb, promising better results may be achieved at different cut-offs. As such, $_{FE}NO_{50}$ 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 for both the whole recruited cohort, and for a subgroup of patients who had atopy.[38] Whilst the overall accuracy was similar in both cohorts (area under the curve 0.69 (95% CI 0.6 to 0.775) and 0.68 (95% CI 0.53 to 0.8) respectively), the cut-off with the highest sum of sensitivity and specificity was higher in non-atopic patients than in atopic patients (32ppb versus 26ppb 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 toward lower cut-offs (see Online Supplement F) for studies which recruited a higher proportion of atopic patients. Reasons for this remain unclear, though 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[36] 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 ppb (SD3.1), compared to 51.25 ppb (SD20.1) 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 ppb (normal), patients should be considered for other diagnoses (rule-out asthma); 20-25 ppb to 50 ppb (elevated), patients should be considered asthmatic and a low dose of ICS prescribed (rule-in asthma); >50 ppb (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 ppb indicate a low likelihood of eosinophilic inflammation and ICS responsiveness, and levels above 50 ppb 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,[36;37;41;51] None reported 100% sensitivity or specificity, despite one study reporting a wide range of cut-offs (25 ppb to 150 ppb). Two studies reported highest sum cut-offs of 20-25 ppb, with one reporting poor sensitivity (53%) and specificity (63%)[41] and one reporting somewhat better values at 83% and 57%[51] respectively. The latter study also reported a cut-off of 50 ppb and showed almost inverse values of 58% and 80% respectively.[51] 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 ppb and 33.9 ppb.[36;37] All four studies are limited in terms of generalisability (three recruited patients with chronic cough who were negative for tests for other conditions[36;37;41], and in one case some patients were already on ICS treatment[36] and one recruited adults and adolescents with the symptoms of asthma[51]). The imperfect reference standard, as previously discussed, may account for some of the differences between index test ($_{FE}NO_{50}$) 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[65-67] patients, COPD[68-72] and diagnosed asthmatics[73]. Given this smaller evidence base, we conclude that the use of $_{FE}NO_{50}$ as a marker of steroid-responsive disease in adult patients with the symptoms of asthma remains very much understudied.

When considering $_{FE}NO_{50}$ 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 $_{FE}NO_{50}$ 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 (Online Supplement G). When study results were ordered according to cut-offs, studies using 20-25 ppb did not consistently deliver good sensitivity (range 36% to 100%). Sensitivities above 80% were only consistently delivered by studies with cut-offs ≤12 ppb, and sensitivities above 70% only consistently provided by studies with cut-offs ≤15 ppb. This perhaps suggests that the cut-off point for "normal" should be set somewhat lower if $_{FE}NO_{50}$ is used in a pathway to diagnose asthma as opposed to steroid-responsive airway disease. Similarly, results

suggested that a higher cut-off than 50ppb may be more likely to consistently deliver high levels of specificity, with specificities ≥90% only consistently being reported in studies with cut-offs >70 ppb.

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 false positive or false negative) may mean it is preferable to sacrifice some sensitivity or specificity to minimise 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 $_{FE}$NO 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 $_{FE}$NO to be assessed taking into account costs, benefits and harms.

Problems with the reference standard currently hamper all studies in this review. One study[48] 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)[26] 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[74] 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 $_{FE}$NO 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 $_{FE}NO_{50}$ 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 $_{FE}NO_{50}$ 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 $_{FE}NO_{50}$ within a pathway remain poorly evidenced.

Acknowledgements

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

Competing interests

Professor Ian Pavord received personal fees from GSK, Astra Zeneca, Almirall, Novartis, Aerocrine, Genentec, Teva, Regeneron and Boerhringer 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 Ingelhelm for educational meetings, all outside this work.

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 [17;18], (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.

Reference List

1. Berry MA, Shaw DE, Green RH, Brightling CE, Wardlaw AJ, Pavord ID et al. The use of exhaled nitric oxide concentration to identify eosinophilic airway inflammation: an observational study in adults with asthma. Clinical & Experimental Allergy 2005; 35(9):1175-1179.

2. Cowan DC, Taylor DR, Peterson LE, Cowan JO, Palmay R, Williamson A et al. Biomarker-based asthma phenotypes of corticosteroid response. J Allergy Clin Immunol 2015; 135(4):877-883.

3. Smith AD, Cowan JO, Brassett KP, Filsell S, McLachlan C, Monti-Sheehan G et al. Exhaled nitric oxide: a predictor of steroid response. Am J Respir Crit Care Med 2005; 172(4):453-459.

4. Bjermer L, Alving K, Diamant Z, Magnussen H, Pavord I, Piacentini G et al. Current evidence and future research needs for FeNO measurement in respiratory diseases. Respir Med 2014; 108(6):830-841.

