

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Accuracy of self-diagnosis in conditions commonly managed in primary care: Diagnostic accuracy systematic review and meta-analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-065748
Article Type:	Original research
Date Submitted by the Author:	15-Jun-2022
Complete List of Authors:	McLellan, Julie; University of Oxford Nuffield Department of Primary Care Health Sciences, Nuffield Department of Primary Health Care Sciences Heneghan, Carl; University of Oxford Nuffield Department of Primary Care Health Sciences, Primary Health Care Roberts, Nia; University of Oxford Bodleian Libraries, Bodleian Health Care Libraries, Pluddemann, Annette; University of Oxford Nuffield Department of Primary Care Health Sciences, Nuffield Department of Primary Care Health Sciences
Keywords:	Adult dermatology < DERMATOLOGY, HIV & AIDS < INFECTIOUS DISEASES, PRIMARY CARE, GYNAECOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **Accuracy of self-diagnosis in conditions commonly managed in primary care:**
4 **Diagnostic accuracy systematic review and meta-analysis**
5
6
7

8 Julie McLellan¹, Carl Heneghan¹, Nia Roberts², Annette Plüddemann¹
9

10 ¹ Nuffield Department of Primary Care Health Sciences, University of Oxford, OX2 6GG
11

12 ² Bodleian Health Care Libraries, University of Oxford, Oxford, UK
13
14
15

16 Corresponding author: Annette Plüddemann, Nuffield Department of Primary Care Health
17 Sciences, University of Oxford, OX2 6GG, annette.plueddemann@phc.ox.ac.uk, +44 (0)1865
18 289353
19

20
21
22 **Keywords:** Diagnostic accuracy systematic review, meta-analysis, self-diagnosis, common
23 conditions in primary care,
24
25
26

27 **Abstract word count:** 299
28

29 **Manuscript word count:** 4073
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Objectives:

To assess the diagnostic accuracy of self-diagnosis compared to a clinical diagnosis for common conditions in primary care

Design:

Systematic review and meta-analysis

Data sources:

Medline, Embase, Cochrane Central Register of Controlled Trials(CENTRAL), Cochrane Database of Systematic Reviews(CDSR), and CINAHL from inception to 25 January 2021

Study selection:

Eligible studies were prospective or retrospective studies comparing the results of self-diagnosis of common conditions in primary care to a relevant clinical diagnosis or laboratory reference standard test performed by a healthcare service provider. Studies that considered self-testing only were excluded.

Data extraction:

Two authors independently extracted data using a predefined data extraction form and assessed risk of bias using QUADAS-2.

Methods & Results:

Searches identified 5,047 records resulting in 18 included studies. Five studies reported on the accuracy of self-diagnosis of vaginal infection, four for common skin conditions, and nine for the self-diagnosis for HIV. For self-diagnosis of vaginal infection and common skin conditions meta-analysis was not appropriate and data were reported narratively. Nine studies, using point-of-care oral fluid tests, reported on the accuracy of self-diagnosis of HIV and data were pooled using bivariate meta-analysis methods. For these nine studies the pooled sensitivity was 92.8% (95%CI, 86% to 96.5%) and specificity was 99.8% (95%CI, 99.1% to 99.9%). Post hoc, the robustness of the pooled findings was tested in a sensitivity analysis only including four studies using laboratory testing as the reference standard. The pooled sensitivity reduced to 87.7% (95%CI 81.4% to 92.2%) and the specificity remained the same. The quality of all 18 included studies was assessed as mixed and overall study methodology was not always well described.

Conclusions and implications of key findings:

Overall, there was a paucity of evidence. The current evidence does not support routine self-diagnosis for common conditions in primary care.

PROSPERO registration number: CRD42018110288

Article Summary

Strengths and limitations of this study

- This systematic review summarises and interprets the available evidence on self-diagnosis of conditions managed in primary care
- This search strategy was extensive including publications identified from databases Medline, Embase, Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), and CINAHL, up to January 2021
- Standard methodology for systematic review of diagnostic accuracy studies was used, including study quality appraisal using QUADAS-2
- There was a paucity of evidence for many common conditions
- Lack of evidence meant meta-analysis was possible only for one condition

Introduction

The workload in primary care continues to increase (1, 2), in part due to global population increases, but also due to age of populations, change in lifestyle and more complex health problems (3). Increased workload has been recognised in the UK as an important factor in working conditions for primary care physicians (General Practitioners, GPs) and strategies are required to manage the workload (4).

One strategy is the potential for self-diagnosis and self-treatment by patients of some commonly occurring conditions. If feasible, this could lead to more rapid diagnosis and treatment reducing the burden on primary care services. The prospect of self-diagnosis is controversial with concerns if results are misinterpreted or patients fail to confirm their findings to a physician (5). In terms of the evidence, the first important question is to assess the diagnostic accuracy of self-diagnosis in the primary care setting. Subsequently, in order to support self-diagnosis, the efficacy needs to be assessed, to inform which conditions can self-diagnosed safely, in which circumstances, and by whom.

Cooke and colleagues recently reported the 30 most commonly managed conditions in primary care in Australia, which has a health landscape broadly comparable with western Europe (6). This list arises from survey data collected between January 2009 and December 2010, which included 194,100 patient encounters from 1,941 GPs.

This systematic review, therefore, aimed to assess the diagnostic accuracy of self-diagnosis compared to a clinical diagnosis for common conditions in primary care by a healthcare provider.

Methods

Types of studies

We included prospective or retrospective studies comparing the results of self-diagnosis of common conditions in primary care to the results of a relevant clinical diagnosis or laboratory reference standard test. We excluded studies with a case-control design due to their high risk of bias (4).

Population: Adults self-diagnosing common conditions in primary care. Common conditions included were broadly based on those reported by Cooke et al (6) and relevant for self-diagnosis (see online supplementary table 1). Studies in children, based in animals or non-human samples were excluded.

Index test: Self-diagnosis, where we defined “self-diagnosis” as a diagnosis made by the patients in the study, including self-evaluation and interpretation of results of rapid tests. Studies that considered self-testing only and not as part of self-diagnosis were excluded, in addition studies assessing accuracy of self-monitoring of an existing condition were excluded.

Reference standard: Clinical diagnosis or laboratory test performed by a healthcare service provider. We excluded studies comparing self-diagnosis with diagnosis by allied health professionals or pharmacists.

1
2
3 Outcome measures: Reported diagnostic accuracy measures (e.g. sensitivity, specificity,
4 likelihood ratios, predictive values, etc.) and primary data for 2x2 tables. We excluded
5 studies reporting only measures of agreement.
6

7 ***Search methods to identify studies***

8
9 The search strategy was based on a combination of terms for self-testing and self-diagnosis,
10 diagnostic accuracy terms (e.g. sensitivity, specificity, etc.) and terms for common conditions
11 in primary care (6) (see online supplementary table 2 for full search strategy).
12

13 We searched the following electronic databases from inception to 25 January 2021: Medline
14 (OvidSP)[1946-present], EMBASE (OvidSP)[1974-present], Cochrane Central Register of
15 Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR) via
16 Cochrane Library, Wiley)[Issue 1 of 12, January 2021] and CINAHL (EBSCOHost)[1982-
17 present]. No restrictions were imposed on study population numbers or language (studies in
18 languages other than English were translated). Letters, narrative reviews, and other non-
19 primary sources were excluded. The reference lists of included studies, plus the first five
20 “similar articles” identified through PubMed for these studies, and reference lists of relevant
21 systematic reviews were used to identify further relevant publications. References were
22 imported into Endnote X9 (7) where duplicates were removed.
23
24
25

26 ***Data collection and analysis***

27 *Selection of studies*

28
29 Two reviewers independently applied the selection criteria to the titles and abstracts of the
30 study reports identified by the searches. Full text of all studies that met the inclusion criteria
31 were reviewed to agree the final list of included studies. Disagreements between reviewers
32 were resolved by discussion and where agreement could not be reached a third reviewer was
33 consulted.
34
35
36

37 *Data extraction and management*

38
39 Two reviewers (JM, AP) independently extracted information from selected studies into a
40 pre-defined data extraction sheet (see online supplementary table 3) and crosschecked the
41 data. Disagreements were resolved by discussion.
42
43

44 *Assessment of methodological quality*

45 We used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) (8) tool to
46 assess methodological quality of included studies, considering risk of bias in four domains
47 (patient selection, index test, reference standard, flow and timing), as well as assessing the
48 applicability of the studies to the review research question. Two reviewers (JM, AP)
49 independently assessed studies’ methodological quality; disagreements were resolved by
50 discussion, or if necessary, by a third reviewer.
51
52
53

54 *Statistical analysis and data synthesis*

55 Data were presented and analysed based on the condition being diagnosed. We compiled
56 summary tables outlining the detailed study information of included studies, including the
57 patient sample, condition, study design, setting, the test under evaluation, the comparator, and
58 conduct of the study. We extracted binary diagnostic accuracy data from all studies and
59 constructed 2×2 tables.
60

Meta-analysis

We used Review Manager (9) to produce paired forest plots to explore the between-study variability of sensitivity and specificity across the included studies. For each study estimate of sensitivity and specificity, corresponding 95% confidence intervals were shown to illustrate the uncertainty related to each study estimate. Where different thresholds were applied these were reported. Where appropriate, we used bivariate meta-analysis methods (10) to generate pooled estimates of sensitivity and specificity. Due to the nature of the data a change was made to the protocol and RStudio (11) was used to generate the model parameters to input into Revman (9).

Investigating heterogeneity

For medical conditions for which data from more than one study was available, and where it was possible to investigate between-study heterogeneity in the results, inclusion of study level characteristics as covariates in meta-analysis and subgroup analyses were considered. These approaches were carried out if there was sufficient data available and sub-group specific pooled estimates were thought to be of clinical relevance.

Investigating reporting bias

Funnel plots used to detect publication bias in reviews of RCTs have been shown to be misleading for diagnostic test accuracy reviews (12, 13). Funnel plots as an assessment of reporting bias were therefore not included in this review.

Patient involvement

Members of the public were part of the Research Programme Committee of the National Institute for Health Research (NIHR) programme grant that funded this study. Updates and details about the study were presented to the committee while the study was ongoing, and the public members provided feedback. This review formed part of the NIHR Evidence Synthesis Working Group (ESWG) and members of the public who were part of the ESGW steering committee commented on the protocol for the study and on updates presented to the steering committee.

The full protocol is provided as online supplementary table 4 and is registered on PROSPERO reference number CRD42018110288.

Results

Figure 1 shows a summary of the search results and the inclusion and exclusion of studies. After removal of duplicates, 5,047 records were identified through database searches, websites and citation searching. This resulted in full texts of 170 articles being assessed for eligibility, and 20 included articles (14-33) reporting results of 18 individual studies (14-20, 22-28, 30-33). These 18 included studies fell into 3 broad groups of commonly managed conditions: Female genital check-up, dermatitis, contact/allergic and viral disease, not otherwise specified. No studies of self-diagnosis were found for any other conditions in primary care.

Most excluded studies only reported on patients' ability to self-test (or self-monitor an existing condition) with the diagnosis being made by a clinician, and did not report diagnostic accuracy of self-diagnosis.

1
2
3 Of the included studies, five reported on the accuracy of self-diagnosis of vaginal infection
4 (11-15), four for common skin conditions (19, 20, 22, 23) and nine for the self-diagnosis for
5 HIV (24-28, 30-33). Online supplementary tables 5 and 6 summarise the characteristics of
6 included studies and characteristics of self-diagnostic (index) and reference tests,
7 respectively. Paired plots of sensitivity and specificity were generated, grouping the studies
8 by the condition to be diagnosed (Figures 2-4). For studies examining the accuracy of self-
9 diagnosis of vaginal infection and in common skin conditions (14-20, 22, 23), meta-analysis
10 was not appropriate due to the between-study heterogeneity and the overall low number of
11 studies, which would make meta-analysis uninformative (34).
12
13

14 ***Self-diagnosis of vaginal infections***

15
16 Five studies assessed the accuracy of self-diagnosis of Bacterial vaginosis and/or infection
17 with *Trichomonas vaginalis* (14-18), with one study assessing the self-diagnosis of *Candida*
18 vaginitis (17) (Figure 2). For Bacterial vaginosis, the accuracy of a vaginal fluid test using a
19 pH strip (for Bacterial vaginosis) was assessed with laboratory testing (Gram staining) as the
20 reference standard in two studies (14, 15) (online supplementary table 5). For the diagnosis
21 of Bacterial vaginosis and/or *T. vaginalis*, a panty liner test kit (VI-SENSE) for vaginal
22 discharge was assessed against a combination of clinical and laboratory assessment as the
23 reference test in one study (16); and a vaginal fluid self-diagnosis kit for women in the
24 military was assessed with clinical and laboratory assessment as the reference test in the
25 second study (17). One study used a vaginal fluid dipstick test for the presence of *T. vaginalis*
26 (OSOM *Trichomonas* rapid test) compared to a laboratory PCR as a reference test (18). For
27 the self-diagnosis of *Candida* vaginosis a military self-testing kit based on a combination of
28 the measurement of pH, amines and the symptom of vaginal itching for self-diagnosis was
29 compared to a combination of clinical and microbiological laboratory assessment in one
30 study (17).
31
32
33

34 ***Bacterial vaginosis; Bacterial vaginosis and/or Trichomonas vaginalis***

35
36 The sensitivity of a self-taken swab applied to a pH test strip for self-diagnosis of Bacterial
37 vaginosis ranged from 0.45 (95%CI, 0.34-0.56) (14) to 0.60 (95%CI, 0.55-0.66) (17) at a pH
38 cut-off of ≥ 4.7 . Notably, Ryan-Wenger et al (17), the study reporting a sensitivity of 0.60,
39 also included symptoms (vaginal itching) and the presence of amines as part of the
40 assessment, which may explain the higher sensitivity. Donders et al (14) assessed a lower cut-
41 off of pH ≥ 4.5 and showed an increased sensitivity of 0.95 (95%CI, 0.88-0.99); however at
42 the expense of specificity. The study on pregnant women (15) assessing the accuracy of pH
43 test strips did not specify the pH cut-off and reported a sensitivity of 0.36 (95%CI, 0.17-0.59)
44 with a specificity of 0.88 (95%CI, 0.82-0.93). The specificity for the pH test strip tests ranged
45 from 0.5 (95%CI, 0.43-0.56) (17) to 0.81 (95%CI, 0.75-0.85) (14) at the pH cut-off of ≥ 4.7 ,
46 decreasing to 0.41 (95%CI, 0.34-0.47) at the lower cut-off of pH ≥ 4.5 (14). Interestingly the
47 low specificity of 0.5 was reported by the study combining the pH test strip with symptoms
48 and the presence of amines (17).
49
50
51

52 The study assessing the vaginal discharge test using a panty liner test kit with an indicator
53 strip incorporated in the liner (16) reported a sensitivity of 0.91 (95%CI, 0.86-0.94) and
54 specificity of 0.81 (95%CI, 0.76-0.86) for the diagnosis of Bacterial vaginosis and/or *T.*
55 *vaginalis* infection.
56

57 ***Trichomonas vaginalis***

1
2
3 One study in Brazil assessed a rapid immunochromatographic *T. vaginalis* test for use at
4 home (18) and reported a sensitivity of 0.68 (95%CI, 0.43-0.87) and specificity of 1.00
5 (95%CI, 0.99-1.00) for self-diagnosis of *T. vaginalis* infection.
6

7 *Candida vaginitis*

8
9 Only one study (17) specifically assessed the diagnostic accuracy of self-diagnosis of
10 *Candida* vaginitis, which formed part of the military self-testing kit, and reported a sensitivity
11 of 0.18 (95%CI, 0.12-0.25) and specificity of 0.89 (95%CI, 0.85-0.92).
12
13
14
15

16 **Self-diagnosis of common skin conditions**

17
18 Four studies assessed the accuracy of self-diagnosis of common skin conditions (19, 20, 22,
19 23) (Figure 3). We included 2 studies that were outside our age inclusion criteria: Berghoi
20 2011 reported patients included had a mean age 17.5 years old, nevertheless these patients
21 would have been 16+ years to qualify as apprentice hairdressers. And Svensson 2002
22 reported the mean age of patients as 40.4 years (no standard deviation), but included patients
23 from age 16.
24

25 *Eczema*

26
27 Two studies assessed the accuracy of self-diagnosis based on a self-evaluated questionnaire
28 of signs and symptoms for the diagnosis of eczema alongside a self-assessment of the
29 presence or absence of eczema based on the questionnaire results, compared to assessment by
30 a clinician (19, 20). Overall, 710 participants were included across the two studies. The
31 reported sensitivity ranged from 0.7 (95%CI, 0.53- 0.84) to 0.87 (95%CI, 0.78-0.93) and
32 specificity ranged from 0.79 (95%CI, 0.70-0.86) to 1.00 (95%CI, 0.99-1.00). The relatively
33 high specificity suggests the potential for patients to use the questionnaire as a tool to confirm
34 that they have eczema and therefore seek healthcare advice; however, it should be noted that
35 there was an unclear risk of bias regarding the patient selection. In Svensson et al (20), the
36 patients were recruited at a dermatology outpatient clinic, where they had been referred to
37 and 113 patients in the study reported having had a diagnosis of eczema in the last 12 months,
38 suggesting a more selected population with a higher pre-test probability. Whilst the setting in
39 the study by Bregnhøj et al. (19) was not reported, it was conducted amongst hairdressers
40 who may have had more experience of eczema either themselves or colleagues being
41 diagnosed with the condition, given the nature of their profession. They may be more aware
42 of the signs and symptoms and may constitute a selected population. Therefore, the
43 diagnostic accuracy of the questionnaire for self-diagnosis may be dependent on the type of
44 population.
45
46
47
48

49 *Skin allergy*

50
51 Two studies assessed the diagnostic accuracy of self-diagnosis to detect an allergic skin
52 reaction including 408 participants across the two studies; one study assessed nickel and/or
53 fragrance allergy (22), and the other nickel allergy alone (23). Both studies used a patch test
54 applied to the arm, which was self-evaluated by participants 2-4 days later. Dermatologists then
55 also evaluated the patch tests as the reference standard. One study recruited participants at
56 hospital dermatology departments (23), while the other recruited through a newspaper
57 advertisement targeted at people with a self-suspected allergy towards fragrance and/or nickel
58 (22). Sensitivity ranged from 0.72 (95%CI, 0.57-0.84) to 0.97 (95%CI, 0.87-1.00) and
59 specificity ranged from 0.87 (95%CI, 0.80-0.92) to 0.91 (95%CI, 0.85-0.95). Elsnér et al.
60

(22) also reported that participants found the information regarding how to apply the test extensive and detailed, the information regarding self-evaluation of the test was limited and should be improved.

Self-diagnosis of HIV

Nine studies were identified that reported the diagnostic accuracy of self-testing and self-diagnosis of HIV (24-28, 30-33). In all studies self-diagnosis was undertaken unsupervised using a rapid point of care (POC) oral fluid test manufactured by OraSure technologies, either OraQuick Inhome intended for lay users or OraQuick Advance intended for professional use. The studies recruited 13103 participants, and all were conducted in African countries except for the phase III trial in the USA by OraSure Technologies (33). The 2019 global HIV prevalence rates for women and men aged 15 to 49 were 0.8% (95%CI, 0.7–1.0) and 0.6% (95%CI, 0.5-0.8) respectively with the overall highest prevalence by country in Eswatini (Africa) at 27.1% (95%CI, 25.4-28.8] (35). All included studies had prevalence rates above the global averages for men and women or, if not reported, were in countries with high prevalence rates. Prevalence rates ranged from 2.12% in the USA study (33) to 22.1% in the Ugandan study (24). The USA study was conducted in 20 clinical sites, 17 identified as high prevalence sites (2.6%) and 3 as low prevalence sites (0.1%). All studies enrolled participants from the general population including the USA where no breakdown of sexual orientation was reported. The reported sensitivity and specificity were similar between studies (Figure 4a); the single study (32) reporting a lower estimate for sensitivity still had a confidence interval that overlapped with half of the other studies. The pooled sensitivity based on all 9 included studies was 92.8% (95%CI, 86% to 96.5%) and the pooled specificity was 99.8% (95%CI, 99.1% to 99.9%). The studies showed low heterogeneity (Figure 4b). The reference standard used in the studies was one of three types: four studies took a venous sample which was sent to a laboratory for testing (28, 30, 32, 33), two studies used a nationally approved algorithm based on a combination of rapid POC tests (24, 25) and three studies used a study based algorithm again based on rapid POC tests (26, 27, 31). In three studies using POC tests for the reference standard, the diagnosis may have been by clinicians, but it was unclear. These studies reported the diagnosis by a research assistant (24) or a counsellor (26, 31). Post hoc, a sensitivity analysis was conducted to test the robustness of the pooled findings by removing the studies using POC tests as the reference standard (including tests where it was unclear whether diagnosis was by a clinician) and only including those studies using laboratory testing. Based on four studies using laboratory testing as the reference standard (28, 30, 32, 33), the pooled sensitivity was 87.7% (95%CI 81.4% to 92.2%) and the pooled specificity was 99.8% (95%CI 98.9% to 99.9%). No data were reported by participant characteristics such as gender, ethnicity, or sexual orientation.

A number of studies (24-26, 30-32) reported on the viability or feasibility of the oral fluid self-test stating that participants found it easy to conduct, but acknowledged that instructions should be adapted to the population using the test, particularly the literacy levels. Furthermore, users must be encouraged to receive a confirmatory test.

Methodological quality of included studies

Assessment of the quality of included studies using the QUADAS-2 framework (8) is presented in Figure 5, which summarises the overall risk of bias and applicability concerns.