5. Dweik RA, Boggs PB, Erzurum SC, Irvin CG, Leigh MW, Lundberg JO et al. An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications. American Journal of Respiratory & Critical Care Medicine 2011; 184(5):602-615.

6. Pavord ID, Hilvering B. Biomarkers and inhaled corticosteroid responsiveness in asthmatic patients. The Journal of allergy and clinical immunology 2015; 135(4):884-885.

7. 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.

8. Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald M et al. Global strategy for asthma management and prevention: GINA executive summary. Eur Respir J 2008; 31(1):143-178.

9. Dweik RA, Boggs PB, Erzurum SC, Irvin CG, Leigh MW, Lundberg JO et al. An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications. Am J Respir Crit Care Med 2011; 184(5):602-615.

10. Moral VP, Gutierrez F, Alvarez Gutierrez P, Clara PC, Barroso NC, Vina AL et al. Guida espanola para el manejo del asma. Arch bronconeumol 2003; 39(Supl 5):3-42.

11. 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(3):210-219.

12. Taylor DR, Taylor DR. Advances in the clinical applications of exhaled nitric oxide measurements. J Breath Res 2012; 6(4):047102.

13. Rodway GW, Choi J, Hoffman LA, Sethi JM, Rodway GW, Choi J et al. Exhaled nitric oxide in the diagnosis and management of asthma: clinical implications. Chronic Respir Dis 2009; 6(1):19-29.

14. Kercsmar C, Kercsmar C. Exhaled nitric oxide in the diagnosis and management of childhood asthma. [Review] [57 refs]. Ther Adv Respir Dis 2010; 4(2):71-82.

15. Barnes PJ, Dweik RA, Gelb AF, Gibson PG, George SC, Grasemann H et al. Exhaled nitric oxide in pulmonary diseases: a comprehensive review. Chest 2010; 138(3):682-692.

16. Dodig S, Richter D, Zrinski-Topic R, Dodig S, Richter D, Zrinski-Topic R. Inflammatory markers in childhood asthma. [Review]. Clinical Chemistry & Laboratory Medicine 2011; 49(4):587-599.

17. Spitale N, Popat N, McIvor A. Update on exhaled nitric oxide in pulmonary disease. Expert Rev Respir Med 2012; 6(1):105-115.

18. Szefler SJ, Wenzel S, Brown R, Erzurum SC, Fahy JV, Hamilton RG et al. Asthma outcomes: biomarkers. [Review]. Journal of Allergy & Clinical Immunology 2012; 129(3 Suppl):S9-23.

19. Feitosa LA, Dornelas de AA, Reinaux CM, Britto MC. Diagnostic accuracy of exhaled nitric oxide in exercise-induced bronchospasm: Systematic review. Revista Portuguesa de Pneumologia 18[4], 198-204.

20. Essat M, Harnan S, Gomersall T, Tappenden P, Wong R, Pavord I et al. Fractional exhaled nitric oxide for the management of asthma in adults: a systematic review. Eur Respir J 2016; 47(3):751-768.

21. Gomersal T, Harnan S, Essat M, Tappenden P, Wong R, Lawson R et al. A systematic review of fractional exhaled nitric oxide in the routine management of childhood asthma. Pediatr Pulmonol 2016; 51(3):316-328.

22. Harnan SE, Tappenden P, Essat M, Gomersall T, Minton J, Wong R 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).

23. 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. January 2015. 2015. 7-9-2016.

24. Asthma: diagnosis and monitoring of asthma in adults, children and young people. Clinical guideline. Methods, evidence and recommendations. Draft for consultation. January 2016.

25. 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 [date accessed: 02/02/2013].

26. National Institute for Health and Care Excellence. Diagnostics Assessment Programme manual. Manchester: 2011.

27. Centre for Reviews and Dissemination. Systematic reviews. CRD's guidance for undertaking reviews in healthcare. https://www.york.ac.uk/crd/guidance/ . 2009.

28. Higgins J, Green S, (editors). Cochrane Handbook for systematic reviews of interventions Version 5.1.0. [Updated March 2011]. 2011.

29. Whiting P, Rutjes A, Westwood M, Mallett S, Deeks J, Reitsma J et al. "QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.". Ann Intern Med 2011; 155(8):529-536.

30. 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(6):493-498.

31. ElHalawani SM, Ly NT, Mahon RT, Amundson DE. Exhaled nitric oxide as a predictor of exercise-induced bronchoconstriction. Chest 2003; 124(2):639-643.

32. Fortuna AM, Feixas T, Gonzalez M, Casan P, Fortuna AM, Feixas T et al. Diagnostic utility of inflammatory biomarkers in asthma: exhaled nitric oxide and induced sputum eosinophil count.[Erratum appears in Respir Med. 2012 Apr;106(4):599]. Respir Med 2007; 101(11):2416-2421.