1
2
3 For patient selection, the risk of bias overall was mixed; where studies were rated unclear in
4 this domain it was because several studies either recruited a selected population with a
5 potentially higher pre-test probability or they did not clearly report the recruitment strategy
6 and whether eligible patients were consecutively recruited. However, with the exception of
7 one study, applicability concerns were low. In several cases, although the population may
8 have been a selected population (e.g. at high risk of HIV infection or skin eczema; or with a
9 prior history of eczema or vaginal infection), it could be argued that these might be the
10 populations where self-diagnosis and/or self-testing may be most relevant. Overall risk of
11 bias regarding the conduct of the index test was low and in most studies participants were
12 blinded to the results of the reference test, though in some studies this was not clearly
13 described. Risk of bias with respect to the reference test was unclear or high in some studies
14 as assessors were either not blinded to the results of the index test or blinding was unclear.
15 For the domain of flow and timing, several studies were judged to be at high risk of bias as
16 the interval between testing was frequently not explicitly reported. In addition differential
17 reference bias was identified as present in some studies and unclear in others, in particular it
18 was unclear in several studies how clinical assessment was conducted or standardised.
19 Overall study procedures were not always clearly described.
20
21
22
23
24
25

26 Discussion

27 With the increasing workload in primary care (1) and the continued development of rapid
28 tests, including those that are intended to be used by patients, we aimed to assess the evidence
29 for the diagnostic accuracy of self-diagnosis. We identified limited evidence on the
30 diagnostic accuracy of self-diagnosis: only 20 publications (reporting data from 18 studies)
31 specifically assessed the accuracy of self-diagnosis, covering three commonly managed
32 conditions, namely vaginal infections, common skin conditions and HIV. Interestingly, no
33 studies of self-diagnosis were found for any other conditions in primary care. It was
34 particularly notable that we did not find any studies assessing the diagnostic accuracy of self-
35 diagnosis of common primary conditions such as upper respiratory tract and urinary tract
36 infections.
37
38
39

40 The evidence for self-diagnosis of vaginal infection suggests a lack of sufficient accuracy to
41 aid self-diagnosis. Tests relying on self-swabs and pH strips showed low sensitivity (below
42 60%), and therefore would not be useful to rule out disease. Although sensitivity improved to
43 95% with increasing pH cut-off, this occurred at the cost of specificity, which dropped from
44 81% to 41%. Therefore, using the test at this cut-off would result in concern for missed
45 diagnoses. The immunochromatography test also showed insufficient sensitivity (68%) to be
46 of use as a rule-out test. The panty liner test, however, had high sensitivity of 91% and may
47 prove to be a useful rule-out test, although this result is limited as it is based on one study
48 judged to be at unclear risk of bias, with a high risk of bias in the flow and timing domain. In
49 terms of their use to confirm the presence of vaginal infection, the highest specificity (100%)
50 was reported for the immunochromatographic test. This may be a useful test to aid self-
51 diagnosis of vaginal infection in systems that currently rely on syndromic management, such
52 as low resource settings, particularly to improve the targeting of antimicrobial prescribing.
53 However, it should be noted this result is also based on one study and requires confirming in
54 a larger study.
55
56
57

58 The allergy patch tests overall showed reasonable accuracy (sensitivity 72-97%; specificity
59 87-91%) and may be useful as an initial self-screening test for patients with a suspected
60

1
2
3 nickel or fragrance allergy; however it could be argued that the main use of these tests might
4 be to safely rule out a contact allergy. The reported diagnostic accuracy suggests the tests are
5 not sufficiently sensitive as a rule-out test, particularly given the relatively wide confidence
6 intervals. The tests may however be useful in settings where access to dermatology services
7 is scarce. The self-diagnosis of eczema using a questionnaire of signs and symptoms showed
8 specificity ranging from 0.79 to 1.00, suggesting this test might be useful as a confirmation
9 test for patients for suspect they have eczema and to then seek treatment and management
10 advice. The selected nature of the patients in the studies (i.e. patients with a previous eczema
11 diagnosis and hairdressers, who may encounter eczema more frequently due to their
12 profession) may overestimate the accuracy of the test; however, it could be argued that these
13 might be the populations in whom the test is most relevant.
14
15

16 We identified nine studies that reported the accuracy of self-diagnosis of HIV (sensitivity
17 93%, specificity 99%). However, the sensitivity is reduced to 88% when only studies using a
18 venous sample and laboratory testing as the reference standard are included. With a
19 sensitivity of 88%, the accuracy data would not support the use of this test as a rule-out test,
20 particularly given the clinical consequences of a false negative test result. Evidence suggests
21 there may be benefits to self-initiated HIV testing, including early identification, increased
22 likelihood for the uptake of HIV prevention interventions, and a reduction in sexual risk
23 behaviours (36), warranting further research, in particular in resource-limited settings where
24 access to testing sites may be a barrier. However, HIV self-testing should also be considered
25 in the context of linkage to care, access to counselling, and adequate regulatory and quality
26 assurance systems (37).
27
28
29

30 The search strategy for this review was broad and extensive with few restrictions resulting in
31 the high number of publications to screen. Whilst it is possible, it is unlikely studies were
32 missed. We are unaware of any other reviews examining this research question. The main
33 limitation was the lack of available evidence for a number of common conditions; studies
34 reporting on self-testing alone were more common, but few studies assessed self-diagnosis,
35 with patients interpreting the test results and making a diagnosis independently. For the three
36 conditions where we identified studies reporting on the diagnostic accuracy of self-diagnosis,
37 there was a paucity of evidence. Many studies were not replicated and included small sample
38 sizes and contained methodological biases that limited the application of the results to
39 practice. For self-diagnosis of vaginal infections, common skin conditions, and HIV further
40 research is required to draw a definitive conclusion on the efficacy of self-diagnosis. For
41 other common conditions in primary care, research is needed on self-diagnosis where this
42 option is available, and studies should go beyond considering self-testing alone and assess the
43 diagnostic accuracy of self-diagnosis. Terminology for self-diagnosis, self-testing and self-
44 screening is overlapping in some cases and needs clarifying. Finally, research is required into
45 the patient's readiness and attitude towards self-diagnosis along with its effect on the
46 patient/physician relationship.
47
48
49

50 The current limited evidence does not support routine self-diagnosis for vaginal infections,
51 common skin conditions and HIV in primary care.
52
53
54

55 **Acknowledgements**

56 The authors would like to thank Hayley Jones for joint authorship of the original protocol,
57 Elizabeth Spencer for assistance with preparation of the protocol, and Emily McFadden for
58 assistance with screening of studies for inclusion.
59
60

Contributors

AP and CH conceived the idea for this review. AP led the research, developed the protocol, and conducted the review (screening and extraction). CH developed the protocol and provided clinical input. NR devised and conducted the search strategy, and approved the manuscript. JM conducted the review (screening, extraction, data analysis). JM, AP and CH jointly prepared the manuscript.

Funding

This project was funded by National Institute for Health Research School for Primary Care Research (NIHR SPCR) [Project Number 390]

Data availability statement

All data are available in published articles.

Competing interests

Julie McLellan reports grants from NIHR School of Primary Care Research (Evidence Synthesis Working Group Project Number 390) and part-funding from the Thames Valley Applied Research Collaborative (ARC), during the conduct of the study; and occasionally receives expenses for teaching Evidence-Based Medicine

Dr. Plüddemann reports grants from NIHR School of Primary Care Research (Evidence Synthesis Working Group Project Number 390), during the conduct of the study; and occasionally receives expenses for teaching Evidence-Based Medicine.

Dr. Heneghan reports receiving expenses and fees for his media work. He has received expenses from the WHO and holds grant funding from the NIHR, the NIHR School of Primary Care Research, The Wellcome Trust and the WHO. He has received financial remuneration from an asbestos case. He has also received income from the publication of a series of toolkit books published by Blackwells. On occasion, he receives expenses for teaching EBM and is also paid for his GP work in NHS out of hours. CEBM jointly runs the EvidenceLive Conference with the BMJ and the Overdiagnosis Conference with some international partners which are based on a non-profit making model.

Ethics Approval

Not applicable.

References

1. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. *Lancet*. 2016;387(10035):2323-30.
2. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*. 2017;7(10):e017902.
3. Orgnaistion WH. Global status report on noncommunicable diseases 2014. 2014.

4. British Medical Association. BMA guidance: Controlling workload in general practice strategy 2020 [Available from: <https://www.bma.org.uk/advice-and-support/gp-practices/managing-workload/controlling-workload-in-general-practice-strategy>] Accessed 18 June 2021.
5. Hynes V. The trend toward self-diagnosis. *CMAJ*. 2013;185(3):E149-E50.
6. Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. *Australian Family Physician*. 2013;42:65-8.
7. The EndNote Team. Endnote. Endnote X9 ed. Philadelphia, PA: Clarivate; 2013.
8. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.
9. Review Manager (RevMan) [Computer program]. Version 5.4. The Cochrane Collaboration. 2020.
10. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.
11. RStudio Team. 2021. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA [Available from: <http://www.rstudio.com/>] Accessed 22 June 2021.
12. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-93.
13. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008;149(12):889-97.
14. Donders GGG, Gonzaga A, Marconi C, Donders F, Michiels T, Eggermont N, et al. Increased vaginal pH in Ugandan women: what does it indicate? *European Journal of Clinical Microbiology & Infectious Diseases*. 2016;35(8):1297-303.
15. Sungkar A, Purwosunu Y, Aziz MF, Pratomo H, Sutrisna B, Sekizawa A. Influence of early self-diagnosis and treatment of bacterial vaginosis on preterm birth rate. *International Journal of Gynaecology & Obstetrics*. 2012;117(3):264-7.
16. Geva A, Bornstein J, Dan M, Shoham HK, Sobel JD. The VI-SENSE-vaginal discharge self-test to facilitate management of vaginal symptoms. *American Journal of Obstetrics & Gynecology*. 2006;195(5):1351-6.
17. Ryan-Wenger NA, Neal JL, Jones AS, Lowe NK. Accuracy of vaginal symptom self-diagnosis algorithms for deployed military women. *Nursing Research*. 2010;59(1):2-10.
18. Jones HE, Lippman SA, Caiaffa-Filho HH, Young T, van de Wijgert JHHM. Performance of a rapid self-test for detection of *Trichomonas vaginalis* in South Africa and Brazil. *Journal of Clinical Microbiology*. 2013;51(3):1037-9.
19. Bregnhøj A, Sosted H, Menne T, Johansen JD. Validation of self-reporting of hand eczema among Danish hairdressing apprentices. *Contact Dermatitis*. 2011;65(3):146-50.
20. Svensson A, Lindberg M, Meding B, Sundberg K, Stenberg B. Self-reported hand eczema: symptom-based reports do not increase the validity of diagnosis. *British Journal of Dermatology*. 2002;147(2):281-4.
21. Elsner P, Bjarnoe H, Aberer W, Brasch J. Evaluation of a self-test device in allergic contact dermatitis. *Dermatitis*. 2009;20 (4):230-1.
22. Elsner P, Bjarnøe H, Aberer W, Brasch J. Evaluation of a Self-Test Device Used in Allergic Contact Dermatitis. *Skin Pharmacol Physiol*. 2015;28(5):257-63.
23. Josefson A, Svensson A, Farm G, Engfeldt M, Meding B. Validation of self-testing as a method to estimate the prevalence of nickel allergy. *Acta Dermato-Venereologica*. 2011;91(5):526-30.

- 1
- 2
- 3
- 4 24. Asiimwe S, Oloya J, Song X, Whalen CC. Accuracy of un-supervised versus
- 5 provider-supervised self-administered HIV testing in Uganda: a randomized implementation
- 6 trial. *AIDS and behavior*. 2014;18(12):2477-84.
- 7 25. Belete W, Deressa T, Feleke A, Menna T, Moshago T, Abdella S, et al. Evaluation of
- 8 diagnostic performance of non-invasive HIV self-testing kit using oral fluid in Addis Ababa,
- 9 Ethiopia: A facility-based cross-sectional study. *PLoS One*. 2019;14(1):e0210866.
- 10 26. Choko AT, Desmond N, Webb EL, Chavula K, Napierala-Mavedzenge S, Gaydos
- 11 CA, et al. The uptake and accuracy of oral kits for HIV self-testing in high HIV prevalence
- 12 setting: a cross-sectional feasibility study in Blantyre, Malawi. *Plos medicine*.
- 13 2011;8(10):e1001102.
- 14 27. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al.
- 15 Uptake, Accuracy, Safety, and Linkage into Care over Two Years of Promoting Annual Self-
- 16 Testing for HIV in Blantyre, Malawi: A Community-Based Prospective Study. *PLoS Med*.
- 17 2015;12(9):e1001873.
- 18 28. Kapaku K, Newman M, Maluzi K, Sigande L, Taegtmeier M, Corbett E, et al.,
- 19 editors. Is OraQuick HIV-self-testing valid among intended users? Analysis from a clinical
- 20 performance study in
- 21 Lusaka, Zambia. 22nd International AIDS Conference; 2017 24-26 July; Paris, France.
- 22 29. UNITAID/PSI HIV Self-Testing Africa (HIV STAR) study. A clinical performance
- 23 study of self-testing using self-collected oral fluid transudate and the OraQuick® HIV Self-
- 24 Test [Available from: [https://hivstar.lshtm.ac.uk/files/2016/11/STAR-Protocol-for-Zambia-](https://hivstar.lshtm.ac.uk/files/2016/11/STAR-Protocol-for-Zambia-Clinical-Performance-Study-web.pdf)
- 25 [Clinical-Performance-Study-web.pdf](https://hivstar.lshtm.ac.uk/files/2016/11/STAR-Protocol-for-Zambia-Clinical-Performance-Study-web.pdf)] Accessed 18 June 2021.
- 26 30. Kurth AE, Cleland CM, Chhun N, Sidle JE, Were E, Naanyu V, et al. Accuracy and
- 27 Acceptability of Oral Fluid HIV Self-Testing in a General Adult Population in Kenya. *AIDS*
- 28 *Behav*. 2016;20(4):870-9.
- 29 31. Martínez Pérez G, Steele SJ, Govender I, Arellano G, Mkwamba A, Hadebe M, et al.
- 30 Supervised oral HIV self-testing is accurate in rural KwaZulu-Natal, South Africa. *Trop Med*
- 31 *Int Health*. 2016;21(6):759-67.
- 32 32. Pant Pai N, Behlim T, Abrahams L, Vadnais C, Shivkumar S, Pillay S, et al. Will an
- 33 unsupervised self-testing strategy for HIV work in health care workers of South Africa? A
- 34 cross sectional pilot feasibility study. *PLoS One*. 2013;8(11):e79772.
- 35 33. OraSure Technologies. Final Advisory Committee Briefing Materials: Available for
- 36 Public Release. OraQuick In-Home HIV Test. Washington (D.C.): Food and Drug
- 37 Administration: Blood Products Advisory Committee; 2012.
- 38 34. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. *Cochrane Handbook*
- 39 *for Systematic Reviews of Diagnostic Test Accuracy*.: The Cochrane Collaboration;
- 40 <http://srdta.cochrane.org>; 2010.
- 41 35. UNAIDS. *AIDSinfo Global factsheets 2019. HIV and AIDS Estimates* [Available
- 42 from <http://aidsinfo.unaids.org/>] Accessed 22 June 2021.
- 43 36. Adebayo OW, Salerno JP. Facilitators, Barriers, and Outcomes of Self-Initiated HIV
- 44 Testing: An Integrative Literature Review. *Res Theory Nurs Pract*. 2019;33(3):275-91.
- 45 37. Njau B, Covin C, Lisasi E, Damian D, Mushi D, Boulle A, et al. A systematic review
- 46 of qualitative evidence on factors enabling and deterring uptake of HIV self-testing in Africa.
- 47 *BMC Public Health*. 2019;19(1):1289.
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

1
2
3
4
5
6
7
8
9
10
11 **Figure 1: Study Selection**
12
13
14

15 **Figure 2: Paired forest plots of sensitivity and specificity for studies of self-diagnosis of vaginal infection (where Donders 2016a used pH threshold ≥ 4.5 , Donders 2016b used pH threshold ≥ 4.7 . Sungkar 2012 represents pooled data from 4 time points)**
16
17
18
19

20
21
22
23 **Figure 3: Paired forest plots of sensitivity and specificity for studies of self-diagnosis of common skin conditions (where Bregnhøj 2011 represents pooled data from recruitment and follow up time points)**
24
25
26
27

28
29 **Figure 4: Studies of self-diagnosis of HIV**
30

- 31 (a) Paired forest plots of sensitivity and specificity (where Choko 2015 represents
32 pooled data from 1- 12 and 13 – 24 months follow up time points)
33
34 (b) Receiver operating characteristic plot HIV self-diagnosis compared with clinical
35 diagnosis or laboratory reference test grouped by reference test type (where size of
36 symbol indicates study size)
37
38
39

40 **Figure 5: QUADAS-2 summary of risk of bias and applicability concerns showing review authors' judgments about each domain for each included study (based on 19 studies (20 data sets))**
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

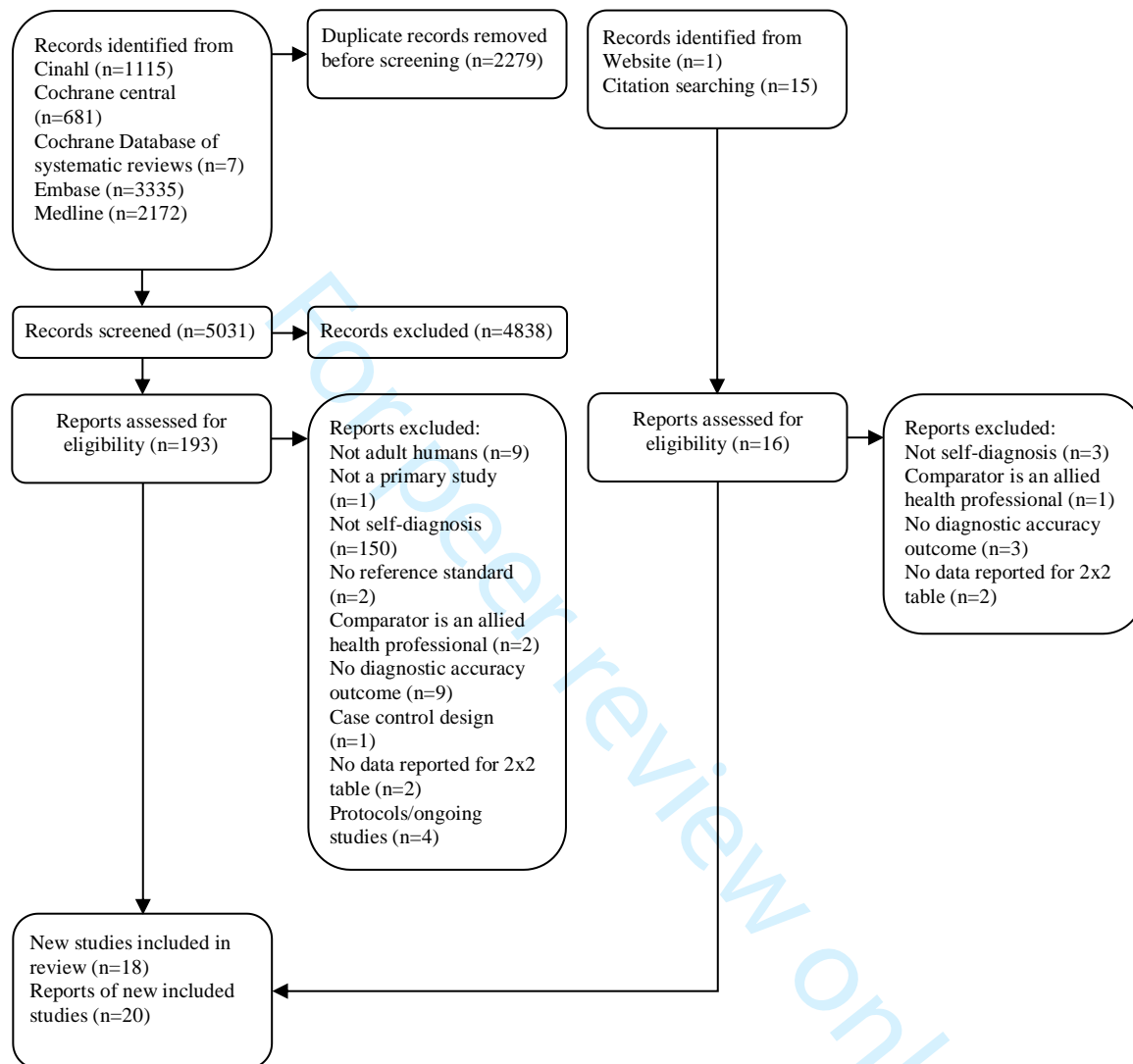
Figure 1: Study Selection

Figure 2: Paired forest plots of sensitivity and specificity for studies of self-diagnosis of vaginal infection (where Donders 2016a used pH threshold ≥ 4.5 , Donders 2016b used pH threshold ≥ 4.7 . Sungkar 2012 represents pooled data from 4 time points)

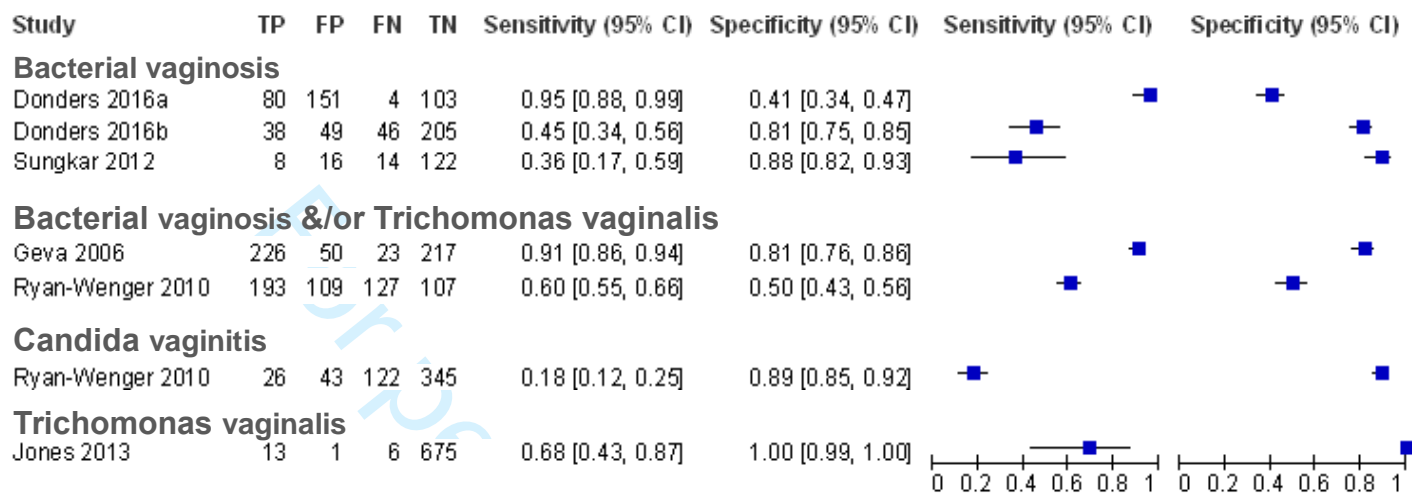


Figure 3: Paired forest plots of sensitivity and specificity for studies of self-diagnosis of common skin conditions (where Bregnhøj 2011 represents pooled data from recruitment and follow up time points)