33. 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(4):599.

34. Fukuhara A, Saito J, Sato S, Sato Y, Nikaido T, Saito K et al. Validation study of asthma screening criteria based on subjective symptoms and fractional exhaled nitric oxide. Ann Allergy Asthma Immunol [(Fukuhara, Saito, Sato, Sato, Nikaido, Saito, Fukuhara-Nakagawa, Inokoshi, Ishii, Tanino, Ishida, Munakata) Department of Pulmonary Medicine, School of Medicine, Fukushima Medical University, Fukushima City, Japan] 2011 107(6):[480-486]

35. Giovannini M, Valli M, Ribuffo V, Melara R, Cappiello G, Businarolo E et al. Relationship between Methacholine Challenge Testing and exhaled nitric oxide in adult patients with suspected bronchial asthma. European Annals of Allergy & Clinical Immunology 2014; 46(3):109-113.

36. 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(11):1350-1355.

37. 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(1):15-19.

38. Katsoulis K, Ganavias L, Michailopoulos P, Bikas C, Dinapogias E, Kontakiotis T et al. Exhaled Nitric Oxide as Screening Tool in Subjects with Suspected Asthma without Reversibility. International Archives of Allergy & Immunology 2013; 162:58-64.

39. Nickels A. Exhaled Nitric Oxide Performance Compared To Methacholine Challenge In Asthma. J Allergy Clin Immunol 2014; 133(2):AB148.

40. 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 Journal 2014; 145(3_MeetingAbstracts):467A.

41. 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(3):816-822.

42. Sastre J, Costa C, del Garc+¡a Potro M, Aguado E, Mahillo I, Fern+índez-Nieto M. Changes in exhaled nitric oxide after inhalation challenge with occupational agents. J Invest Allergol Clin Immunol 2013; 23(6):421-427.

43. Sato S, Saito J, Sato Y, Ishii T, Xintao W, Tanino Y et al. Clinical usefulness of fractional exhaled nitric oxide for diagnosing prolonged cough. Respir Med 2008; 102(10):1452-1459.

44. 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(2):158-165.

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: FENO <= 16 ppb better than FENO <= 12 ppb to rule out mild and moderate to severe asthma. Respir Res 2009; 10.

46. 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.

47. 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(2):209-216.

48. Schneider A, Faderl B, Schwarzbach J, Welker L, Karsch-V+Âlk M, J+Ârres RA. Prognostic value of bronchial provocation and FENO measurement for asthma diagnosisΓÇôResults of a delayed type of diagnostic study. Respir Med 2014; 108(1):34-40.

49. Heffler E, Guida G, Marsico P, Bergia R, Bommarito L, Ferrero N et al. Exhaled nitric oxide as a diagnostic test for asthma in rhinitic patients with asthmatic symptoms. Respir Med 2006; 100(11):1981-1987.

50. Smith AD, Cowan JO, Brassett KP, Filsell S, McLachlan C, Monti-Sheehan G et al. Exhaled nitric oxide: a predictor of steroid response. American Journal of Respiratory & Critical Care Medicine 2005; 172(4):453-459.

51. 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(11):1629-1634.

52. Pedrosa M, Cancelliere N, Barranco P, Lopez-Carrasco V, Quirce S. Usefulness of Exhaled Nitric Oxide for Diagnosing Asthma. J Asthma 2010; 47(7):817-821.

53. Pedrosa M, Cancelliere N, Barranco P, San JB, Quirce S. Usefulness of exhaled nitric oxide in asthma diagnosis. J Allergy Clin Immunol 2010; 125(2 SUPPL. 1):AB7.

54. 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(4):411-416.

55. Smith AD, Cowan JO, Filsell S, McLachlan C, Monti-Sheehan G, Jackson P et al. Diagnosing asthma: comparisons between exhaled nitric oxide measurements and conventional tests. American Journal of Respiratory & Critical Care Medicine 2004; 169(4):473-478.

56. 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(2):119-126.

57. 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; 129(2 SUPPL. 1):AB3.

58. Zhang YM, Lin JT, Su N, Chen X, Liu GL, Yu HX et al. [Values of fractional exhaled nitric oxide in the diagnosis of chronic cough]. [Chinese]. Chung-Hua i Hsueh Tsa Chih [Chinese Medical Journal] 2011; 91(18):1254-1258.

59. Zhang Y. The values of fractional exhaled nitric oxide in the diagnosis and treatment of chronic cough. Respirology 2011; 16:213.

60. 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.

61. Pizzimenti S, Heffler E, Piccioni P, Bugiani M, Migliore E, Guida G et al. Usefulness of exhaled nitric oxide (FeNO) measured by a portable analyzer to diagnose Cough Variant asthma in a clinical setting of chronic cough. Allergy Eur J Allergy Clin Immunol 2009; 64:395.

62. 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 MeetingAbstracts).

63. Brannan JD, Adoni H, Daw L, Huang HC, Hurwitz M, Figurski D. Fraction exhaled no in patients reffered to pulmonary function laboratory (PFLAB) for mannitol challenge. Respirology 2013; 18:43.

64. Chancafe-Morgan J, Ramos-Quispe Y, Gomez-Garcia R, Vargas-Espinal J, Puente-Maestu L. Validity of the fractional exhaled nitric oxide (FE_{NO}) for identification of bronchial hyperresponsiveness in a pulmonary function laboratory. European Respiratory Society Annual Congress, Barcelona, Spain 2013 . 2013. 4-10-2013.

65. Szefler SJ, Martin RJ, King TS, Boushey HA, Cherniack RM, Chinchilli VM et al. Significant variability in response to inhaled corticosteroids for persistent asthma. Journal of Allergy & Clinical Immunology 2002; 109(3):410-418.

66. Knuffman JE, Sorkness CA, Lemanske RF, Mauger DT, Boehmer SJ, Martinez FD et al. Phenotypic predictors of long-term response to inhaled corticosteroid and leukotriene modifier therapies in pediatric asthma. J Allergy Clin Immunol 2009; 123(2):411-416.

67. Klaassen EM, van de Kant KD, Jobsis Q, Hovig ST, van Schayck CP, Rijkers GT et al. Symptoms, but not a biomarker response to inhaled corticosteroids, predict asthma in preschool children with recurrent wheeze. Mediators Inflamm 2012; 2012:162571.

68. 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(7):816-824.

69. Antus B. [Role of exhaled nitric oxide in predicting steroid response in chronic obstructive pulmonary disease]. Orv Hetil 2010; 151(51):2083-2088.

70. 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):261-270.

71. Eosinophilic airway disorders. Copyright-® 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.; 2006.

72. Dummer JF, Epton MJ, Cowan JO, Cook JM, Condliffe R, Landhuis CE et al. Predicting corticosteroid response in chronic obstructive pulmonary disease using exhaled nitric oxide. Am J Respir Crit Care Med 2009; 180(9):846-852.

73. Malinovschi A, Van Muylem A, Michils A. Both intermediate and high exhaled nitric oxide levels predict improvement in asthma control after new-onset of inhaled corticosteroids. Allergy European Journal of Allergy and Clinical Immunology 2013; 68(Suppl. s97):164.

74. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003; 138(1):W1-12.

Legends to Figures

Figure 1 **Generic asthma diagnostic pathway**

Figure 2 **PRISMA Flow Diagram (adapted from http://www.prisma-statement.org/) for the review of diagnostic accuracy studies.**

Figure 3 **Author's judgement of risk of bias of included studies, according to adapted version of QUADAS II.[22]**

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 $_{FE}NO_{50}$ ppb (SD)**	**Corticosteroid use during testing**	**Smokers**	**Atopy**
POSITION A: ASTHMA SYMPTOMS, NO PREVIOUS TESTS								
i) Reference standard of pathway including AHR & ART								
Cordeiro 2011[56] Netherlands	Asthma symptoms All ages	R 114/114 (100%)	median (range) A: 39 (7 to 83); NA: 38 (7	NR	Median (range) A: 44 (6 to 290); NA: 17 (5 to 45)	ICS or OCS in previous 6 weeks excluded	10%	71%

Study author, year, country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1 % (SD)	Mean $_{FE}NO_{50}$ ppb (SD)	Corticosteroid use during testing	Smokers	Atopy
			to 87)					
Fortuna, 2007[32; 33] 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[34] Japan	Asthma symptoms	P 61/97 (63%)	55.6 (17 to 81)	96% (95% Cl, 90.1-102.0)	74.5 (95% Cl, 56.2 to 92.8)	ICS or OCS excluded	Current or Ex: 31%	23%
Schneider 2013[47] Germany	Asthma symptoms	P, C 393/400 (98%)	43.2 (16.3)*	104.8 (17.1)*	29.9 (33.2)*	ICS use stopped 12 hours before	Current: 10%; Ex: 35%	NR
Schneider 2009[45; 46] Germany	Asthma symptoms	P, C 160/160 (100%)	43.9 (16.3)*	95.8 (20.0)*	31.5 (36.0)*	ICS use stopped 12 hours before	Current and ex 54%	NR
Smith 2005[50] New Zealand	Asthma symptoms Adults and adolescents	P, C 52/60 (87%)	40.5 (14 to 71)	97.8 (14.2)	Range 6.3 to 242.0[51]	ICS or OCS in previous 4 weeks excluded	Current 6% Ex 19%	77%[51]
Smith 2004[55] New Zealand	Asthma symptoms All ages	P, C 44/51 (86%)	A: 41.6 (9 to 72) NA: 31.8 (9 to	102.9 (18.0)*	28.8 (28.6)*	ICS or OCS in previous 4 weeks excluded	Ex 11%	NR