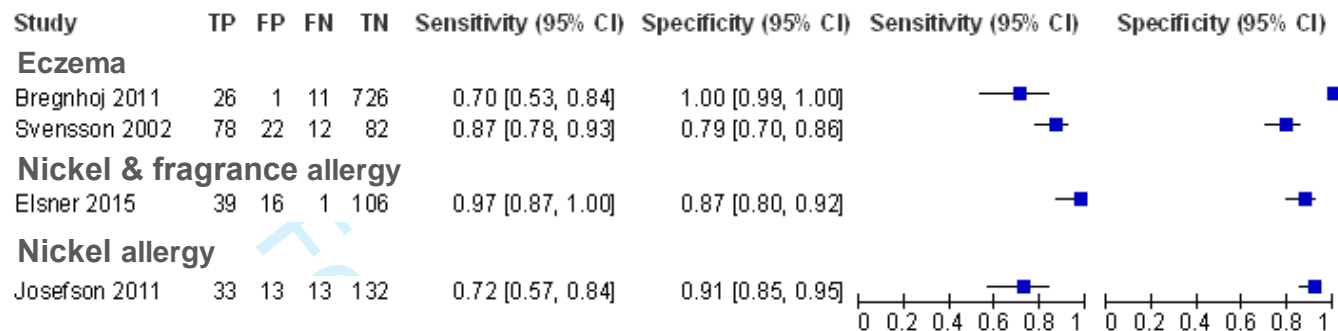
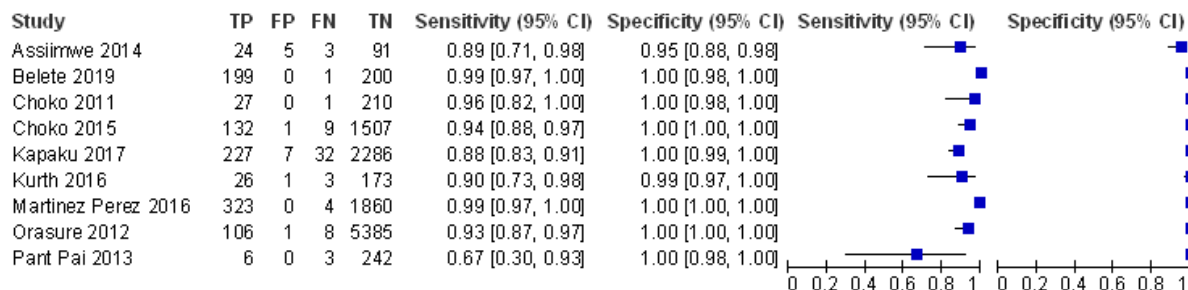
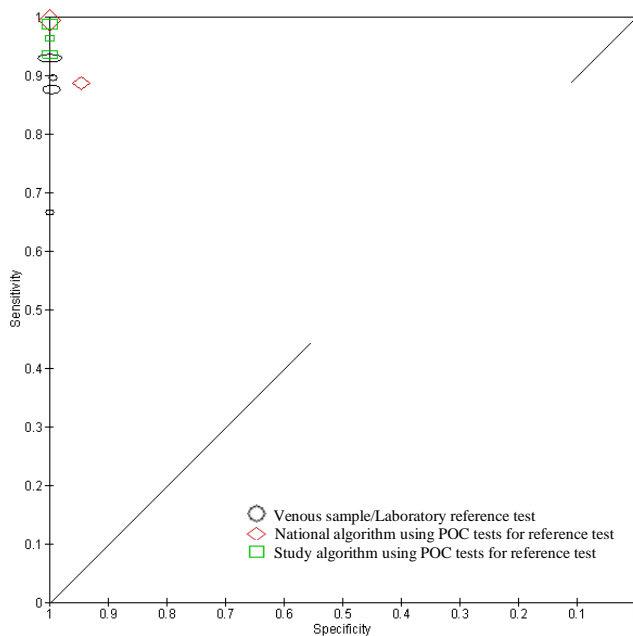


Figure 4: Studies of self-diagnosis of HIV

(a) Paired forest plots of sensitivity and specificity (where Choko 2015 represents pooled data from 1- 12 and 13 – 24 months follow up time points)



(b) Receiver operating characteristic plot HIV self-diagnosis compared with clinical diagnosis or laboratory reference test grouped by reference test type (where size of symbol indicates study size)



All studies (n=9):

Sensitivity: 92.8% (95%CI 86% to 96.5%); Specificity: 99.8% (95%CI 99.1% to 99.9%)

Sensitivity analysis retaining only studies with venous/laboratory reference test (n=4):

Sensitivity: 87.7% (95%CI 81.4% to 92.2%); Specificity: 99.8% (95%CI 98.9% to 99.9%)

Figure 5: QUADAS-2 summary of risk of bias and applicability concerns showing review authors' judgments about each domain for each included study (based on 19 studies (20 data sets))

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Assiimwe 2014	+	+	?	+	+	+	+
Belete 2019	+	?	?	+	+	+	+
Bregnhøj 2011	+	?	?	+	?	+	+
Choko 2011	+	+	-	+	+	+	+
Choko 2015	+	+	-	?	+	+	+
Donders 2016a	?	+	+	-	+	+	+
Donders 2016b	?	+	+	-	+	+	+
Elsner 2015	?	?	-	+	+	+	+
Geva 2006	?	?	+	-	+	+	+
Jones 2013	-	+	?	+	+	+	+
Josefson 2011	+	+	+	-	+	+	+
Kapaku 2017	+	+	?	+	+	+	+
Kurth 2016	?	+	+	-	+	+	+
Martinez Perez 2016	-	?	-	+	+	+	+
Orasure 2012	-	+	?	?	+	+	+
Pant Pai 2013	-	+	?	+	+	+	+
Ryan-Wenger 2010	?	+	-	-	+	+	+
Sungkar 2012	+	+	-	-	+	+	+
Svensson 2002	?	+	+	-	+	+	+

High
 Unclear
 Low

SUPPLEMENTARY MATERIAL

Supplementary Table 1: Commonly managed conditions in primary care

As reported by Cooke et al 2013 (Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. Australian Family Physician. 2013; 42:65-8)

1	Hypertension
2	Immunisation/vaccination: all
3	Acute upper respiratory tract infection
4	Depression
5	Diabetes: nongestational
6	Lipid disorders
7	General check-up
8	Osteoarthritis
9	Back complaint
10	Prescription
11	Oesophagus disease
12	Female genital check-up
13	Acute bronchitis/bronchiolitis
14	Asthma
15	Anxiety
16	Test results
17	Urinary tract infection
18	Dermatitis, contact/allergic
19	Pregnancy
20	Sleep disturbance
21	Sinusitis acute/chronic
22	Gastroenteritis
23	Vitamin/nutritional deficiency

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

24	Malignant neoplasm of skin
25	Abnormal test results
26	Atrial fibrillation/flutter
27	Oral contraception
28	Solar keratosis/sunburn
29	Ischaemic heart disease
30	Viral disease, not otherwise specified

For peer review only

Supplementary Table 2: Search strategy

Medline	
#	Searches
1	Diagnostic Self Evaluation/
2	((self* adj diagnos*) or selfdiagnos*).ti,ab.
3	(self* and diagnos*).ti.
4	((self* adj test*) or selftest*).ti,ab.
5	(home adj3 diagnos*).ti,ab.
6	((selfreport* or self-report*) and diagnos*).ti.
7	((selfreport* or self-report*) adj5 diagnos*).ti,ab.
8	(diagnos* and (selftreat* or self-treat*)).ti,ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	respiratory tract infections/ or common cold/ or influenza, human/ or laryngitis/ or exp pharyngitis/ or rhinitis/ or exp sinusitis/ or exp supraglottitis/ or tracheitis/ or exp otitis media/
11	((respirat* or airway*) adj2 infection*) or (common cold or influenza or flu or pharyngitis or laryngitis or tonsillitis or sore throat or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or otitis media)).ti,ab.
12	exp Depressive Disorder/ or Depression/
13	(depress* or (mood adj2 disorder*)).ti,ab.
14	diabetes mellitus/ or exp diabetes mellitus, type 1/ or exp diabetes mellitus, type 2/
15	diabet*.ti,ab.
16	dyslipidemias/ or exp hyperlipidemias/
17	(dyslipid?emia? or hyperlipid?emia? or hypercholesterol?emia? or cholesterol* or triglyceride*).ti,ab.
18	hypertension/
19	(hypertens* or high blood pressure).ti,ab.
20	exp Osteoarthritis/
21	(osteoarthritis or osteo-arthritis or (degenerat* adj2 arthritis)).ti,ab.
22	back pain/ or low back pain/
23	Sciatica/
24	Intervertebral Disc Displacement/
25	((back adj2 (pain or problem? or disorder?)) or slipped disc* or sciatica).ti,ab.
26	Esophageal Diseases/ or esophageal motility disorders/ or exp gastroesophageal reflux/ or Barrett Esophagus/
27	((oesophag* or esophag*) adj2 disease).ti,ab.
28	((oesophag* or esophag* or gastr*) adj2 reflux) or heartburn or heart burn).ti,ab.
29	((barrett* or globus) adj2 (esophag* or oesophag*)).ti,ab.
30	exp Asthma/
31	asthma*.ti,ab.

32	exp Anxiety Disorders/ or Anxiety/
33	(anxiety or phobia*).ti,ab.
34	exp Urinary Tract Infections/
35	exp Cystitis/
36	((urin* adj3 infection*) or bacteriuria or pyuria or cystitis).ti,ab.
37	exp Vaginitis/
38	(vaginitis or vaginosis).ti,ab.
39	dermatitis/ or exp dermatitis, contact/ or eczema/
40	(dermatitis or eczema).ti,ab.
41	exp Sleep Wake Disorders/
42	exp Sleep Apnea Syndromes/
43	((sleep adj2 (disorder* or disturbance*)) or insomnia*).ti,ab.
44	(sleep adj2 (apnea or apnoea)).ti,ab.
45	restless leg*.ti,ab.
46	Gastroenteritis/ or diarrhea/ or vomiting/
47	(gastroenteritis or (stomach adj2 (bug? or upset))).ti,ab.
48	(diarrhoea or diarrhea or food poisoning).ti,ab.
49	Sunburn/
50	(sunburn or solar keratosis).ti,ab.
51	or/10-50
52	9 and 51
53	exp "Sensitivity and Specificity"/
54	exp "REPRODUCIBILITY OF RESULTS"/
55	(sensitiv* or specific* or predict* or accura* or valid* or reproduc*).ti,ab.
56	53 or 54 or 55
57	52 and 56
Embase	
#	Searches
1	*Self Evaluation/
2	((self* adj diagnos*) or selfdiagnos*).ti,ab.
3	(self* and diagnos*).ti.
4	((self* adj test*) or selftest*).ti,ab.
5	(home adj3 diagnos*).ti,ab.
6	((selfreport* or self-report*) and diagnos*).ti.
7	((selfreport* or self-report*) adj5 diagnos*).ti,ab.
8	(diagnos* and (selftreat* or self-treat*)).ti,ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	respiratory tract infection/ or upper respiratory tract infection/ or exp respiratory tract inflammation/ or influenza/ or otitis media/
11	((respirat* or airway*) adj2 infection*) or (common cold or influenza or flu or pharyngitis or laryngitis or tonsillitis or sore throat or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or otitis media).ti,ab.

12	exp Depression/
13	(depress* or (mood adj2 disorder*)).ti,ab.
14	diabetes mellitus/ or exp diabetes mellitus, type 1/ or exp diabetes mellitus, type 2/
15	diabet*.ti,ab.
16	dyslipidemia/ or exp hyperlipidemia/
17	(dyslipid?emia? or hyperlipid?emia? or hypercholesterol?emia? or cholesterol* or triglyceride*).ti,ab.
18	hypertension/
19	(hypertens* or high blood pressure).ti,ab.
20	exp Osteoarthritis/
21	(osteoarthritis or osteo-arthritis or (degenerat* adj2 arthritis)).ti,ab.
22	backache/ or discogenic pain/ or low back pain/
23	Sciatica/
24	intervertebral disk hernia/ or intervertebral disk disease/
25	((back adj2 (pain or problem? or disorder?)) or slipped disc* or sciatica).ti,ab.
26	exp gastroesophageal reflux/ or barrett esophagus/ or esophagus disease/
27	((oesophag* or esophag*) adj2 disease).ti,ab.
28	((oesophag* or esophag* or gastr*) adj2 reflux) or heartburn or heart burn).ti,ab.
29	((barrett* or globus) adj2 (esophag* or oesophag*)).ti,ab.
30	exp Asthma/
31	asthma*.ti,ab.
32	exp Anxiety Disorder/
33	(anxiety or phobia*).ti,ab.
34	exp Urinary Tract Infection/
35	Cystitis/
36	((urin* adj3 infection*) or bacteriuria or pyuria or cystitis).ti,ab.
37	vagina discharge/ or exp vaginitis/
38	(vaginitis or vaginosis).ti,ab.
39	dermatitis/ or contact dermatitis/ or exp eczema/
40	(dermatitis or eczema).ti,ab.
41	sleep disorder/ or exp insomnia/ or periodic limb movement disorder/
42	exp sleep disordered breathing/
43	((sleep adj2 (disorder* or disturbance*)) or insomnia*).ti,ab.
44	(sleep adj2 (apnea or apnoea)).ti,ab.
45	restless leg*.ti,ab.
46	Gastroenteritis/ or diarrhea/ or vomiting/
47	(gastroenteritis or (stomach adj2 (bug? or upset))).ti,ab.
48	(diarrhoea or diarrhea or food poisoning).ti,ab.
49	Sunburn/
50	(sunburn or solar keratosis).ti,ab.
51	or/10-50

52	9 and 51
53	"Sensitivity and Specificity"/
54	diagnostic accuracy/ or diagnostic test accuracy study/
55	predictive validity/ or predictive value/ or reproducibility/
56	(sensitiv* or specific* or predict* or accura* or valid* or reproduc*).ti,ab.
57	53 or 54 or 55 or 56
58	52 and 57
Cochrane	
ID	Search
#1	MeSH descriptor: [Diagnostic Self Evaluation] explode all trees
#2	((self* AND diagnos*) or selfdiagnos*):ti OR ((self* NEXT diagnos*) or selfdiagnos*):ti,ab,kw OR ((self* NEXT test*) or selftest*):ti,ab,kw OR (home NEAR/3 diagnos*):ti,ab,kw
#3	((selfreport* or self-report*) and diagnos*):ti OR ((selfreport* or self-report*) NEAR diagnos*):ti,ab,kw OR ((diagnos* and (selftreat* or self-treat*)):ti,ab,kw
#4	#1 OR #2 OR #3
#5	MeSH descriptor: [Sensitivity and Specificity] explode all trees
#6	MeSH descriptor: [Reproducibility of Results] explode all trees
#7	(sensitiv* or specific* or predict* or accura* or valid* or reproduc*).ti,ab,kw
#8	#5 OR #6 OR #7
#9	#4 AND #8
Cinahl	
#	Query
S39	S3 AND S35 AND S38
S38	S36 OR S37
S37	TX sensitiv* or specific* or predict* or accura* or valid* or reproduc*
S36	(MH "Sensitivity and Specificity") OR (MH "Predictive Value of Tests")
S35	(S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34)
S34	TI (sunburn or "solar keratosis") OR AB (sunburn or "solar keratosis")
S33	(MH "Sunburn")
S32	TI ((gastroenteritis or (stomach N2 (bug or bugs or upset)))) OR AB ((gastroenteritis or (stomach N2 (bug or bugs or upset)))) OR TI (diarrhoea or diarrhea or "food poisoning") OR AB (diarrhoea or diarrhea or "food poisoning")
S31	(MH "Gastroenteritis") OR (MH "Vomiting") OR (MH "Diarrhea")
S30	TI (((sleep N2 (disorder* or disturbance*)) or insomnia*)) OR AB (((sleep N2 (disorder* or disturbance*)) or insomnia*)) OR TI ((sleep N2 (apnea or apnoea))) OR AB ((sleep N2 (apnea or apnoea))) OR TI "restless legs" OR AB "restless legs"
S29	(MH "Sleep Disorders, Intrinsic+")
S28	TI (dermatitis or eczema) OR AB (dermatitis or eczema)
S27	(MH "Dermatitis") OR (MH "Dermatitis, Contact+") OR (MH "Eczema")
S26	TI (vaginitis or vaginosis) OR AB (vaginitis or vaginosis)

S25	(MH "Vaginitis+")
S24	TI (((urin* N3 infection*) or bacteriuria or pyuria or cystitis)) OR AB (((urin* N3 infection*) or bacteriuria or pyuria or cystitis))
S23	(MH "Urinary Tract Infections+") OR (MH "Cystitis")
S22	TI asthma* OR AB asthma*
S21	(MH "Asthma+")
S20	TI (((oesophag* or esophag*) N2 disease)) OR (((oesophag* or esophag*) N2 disease)) OR TI ((((oesophag* or esophag* or gastr*) N2 reflux) or heartburn or "heart burn")) OR AB ((((oesophag* or esophag* or gastr*) N2 reflux) or heartburn or "heart burn")) OR (((barrett* or globus) N2 (esophag* or oesophag*))) OR (((barrett* or globus) N2 (esophag* or oesophag*)))
S19	(MH "Gastroesophageal Reflux") OR (MH "Esophageal Motility Disorders") OR (MH "Esophageal Diseases") OR (MH "Barrett Esophagus")
S18	TI (((back N2 (pain or problem* or disorder*)) or "slipped disc*" or sciatica)) OR AB (((back N2 (pain or problem* or disorder*)) or "slipped disc*" or sciatica))
S17	(MH "Low Back Pain") OR (MH "Back Pain") OR (MH "Sciatica") OR (MH "Intervertebral Disk Displacement")
S16	TI ((osteoarthritis or osteo-arthritis or (degenerat* N2 arthritis))) OR AB ((osteoarthritis or osteo-arthritis or (degenerat* N2 arthritis)))
S15	(MH "Osteoarthritis+")
S14	TI (hypertens* or "high blood pressure") OR AB (hypertens* or "high blood pressure")
S13	(MH "Hypertension")
S12	TI (dyslipidemia* or hyperlipidemia* or hypercholesterolemia* or dyslipidaemia* or hyperlipidaemia* or hypercholesterolaemia* or cholesterol* or triglyceride*) OR AB (dyslipidemia* or hyperlipidemia* or hypercholesterolemia* or dyslipidaemia* or hyperlipidaemia* or hypercholesterolaemia* or cholesterol* or triglyceride*)
S11	(MH "Hyperlipidemia+")
S10	TI diabet* OR AB diabet*
S9	(MH "Diabetes Mellitus") OR (MH "Diabetes Mellitus, Type 1") OR (MH "Diabetes Mellitus, Type 2")
S8	TI ((depress* OR anxiety or (mood N2 disorder*))) OR AB ((depress* OR anxiety or (mood N2 disorder*)))
S7	(MH "Depression") OR (MH "Anxiety")
S6	TI ((((respirat* or airway*) N2 infection*) or ("common cold" or influenza or flu or pharyngitis or laryngitis or tonsillitis or "sore throat" or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or "otitis media"))) OR AB ((((respirat* or airway*) N2 infection*) or ("common cold" or influenza or flu or pharyngitis or laryngitis or tonsillitis or "sore throat" or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or "otitis media")))
S5	(MH "Otitis Media+")
S4	(MH "Respiratory Tract Infections") OR (MH "Common Cold") OR (MH "Influenza") OR (MH "Influenza, Human+") OR (MH "Laryngitis+") OR (MH "Pharyngitis") OR (MH "Rhinitis+") OR (MH "Sinusitis+") OR (MH "Tonsillitis+")

S3	S1 OR S2
S2	TI (((self* AND diagnos*) or selfdiagnos*)) OR AB (((self* N1 diagnos*) or selfdiagnos*)) OR TI (((self* N1 test*) or selftest*)) OR AB (((self* N1 test*) or selftest*)) OR TI (home N3 diagnos*) OR AB (home N3 diagnos*) OR TI (((selfreport* or self-report*) and diagnos*)) OR AB (((selfreport* or self-report*) N5 diagnos*)) OR TI ((diagnos* and (selftreat* or self-treat*))) OR AB ((diagnos* and (selftreat* or self-treat*)))
S1	(MH "Self Diagnosis")

For peer review only

Supplementary table 3: Data items included in extraction sheet (where available)

Study identification - author, year, location
Study research question
Study design and setting
Study funding source
Target condition definition/diagnostic criteria
Participant characteristics and numbers, including exclusions
Index test
Reference standard
Flow of participants through study including losses to follow-up
Patient presentation and prior testing
Conduct of the study including timing of the tests, and information on masking
Absolute counts of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) diagnoses.
Statistical analyses that were performed, including whether all participants were included in analyses
Additional summary information on participant preference, timing, or cost, as available.

Supplementary table 4: Protocol

Accuracy of self-diagnosis in conditions commonly managed in primary care: diagnostic accuracy review

Annette Plüddemann¹, Hayley E Jones², Carl Heneghan¹

1. Nuffield Department of Primary Care Health Sciences, University of Oxford, OX2 6GG
2. Population Health Sciences, Bristol Medical School, University of Bristol, BS8 2PS

Background and rationale

A wide range of conditions present to primary care, some acute, some chronic. As a consequence primary care is facing increasing workload (Hobbs et al 2016) that may become unmanageable.

Some common conditions in primary care, therefore, have the potential to be self-diagnosed and self-treated by the patients themselves. This offers superior convenience for individuals, swifter diagnosis and treatment where relevant, reduced costs for health service providers and potentially reduce the burden on primary care services. Self-diagnosis may apply to initial diagnosis, or to diagnosing an exacerbation of an ongoing condition, such as an exacerbation of chronic obstructive pulmonary disease (COPD). In order to support self-diagnosis where safe and appropriate, the efficacy of self-diagnosis needs to be assessed, and this information used to make evidence-based decisions on who can self-diagnose safely.

Where there is an available comparison with diagnosis by a healthcare professional, it is possible to assess the accuracy of self-diagnosis. A further comparison could potentially be made with diagnosis by an allied healthcare professional such as pharmacists, but this is outside the scope of this current review. Sometimes, self-diagnosis involves using a diagnostic test (any type of medical test used to help diagnose or detect disease). For conditions where this is available and appropriate, the accuracy of these tests also informs the safety and efficacy of self-diagnosis by the patient.

Previous studies on the safety and accuracy of self-diagnosis of conditions commonly managed in primary care setting include self-diagnosis of urinary tract infection (Donofrio & Weiner 2013), high blood pressure (Tormo et al 2000) and depression (Hedayati et al 2006).

This review aims to identify, appraise and summarise the available evidence on self-diagnosis in common conditions in primary care. Cooke and colleagues recently reported the 30 most commonly managed conditions in primary care in Australia, which has a health landscape broadly comparable with western Europe (Cooke et al 2013). This list arises from survey data collected between January 2009 and December 2010, which included 194,100 patient encounters from 1,941 GPs. The most commonly managed conditions included some with the potential for self-diagnosis, e.g. urinary tract infection, as well as some that would be unsuitable for self-diagnosis, such as "general check-up". We base our review on conditions

1
2
3 from this list that are relevant for self-diagnosis. We may review infectious diseases and non-
4 communicable diseases separately.
5

6 7 **Objectives**

8 ***Primary objective***

9
10 Our primary objective is to summarize the accuracy of self-diagnosis of common conditions in
11 primary care, compared with diagnosis by a healthcare provider.
12

13 ***Secondary objective***

14 To summarise any associated relevant information relating to self-diagnosis of common
15 conditions in primary care, such as information on patient preference, timing, or cost (only
16 using information from studies we include for accuracy data). Where there is substantial
17 qualitative information reported, this will only be summarised briefly; detailed qualitative
18 approaches will not be used.
19
20
21

22 23 **Methods**

24 **Criteria for considering studies for this review**

25 ***Types of studies***

26
27 Prospective or retrospective studies comparing the results of self-diagnosis of common self-
28 limiting conditions in primary care by free-living individuals, to the results of a reference
29 standard test performed by a healthcare service provider, will be included. Studies with a
30 case-control design will be excluded. In case of duplicate publications we will include the
31 study report with the highest methodological quality. There will be no language restrictions.
32

33 We will exclude studies comparing self-diagnosis with diagnosis by allied health professionals
34 such as a pharmacists.
35

36 ***Participants***

37 Adults (≥ 18 years of age) self-diagnosing conditions common in primary care.
38

39 ***Index tests***

40 Index tests will be the self-testing or self-diagnosis of relevant conditions, compared with
41 diagnosis by a healthcare practitioner.
42

43 ***Comparator tests***

44 Comparator tests will comprise diagnosis by a healthcare practitioner.
45

46 ***Outcome measures***

47 Diagnostic accuracy measures (e.g. sensitivity, specificity, likelihood ratios, predictive values,
48 etc.) and primary data for 2x2 tables. Studies reporting only measures of agreement will be
49 excluded.
50
51
52
53
54
55
56
57
58
59
60

Search methods to identify studies

Electronic searches

The search strategy will be developed in consultation with a healthcare librarian experienced with supporting systematic reviews. No language restrictions will be applied. The search strategy will use multiple electronic databases, from inception onwards including:

Medline

EMBASE

Cochrane Central Register of Controlled Trials (CENTRAL)

Trip database

Web of Science for conference proceedings, dissertations, and theses

World Health Organization International Clinical Trials Registry Platform (ICTRP),

ClinicalTrials.gov

Database of Abstracts of Reviews of Effect (DARE)

We will also search Science Citation Index Expanded for study reports that cite the included studies.

The search may use relevant filters, but in order to maximise sensitivity, will not be limited to these. The reference lists of relevant studies will be examined and additional tools such as the “related articles” feature in PubMed will also be used to identify relevant publications.

Data collection and analysis

Selection of studies

Two reviewers will independently apply the selection criteria to the titles and abstracts of the study reports identified by the searches. If the decision to exclude a study cannot be made on the basis of the title and the abstract, the full study report will be retrieved for inclusion assessment. The final decision on inclusion will be based on the full study report.

Disagreements between reviewers will be resolved by discussion, or if necessary by a third reviewer. Study identification will be summarised in a PRISMA flow diagram.

Data extraction and management

Two reviewers will independently extract information from selected studies into a data extraction sheet. Disagreements will be resolved by discussion, or if necessary with the help of a third reviewer.

Where this is insufficient (or unclear) information, where there is an email address provided, the authors will be contacted via email for clarification. Where data is not available for completion of 2x2 tables, the studies will be excluded from the analysis.

Data to be extracted

The following information will be extracted from the included studies, where available:

Study identification - author, year, location

Study research question

Study design and setting

Target condition definition/diagnostic criteria

Participant characteristics and numbers, including exclusions

Index test

Reference standard

Flow of participants through study including losses to follow-up

Patient presentation and prior testing

Conduct of the study including timing of the tests, and information on masking

Absolute counts of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) diagnoses.

Statistical analyses that were performed, including whether all participants were included in analyses

Additional summary information on participant preference, timing, or cost, as available.

Assessment of methodological quality

To assess methodological quality, we will use the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (Whiting et al 2011). Two reviewers will independently assess studies' methodological quality; disagreements will be resolved by discussion, or if necessary, by a third reviewer. The QUADAS-2 tool facilitates assessment of bias in four areas: patient selection; index test; reference standard; flow and timing; and also facilitates assessment of applicability of the studies to the review research question.

The data will be presented in a tables showing risk of bias and applicability within each domain assessed for each study. These data will be considered in relation to interpreting the results of the studies.

Statistical analysis and data synthesis

Analyses will be conducted for each category of condition specified. Summary tables will detail study information including the patient sample, condition, study design, the test under evaluation, and the comparator.

Meta-analysis

For each test, RevMan will be used to produce paired forest plots to explore the between-study variability of sensitivity and specificity across the included studies. For each study estimate of sensitivity and specificity, corresponding 95% confidence intervals will be shown

1
2
3 to illustrate the uncertainty related to each study estimate. If accuracy has been reported at
4 multiple common thresholds, forest plots will be sub-grouped on threshold.
5

6 Bivariate meta-analysis methods (Reitsma et al 2005) will be used to generate pooled
7 estimates of sensitivity and specificity where sufficient data is available for each test or
8 condition. These will be plotted with 95% confidence and prediction ellipses in Receiver
9 Operating Characteristic (ROC) space. Where appropriate, summary ROC curves will also be
10 plotted, drawing on the equivalence of the bivariate method and the hierarchical summary
11 ROC meta-analysis model (Rutter and Gatsonis 2001; Harbord et al 2007). For these
12 analyses, we will use WinBUGS or the metandi command in Stata, as appropriate, and feed
13 parameters directly into Revman to produce Cochrane-standardised output.
14

15 Where appropriate, meta-analysis models that include multiple thresholds will be employed
16 (e.g. Steinhauser et al 2016 or similar).
17

18 *Investigating heterogeneity*

19

20 For medical conditions for which data from more than one study are available, it may be
21 possible to investigate heterogeneity in the results. Two approaches will be used to explore
22 the sources of between-study heterogeneity: 1) inclusion of study level characteristics as
23 covariates in the bivariate model (meta-regression) 2) carrying out sub-group analyses.
24 These approaches will only be carried out if there is sufficient data available and sub-group
25 specific pooled estimates are thought to be of clinical relevance. Any meta-regressions will be
26 carried out using WinBUGS or the xtmelogit command in Stata.
27

28 *Sensitivity analyses*

29

30 If there appear to be any outliers in the data, these studies will be removed from the analysis
31 to evaluate the impact on the overall pooled estimates.
32

33 *Investigating reporting bias*

34

35 Funnel plots used to detect publication bias in reviews of RCTs have been shown to be
36 misleading for diagnostic test accuracy reviews (Deeks et al 2005; Leeflang et al 2008).
37 Funnel plots as an assessment of reporting bias will therefore not be included in this review.
38 Publication bias will be assessed using Deek's test, as recommended by the Cochrane
39 Handbook, where data allows (Deeks et al 2005).
40

41 **Funding**

42

43 This project is funded by National Institute for Health Research School for Primary Care
44 Research (NIHR SPCR) [ProjectNumber 390]
45
46

47 **Declarations of conflict of interest**

48

49 Dr. Plüddemann reports grants from NIHR, grants from NIHR School of Primary Care
50 Research, during the conduct of the study; and occasionally receives expenses for teaching
51 Evidence-Based Medicine. Dr. Heneghan reports receiving expenses and fees for his media
52 work. He has received expenses from the WHO and holds grant funding from the NIHR, the
53 NIHR School of Primary Care Research, The Wellcome Trust and the WHO. He has received
54 financial remuneration from an asbestos case. He has also received income from the
55
56
57
58
59
60

1
2
3 publication of a series of toolkit books published by Blackwells. On occasion, he receives
4 expenses for teaching EBM and is also paid for his GP work in NHS out of hours. CEBM
5 jointly runs the EvidenceLive Conference with the BMJ and the Overdiagnosis Conference
6 with some international partners which are based on a non-profit making model.
7

8 **References**

9
10
11 Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and
12 publication frequency. *Aust Fam Physician*. 2013;42(1-2):65-8.

13
14 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample
15 size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*.
16 2005;58(9):882-93

17
18 Donofrio JC, Weiner SG. Female patient self-diagnosis compared with emergency physician
19 diagnosis of urinary tract infection. *J Emerg Med*. 2013;45(6):969-73. doi:
20 10.1016/j.jemermed.2013.07.019

21
22 Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-
23 analysis of diagnostic accuracy studies. *Biostatistics*. 2007 Apr;8(2):239-51.

24
25 Hedayati SS, Bosworth HB, Kuchibhatla M, Kimmel PL, Szczech LA. The predictive value of
26 self-report scales compared with physician diagnosis of depression in hemodialysis patients.
27 *Kidney Int*. 2006 May;69(9):1662-8.

28
29 Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, Salisbury C;
30 National Institute for Health Research School for Primary Care Research. Clinical workload in
31 UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14.
32 *The Lancet*; 387; 10035: 2323-2330
33

34
35 Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy
36 Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008;
37 16;149(12): 889-97

38
39 Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic
40 test accuracy evaluations. *Stat Med*. 2001 Oct 15;20(19):2865-84.

41
42 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate
43 analysis of sensitivity and specificity produces informative summary measures in diagnostic
44 reviews. *J Clin Epidemiol*. 2005 Oct;58(10):982-90

45
46 Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic
47 Cochrane Centre, The Cochrane Collaboration, 2014

48
49 Steinhauser S, Schumacher M, Rucker G. Modelling multiple thresholds in meta-analysis of
50 diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016 Aug 12;16(1):97. doi:
51 10.1186/s12874-016-0196-1.

52
53 Tormo MJ, Navarro C, Chirlaque MD, Barber X. Validation of self diagnosis of high blood
54 pressure in a sample of the Spanish EPIC cohort: overall agreement and predictive values.
55 EPIC Group of Spain. *J Epidemiol Community Health*. 2000 Mar;54(3):221-6.

Supplementary Table 5: Characteristics of included studies

Disease/condition	Author, year	Design	Setting	Country	Study duration	Number of participants	Mean age* (years)	Female (%)
Vaginal infection	Donders 2016	Prospective diagnostic accuracy study	Birth control, general gynaecology, infertility and prenatal clinics	Uganda	N/A	360	28.3	100
	Sungkar 2012	RCT (intervention arm only)	Prenatal clinics or hospitals	Indonesia	24 weeks	176	28	100
	Geva 2006	Prospective diagnostic accuracy study	Gynaecologic clinics	Israel	N/A	593	18 – 60 range	100
	Ryan-Wenger 2010	Prospective diagnostic accuracy study	Military clinics	USA	N/A	546	25.7	100
	Jones 2013	RCT (both arms)	Clinics or home	Brazil	N/A	695	18 – 40 range	100
Common skin condition	Bregnhøj 2011	Prospective cohort	Not reported	Denmark	18 months	502	17.5	95
	Svensson 2002	Prospective diagnostic accuracy study	Dermatology outpatient clinics	Sweden	N/A	208	40.4	50
	Elsner 2015	Prospective diagnostic accuracy study	Not reported	Germany & Austria	N/A	165	≥18 years	81
	Josefson 2011	Prospective diagnostic accuracy study	Hospital dermatology departments	Sweden	N/A	243	44	69
HIV	Assiimwe 2014	RCT (unsupervised arm only)	Home	Uganda	N/A	123	28 (23–32), median (IQR)	47
	Belete 2019	Cross-sectional	Public health facilities	Ethiopia	N/A	400	29 (17.7–40.3), median (IQR)	61
	Choko 2011	Cross-sectional	Home	Malawi	N/A	241	27 (NR) median (IQR)	52
	Choko 2015	RCT (intervention arm only)	Home	Malawi	2 years	2370	NR	NR

	Kapaku 2017	Cohort	Home/Voluntary counselling & testing facilities	Zambia	1 year	2572	26 (21–35), median (IQR)	59
	Kurth 2016	Cross-sectional	NR (not at home)	Kenya	N/A	240**	36	33
	Martinez Perez 2016	Cross-sectional	Health care clinics/HIV testing sites	South Africa	N/A	2205	Male: 27 (11-36) Female: 28 (22-36) median (IQR)	66
	Orasure 2012	Cross-sectional	Study site	USA	N/A	5798	NR	50
	Pant Pai 2013	Cross-sectional	Hospital	South Africa	N/A	251	≥18 years	79

RCT – randomised control trial; N/A – not applicable; NR – not reported, * unless otherwise indicated, ** Subset (N=113) that used laboratory reference standard included in systematic review and pooled analysis.

Supplementary Table 6: Characteristics of self-diagnosis (index) tests and reference standard tests

Disease/condition	Author, year	Test for	Self-diagnosis test (index)	Self-diagnosis test threshold	Reference test	Reference test threshold	Interval between index & reference test	Data collection points
Vaginal infection	Donders 2016a	Bacterial vaginosis	Vaginal fluid test using pH strip	+/- (≥ 4.5 pH)	Air dried vaginal fluid test using gram staining (Nugent score)	+/- (≥ 4.5 pH, Nugent score 7-10)	NR	N/A
	Donders 2016b			+/- (≥ 4.7 pH)	Assessment by central laboratory	+/- (≥ 4.7 pH, Nugent score 7-10)		
	Sungkar 2012	Bacterial vaginosis	Vaginal fluid test using pH strip	+/-	Vaginal fluid test using microbiological gram staining test (Kopeloff modified Gram stain) Prepared by midwives for laboratory assessment	+/-	NR	Baseline, 16 – 18, 18 – 20, 20 – 22, 22 – 24 weeks
	Geva 2006	Bacterial vaginosis and/or <i>Trichomonas vaginalis</i>	Vaginal discharge test using panty liner test kit (VI-SENSE)	+/- (based on strip colour, no level reported*)	Clinical diagnosis plus vaginal wall swabs testing pH (nitrazine paper), amine, culture (InPouch TV, BioMed Diagnostic) and gram staining (Nugent score). Assessment by board certified gynaecologists and central laboratory.	BV: >3: (a) homogeneous discharge, (b) pH value >4.5, (c) release of fishy odor (KOH was added to the vaginal discharge, and (d) presence of clue cells; or 7+ Nugent score of gram stain TV: +/- by culture	Within 6 hours	N/A
	Ryan-Wenger 2010	Bacterial vaginosis and/or <i>Trichomonas vaginalis</i>	Vaginal fluid test using Women in the military self-diagnosis kit	+/- (based on ≥ 4.7 pH, amines,	Clinical interview plus vaginal fluid test for pH (nitrazine paper, amines (FemExam card), whiff, wet mount	+/-	None	N/A

			(includes FemExam card)	vaginal itching)	microscopy (Affirm VPIII Microbial Identification Test) Assessment by women's health nurse practitioner plus in clinic microscopy testing			
	Ryan-Wenger 2010	<i>Candida</i> vaginitis	Vaginal fluid test using Women in the military self-diagnosis kit (includes FemExam card)	+/- (based on ≥ 4.7 pH, amines, vaginal itching)	Clinical interview plus vaginal fluid test for pH (nitrazine paper, amines (FemExam card), whiff, wet mount microscopy (Affirm VPIII Microbial Identification Test) Assessment by women's health nurse practitioner plus in clinic microscopy testing	+/-	None	N/A
	Jones 2013	<i>Trichomonas vaginalis</i>	Vaginal fluid testing using dipstick test (OSOM Trichomonas rapid test)	+/- (two red lines)	Vaginal fluid test using PCR test Assessment at central laboratory	+/-	None	N/A
Common skin condition	Bregnhøj 2011	Eczema	Questionnaire on presence of eczema	Positive response	Hand Eczema Severity Index (HECSI) Interpretation by clinician	+ve for presence of eczema	Same day	Inclusion & 18 month follow-up
	Svensson 2002	Hand eczema	Questionnaire on presence of eczema	Positive response	Hand examination for erythema, papules, vesicles, scaling, fissures, lichenification and hyperkeratotic areas. Assessment by experienced dermatologist	+ve if erythema and papules / vesicles OR erythema and scaling and fissures / lichenification	None	N/A
	Elsner 2015	Allergic reaction to nickel and fragrance	Irritant reaction to surgical tapes on upper arm after 48 hours	+/-	Irritant reaction to surgical tapes on upper arm after 48 hours Assessment by clinician	+/-	None	N/A
	Josefson 2011	Allergic reaction to nickel	Medical plaster patches (Nixema) on upper arm with	+/-	Medical plaster patches (Finn	+/-	Same day	N/A

			readings on days 3-4		chambers® on Scanpor® tape or IQ Ultra® Chambers) on back with readings on days 3-4 and/or day 7 Assessment by dermatologist			
HIV	Assiimwe 2014	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® In-Home Rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Finger pick blood test. Nationally approved algorithm of POC rapid HIV tests (Determine (Abbot Laboratories), STAT-PAK (Chembio Diagnostic Systems Inc) and Unigold (Trinity Biotech plc) as a tiebreaker). Assessment by research assistants.	+/-	12 -72 hours	N/A
	Belete 2019	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® Rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Finger pick blood test. Nationally approved algorithm of POC rapid HIV tests (Wanti (Beijing), Unigold (Trinity Biotech plc), Vikia). Assessment by health professional.	+/-	Same time	N/A
	Choko 2011	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Finger pick blood test. Algorithm of POC rapid HIV tests (Determine (Abbot Laboratories), Unigold (Trinity Biotech plc) and SD Bioline HIV i/II (Standard Diagnostics, Inc.) as a tiebreaker). Assessment by counsellor.	+/-	Same time	N/A
	Choko 2015	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Finger pick blood test. Two parallel POC rapid HIV tests (Determine (Abbot Laboratories), Unigold (Trinity Biotech plc)). Assessment by nurse.	+/-	Approx. 1 week	1 – 12, 13 – 24 months (max. 1 test in 12 months)

	Kapaku 2017	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Venous blood test (EDTA tube) Testing in certified central laboratory (Abbott Architect HIV1 Ag/Ab combo assay, positive results confirmed by BioRad GS HIV combo Ag/Ab assay)	+/-	Within 8 hours	Once in 12 month study period.
	Kurth 2016	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Venous blood test. ELISA testing by single person in certified central laboratory (Vironostika HIV Uni-Form II Ag/Ab (bioMe´rieux Inc.))	+/-	Within 8 hours	N/A
	Martinez Perez 2016	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Finger pick blood test. POC rapid HIV tests (Determine (Abbot Laboratories), confirmatory Unigold (Trinity Biotech plc)). Assessment by HIV counsellor.	+/-	Same time	N/A
	Orasure 2012	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® In-Home Rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Venous blood test. FDA approved serum EIA and Western blot in FDA approved laboratory.	+/-	Unclear	N/A
	Pant Pai 2013	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Venous blood test. ELISA with p24 Antigen tests in reference laboratories all within 24 hours (Architect HIV Ag/Ab combo (Abbott Laboratories), positive results confirmed by Western Blot)	+/-	Same time	N/A

HIV - Human immunodeficiency virus, POC – Point of care, max. – maximum, FDA – Food and Drug Administration, USA, N/A – not applicable; NR – not reported; KOH – potassium hydroxide; PCR – polymerase chain reaction; * whilst no reported level was given for VI-SENSE, it was reported that the polymer used in this product had a range of 4.3 – 5.1 pH



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title page (pg 1)
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction (pg 4)
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction (pg 2)
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods, Types of studies (pg 4) & Statistical analysis and data analysis (pg 5)
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods, Search methods to identify studies (pg 4)
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Supplementary table 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods, Selection of studies (pg 5). No automation tools used.
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods, Data extraction and management (pg 5). No automation tools used.
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Supplementary table 2. Methods, Data extraction and



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
			management (pg 5).
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Supplementary table 2. Methods, Data extraction and management (pg 5).
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods, Assessment of methodological quality (pg 5). No automation tools used.
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods, Statistical analysis and data synthesis (pg 5) Supplementary table 2
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods, Statistical analysis and data synthesis (pg 5)
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A Studies included providing 2x2 published data.
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods, Statistical analysis and data synthesis (pg 5)
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods, meta-analysis (pg 5)
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods, Investigating



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
			heterogeneity (pg 5 & 6)
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Results, Self-diagnosis of HIV (pg 9)
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not examined.
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods, meta-analysis (pg 5)
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Results (pg 6) Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A. Rather 2 studies were included that were borderline exclusion (pg 7 & 8)
Study characteristics	17	Cite each included study and present its characteristics.	Results (pg 6 – 9). Supplementary tables 4 & 5
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Figure 5
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Figures 2 - 4
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results (pg 6-9)
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results, Self-diagnosis of HIV (pg 8 & 9). Figure 4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results, Self-diagnosis of HIV (pg 8 & 9). Figure 4
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Results, Self-



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
			diagnosis of HIV (pg 9). Figure 4
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Discussion (pg 10 & 11)
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion (pg 10 & 11)
	23b	Discuss any limitations of the evidence included in the review.	Discussion (pg 10 & 11)
	23c	Discuss any limitations of the review processes used.	Discussion (pg 10 & 11)
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion (pg 11)
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Abstract (pg 2). Methods (pg 6)
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Supplementary table 3
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Methods, meta-analysis (pg 5). Results, self-diagnosis of HIV (post-hoc analysis, pg 9)
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Funding statement (pg 11)
Competing interests	26	Declare any competing interests of review authors.	Competing interests statement (pg 11)
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data availability statement (pg 11)



PRISMA 2020 Checklist

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71
 For more information, visit: <http://www.prisma-statement.org/>

For peer review only

BMJ Open

Accuracy of self-diagnosis in conditions commonly managed in primary care: Diagnostic accuracy systematic review and meta-analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-065748.R1
Article Type:	Original research
Date Submitted by the Author:	26-Sep-2022
Complete List of Authors:	McLellan, Julie; University of Oxford Nuffield Department of Primary Care Health Sciences, Nuffield Department of Primary Health Care Sciences Heneghan, Carl; University of Oxford Nuffield Department of Primary Care Health Sciences, Primary Health Care Roberts, Nia; University of Oxford Bodleian Libraries, Bodleian Health Care Libraries, Pluddemann, Annette; University of Oxford Nuffield Department of Primary Care Health Sciences, Nuffield Department of Primary Care Health Sciences
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Dermatology, HIV/AIDS, Obstetrics and gynaecology
Keywords:	Adult dermatology < DERMATOLOGY, HIV & AIDS < INFECTIOUS DISEASES, PRIMARY CARE, GYNAECOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **Accuracy of self-diagnosis in conditions commonly managed in primary care:**
4 **Diagnostic accuracy systematic review and meta-analysis**
5
6
7

8 Julie McLellan¹, Carl Heneghan¹, Nia Roberts², Annette Plüddemann¹
9

10 ¹ Nuffield Department of Primary Care Health Sciences, University of Oxford, OX2 6GG
11

12 ² Bodleian Health Care Libraries, University of Oxford, Oxford, UK
13
14
15

16 Corresponding author: Annette Plüddemann, Nuffield Department of Primary Care Health
17 Sciences, University of Oxford, OX2 6GG, annette.plueddemann@phc.ox.ac.uk, +44 (0)1865
18 289353
19

20
21
22 **Keywords:** Diagnostic accuracy systematic review, meta-analysis, self-diagnosis, common
23 conditions in primary care,
24
25
26

27 **Abstract word count:** 299
28

29 **Manuscript word count:**
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Objectives:

To assess the diagnostic accuracy of self-diagnosis compared to a clinical diagnosis for common conditions in primary care

Design:

Systematic review. Meta-analysis

Data sources:

Medline, Embase, Cochrane CENTRAL, Cochrane CDSR, and CINAHL from inception to 25 January 2021

Study selection:

Eligible studies were prospective or retrospective studies comparing the results of self-diagnosis of common conditions in primary care to a relevant clinical diagnosis or laboratory reference standard test performed by a healthcare service provider. Studies that considered self-testing only were excluded.