Study author, year, country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1% (SD)	Mean $_{FE}NO_{50}$ ppb (SD)	Corticosteroid use during testing	Smokers	Atopy
			64)					
ii) Reference standard of AHR								
Backer 2014[54] Denmark	Asthma symptoms 10 to 45 years old	R 141/348 (40.5%)	28.3 (9.8)	97%	NR	NR	Current: 4% Ex: 19%	90%
Giovannini 2014[35] 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[51]** New Zealand	Asthma symptoms Adults and adolescents	P, C 52/60 (87%)	40.5 (14 to 71)	97.8 (14.2)	Range 6.3 to 242.0[51]	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[48] 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 hours before	Current: 10%; Ex: 35%	NR
POSITION A: ASTHMA SYMPTOMS PLUS ONE ADDITIONAL SYMPTOM.								
v) Reference standard of pathway including AHR & ART								
Heffler 2006[49] Italy	Rhinitis and asthma symptoms Adults and adolescents	P, C 48/48 (100%)	40.08 (NR)	89.2 (95% CI 80.1 to 98.4)	59.7 (95% CI 50.2 to 89)	NR	0	92%
POSITION C : NEGATIVE ART								
vi) Reference standard of AHR								
Katsouli	Negative	P	Medi	Media	Median	ICS users	Recent	45.5%

Study author, year, country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1 % (SD)	Mean $_{FE}NO_{50}$ ppb (SD)	Corticosteroid use during testing	Smokers	Atopy
s 2013[38] Greece	ART	112/112 (100%)	an 25 years (IQR 22 to 37)	n 89% (IQR 83 to 99%)	20.5 (IQR 12 to 34)	excluded	ex-smokers excluded	
POSITION C OR D: REFERRED FOR AHR								
vii) Reference standard of AHR								
Brannan 2013[63] Australia	Unclear, referred for mannitol challenge Unclear age group	R 401/401 (100%)	NR	NR	NR	NR	NR	NR
Chancafe-Morgan 2013[64] 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[39; 40]	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[45; 46] Germany	Normal spirometry	P, C 101/101 (100%)	NR	NR	NR	ICS use stopped 12 hours 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[36] 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[37	No response	R	49 (14)	91.8 (15.3)	mean rank 47	NR	0%	NR

Study author, year, country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1 % (SD)	Mean $_{FE}NO_{50}$ ppb (SD)	Corticosteroid use during testing	Smokers	Atopy
] Taiwan	to GORD or UACS	81/114 (71%)						
Prieto 2009[41] Spain	Normal chest radiograph, CT scan, spirometry	P 43/43 (100%)	48 (95% CI 43 to 52)	113.2 (95% CI 108.0 to 118.3)	GM Mean (95% CI): ICS responders 23.2 (17.5 to 30.7); non-responders 18.6 (14.7 to 24.0)	ICS or OCS excluded	0%	100%
x) Reference standard including ART and AHR								
Sato 2008[43] Japan	Normal chest radiograph or CT scan	P, C 71/71 (100%)	NR	NR	NR	ICS or OCS excluded	75%	NR
Zhang 2011[58-60] China	Normal chest radiograph Age group unclear	U 106/106 (100%)	NR	NR	NR	NR	NR	NR
xi) Other studies								
Arora 2006[30] 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[57] Spain	Failed MCT All ages	P, C 30/30 (100%)	37.3 (13 to 69)	NR	NR	NR	NR	NR
El Halawani	Suspected EIB	P, C 49/50 (98%)	27.9 (SD: NR)	NR	EIB +ve, 41; EIB	ICS or OCS excluded	0%	0%

Study author, year, country	Population (adults unless otherwise stated)	Study design, N (analysed/recruited)	Mean age in years (SD or range)	FEV1% (SD)	Mean $_{FE}NO_{50}$ ppb (SD)	Corticosteroid use during testing	Smokers	Atopy
2003[31] USA					-ve: 25.6			
Mathew 2011[62] UK	Normal spirometry and negative ART	P 84/84 (100%)	NR	NR	NR	NR	NR	NR
Pedrosa 2010[52; 53] Spain	Normal spirometry or negative ART Adults and adolescents,	P, C 114/115 (99%)	34 (13)	104.29 (14.95)	34	NR	current: 15% ex: 10%	87%
Pizzimenti 2009[61]	Chronic cough Unclear age group	P, C 156/156 (100%)	NR	NR	34.1 (95% Cl: 28.5 to 39.5)	NR	9%	47%
Sastre 2013[42] Spain	Suspected occupational asthma (various)	NR 68/68 (100%)	39.5 (10.3)	NR	Median 23 (IQR15.0 to 32.9)	ICS stopped 1 week before	Current: 21% Ex: 11%	66.2%
Schleich, 2012[44] Belgium	Normal spirometry or negative ART	P, C 174/237 (73%)	41 (16)	97 (13)	Median 17 range (4 to 271)	ICS excluded	34%	48%