Data extraction:

Two authors independently extracted data using a predefined data extraction form and assessed risk of bias using QUADAS-2.

Methods & Results:

5,047 records identified 18 studies for inclusion covering the self-diagnosis of three common conditions: vaginal infection (five studies), common skin conditions (four studies) and HIV (nine studies). No studies were found for any other condition. For self-diagnosis of vaginal infection and common skin conditions meta-analysis was not appropriate and data were reported narratively. Nine studies, using point-of-care oral fluid tests, reported on the accuracy of self-diagnosis of HIV and data were pooled using bivariate meta-analysis methods. For these nine studies the pooled sensitivity was 92.8% (95%CI, 86% to 96.5%) and specificity was 99.8% (95%CI, 99.1% to 99.9%). Post hoc, the robustness of the pooled findings was tested in a sensitivity analysis only including four studies using laboratory testing as the reference standard. The pooled sensitivity reduced to 87.7% (95%CI 81.4% to 92.2%) and the specificity remained the same. The quality of all 18 included studies was assessed as mixed and overall study methodology was not always well described.

Conclusions and implications of key findings:

Overall, there was a paucity of evidence. The current evidence does not support routine self-diagnosis for vaginal infections, common skin conditions and HIV in primary care.

PROSPERO registration number: CRD42018110288

Article Summary

Strengths and limitations of this study

- This systematic review summarises and interprets the available evidence on self-diagnosis of conditions managed in primary care
- This search strategy was extensive including publications identified from databases Medline, Embase, Cochrane Central Register of Controlled Trials (CENTRAL), Cochrane Database of Systematic Reviews (CDSR), and CINAHL, up to January 2021
- Standard methodology for systematic review of diagnostic accuracy studies was used, including study quality appraisal using QUADAS-2
- There was a paucity of evidence for many common conditions
- Lack of evidence meant meta-analysis was possible only for one condition

Introduction

The workload in primary care continues to increase (1, 2), in part due to global population increases, but also due to age of populations, change in lifestyle and more complex health problems (3). Increased workload has been recognised in the UK as an important factor in working conditions for primary care physicians (General Practitioners, GPs) and strategies are required to manage the workload (4).

One strategy is the potential for self-diagnosis and self-treatment by patients of some commonly occurring conditions. If feasible, this could lead to more rapid diagnosis and treatment reducing the burden on primary care services. The prospect of self-diagnosis is controversial with concerns if results are misinterpreted or patients fail to confirm their findings to a physician (5). In terms of the evidence, the first important question is to assess the diagnostic accuracy of self-diagnosis in the primary care setting. Subsequently, in order to support self-diagnosis, the efficacy needs to be assessed, to inform which conditions can self-diagnosed safely, in which circumstances, and by whom.

Cooke and colleagues recently reported the 30 most commonly managed conditions in primary care in Australia, which has a health landscape broadly comparable with western Europe (6). This list arises from survey data collected between January 2009 and December 2010, which included 194,100 patient encounters from 1,941 GPs.

This systematic review, therefore, aimed to assess the diagnostic accuracy of self-diagnosis compared to a clinical diagnosis for common conditions in primary care by a healthcare provider.

Methods

Types of studies

We included prospective or retrospective studies comparing the results of self-diagnosis of common conditions in primary care to the results of a relevant clinical diagnosis or laboratory reference standard test. We excluded studies with a case-control design due to their high risk of bias (4).

Population: Adults self-diagnosing common conditions in primary care. Common conditions included were broadly based on those reported by Cooke et al (6) and relevant for self-diagnosis (see online supplementary table 1). Studies in children, based in animals or non-human samples were excluded.

Index test: Self-diagnosis, where we defined “self-diagnosis” as a diagnosis made by the patients in the study, including self-evaluation and interpretation of results of rapid tests. Studies that considered self-testing only and not as part of self-diagnosis were excluded, in addition studies assessing accuracy of self-monitoring of an existing condition were excluded.

Reference standard: Clinical diagnosis or laboratory test performed by a healthcare service provider. We excluded studies comparing self-diagnosis with diagnosis by allied health professionals or pharmacists.

1
2
3 Outcome measures: Reported diagnostic accuracy measures (e.g. sensitivity, specificity,
4 likelihood ratios, predictive values, etc.) and primary data for 2x2 tables. We excluded
5 studies reporting only measures of agreement.
6

7 ***Search methods to identify studies***

8
9 The search strategy was based on a combination of terms for self-testing and self-diagnosis,
10 diagnostic accuracy terms (e.g. sensitivity, specificity, etc.) and terms for common conditions
11 in primary care (6) (see online supplementary table 2 for full search strategy).
12

13 We searched the following electronic databases from inception to 25 January 2021: Medline
14 (OvidSP)[1946-present], EMBASE (OvidSP)[1974-present], Cochrane Central Register of
15 Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR) via
16 Cochrane Library, Wiley)[Issue 1 of 12, January 2021] and CINAHL (EBSCOHost)[1982-
17 present]. No restrictions were imposed on study population numbers or language (studies in
18 languages other than English were translated). Letters, narrative reviews, and other non-
19 primary sources were excluded. The reference lists of included studies, plus the first five
20 “similar articles” identified through PubMed for these studies, and reference lists of relevant
21 systematic reviews were used to identify further relevant publications. References were
22 imported into Endnote X9 (7) where duplicates were removed.
23
24
25

26 ***Data collection and analysis***

27 *Selection of studies*

28
29 Two reviewers independently applied the selection criteria to the titles and abstracts of the
30 study reports identified by the searches. Full text of all studies that met the inclusion criteria
31 were reviewed to agree the final list of included studies. Disagreements between reviewers
32 were resolved by discussion and where agreement could not be reached a third reviewer was
33 consulted.
34
35
36

37 *Data extraction and management*

38
39 Two reviewers (JM, AP) independently extracted information from selected studies into a
40 pre-defined data extraction sheet (see online supplementary table 3) and crosschecked the
41 data. Disagreements were resolved by discussion.
42
43

44 *Assessment of methodological quality*

45 We used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) (8) tool to
46 assess methodological quality of included studies. This considered the risk of bias in four
47 domains (patient selection, index test, reference standard, flow and timing), as well as
48 assessing the applicability (for the first three domains) of the studies to the review research
49 question. Studies were assessed as low, high or unclear risk of bias/concerns regarding
50 applicability for each domain. Two reviewers (JM, AP) independently assessed studies’
51 methodological quality; disagreements were resolved by discussion, or if necessary, by a
52 third reviewer. The results of the QUADAS-2 assessment were presented in a summary table.
53
54
55

56 *Statistical analysis and data synthesis*

57 Data were presented and analysed based on the condition being diagnosed. We compiled
58 summary tables outlining the detailed study information of included studies, including the
59 patient sample, condition, study design, setting, the test under evaluation, the comparator, and
60

1
2
3 conduct of the study. We extracted binary diagnostic accuracy data from all studies and
4 constructed 2×2 tables.
5

6 *Meta-analysis*

7
8 We used Review Manager (9) to produce paired forest plots to explore the between-study
9 variability of sensitivity and specificity across the included studies. For each study estimate
10 of sensitivity and specificity, corresponding 95% confidence intervals were shown to
11 illustrate the uncertainty related to each study estimate. Where different thresholds were
12 applied these were reported. Where appropriate, we used bivariate meta-analysis methods
13 (10) to generate pooled estimates of sensitivity and specificity. Due to the nature of the data a
14 change was made to the protocol and RStudio (11) was used to generate the model
15 parameters to input into Revman (9).
16
17

18 *Investigating heterogeneity*

19
20 For medical conditions for which data from more than one study was available, and where it
21 was possible to investigate between-study heterogeneity in the results, inclusion of study
22 level characteristics as covariates in meta-analysis and subgroup analyses were considered.
23 These approaches were carried out if there was sufficient data available and sub-group
24 specific pooled estimates were thought to be of clinical relevance.
25
26

27 *Investigating reporting bias*

28
29 Funnel plots used to detect publication bias in reviews of RCTs have been shown to be
30 misleading for diagnostic test accuracy reviews (12, 13). Funnel plots as an assessment of
31 reporting bias were therefore not included in this review.
32

33 *Patient involvement*

34
35 Members of the public were part of the Research Programme Committee of the National
36 Institute for Health Research (NIHR) programme grant that funded this study. Updates and
37 details about the study were presented to the committee while the study was ongoing, and the
38 public members provided feedback. This review formed part of the NIHR Evidence Synthesis
39 Working Group (ESWG) and members of the public who were part of the ESGW steering
40 committee commented on the protocol for the study and on updates presented to the steering
41 committee.
42
43

44 The full protocol is provided as online supplementary table 4 and is registered on
45 PROSPERO reference number CRD42018110288.
46

47 **Results**

48
49 Figure 1 shows a summary of the search results and the inclusion and exclusion of studies.
50 After removal of duplicates, 5,047 records were identified through database searches,
51 websites and citation searching. This resulted in full texts of 170 articles being assessed for
52 eligibility, and 20 included articles (14-33) reporting results of 18 individual studies (14-20,
53 22-28, 30-33). These 18 included studies fell into 3 broad groups of commonly managed
54 conditions: Female genital infection, dermatitis, contact/allergic and viral disease, not
55 otherwise specified. No studies of self-diagnosis were found for any other conditions in
56 primary care.
57
58
59
60

1
2
3 Most excluded studies only reported on patients' ability to self-test (or self-monitor an
4 existing condition) with the diagnosis being made by a clinician, and did not report diagnostic
5 accuracy of self-diagnosis.
6

7
8 Of the included studies, five reported on the accuracy of self-diagnosis of vaginal infection
9 (11-15), four for common skin conditions (19, 20, 22, 23) and nine for the self-diagnosis for
10 HIV (24-28, 30-33). Online supplementary tables 5 and 6 summarise the characteristics of
11 included studies and characteristics of self-diagnostic (index) and reference tests,
12 respectively. Paired plots of sensitivity and specificity were generated, grouping the studies
13 by the condition to be diagnosed (Figures 2-4). For studies examining the accuracy of self-
14 diagnosis of vaginal infection and in common skin conditions (14-20, 22, 23), meta-analysis
15 was not appropriate due to the between-study heterogeneity and the overall low number of
16 studies, which would make meta-analysis uninformative (34).
17

18 ***Self-diagnosis of vaginal infections***

19
20
21 Five studies assessed the accuracy of self-diagnosis of Bacterial vaginosis and/or infection
22 with *Trichomonas vaginalis* (14-18), with one study assessing the self-diagnosis of *Candida*
23 vaginitis (17) (Figure 2). For Bacterial vaginosis, the accuracy of a vaginal fluid test using a
24 pH strip (for Bacterial vaginosis) was assessed with laboratory testing (Gram staining) as the
25 reference standard in two studies (14, 15) (online supplementary table 5). For the diagnosis
26 of Bacterial vaginosis and/or *T. vaginalis*, a panty liner test kit (VI-SENSE) for vaginal
27 discharge was assessed against a combination of clinical and laboratory assessment as the
28 reference test in one study (16); and a vaginal fluid self-diagnosis kit for women in the
29 military was assessed with clinical and laboratory assessment as the reference test in the
30 second study (17). One study used a vaginal fluid dipstick test for the presence of *T. vaginalis*
31 (OSOM *Trichomonas* rapid test) compared to a laboratory PCR as a reference test (18). For
32 the self-diagnosis of *Candida* vaginosis a military self-testing kit based on a combination of
33 the measurement of pH, amines and the symptom of vaginal itching for self-diagnosis was
34 compared to a combination of clinical and microbiological laboratory assessment in one
35 study (17).
36
37
38

39 ***Bacterial vaginosis; Bacterial vaginosis and/or Trichomonas vaginalis***

40
41 The sensitivity of a self-taken swab applied to a pH test strip for self-diagnosis of Bacterial
42 vaginosis ranged from 0.45 (95%CI, 0.34-0.56) (14) to 0.60 (95%CI, 0.55-0.66) (17) at a pH
43 cut-off of ≥ 4.7 . Notably, Ryan-Wenger et al (17), the study reporting a sensitivity of 0.60,
44 also included symptoms (vaginal itching) and the presence of amines as part of the
45 assessment, which may explain the higher sensitivity. Donders et al (14) assessed a lower cut-
46 off of pH ≥ 4.5 and showed an increased sensitivity of 0.95 (95%CI, 0.88-0.99); however at
47 the expense of specificity. The study on pregnant women (15) assessing the accuracy of pH
48 test strips did not specify the pH cut-off and reported a sensitivity of 0.36 (95%CI, 0.17-0.59)
49 with a specificity of 0.88 (95%CI, 0.82-0.93). The specificity for the pH test strip tests ranged
50 from 0.5 (95%CI, 0.43-0.56) (17) to 0.81 (95%CI, 0.75-0.85) (14) at the pH cut-off of ≥ 4.7 ,
51 decreasing to 0.41 (95%CI, 0.34-0.47) at the lower cut-off of pH ≥ 4.5 (14). Interestingly the
52 low specificity of 0.5 was reported by the study combining the pH test strip with symptoms
53 and the presence of amines (17).
54
55
56

57 The study assessing the vaginal discharge test using a panty liner test kit with an indicator
58 strip incorporated in the liner (16) reported a sensitivity of 0.91 (95%CI, 0.86-0.94) and
59 specificity of 0.81 (95%CI, 0.76-0.86) for the diagnosis of Bacterial vaginosis and/or *T.*
60 *vaginalis* infection.

Trichomonas vaginalis

One study in Brazil assessed a rapid immunochromatographic *T. vaginalis* test for use at home (18) and reported a sensitivity of 0.68 (95%CI, 0.43-0.87) and specificity of 1.00 (95%CI, 0.99-1.00) for self-diagnosis of *T. vaginalis* infection.

Candida vaginitis

Only one study (17) specifically assessed the diagnostic accuracy of self-diagnosis of *Candida* vaginitis, which formed part of the military self-testing kit, and reported a sensitivity of 0.18 (95%CI, 0.12-0.25) and specificity of 0.89 (95%CI, 0.85-0.92).

Self-diagnosis of common skin conditions

Four studies assessed the accuracy of self-diagnosis of common skin conditions (19, 20, 22, 23) (Figure 3). We included 2 studies that were outside our age inclusion criteria: Berghoi 2011 reported patients included had a mean age 17.5 years old, nevertheless these patients would have been 16+ years to qualify as apprentice hairdressers. And Svensson 2002 reported the mean age of patients as 40.4 years (no standard deviation), but included patients from age 16.

Eczema

Two studies assessed the accuracy of self-diagnosis based on a self-evaluated questionnaire of signs and symptoms for the diagnosis of eczema alongside a self-assessment of the presence or absence of eczema based on the questionnaire results, compared to assessment by a clinician (19, 20). Overall, 710 participants were included across the two studies. The reported sensitivity ranged from 0.7 (95%CI, 0.53- 0.84) to 0.87 (95%CI, 0.78-0.93) and specificity ranged from 0.79 (95%CI, 0.70-0.86) to 1.00 (95%CI, 0.99-1.00). The relatively high specificity suggests the potential for patients to use the questionnaire as a tool to confirm that they have eczema and therefore seek healthcare advice; however, it should be noted that there was an unclear risk of bias regarding the patient selection. In Svensson et al (20), the patients were recruited at a dermatology outpatient clinic, where they had been referred to and 113 patients in the study reported having had a diagnosis of eczema in the last 12 months, suggesting a more selected population with a higher pre-test probability. Whilst the setting in the study by Bregnhøj et al. (19) was not reported, it was conducted amongst hairdressers who may have had more experience of eczema either themselves or colleagues being diagnosed with the condition, given the nature of their profession. They may be more aware of the signs and symptoms and may constitute a selected population. Therefore, the diagnostic accuracy of the questionnaire for self-diagnosis may be dependent on the type of population.

Skin allergy

Two studies assessed the diagnostic accuracy of self-diagnosis to detect an allergic skin reaction including 408 participants across the two studies; one study assessed nickel and/or fragrance allergy (22), and the other nickel allergy alone (23). Both studies used a patch test applied to the arm, which was self-evaluated by participants 2-4 days later. Dermatologists then also evaluated the patch tests as the reference standard. One study recruited participants at hospital dermatology departments (23), while the other recruited through a newspaper advertisement targeted at people with a self-suspected allergy towards fragrance and/or nickel

1
2
3 (22). Sensitivity ranged from 0.72 (95%CI, 0.57-0.84) to 0.97 (95%CI, 0.87-1.00) and
4 specificity ranged from 0.87 (95%CI, 0.80-0.92) to 0.91 (95%CI, 0.85-0.95). Elsner et al.
5 (22) also reported that participants found the information regarding how to apply the test
6 extensive and detailed, the information regarding self-evaluation of the test was limited and
7 should be improved.
8
9

10 11 ***Self-diagnosis of HIV***

12
13
14 Nine studies were identified that reported the diagnostic accuracy of self-testing and self-
15 diagnosis of HIV (24-28, 30-33). In all studies self-diagnosis was undertaken unsupervised
16 using a rapid point of care (POC) oral fluid test manufactured by OraSure technologies, either
17 OraQuick Inhome intended for lay users or OraQuick Advance intended for professional use.
18 The studies recruited 13103 participants, and all were conducted in African countries except
19 for the phase III trial in the USA by OraSure Technologies (33). The 2019 global HIV
20 prevalence rates for women and men aged 15 to 49 were 0.8% (95%CI, 0.7–1.0) and 0.6%
21 (95%CI, 0.5-0.8) respectively with the overall highest prevalence by country in Eswatini
22 (Africa) at 27.1% (95%CI, 25.4-28.8] (35). All included studies had prevalence rates above
23 the global averages for men and women or, if not reported, were in countries with high
24 prevalence rates. Prevalence rates ranged from 2.12% in the USA study (33) to 22.1% in the
25 Ugandan study (24). The USA study was conducted in 20 clinical sites, 17 identified as high
26 prevalence sites (2.6%) and 3 as low prevalence sites (0.1%). All studies enrolled participants
27 from the general population including the USA where no breakdown of sexual orientation
28 was reported. The reported sensitivity and specificity were similar between studies (Figure
29 4a); the single study (32) reporting a lower estimate for sensitivity still had a confidence
30 interval that overlapped with half of the other studies. The pooled sensitivity based on all 9
31 included studies was 92.8% (95%CI, 86% to 96.5%) and the pooled specificity was 99.8%
32 (95%CI, 99.1% to 99.9%). The studies showed low heterogeneity (Figure 4b). The reference
33 standard used in the studies was one of three types: four studies took a venous sample which
34 was sent to a laboratory for testing (28, 30, 32, 33), two studies used a nationally approved
35 algorithm based on a combination of rapid POC tests (24, 25) and three studies used a study
36 based algorithm again based on rapid POC tests (26, 27, 31). In three studies using POC tests
37 for the reference standard, the diagnosis may have been by clinicians, but it was unclear.
38 These studies reported the diagnosis by a research assistant (24) or a counsellor (26, 31). Post
39 hoc, a sensitivity analysis was conducted to test the robustness of the pooled findings by
40 removing the studies using POC tests as the reference standard (including tests where it was
41 unclear whether diagnosis was by a clinician) and only including those studies using
42 laboratory testing. Based on four studies using laboratory testing as the reference standard
43 (28, 30, 32, 33), the pooled sensitivity was 87.7% (95%CI 81.4% to 92.2%) and the pooled
44 specificity was 99.8% (95%CI 98.9% to 99.9%). No data were reported by participant
45 characteristics such as gender, ethnicity, or sexual orientation.
46
47
48
49
50

51 A number of studies (24-26, 30-32) reported on the viability or feasibility of the oral fluid
52 self-test stating that participants found it easy to conduct, but acknowledged that instructions
53 should be adapted to the population using the test, particularly the literacy levels.
54 Furthermore, users must be encouraged to receive a confirmatory test.
55
56
57
58