N, number; SD, standard deviation; FEV1%, Forced expiratory volume in one second percent predicted; $FeNO_{50}$, fraction of exhaled nitric oxide measure at a flow rate of 50mL/s; ppb, 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 hyper-responsiveness; MCT, methacholine challenge test; CI, confidence interval; NR, not reported; GP, general practitioner; IQR, inter-quartile range; ICS, inhaled corticosteroid; EIB+ve, patients who tested positive for exercise induced bronchoconstriction; EIB-ve, patients who tested negative for exercise induced bronchoconstriction

* Reviewer calculated

** Reanalysis of Smith 2005[50]
*** 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.

				Highest sum of sens and spec or pre-specified cut-off					**Rule-out**					**Rule-in**				
Study author, year	**Device**	**Reference standard**	**Derived or pre-specified**	**$FeNO_{50}$ Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**
POSITION A: ASTHMA SYMPTOMS, NO PREVIOUS TESTS																		
i) Reference standard of pathway including AHR & ART																		
Cordeiro 2011[56]	Niox-Flex	ART (BDR), AHR (histamine)	D	27	78	92	85	88	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
	Niox-Flex; Airway reversibility		D	27	87	90	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Fortuna, 2007[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
Fukuhara, 2011[34]	NA623 (Chest MI, Tokyo, Japan)	At least 2 of: induced sputum eosinophilia, AHR, 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
Schneider 2013[47]	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

				Highest sum of sens and spec or pre-specified cut-off					Rule-out					Rule-in				
Study author, year	**Device**	**Reference standard**	**Derived or pre-specified**	**$FE_{NO_{50}}$ Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**
Schneider 2009[45;46]	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[50]	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[55]	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	NR
ii) Reference standard of AHR																		
Backer 2014[54]	NIOX MINO	AHR (mannitol)	P	25	36.2	86.8	65.6	66.1	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Giovannini 2014[35]	HypAir FeNO;	AHR (MCT)	P	30	14.3	100	100	53.9	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
iii) Reference standard of ART																		
De La Barra 2011[51]	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[48]	NIOX MINO	Spirometry, ART (BDR), AHR (MCT) and long term follow-up of	D	26	47	73.1	39.8	78.4	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR

				Highest sum of sens and spec or pre-specified cut-off					Rule-out					Rule-in				
Study author, year	**Device**	**Reference standard**	**Derived or pre-specified**	**FE_{NO} Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**
		patients and GPs																
POSITION A: ASTHMA SYMPTOMS PLUS ONE ADDITIONAL SYMPTOM.																		
v) Reference standard of pathway including AHR & ART																		
Heffler 2006[49]	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[38]	NIOX MINO	AHR (MCT)	D*	32	47	85	70.1	68.1	10	81	39	49.9	73.2	30	49	82	67.1	68.2
			D**	26	55	85	NR	NR	10	90	10	NR	NR	30	48	85	NR	NR
POSITION C OR D: REFERRED FOR AHR																		
vii) Reference standard of AHR																		
Brannan 2013[63]	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[64]	NR	AHR (MCT)	P	35	75	83.3	75	83.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Nickels 2014[39;40]	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[45;46]	NIOX MINO	AHR (MCT)	D	46	35	90	NR	NR	16	78	45	NR	NR	NR	NR	NR	NR	NR
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[36]	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[37]	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

				Highest sum of sens and spec or pre-specified cut-off					Rule-out					Rule-in				
Study author, year	**Device**	**Reference standard**	**Derived or pre-specified**	**FE_{NO} Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**
Prieto 2009[41]	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[43]	Chemi-luminescence 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[58]	NIOX MINO	Sputum cell counts, spirometry, AHR, 24-h esophageal 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																		
Arora 2006[30]	Niox	AHR (Histamine)	D	>17	63	58.8	86.1	28.2	>6	96.4	0	79.6	0.000	>46	16.7	100	100	22.8
Bobolea 2012[57]	NIOX MINO	AHR (Adenosine)	P	30***	100	29.2	26	100	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
El Halawani 2003[31]	Sievers 280A	Exercise challenge	D	12***	100	31	19.4	100	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Mathew 2011[62]	NR	AHR (MCT)	P	NR	10	67.2	8.7	70.5	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Pedrosa 2010[52]	NIOX MINO	AHR (MCT)	D	40	74.3	72.5	54.1	86.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR

				Highest sum of sens and spec or pre-specified cut-off					Rule-out					Rule-in				
Study author, year	**Device**	**Reference standard**	**Derived or pre-specified**	**$FeNO_{50}$ Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**	**Cut-off**	**Sens**	**Spec**	**PPV**	**NPV**
Pizzimenti 2009[61]	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[42]	NIOX MINO	Specific inhalation challenge	D	25	60	80	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Schleich, 2012[44]	Niox	AHR (MCT)	D	34	35	95	87.8	62.4	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
	Niox FEV1 ≤101%		D	34	24.4	98.9	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR

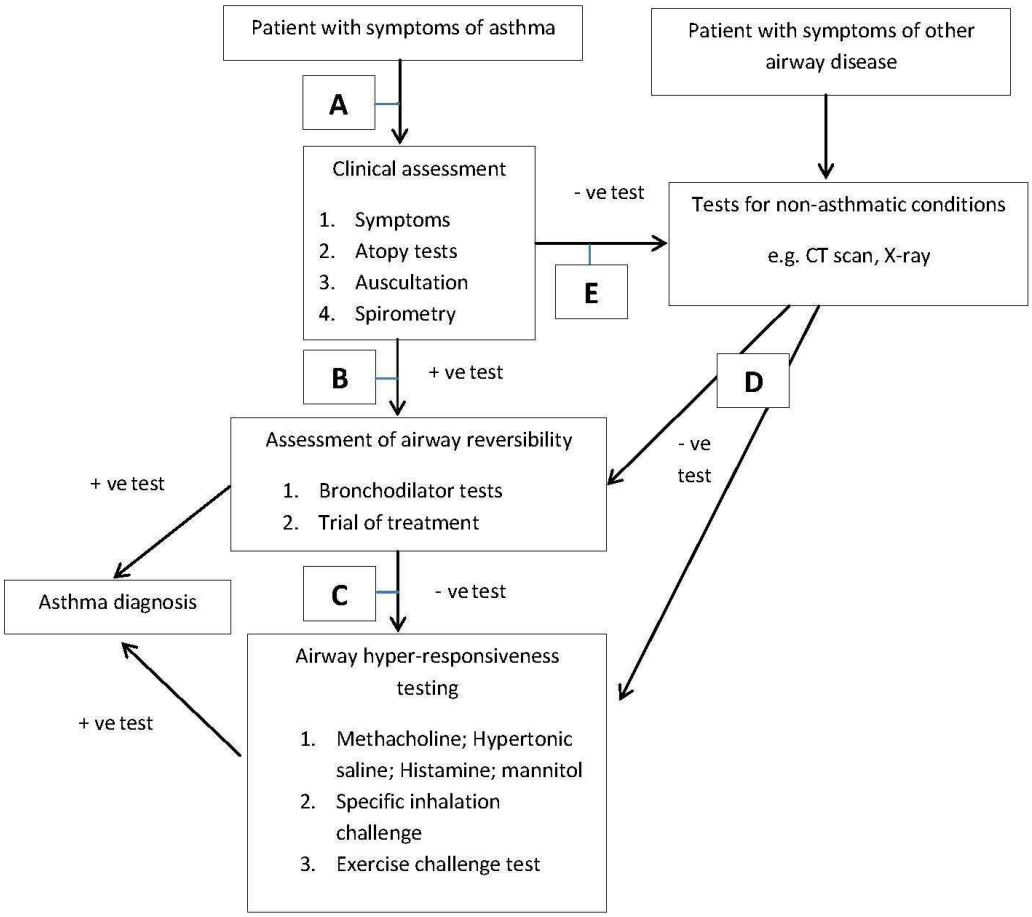
N, number analysed; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value; $FeNO_{50}$, fractional exhaled nitric oxide measure at a flow rate of 50 mL/s; FEV1, forced expiratory volume in first second; FVC, forced vital capacity; BDR, Bronchodilator reversibility 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

Figure 1 Generic asthma diagnostic pathway



Based on the BTS/SIGN guidelines (2012).[25] Additional tests used internationally include: whole body plethysmography, sputum eosinophils.

A, B, C, D: potential positions for the use of $_{FE}$NO 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.

Figure 2 PRISMA Flow Diagram (adapted from http://www.prisma-statement.org/) for the review of diagnostic accuracy studies.