59 ***Methodological quality of included studies***

60

1
2
3 Assessment of the quality of included studies using the QUADAS-2 framework (8) is
4 presented in Figure 5, which summarises the overall risk of bias and applicability concerns.
5 For patient selection, the risk of bias overall was mixed; where studies were rated unclear in
6 this domain it was because several studies either recruited a selected population with a
7 potentially higher pre-test probability or they did not clearly report the recruitment strategy
8 and whether eligible patients were consecutively recruited. However, with the exception of
9 one study, applicability concerns were low. In several cases, although the population may
10 have been a selected population (e.g. at high risk of HIV infection or skin eczema; or with a
11 prior history of eczema or vaginal infection), it could be argued that these might be the
12 populations where self-diagnosis and/or self-testing may be most relevant. Overall risk of
13 bias regarding the conduct of the index test was low and in most studies participants were
14 blinded to the results of the reference test, though in some studies this was not clearly
15 described. Risk of bias with respect to the reference test was unclear or high in some studies
16 as assessors were either not blinded to the results of the index test or blinding was unclear.
17 For the domain of flow and timing, several studies were judged to be at high risk of bias as
18 the interval between testing was frequently not explicitly reported. In addition differential
19 reference bias was identified as present in some studies and unclear in others, in particular it
20 was unclear in several studies how clinical assessment was conducted or standardised.
21 Overall study procedures were not always clearly described.
22
23
24
25
26
27

28 Discussion

29
30 With the increasing workload in primary care (1) and the continued development of rapid
31 tests, including those that are intended to be used by patients, we aimed to assess the evidence
32 for the diagnostic accuracy of self-diagnosis. We identified limited evidence on the
33 diagnostic accuracy of self-diagnosis: only 20 publications (reporting data from 18 studies)
34 specifically assessed the accuracy of self-diagnosis, covering three commonly managed
35 conditions, namely vaginal infections, common skin conditions and HIV. Interestingly, no
36 studies of self-diagnosis were found for any other conditions in primary care. It was
37 particularly notable that we did not find any studies assessing the diagnostic accuracy of self-
38 diagnosis of common primary conditions such as upper respiratory tract and urinary tract
39 infections. As technology develops potentially enabling increased self-diagnosis in primary
40 care, we would expect future reviews examining this research question to include more
41 common conditions. In particular, we would expect to include studies for self-diagnosis of
42 COVID-19 following the rapid development of tests during the COVID-19 pandemic.
43
44
45

46 The evidence for self-diagnosis of vaginal infection suggests a lack of sufficient accuracy to
47 aid self-diagnosis. Tests relying on self-swabs and pH strips showed low sensitivity (below
48 60%), and therefore would not be useful to rule out disease. Although sensitivity improved to
49 95% with increasing pH cut-off, this occurred at the cost of specificity, which dropped from
50 81% to 41%. Using the test at this cut-off would, therefore, result in concern for missed
51 diagnoses. The immunochromatography test also showed insufficient sensitivity (68%) to be
52 of use as a rule-out test. The panty liner test, however, had high sensitivity of 91% and may
53 prove to be a useful rule-out test, although this result is limited as it is based on one study
54 judged to be at unclear risk of bias, with a high risk of bias in the flow and timing domain. In
55 terms of their use to confirm the presence of vaginal infection, the highest specificity (100%)
56 was reported for the immunochromatographic test. This may be a useful test to aid self-
57 diagnosis of vaginal infection in systems that currently rely on syndromic management, such
58 as low resource settings, particularly to improve the targeting of antimicrobial prescribing.
59
60

1
2
3 However, it should be noted this result is also based on one study and requires confirming in
4 a larger study.
5

6 The allergy patch tests overall showed reasonable accuracy (sensitivity 72-97%; specificity
7 87-91%) and may be useful as an initial self-screening test for patients with a suspected
8 nickel or fragrance allergy; however it could be argued that the main use of these tests might
9 be to safely rule out a contact allergy. The reported diagnostic accuracy suggests the tests are
10 not sufficiently sensitive as a rule-out test, particularly given the relatively wide confidence
11 intervals. The tests may however be useful in settings where access to dermatology services
12 is scarce. The self-diagnosis of eczema using a questionnaire of signs and symptoms showed
13 specificity ranging from 0.79 to 1.00, suggesting this test might be useful as a confirmation
14 test for patients for suspect they have eczema and to then seek treatment and management
15 advice. The selected nature of the patients in the studies (i.e. patients with a previous eczema
16 diagnosis and hairdressers, who may encounter eczema more frequently due to their
17 profession) may overestimate the accuracy of the test; however, it could be argued that these
18 might be the populations in whom the test is most relevant.
19
20
21

22 We identified nine studies that reported the accuracy of self-diagnosis of HIV (sensitivity
23 93%, specificity 99%). However, the sensitivity is reduced to 88% when only studies using a
24 venous sample and laboratory testing as the reference standard are included. With a
25 sensitivity of 88%, the accuracy data would not support the use of this test as a rule-out test,
26 particularly given the clinical consequences of a false negative test result. Evidence suggests
27 there may be benefits to self-initiated HIV testing, including early identification, increased
28 likelihood for the uptake of HIV prevention interventions, and a reduction in sexual risk
29 behaviours (36), warranting further research, in particular in resource-limited settings where
30 access to testing sites may be a barrier. However, HIV self-testing should also be considered
31 in the context of linkage to care, access to counselling, and adequate regulatory and quality
32 assurance systems (37).
33
34
35

36 The search strategy for this review was broad and extensive with few restrictions resulting in
37 the high number of publications to screen. Whilst it is possible, it is unlikely, studies were
38 missed. We are unaware of any other reviews examining this research question. The main
39 limitation was the lack of available evidence for a number of common conditions; studies
40 reporting on self-testing alone were more common, but few studies assessed self-diagnosis,
41 with patients interpreting the test results and making a diagnosis independently. For the three
42 conditions where we identified studies reporting on the diagnostic accuracy of self-diagnosis,
43 there was a paucity of evidence. Many studies were not replicated and included small sample
44 sizes and contained methodological biases that limited the application of the results to
45 practice. For self-diagnosis of vaginal infections, common skin conditions, and HIV further
46 research is required to draw a definitive conclusion on the efficacy of self-diagnosis. For
47 other common conditions in primary care, research is needed on self-diagnosis where this
48 option is available, and studies should go beyond considering self-testing alone and also
49 assess the diagnostic accuracy of self-diagnosis. Terminology for self-diagnosis, self-testing
50 and self-screening is overlapping in some cases and needs clarifying. Finally, research is
51 required into the patient's readiness and attitude towards self-diagnosis along with its effect
52 on the patient/physician relationship.
53
54
55

56 The current limited evidence does not support routine self-diagnosis for vaginal infections,
57 common skin conditions and HIV in primary care.
58
59
60

Acknowledgements

The authors would like to thank Hayley Jones for joint authorship of the original protocol, Elizabeth Spencer for assistance with preparation of the protocol, and Emily McFadden for assistance with screening of studies for inclusion.

Contributors

AP and CH conceived the idea for this review. AP led the research, developed the protocol, and conducted the review (screening and extraction). CH developed the protocol and provided clinical input. NR devised and conducted the search strategy, and approved the manuscript. JM conducted the review (screening, extraction, data analysis). JM, AP and CH jointly prepared the manuscript.

Funding

This project was funded by National Institute for Health Research School for Primary Care Research (NIHR SPCR) [Project Number 390]

Data availability statement

All data are available in published articles.

Competing interests

Julie McLellan reports grants from NIHR School of Primary Care Research (Evidence Synthesis Working Group Project Number 390) and part-funding from the Thames Valley Applied Research Collaborative (ARC), during the conduct of the study; and occasionally receives expenses for teaching Evidence-Based Medicine

Dr. Plüddemann reports grants from NIHR School of Primary Care Research (Evidence Synthesis Working Group Project Number 390), during the conduct of the study; and occasionally receives expenses for teaching Evidence-Based Medicine.

Dr. Heneghan reports receiving expenses and fees for his media work. He has received expenses from the WHO and holds grant funding from the NIHR, the NIHR School of Primary Care Research, The Wellcome Trust and the WHO. He has received financial remuneration from an asbestos case. He has also received income from the publication of a series of toolkit books published by Blackwells. On occasion, he receives expenses for teaching EBM and is also paid for his GP work in NHS out of hours. CEBM jointly runs the EvidenceLive Conference with the BMJ and the Overdiagnosis Conference with some international partners which are based on a non-profit making model.

References

1. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. *Lancet*. 2016;387(10035):2323-30.
2. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*. 2017;7(10):e017902.

3. Orgnaistion WH. Global status report on noncommunicable diseases 2014. 2014.
4. British Medical Association. BMA guidance: Controlling workload in general practice strategy 2020 [Available from: <https://www.bma.org.uk/advice-and-support/gp-practices/managing-workload/controlling-workload-in-general-practice-strategy>] Accessed 18 June 2021.
5. Hynes V. The trend toward self-diagnosis. *CMAJ*. 2013;185(3):E149-E50.
6. Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. *Australian Family Physician*. 2013;42:65-8.
7. The EndNote Team. Endnote. Endnote X9 ed. Philadelphia, PA: Clarivate; 2013.
8. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.
9. Review Manager (RevMan) [Computer program]. Version 5.4. The Cochrane Collaboration. 2020.
10. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.
11. RStudio Team. 2021. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA [Available from: <http://www.rstudio.com/>] Accessed 22 June 2021.
12. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-93.
13. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008;149(12):889-97.
14. Donders GGG, Gonzaga A, Marconi C, Donders F, Michiels T, Eggermont N, et al. Increased vaginal pH in Ugandan women: what does it indicate? *European Journal of Clinical Microbiology & Infectious Diseases*. 2016;35(8):1297-303.
15. Sungkar A, Purwosunu Y, Aziz MF, Pratomo H, Sutrisna B, Sekizawa A. Influence of early self-diagnosis and treatment of bacterial vaginosis on preterm birth rate. *International Journal of Gynaecology & Obstetrics*. 2012;117(3):264-7.
16. Geva A, Bornstein J, Dan M, Shoham HK, Sobel JD. The VI-SENSE-vaginal discharge self-test to facilitate management of vaginal symptoms. *American Journal of Obstetrics & Gynecology*. 2006;195(5):1351-6.
17. Ryan-Wenger NA, Neal JL, Jones AS, Lowe NK. Accuracy of vaginal symptom self-diagnosis algorithms for deployed military women. *Nursing Research*. 2010;59(1):2-10.
18. Jones HE, Lippman SA, Caiaffa-Filho HH, Young T, van de Wijgert JHHM. Performance of a rapid self-test for detection of *Trichomonas vaginalis* in South Africa and Brazil. *Journal of Clinical Microbiology*. 2013;51(3):1037-9.
19. Bregnhøj A, Sosted H, Menne T, Johansen JD. Validation of self-reporting of hand eczema among Danish hairdressing apprentices. *Contact Dermatitis*. 2011;65(3):146-50.
20. Svensson A, Lindberg M, Meding B, Sundberg K, Stenberg B. Self-reported hand eczema: symptom-based reports do not increase the validity of diagnosis. *British Journal of Dermatology*. 2002;147(2):281-4.
21. Elsner P, Bjarnoe H, Aberer W, Brasch J. Evaluation of a self-test device in allergic contact dermatitis. *Dermatitis*. 2009;20 (4):230-1.
22. Elsner P, Bjarnøe H, Aberer W, Brasch J. Evaluation of a Self-Test Device Used in Allergic Contact Dermatitis. *Skin Pharmacol Physiol*. 2015;28(5):257-63.
23. Josefson A, Svensson A, Farm G, Engfeldt M, Meding B. Validation of self-testing as a method to estimate the prevalence of nickel allergy. *Acta Dermato-Venereologica*. 2011;91(5):526-30.

- 1
- 2
- 3
- 4 24. Asiimwe S, Oloya J, Song X, Whalen CC. Accuracy of un-supervised versus
- 5 provider-supervised self-administered HIV testing in Uganda: a randomized implementation
- 6 trial. *AIDS and behavior*. 2014;18(12):2477-84.
- 7 25. Belete W, Deressa T, Feleke A, Menna T, Moshago T, Abdella S, et al. Evaluation of
- 8 diagnostic performance of non-invasive HIV self-testing kit using oral fluid in Addis Ababa,
- 9 Ethiopia: A facility-based cross-sectional study. *PLoS One*. 2019;14(1):e0210866.
- 10 26. Choko AT, Desmond N, Webb EL, Chavula K, Napierala-Mavedzenge S, Gaydos
- 11 CA, et al. The uptake and accuracy of oral kits for HIV self-testing in high HIV prevalence
- 12 setting: a cross-sectional feasibility study in Blantyre, Malawi. *Plos medicine*.
- 13 2011;8(10):e1001102.
- 14 27. Choko AT, MacPherson P, Webb EL, Willey BA, Feasy H, Sambakunsi R, et al.
- 15 Uptake, Accuracy, Safety, and Linkage into Care over Two Years of Promoting Annual Self-
- 16 Testing for HIV in Blantyre, Malawi: A Community-Based Prospective Study. *PLoS Med*.
- 17 2015;12(9):e1001873.
- 18 28. Kapaku K, Newman M, Maluzi K, Sigande L, Taegtmeier M, Corbett E, et al.,
- 19 editors. Is OraQuick HIV-self-testing valid among intended users? Analysis from a clinical
- 20 performance study in
- 21 Lusaka, Zambia. 22nd International AIDS Conference; 2017 24-26 July; Paris, France.
- 22 29. UNITAID/PSI HIV Self-Testing Africa (HIV STAR) study. A clinical performance
- 23 study of self-testing using self-collected oral fluid transudate and the OraQuick® HIV Self-
- 24 Test [Available from: [https://hivstar.lshtm.ac.uk/files/2016/11/STAR-Protocol-for-Zambia-](https://hivstar.lshtm.ac.uk/files/2016/11/STAR-Protocol-for-Zambia-Clinical-Performance-Study-web.pdf)
- 25 [Clinical-Performance-Study-web.pdf](https://hivstar.lshtm.ac.uk/files/2016/11/STAR-Protocol-for-Zambia-Clinical-Performance-Study-web.pdf)] Accessed 18 June 2021.
- 26 30. Kurth AE, Cleland CM, Chhun N, Sidle JE, Were E, Naanyu V, et al. Accuracy and
- 27 Acceptability of Oral Fluid HIV Self-Testing in a General Adult Population in Kenya. *AIDS*
- 28 *Behav*. 2016;20(4):870-9.
- 29 31. Martínez Pérez G, Steele SJ, Govender I, Arellano G, Mkwamba A, Hadebe M, et al.
- 30 Supervised oral HIV self-testing is accurate in rural KwaZulu-Natal, South Africa. *Trop Med*
- 31 *Int Health*. 2016;21(6):759-67.
- 32 32. Pant Pai N, Behlim T, Abrahams L, Vadnais C, Shivkumar S, Pillay S, et al. Will an
- 33 unsupervised self-testing strategy for HIV work in health care workers of South Africa? A
- 34 cross sectional pilot feasibility study. *PLoS One*. 2013;8(11):e79772.
- 35 33. OraSure Technologies. Final Advisory Committee Briefing Materials: Available for
- 36 Public Release. OraQuick In-Home HIV Test. Washington (D.C.): Food and Drug
- 37 Administration: Blood Products Advisory Committee; 2012.
- 38 34. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. *Cochrane Handbook*
- 39 *for Systematic Reviews of Diagnostic Test Accuracy*.: The Cochrane Collaboration;
- 40 <http://srdta.cochrane.org>; 2010.
- 41 35. UNAIDS. *AIDSinfo Global factsheets 2019. HIV and AIDS Estimates* [Available
- 42 from <http://aidsinfo.unaids.org/>] Accessed 22 June 2021.
- 43 36. Adebayo OW, Salerno JP. Facilitators, Barriers, and Outcomes of Self-Initiated HIV
- 44 Testing: An Integrative Literature Review. *Res Theory Nurs Pract*. 2019;33(3):275-91.
- 45 37. Njau B, Covin C, Lisasi E, Damian D, Mushi D, Boulle A, et al. A systematic review
- 46 of qualitative evidence on factors enabling and deterring uptake of HIV self-testing in Africa.
- 47 *BMC Public Health*. 2019;19(1):1289.
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

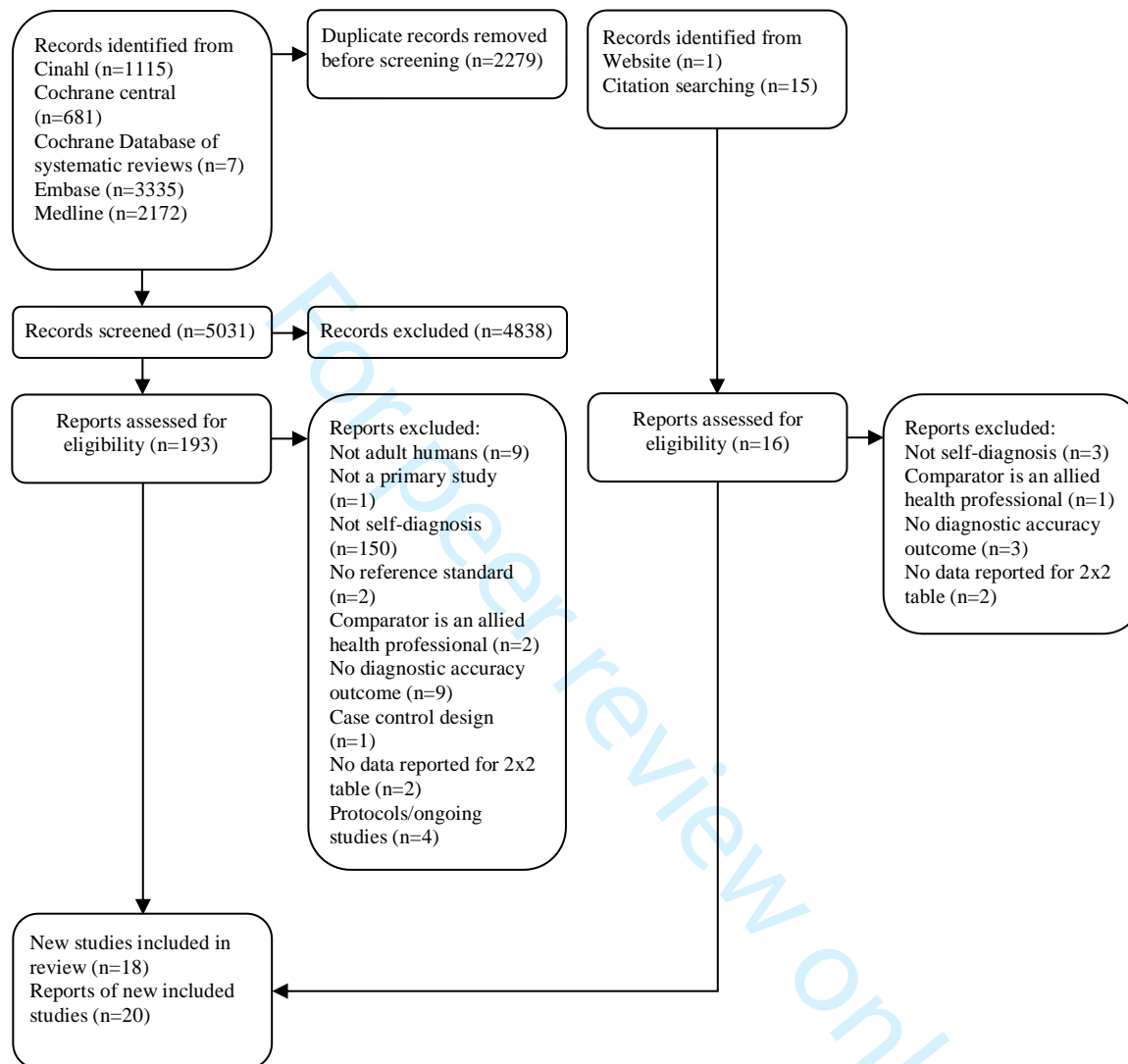
Figure 1: Study Selection

Figure 2: Paired forest plots of sensitivity and specificity for studies of self-diagnosis of vaginal infection (where Donders 2016a used pH threshold ≥ 4.5 , Donders 2016b used pH threshold ≥ 4.7 . Sungkar 2012 represents pooled data from 4 time points)

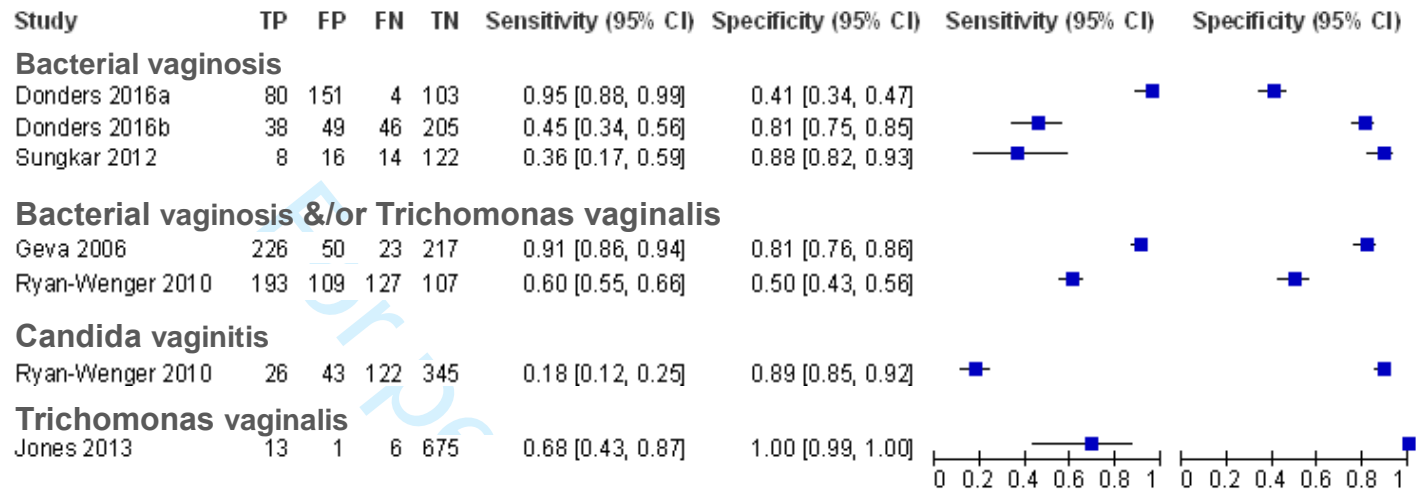


Figure 3: Paired forest plots of sensitivity and specificity for studies of self-diagnosis of common skin conditions (where Bregnhøj 2011 represents pooled data from recruitment and follow up time points)