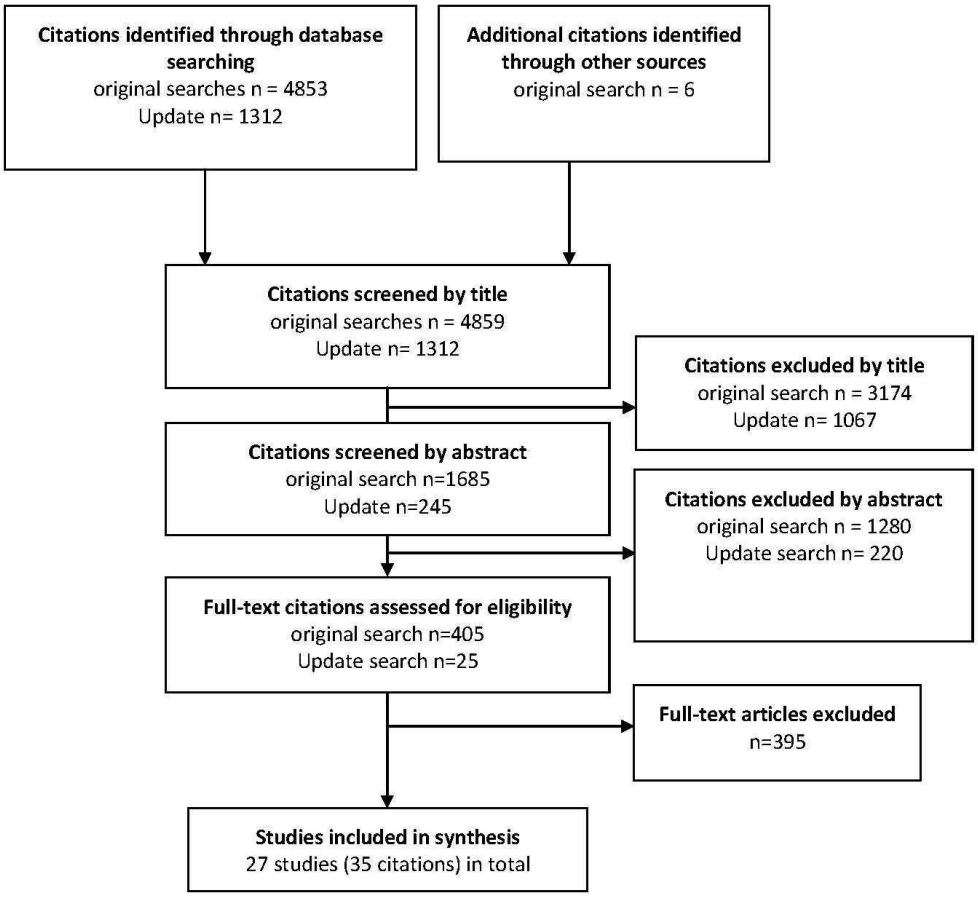


Figure 3 Author's judgement of risk of bias of included studies, according to adapted version of QUADAS II.[29]

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[56]	unclear	high	High	high
Fortuna, 2007[32;33]	low	low	Unclear	unclear
Fukuhara 2011[34]	unclear	low	Unclear	unclear
Schneider 2013[47]	low	high	low	high
Schneider 2009[45;46]	low	high	low	unclear
Smith 2005[48]	low	high	unclear	unclear
Smith 2004[55]	low	high	unclear	unclear
ii) Reference standard of AHR				
Backer 2014[54]	High	unclear	Unclear	unclear
Giovannini 2014[35]	high	low	Unclear	unclear
iii) Reference standard of ART				
De La Barra 2011[51]	low	high	unclear	unclear
iv) Reference standard of pathway plus long term follow-up				
Schneider 2014[48]	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[49]	low	high	Unclear	unclear
Position C: negative ART				
vi) Reference standard of AHR				
Katsoulis 2013[38]	high	high	unclear	low
Position C or D: referred for AHR				
vii) Reference standard of AHR				
Brannan 2013[63]	high	unclear	unclear	unclear
Chancafe-Morgan 2013[64]	high	unclear	unclear	unclear
Nickels 2014[39;40]	high	high	high	low
Position E: negative for clinical assessment (e.g. normal spirometry)				
viii) Reference standard of AHR				
Schneider 2009[45;46]	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[36]	high	unclear	unclear	unclear
Hsu 2013[37]	high	unclear	unclear	unclear

Prieto 2009[41]	unclear	high	low	unclear
x) Reference standard including ART and AHR				
Sato 2008[43]	low	high	unclear	unclear
Zhang 2011[58-60]	unclear	high	unclear	unclear
xi) Other studies				
Arora 2006[30]	high	high	low	unclear
Bobolea 2012[57]	low	low	unclear	unclear
El Halawani 2003[31]	high	high	unclear	unclear
Mathew 2011[62]	unclear	low	unclear	unclear
Pedrosa 2010[52;53]	low	high	unclear	low
Pizzimenti 2009[61]	high	high	unclear	unclear
Sastre 2013[42]	low	high	unclear	low
Schleich, 2012[44]	low	high	unclear	unclear