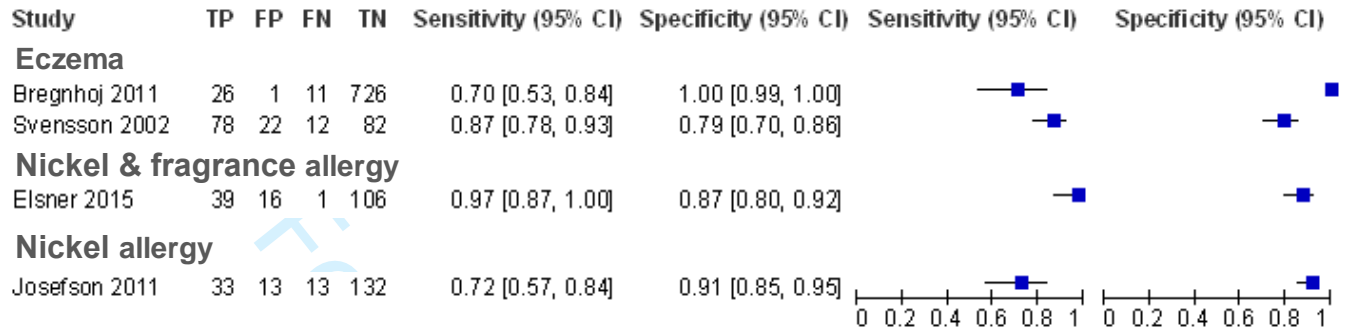
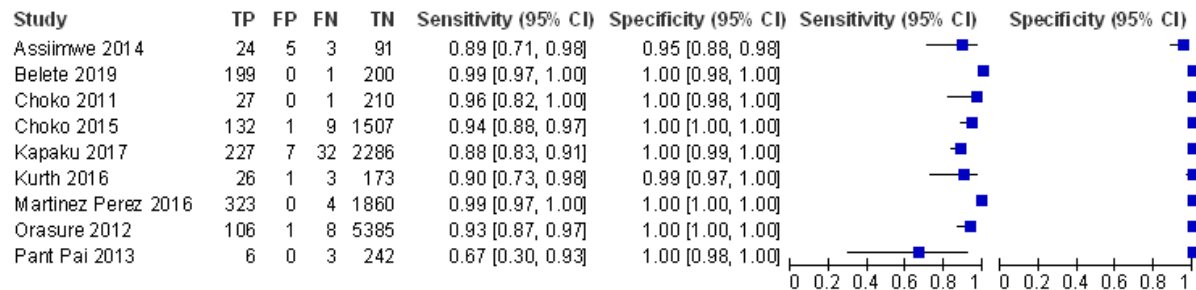
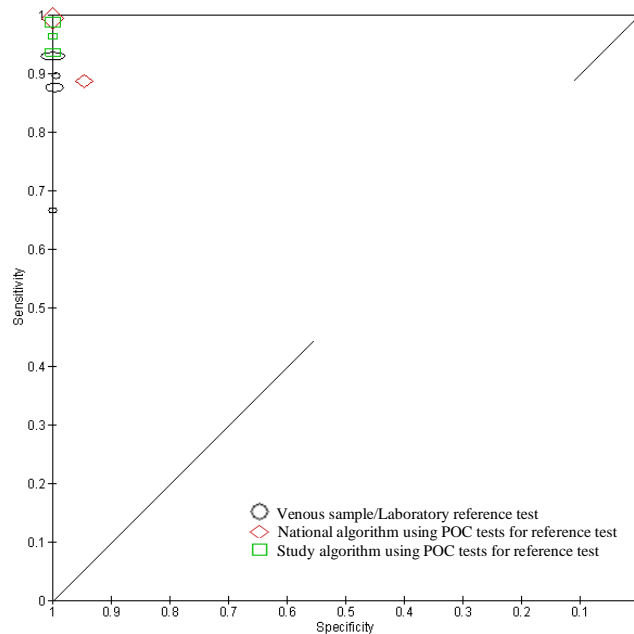


Figure 4: Studies of self-diagnosis of HIV

(a) Paired forest plots of sensitivity and specificity (where Choko 2015 represents pooled data from 1- 12 and 13 – 24 months follow up time points)



(b) Receiver operating characteristic plot HIV self-diagnosis compared with clinical diagnosis or laboratory reference test grouped by reference test type (where size of symbol indicates study size)



All studies (n=9):

Sensitivity: 92.8% (95% CI 86% to 96.5%); Specificity: 99.8% (95% CI 99.1% to 99.9%)

Sensitivity analysis retaining only studies with venous/laboratory reference test (n=4):

Sensitivity: 87.7% (95% CI 81.4% to 92.2%); Specificity: 99.8% (95% CI 98.9% to 99.9%)

Figure 5: QUADAS-2 summary of risk of bias and applicability concerns showing review authors' judgments about each domain for each included study (based on 19 studies (20 data sets))

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Assiimwe 2014	+	+	?	+	+	+	+
Belete 2019	+	?	?	+	+	+	+
Bregnhøj 2011	+	?	?	+	?	+	+
Choko 2011	+	+	-	+	+	+	+
Choko 2015	+	+	-	?	+	+	+
Donders 2016a	?	+	+	-	+	+	+
Donders 2016b	?	+	+	-	+	+	+
Elsner 2015	?	?	-	+	+	+	+
Geva 2006	?	?	+	-	+	+	+
Jones 2013	-	+	?	+	+	+	+
Josefson 2011	+	+	+	-	+	+	+
Kapaku 2017	+	+	?	+	+	+	+
Kurth 2016	?	+	+	-	+	+	+
Martinez Perez 2016	-	?	-	+	+	+	+
Orasure 2012	-	+	?	?	+	+	+
Pant Pai 2013	-	+	?	+	+	+	+
Ryan-Wenger 2010	?	+	-	-	+	+	+
Sungkar 2012	+	+	-	-	+	+	+
Svensson 2002	?	+	+	-	+	+	+

- High
 ? Unclear
 + Low

only

SUPPLEMENTARY MATERIAL

Supplementary Table 1: Commonly managed conditions in primary care

As reported by Cooke et al 2013 (Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. Australian Family Physician. 2013; 42:65-8)

1	Hypertension
2	Immunisation/vaccination: all
3	Acute upper respiratory tract infection
4	Depression
5	Diabetes: nongestational
6	Lipid disorders
7	General check-up
8	Osteoarthritis
9	Back complaint
10	Prescription
11	Oesophagus disease
12	Female genital check-up
13	Acute bronchitis/bronchiolitis
14	Asthma
15	Anxiety
16	Test results
17	Urinary tract infection
18	Dermatitis, contact/allergic
19	Pregnancy
20	Sleep disturbance
21	Sinusitis acute/chronic
22	Gastroenteritis
23	Vitamin/nutritional deficiency

24	Malignant neoplasm of skin
25	Abnormal test results
26	Atrial fibrillation/flutter
27	Oral contraception
28	Solar keratosis/sunburn
29	Ischaemic heart disease
30	Viral disease, not otherwise specified

Supplementary Table 2: Search strategy

Medline	
#	Searches
1	Diagnostic Self Evaluation/
2	((self* adj diagnos*) or selfdiagnos*).ti,ab.
3	(self* and diagnos*).ti.
4	((self* adj test*) or selftest*).ti,ab.
5	(home adj3 diagnos*).ti,ab.
6	((selfreport* or self-report*) and diagnos*).ti.
7	((selfreport* or self-report*) adj5 diagnos*).ti,ab.
8	(diagnos* and (selftreat* or self-treat*)).ti,ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	respiratory tract infections/ or common cold/ or influenza, human/ or laryngitis/ or exp pharyngitis/ or rhinitis/ or exp sinusitis/ or exp supraglottitis/ or tracheitis/ or exp otitis media/
11	((((respirat* or airway*) adj2 infection*) or (common cold or influenza or flu or pharyngitis or laryngitis or tonsillitis or sore throat or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or otitis media)).ti,ab.
12	exp Depressive Disorder/ or Depression/
13	(depress* or (mood adj2 disorder*)).ti,ab.
14	diabetes mellitus/ or exp diabetes mellitus, type 1/ or exp diabetes mellitus, type 2/
15	diabet*.ti,ab.
16	dyslipidemias/ or exp hyperlipidemias/
17	(dyslipid?emia? or hyperlipid?emia? or hypercholesterol?emia? or cholesterol* or triglyceride*).ti,ab.
18	hypertension/
19	(hypertens* or high blood pressure).ti,ab.
20	exp Osteoarthritis/
21	(osteoarthritis or osteo-arthritis or (degenerat* adj2 arthritis)).ti,ab.
22	back pain/ or low back pain/
23	Sciatica/
24	Intervertebral Disc Displacement/
25	((back adj2 (pain or problem? or disorder?)) or slipped disc* or sciatica).ti,ab.
26	Esophageal Diseases/ or esophageal motility disorders/ or exp gastroesophageal reflux/ or Barrett Esophagus/
27	((oesophag* or esophag*) adj2 disease).ti,ab.
28	((((oesophag* or esophag* or gastr*) adj2 reflux) or heartburn or heart burn).ti,ab.
29	((barrett* or globus) adj2 (esophag* or oesophag*)).ti,ab.
30	exp Asthma/
31	asthma*.ti,ab.

32	exp Anxiety Disorders/ or Anxiety/
33	(anxiety or phobia*).ti,ab.
34	exp Urinary Tract Infections/
35	exp Cystitis/
36	((urin* adj3 infection*) or bacteriuria or pyuria or cystitis).ti,ab.
37	exp Vaginitis/
38	(vaginitis or vaginosis).ti,ab.
39	dermatitis/ or exp dermatitis, contact/ or eczema/
40	(dermatitis or eczema).ti,ab.
41	exp Sleep Wake Disorders/
42	exp Sleep Apnea Syndromes/
43	((sleep adj2 (disorder* or disturbance*)) or insomnia*).ti,ab.
44	(sleep adj2 (apnea or apnoea)).ti,ab.
45	restless leg*.ti,ab.
46	Gastroenteritis/ or diarrhea/ or vomiting/
47	(gastroenteritis or (stomach adj2 (bug? or upset))).ti,ab.
48	(diarrhoea or diarrhea or food poisoning).ti,ab.
49	Sunburn/
50	(sunburn or solar keratosis).ti,ab.
51	or/10-50
52	9 and 51
53	exp "Sensitivity and Specificity"/
54	exp "REPRODUCIBILITY OF RESULTS"/
55	(sensitiv* or specific* or predict* or accura* or valid* or reproduc*).ti,ab.
56	53 or 54 or 55
57	52 and 56
Embase	
#	Searches
1	*Self Evaluation/
2	((self* adj diagnos*) or selfdiagnos*).ti,ab.
3	(self* and diagnos*).ti.
4	((self* adj test*) or selftest*).ti,ab.
5	(home adj3 diagnos*).ti,ab.
6	((selfreport* or self-report*) and diagnos*).ti.
7	((selfreport* or self-report*) adj5 diagnos*).ti,ab.
8	(diagnos* and (selftreat* or self-treat*).ti,ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	respiratory tract infection/ or upper respiratory tract infection/ or exp respiratory tract inflammation/ or influenza/ or otitis media/
11	((respirat* or airway*) adj2 infection*) or (common cold or influenza or flu or pharyngitis or laryngitis or tonsillitis or sore throat or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or otitis media).ti,ab.

12	exp Depression/
13	(depress* or (mood adj2 disorder*)).ti,ab.
14	diabetes mellitus/ or exp diabetes mellitus, type 1/ or exp diabetes mellitus, type 2/
15	diabet*.ti,ab.
16	dyslipidemia/ or exp hyperlipidemia/
17	(dyslipid?emia? or hyperlipid?emia? or hypercholesterol?emia? or cholesterol* or triglyceride*).ti,ab.
18	hypertension/
19	(hypertens* or high blood pressure).ti,ab.
20	exp Osteoarthritis/
21	(osteoarthritis or osteo-arthritis or (degenerat* adj2 arthritis)).ti,ab.
22	backache/ or discogenic pain/ or low back pain/
23	Sciatica/
24	intervertebral disk hernia/ or intervertebral disk disease/
25	((back adj2 (pain or problem? or disorder?)) or slipped disc* or sciatica).ti,ab.
26	exp gastroesophageal reflux/ or barrett esophagus/ or esophagus disease/
27	((oesophag* or esophag*) adj2 disease).ti,ab.
28	((oesophag* or esophag* or gastr*) adj2 reflux) or heartburn or heart burn).ti,ab.
29	((barrett* or globus) adj2 (esophag* or oesophag*)).ti,ab.
30	exp Asthma/
31	asthma*.ti,ab.
32	exp Anxiety Disorder/
33	(anxiety or phobia*).ti,ab.
34	exp Urinary Tract Infection/
35	Cystitis/
36	((urin* adj3 infection*) or bacteriuria or pyuria or cystitis).ti,ab.
37	vagina discharge/ or exp vaginitis/
38	(vaginitis or vaginosis).ti,ab.
39	dermatitis/ or contact dermatitis/ or exp eczema/
40	(dermatitis or eczema).ti,ab.
41	sleep disorder/ or exp insomnia/ or periodic limb movement disorder/
42	exp sleep disordered breathing/
43	((sleep adj2 (disorder* or disturbance*)) or insomnia*).ti,ab.
44	(sleep adj2 (apnea or apnoea)).ti,ab.
45	restless leg*.ti,ab.
46	Gastroenteritis/ or diarrhea/ or vomiting/
47	(gastroenteritis or (stomach adj2 (bug? or upset))).ti,ab.
48	(diarrhoea or diarrhea or food poisoning).ti,ab.
49	Sunburn/
50	(sunburn or solar keratosis).ti,ab.
51	or/10-50

52	9 and 51
53	"Sensitivity and Specificity"/
54	diagnostic accuracy/ or diagnostic test accuracy study/
55	predictive validity/ or predictive value/ or reproducibility/
56	(sensitiv* or specific* or predict* or accura* or valid* or reproduc*).ti,ab.
57	53 or 54 or 55 or 56
58	52 and 57
Cochrane	
ID	Search
#1	MeSH descriptor: [Diagnostic Self Evaluation] explode all trees
#2	((self* AND diagnos*) or selfdiagnos*):ti OR ((self* NEXT diagnos*) or selfdiagnos*):ti,ab,kw OR ((self* NEXT test*) or selftest*):ti,ab,kw OR (home NEAR/3 diagnos*):ti,ab,kw
#3	((selfreport* or self-report*) and diagnos*):ti OR ((selfreport* or self-report*) NEAR diagnos*):ti,ab,kw OR ((diagnos* and (selftreat* or self-treat*)):ti,ab,kw
#4	#1 OR #2 OR #3
#5	MeSH descriptor: [Sensitivity and Specificity] explode all trees
#6	MeSH descriptor: [Reproducibility of Results] explode all trees
#7	(sensitiv* or specific* or predict* or accura* or valid* or reproduc*).ti,ab,kw
#8	#5 OR #6 OR #7
#9	#4 AND #8
Cinahl	
#	Query
S39	S3 AND S35 AND S38
S38	S36 OR S37
S37	TX sensitiv* or specific* or predict* or accura* or valid* or reproduc*
S36	(MH "Sensitivity and Specificity") OR (MH "Predictive Value of Tests")
S35	(S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34)
S34	TI (sunburn or "solar keratosis") OR AB (sunburn or "solar keratosis")
S33	(MH "Sunburn")
S32	TI ((gastroenteritis or (stomach N2 (bug or bugs or upset)))) OR AB ((gastroenteritis or (stomach N2 (bug or bugs or upset)))) OR TI (diarrhoea or diarrhea or "food poisoning") OR AB (diarrhoea or diarrhea or "food poisoning")
S31	(MH "Gastroenteritis") OR (MH "Vomiting") OR (MH "Diarrhea")
S30	TI (((sleep N2 (disorder* or disturbance*)) or insomnia*)) OR AB (((sleep N2 (disorder* or disturbance*)) or insomnia*)) OR TI ((sleep N2 (apnea or apnoea))) OR AB ((sleep N2 (apnea or apnoea))) OR TI "restless legs" OR AB "restless legs"
S29	(MH "Sleep Disorders, Intrinsic")
S28	TI (dermatitis or eczema) OR AB (dermatitis or eczema)
S27	(MH "Dermatitis") OR (MH "Dermatitis, Contact+") OR (MH "Eczema")
S26	TI (vaginitis or vaginosis) OR AB (vaginitis or vaginosis)

S25	(MH "Vaginitis+")
S24	TI (((urin* N3 infection*) or bacteriuria or pyuria or cystitis)) OR AB (((urin* N3 infection*) or bacteriuria or pyuria or cystitis))
S23	(MH "Urinary Tract Infections+") OR (MH "Cystitis")
S22	TI asthma* OR AB asthma*
S21	(MH "Asthma+")
S20	TI (((oesophag* or esophag*) N2 disease)) OR (((oesophag* or esophag*) N2 disease)) OR TI ((((oesophag* or esophag* or gastr*) N2 reflux) or heartburn or "heart burn")) OR AB ((((oesophag* or esophag* or gastr*) N2 reflux) or heartburn or "heart burn")) OR (((barrett* or globus) N2 (esophag* or oesophag*))) OR (((barrett* or globus) N2 (esophag* or oesophag*)))
S19	(MH "Gastroesophageal Reflux") OR (MH "Esophageal Motility Disorders") OR (MH "Esophageal Diseases") OR (MH "Barrett Esophagus")
S18	TI (((back N2 (pain or problem* or disorder*)) or "slipped disc*" or sciatica)) OR AB (((back N2 (pain or problem* or disorder*)) or "slipped disc*" or sciatica))
S17	(MH "Low Back Pain") OR (MH "Back Pain") OR (MH "Sciatica") OR (MH "Intervertebral Disk Displacement")
S16	TI ((osteoarthritis or osteo-arthritis or (degenerat* N2 arthritis))) OR AB ((osteoarthritis or osteo-arthritis or (degenerat* N2 arthritis)))
S15	(MH "Osteoarthritis+")
S14	TI (hypertens* or "high blood pressure") OR AB (hypertens* or "high blood pressure")
S13	(MH "Hypertension")
S12	TI (dyslipidemia* or hyperlipidemia* or hypercholesterolemia* or dyslipidaemia* or hyperlipidaemia* or hypercholesterolaemia* or cholesterol* or triglyceride*) OR AB (dyslipidemia* or hyperlipidemia* or hypercholesterolemia* or dyslipidaemia* or hyperlipidaemia* or hypercholesterolaemia* or cholesterol* or triglyceride*)
S11	(MH "Hyperlipidemia+")
S10	TI diabet* OR AB diabet*
S9	(MH "Diabetes Mellitus") OR (MH "Diabetes Mellitus, Type 1") OR (MH "Diabetes Mellitus, Type 2")
S8	TI ((depress* OR anxiety or (mood N2 disorder*))) OR AB ((depress* OR anxiety or (mood N2 disorder*)))
S7	(MH "Depression") OR (MH "Anxiety")
S6	TI ((((respirat* or airway*) N2 infection*) or ("common cold" or influenza or flu or pharyngitis or laryngitis or tonsillitis or "sore throat" or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or "otitis media"))) OR AB ((((respirat* or airway*) N2 infection*) or ("common cold" or influenza or flu or pharyngitis or laryngitis or tonsillitis or "sore throat" or rhinitis or nasopharyngitis or nasolaryngitis or sinusitis or supraglottitis or epiglottitis or tracheitis or urti or "otitis media")))
S5	(MH "Otitis Media+")
S4	(MH "Respiratory Tract Infections") OR (MH "Common Cold") OR (MH "Influenza") OR (MH "Influenza, Human+") OR (MH "Laryngitis+") OR (MH "Pharyngitis") OR (MH "Rhinitis+") OR (MH "Sinusitis+") OR (MH "Tonsillitis+")

S3	S1 OR S2
S2	TI (((self* AND diagnos*) or selfdiagnos*)) OR AB (((self* N1 diagnos*) or selfdiagnos*)) OR TI (((self* N1 test*) or selftest*)) OR AB (((self* N1 test*) or selftest*)) OR TI (home N3 diagnos*) OR AB (home N3 diagnos*) OR TI (((selfreport* or self-report*) and diagnos*)) OR AB (((selfreport* or self-report*) N5 diagnos*)) OR TI ((diagnos* and (selftreat* or self-treat*))) OR AB ((diagnos* and (selftreat* or self-treat*)))
S1	(MH "Self Diagnosis")

For peer review only

Supplementary table 3: Data items included in extraction sheet (where available)

Study identification - author, year, location
Study research question
Study design and setting
Study funding source
Target condition definition/diagnostic criteria
Participant characteristics and numbers, including exclusions
Index test
Reference standard
Flow of participants through study including losses to follow-up
Patient presentation and prior testing
Conduct of the study including timing of the tests, and information on masking
Absolute counts of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) diagnoses.
Statistical analyses that were performed, including whether all participants were included in analyses
Additional summary information on participant preference, timing, or cost, as available.

Supplementary table 4: Protocol

Accuracy of self-diagnosis in conditions commonly managed in primary care: diagnostic accuracy review

Annette Plüddemann¹, Hayley E Jones², Carl Heneghan¹

1. Nuffield Department of Primary Care Health Sciences, University of Oxford, OX2 6GG
2. Population Health Sciences, Bristol Medical School, University of Bristol, BS8 2PS

Background and rationale

A wide range of conditions present to primary care, some acute, some chronic. As a consequence primary care is facing increasing workload (Hobbs et al 2016) that may become unmanageable.

Some common conditions in primary care, therefore, have the potential to be self-diagnosed and self-treated by the patients themselves. This offers superior convenience for individuals, swifter diagnosis and treatment where relevant, reduced costs for health service providers and potentially reduce the burden on primary care services. Self-diagnosis may apply to initial diagnosis, or to diagnosing an exacerbation of an ongoing condition, such as an exacerbation of chronic obstructive pulmonary disease (COPD). In order to support self-diagnosis where safe and appropriate, the efficacy of self-diagnosis needs to be assessed, and this information used to make evidence-based decisions on who can self-diagnose safely.

Where there is an available comparison with diagnosis by a healthcare professional, it is possible to assess the accuracy of self-diagnosis. A further comparison could potentially be made with diagnosis by an allied healthcare professional such as pharmacists, but this is outside the scope of this current review. Sometimes, self-diagnosis involves using a diagnostic test (any type of medical test used to help diagnose or detect disease). For conditions where this is available and appropriate, the accuracy of these tests also informs the safety and efficacy of self-diagnosis by the patient.

Previous studies on the safety and accuracy of self-diagnosis of conditions commonly managed in primary care setting include self-diagnosis of urinary tract infection (Donofrio & Weiner 2013), high blood pressure (Tormo et al 2000) and depression (Hedayati et al 2006).

This review aims to identify, appraise and summarise the available evidence on self-diagnosis in common conditions in primary care. Cooke and colleagues recently reported the 30 most commonly managed conditions in primary care in Australia, which has a health landscape broadly comparable with western Europe (Cooke et al 2013). This list arises from survey data collected between January 2009 and December 2010, which included 194,100 patient encounters from 1,941 GPs. The most commonly managed conditions included some with the potential for self-diagnosis, e.g. urinary tract infection, as well as some that would be unsuitable for self-diagnosis, such as "general check-up". We base our review on conditions

1
2
3 from this list that are relevant for self-diagnosis. We may review infectious diseases and non-
4 communicable diseases separately.
5

6 7 **Objectives**

8 9 ***Primary objective***

10
11 Our primary objective is to summarize the accuracy of self-diagnosis of common conditions in
12 primary care, compared with diagnosis by a healthcare provider.

13 14 ***Secondary objective***

15
16 To summarise any associated relevant information relating to self-diagnosis of common
17 conditions in primary care, such as information on patient preference, timing, or cost (only
18 using information from studies we include for accuracy data). Where there is substantial
19 qualitative information reported, this will only be summarised briefly; detailed qualitative
20 approaches will not be used.
21

22 23 **Methods**

24 25 **Criteria for considering studies for this review**

26 27 ***Types of studies***

28
29 Prospective or retrospective studies comparing the results of self-diagnosis of common self-
30 limiting conditions in primary care by free-living individuals, to the results of a reference
31 standard test performed by a healthcare service provider, will be included. Studies with a
32 case-control design will be excluded. In case of duplicate publications we will include the
33 study report with the highest methodological quality. There will be no language restrictions.

34
35 We will exclude studies comparing self-diagnosis with diagnosis by allied health professionals
36 such as a pharmacists.

37 38 ***Participants***

39
40 Adults (≥ 18 years of age) self-diagnosing conditions common in primary care.

41 42 ***Index tests***

43
44 Index tests will be the self-testing or self-diagnosis of relevant conditions, compared with
45 diagnosis by a healthcare practitioner.

46 47 ***Comparator tests***

48
49 Comparator tests will comprise diagnosis by a healthcare practitioner.

50 51 ***Outcome measures***

52
53 Diagnostic accuracy measures (e.g. sensitivity, specificity, likelihood ratios, predictive values,
54 etc.) and primary data for 2x2 tables. Studies reporting only measures of agreement will be
55 excluded.
56
57
58
59
60

Search methods to identify studies

Electronic searches

The search strategy will be developed in consultation with a healthcare librarian experienced with supporting systematic reviews. No language restrictions will be applied. The search strategy will use multiple electronic databases, from inception onwards including:

Medline

EMBASE

Cochrane Central Register of Controlled Trials (CENTRAL)

Trip database

Web of Science for conference proceedings, dissertations, and theses

World Health Organization International Clinical Trials Registry Platform (ICTRP),

ClinicalTrials.gov

Database of Abstracts of Reviews of Effect (DARE)

We will also search Science Citation Index Expanded for study reports that cite the included studies.

The search may use relevant filters, but in order to maximise sensitivity, will not be limited to these. The reference lists of relevant studies will be examined and additional tools such as the “related articles” feature in PubMed will also be used to identify relevant publications.

Data collection and analysis

Selection of studies

Two reviewers will independently apply the selection criteria to the titles and abstracts of the study reports identified by the searches. If the decision to exclude a study cannot be made on the basis of the title and the abstract, the full study report will be retrieved for inclusion assessment. The final decision on inclusion will be based on the full study report.

Disagreements between reviewers will be resolved by discussion, or if necessary by a third reviewer. Study identification will be summarised in a PRISMA flow diagram.

Data extraction and management

Two reviewers will independently extract information from selected studies into a data extraction sheet. Disagreements will be resolved by discussion, or if necessary with the help of a third reviewer.

Where this is insufficient (or unclear) information, where there is an email address provided, the authors will be contacted via email for clarification. Where data is not available for completion of 2x2 tables, the studies will be excluded from the analysis.

Data to be extracted

The following information will be extracted from the included studies, where available:

Study identification - author, year, location

Study research question

Study design and setting

Target condition definition/diagnostic criteria

Participant characteristics and numbers, including exclusions

Index test

Reference standard

Flow of participants through study including losses to follow-up

Patient presentation and prior testing

Conduct of the study including timing of the tests, and information on masking

Absolute counts of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) diagnoses.

Statistical analyses that were performed, including whether all participants were included in analyses

Additional summary information on participant preference, timing, or cost, as available.

Assessment of methodological quality

To assess methodological quality, we will use the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (Whiting et al 2011). Two reviewers will independently assess studies' methodological quality; disagreements will be resolved by discussion, or if necessary, by a third reviewer. The QUADAS-2 tool facilitates assessment of bias in four areas: patient selection; index test; reference standard; flow and timing; and also facilitates assessment of applicability of the studies to the review research question.

The data will be presented in a tables showing risk of bias and applicability within each domain assessed for each study. These data will be considered in relation to interpreting the results of the studies.

Statistical analysis and data synthesis

Analyses will be conducted for each category of condition specified. Summary tables will detail study information including the patient sample, condition, study design, the test under evaluation, and the comparator.

Meta-analysis

For each test, RevMan will be used to produce paired forest plots to explore the between-study variability of sensitivity and specificity across the included studies. For each study estimate of sensitivity and specificity, corresponding 95% confidence intervals will be shown

1
2
3 to illustrate the uncertainty related to each study estimate. If accuracy has been reported at
4 multiple common thresholds, forest plots will be sub-grouped on threshold.
5

6 Bivariate meta-analysis methods (Reitsma et al 2005) will be used to generate pooled
7 estimates of sensitivity and specificity where sufficient data is available for each test or
8 condition. These will be plotted with 95% confidence and prediction ellipses in Receiver
9 Operating Characteristic (ROC) space. Where appropriate, summary ROC curves will also be
10 plotted, drawing on the equivalence of the bivariate method and the hierarchical summary
11 ROC meta-analysis model (Rutter and Gatsonis 2001; Harbord et al 2007). For these
12 analyses, we will use WinBUGS or the metandi command in Stata, as appropriate, and feed
13 parameters directly into Revman to produce Cochrane-standardised output.
14

15 Where appropriate, meta-analysis models that include multiple thresholds will be employed
16 (e.g. Steinhauser et al 2016 or similar).
17

18 *Investigating heterogeneity*

19
20 For medical conditions for which data from more than one study are available, it may be
21 possible to investigate heterogeneity in the results. Two approaches will be used to explore
22 the sources of between-study heterogeneity: 1) inclusion of study level characteristics as
23 covariates in the bivariate model (meta-regression) 2) carrying out sub-group analyses.
24 These approaches will only be carried out if there is sufficient data available and sub-group
25 specific pooled estimates are thought to be of clinical relevance. Any meta-regressions will be
26 carried out using WinBUGS or the xtmelogit command in Stata.
27

28 *Sensitivity analyses*

29
30 If there appear to be any outliers in the data, these studies will be removed from the analysis
31 to evaluate the impact on the overall pooled estimates.
32

33 *Investigating reporting bias*

34
35 Funnel plots used to detect publication bias in reviews of RCTs have been shown to be
36 misleading for diagnostic test accuracy reviews (Deeks et al 2005; Leeflang et al 2008).
37 Funnel plots as an assessment of reporting bias will therefore not be included in this review.
38 Publication bias will be assessed using Deek's test, as recommended by the Cochrane
39 Handbook, where data allows (Deeks et al 2005).
40

41 **Funding**

42
43 This project is funded by National Institute for Health Research School for Primary Care
44 Research (NIHR SPCR) [ProjectNumber 390]
45
46

47 **Declarations of conflict of interest**

48
49 Dr. Plüddemann reports grants from NIHR, grants from NIHR School of Primary Care
50 Research, during the conduct of the study; and occasionally receives expenses for teaching
51 Evidence-Based Medicine. Dr. Heneghan reports receiving expenses and fees for his media
52 work. He has received expenses from the WHO and holds grant funding from the NIHR, the
53 NIHR School of Primary Care Research, The Wellcome Trust and the WHO. He has received
54 financial remuneration from an asbestos case. He has also received income from the
55
56
57
58
59
60

publication of a series of toolkit books published by Blackwells. On occasion, he receives expenses for teaching EBM and is also paid for his GP work in NHS out of hours. CEBM jointly runs the EvidenceLive Conference with the BMJ and the Overdiagnosis Conference with some international partners which are based on a non-profit making model.

References

Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. *Aust Fam Physician*. 2013;42(1-2):65-8.

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-93

Donofrio JC, Weiner SG. Female patient self-diagnosis compared with emergency physician diagnosis of urinary tract infection. *J Emerg Med*. 2013;45(6):969-73. doi: 10.1016/j.jemermed.2013.07.019

Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics*. 2007 Apr;8(2):239-51.

Hedayati SS, Bosworth HB, Kuchibhatla M, Kimmel PL, Szczech LA. The predictive value of self-report scales compared with physician diagnosis of depression in hemodialysis patients. *Kidney Int*. 2006 May;69(9):1662-8.

Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, Salisbury C; National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *The Lancet*; 387; 10035: 2323-2330

Leeftang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med*. 2008; 149(12): 889-97

Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001 Oct 15;20(19):2865-84.

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005 Oct;58(10):982-90

Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014

Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Med Res Methodol*. 2016 Aug 12;16(1):97. doi: 10.1186/s12874-016-0196-1.

Tormo MJ, Navarro C, Chirlaque MD, Barber X. Validation of self diagnosis of high blood pressure in a sample of the Spanish EPIC cohort: overall agreement and predictive values. EPIC Group of Spain. *J Epidemiol Community Health*. 2000 Mar;54(3):221-6.

Supplementary Table 5: Characteristics of included studies

Disease/ condition	Author, year	Design	Setting	Country	Study duration	Number of participants	Mean age* (years)	Female (%)
Vaginal infection	Donders 2016	Prospective diagnostic accuracy study	Birth control, general gynaecology, infertility and prenatal clinics	Uganda	N/A	360	28.3	100
	Sungkar 2012	RCT (intervention arm only)	Prenatal clinics or hospitals	Indonesia	24 weeks	176	28	100
	Geva 2006	Prospective diagnostic accuracy study	Gynaecologic clinics	Israel	N/A	593	18 – 60 range	100
	Ryan-Wenger 2010	Prospective diagnostic accuracy study	Military clinics	USA	N/A	546	25.7	100
	Jones 2013	RCT (both arms)	Clinics or home	Brazil	N/A	695	18 – 40 range	100
Common skin condition	Bregnhøj 2011	Prospective cohort	Not reported	Denmark	18 months	502	17.5	95
	Svensson 2002	Prospective diagnostic accuracy study	Dermatology outpatient clinics	Sweden	N/A	208	40.4	50
	Elsner 2015	Prospective diagnostic accuracy study	Not reported	Germany & Austria	N/A	165	≥18 years	81
	Josefson 2011	Prospective diagnostic accuracy study	Hospital dermatology departments	Sweden	N/A	243	44	69
HIV	Assiimwe 2014	RCT (unsupervised arm only)	Home	Uganda	N/A	123	28 (23–32), median (IQR)	47
	Belete 2019	Cross-sectional	Public health facilities	Ethiopia	N/A	400	29 (17.7–40.3), median (IQR)	61
	Choko 2011	Cross-sectional	Home	Malawi	N/A	241	27 (NR) median (IQR)	52
	Choko 2015	RCT (intervention arm only)	Home	Malawi	2 years	2370	NR	NR

	Kapaku 2017	Cohort	Home/Voluntary counselling & testing facilities	Zambia	1 year	2572	26 (21–35), median (IQR)	59
	Kurth 2016	Cross-sectional	NR (not at home)	Kenya	N/A	240**	36	33
	Martinez Perez 2016	Cross-sectional	Health care clinics/HIV testing sites	South Africa	N/A	2205	Male: 27 (11-36) Female: 28 (22-36) median (IQR)	66
	Orasure 2012	Cross-sectional	Study site	USA	N/A	5798	NR	50
	Pant Pai 2013	Cross-sectional	Hospital	South Africa	N/A	251	≥18 years	79

RCT – randomised control trial; N/A – not applicable; NR – not reported, * unless otherwise indicated, ** Subset (N=113) that used laboratory reference standard included in systematic review and pooled analysis.

Supplementary Table 6: Characteristics of self-diagnosis (index) tests and reference standard tests

Disease/condition	Author, year	Test for	Self-diagnosis test (index)	Self-diagnosis test threshold	Reference test	Reference test threshold	Interval between index & reference test	Data collection points
Vaginal infection	Donders 2016a	Bacterial vaginosis	Vaginal fluid test using pH strip	+/- (≥ 4.5 pH)	Air dried vaginal fluid test using gram staining (Nugent score)	+/- (≥ 4.5 pH, Nugent score 7-10)	NR	N/A
	Donders 2016b			+/- (≥ 4.7 pH)	Assessment by central laboratory	+/- (≥ 4.7 pH, Nugent score 7-10)		
	Sungkar 2012	Bacterial vaginosis	Vaginal fluid test using pH strip	+/-	Vaginal fluid test using microbiological gram staining test (Kopeloff modified Gram stain) Prepared by midwives for laboratory assessment	+/-	NR	Baseline, 16 – 18, 18 – 20, 20 – 22, 22 – 24 weeks
	Geva 2006	Bacterial vaginosis and/or <i>Trichomonas vaginalis</i>	Vaginal discharge test using panty liner test kit (VI-SENSE)	+/- (based on strip colour, no level reported*)	Clinical diagnosis plus vaginal wall swabs testing pH (nitrazine paper), amine, culture (InPouch TV, BioMed Diagnostic) and gram staining (Nugent score). Assessment by board certified gynaecologists and central laboratory.	BV: >3: (a) homogeneous discharge, (b) pH value >4.5, (c) release of fishy odor (KOH was added to the vaginal discharge, and (d) presence of clue cells; or 7+ Nugent score of gram stain TV: +/- by culture	Within 6 hours	N/A
	Ryan-Wenger 2010	Bacterial vaginosis and/or <i>Trichomonas vaginalis</i>	Vaginal fluid test using Women in the military self-diagnosis kit	+/- (based on ≥ 4.7 pH, amines,	Clinical interview plus vaginal fluid test for pH (nitrazine paper, amines (FemExam card), whiff, wet mount	+/-	None	N/A

			(includes FemExam card)	vaginal itching)	microscopy (Affirm VPIII Microbial Identification Test) Assessment by women's health nurse practitioner plus in clinic microscopy testing			
	Ryan-Wenger 2010	<i>Candida</i> vaginitis	Vaginal fluid test using Women in the military self-diagnosis kit (includes FemExam card)	+/- (based on ≥ 4.7 pH, amines, vaginal itching)	Clinical interview plus vaginal fluid test for pH (nitrazine paper, amines (FemExam card), whiff, wet mount microscopy (Affirm VPIII Microbial Identification Test) Assessment by women's health nurse practitioner plus in clinic microscopy testing	+/-	None	N/A
	Jones 2013	<i>Trichomonas vaginalis</i>	Vaginal fluid testing using dipstick test (OSOM Trichomonas rapid test)	+/- (two red lines)	Vaginal fluid test using PCR test Assessment at central laboratory	+/-	None	N/A
Common skin condition	Bregnhøj 2011	Eczema	Questionnaire on presence of eczema	Positive response	Hand Eczema Severity Index (HECSI) Interpretation by clinician	+ve for presence of eczema	Same day	Inclusion & 18 month follow-up
	Svensson 2002	Hand eczema	Questionnaire on presence of eczema	Positive response	Hand examination for erythema, papules, vesicles, scaling, fissures, lichenification and hyperkeratotic areas. Assessment by experienced dermatologist	+ve if erythema and papules / vesicles OR erythema and scaling and fissures / lichenification	None	N/A
	Elsner 2015	Allergic reaction to nickel and fragrance	Irritant reaction to surgical tapes on upper arm after 48 hours	+/-	Irritant reaction to surgical tapes on upper arm after 48 hours Assessment by clinician	+/-	None	N/A
	Josefson 2011	Allergic reaction to nickel	Medical plaster patches (Nixema) on upper arm with	+/-	Medical plaster patches (Finn	+/-	Same day	N/A

			readings on days 3-4		chambers® on Scanpor® tape or IQ Ultra® Chambers) on back with readings on days 3-4 and/or day 7 Assessment by dermatologist			
HIV	Assiimwe 2014	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® In-Home Rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Finger pick blood test. Nationally approved algorithm of POC rapid HIV tests (Determine (Abbot Laboratories), STAT-PAK (Chembio Diagnostic Systems Inc) and Unigold (Trinity Biotech plc) as a tiebreaker). Assessment by research assistants.	+/-	12 -72 hours	N/A
	Belete 2019	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® Rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Finger pick blood test. Nationally approved algorithm of POC rapid HIV tests (Wanti (Beijing), Unigold (Trinity Biotech plc), Vikia). Assessment by health professional.	+/-	Same time	N/A
	Choko 2011	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Finger pick blood test. Algorithm of POC rapid HIV tests (Determine (Abbot Laboratories), Unigold (Trinity Biotech plc) and SD Bioline HIV i/II (Standard Diagnostics, Inc.) as a tiebreaker). Assessment by counsellor.	+/-	Same time	N/A
	Choko 2015	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Finger pick blood test. Two parallel POC rapid HIV tests (Determine (Abbot Laboratories), Unigold (Trinity Biotech plc)). Assessment by nurse.	+/-	Approx. 1 week	1 – 12, 13 – 24 months (max. 1 test in 12 months)

	Kapaku 2017	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Venous blood test (EDTA tube) Testing in certified central laboratory (Abbott Architect HIV1 Ag/Ab combo assay, positive results confirmed by BioRad GS HIV combo Ag/Ab assay)	+/-	Within 8 hours	Once in 12 month study period.
	Kurth 2016	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Venous blood test. ELISA testing by single person in certified central laboratory (Vironostika HIV Uni-Form II Ag/Ab (bioMe´rieux Inc.))	+/-	Within 8 hours	N/A
	Martinez Perez 2016	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick ADVANCE® Rapid HIV-1/2 Antibody Test	+/-	Finger pick blood test. POC rapid HIV tests (Determine (Abbot Laboratories), confirmatory Unigold (Trinity Biotech plc)). Assessment by HIV counsellor.	+/-	Same time	N/A
	Orasure 2012	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® In-Home Rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Venous blood test. FDA approved serum EIA and Western blot in FDA approved laboratory.	+/-	Unclear	N/A
	Pant Pai 2013	Antibodies for HIV1 and HIV2	Oral fluid test using OraQuick® rapid HIV-1/2 Antibody Test (Orasure Technologies)	+/-	Venous blood test. ELISA with p24 Antigen tests in reference laboratories all within 24 hours (Architect HIV Ag/Ab combo (Abbott Laboratories), positive results confirmed by Western Blot)	+/-	Same time	N/A

HIV - Human immunodeficiency virus, POC – Point of care, max. – maximum, FDA – Food and Drug Administration, USA, N/A – not applicable; NR – not reported; KOH – potassium hydroxide; PCR – polymerase chain reaction; * whilst no reported level was given for VI-SENSE, it was reported that the polymer used in this product had a range of 4.3 – 5.1 pH



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title page (pg 1)
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction (pg 4)
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction (pg 2)
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods, Types of studies (pg 4) & Statistical analysis and data analysis (pg 5)
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods, Search methods to identify studies (pg 4)
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Supplementary table 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods, Selection of studies (pg 5). No automation tools used.
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods, Data extraction and management (pg 5). No automation tools used.
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Supplementary table 2. Methods, Data extraction and



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
			management (pg 5).
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Supplementary table 2. Methods, Data extraction and management (pg 5).
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods, Assessment of methodological quality (pg 5). No automation tools used.
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods, Statistical analysis and data synthesis (pg 5) Supplementary table 2
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods, Statistical analysis and data synthesis (pg 5)
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A Studies included providing 2x2 published data.
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods, Statistical analysis and data synthesis (pg 5)
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods, meta-analysis (pg 5)
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods, Investigating



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
			heterogeneity (pg 5 & 6)
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Results, Self-diagnosis of HIV (pg 9)
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not examined.
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods, meta-analysis (pg 5)
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Results (pg 6) Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A. Rather 2 studies were included that were borderline exclusion (pg 7 & 8)
Study characteristics	17	Cite each included study and present its characteristics.	Results (pg 6 – 9). Supplementary tables 4 & 5
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Figure 5
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Figures 2 - 4
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results (pg 6-9)
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results, Self-diagnosis of HIV (pg 8 & 9). Figure 4
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results, Self-diagnosis of HIV (pg 8 & 9). Figure 4
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Results, Self-



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
			diagnosis of HIV (pg 9). Figure 4
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Discussion (pg 10 & 11)
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion (pg 10 & 11)
	23b	Discuss any limitations of the evidence included in the review.	Discussion (pg 10 & 11)
	23c	Discuss any limitations of the review processes used.	Discussion (pg 10 & 11)
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion (pg 11)
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Abstract (pg 2). Methods (pg 6)
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Supplementary table 3
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Methods, meta-analysis (pg 5). Results, self-diagnosis of HIV (post-hoc analysis, pg 9)
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Funding statement (pg 11)
Competing interests	26	Declare any competing interests of review authors.	Competing interests statement (pg 11)
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data availability statement (pg 11)



PRISMA 2020 Checklist

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71
For more information, visit: <http://www.prisma-statement.org/>

For peer review only