

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 (<u>http://bmjopen.bmj.com</u>).

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

**BMJ** Open

# **BMJ Open**

# Over and undertesting in primary care: a systematic review and meta-analysis.

Journal:	BMJ Open
Manuscript ID	bmjopen-2017-018557
Article Type:	Research
Date Submitted by the Author:	06-Jul-2017
Complete List of Authors:	O'Sullivan, Jack; Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences Albasri, Ali Nicholson, B; University of Oxford, Perera, Rafael; University of Oxford, Primary Health Care Aronson, Jeffrey; University of Oxford, Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences Roberts, Nia; University of Oxford, UK, Bodleian Health Care Libraries, Heneghan, Carl; Oxford University, Primary Health Care
<b>Primary Subject Heading</b> :	Epidemiology
Secondary Subject Heading:	Diagnostics, General practice / Family practice
Keywords:	PRIMARY CARE, RADIOLOGY & IMAGING, EPIDEMIOLOGY, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE<sup>™</sup> Manuscripts

2/

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

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

# Over and undertesting in primary care: a systematic review and meta-analysis.

O'Sullivan J<sup>1</sup>, Albasri A<sup>1</sup>, Nicholson B<sup>1</sup>, Perera R<sup>1</sup>, Aronson J<sup>1</sup>, Roberts N<sup>2</sup>, Heneghan C<sup>1</sup>

<sup>1</sup> Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Science, University of Oxford, UK

<sup>2</sup> Bodleian Health Care Libraries, University of Oxford.

Jack W O'Sullivan, Clinical Researcher, jack.osullivan@phc.ox.ac.uk Ali Albasri, Clinical Researcher, ali.albasri@gtc.ox.ac.uk Brian D Nicholson, Clinical Researcher, brian.nicholson@phc.ox.ac.uk Rafael Perera, Professor of Medical Statistics, rafael.perera@phc.ox.ac.uk Jeffrey Aronson, Reader in Evidence-Based Medicine, jeffrey.aronson@phc.ox.ac.uk Nia Roberts, Medical Librarian, nia.roberts@bodleian.ox.ac.uk Carl Heneghan, Professor of Evidence-Based Medicine, carl.heneghan@phc.ox.ac.uk

Correspondence to: Dr Jack O'Sullivan

Centre for Evidence-Based Medicine Nuffield Department of Primary Care Health Sciences Radcliffe Observatory Quarter, Oxford, OX2 6GG

#### Abstract

#### Background

Health systems are currently subject to unprecedented financial strains. As most patient care is provided in primary care, inappropriate test use wastes finite health resources (overuse) and delays diagnoses and treatment (underuse).

#### Objective

To identify over and under use of diagnostic tests in primary care.

#### Design

Systematic review and meta-analysis.

#### Data sources and eligibility criteria

We searched MEDLINE and EMBASE from January 1999 to February 2016 for studies that measured the inappropriateness of any diagnostic test (measured against a national or international guideline) ordered for adult patients in primary care.

#### Results

We included 206,601 patients from 55 observational studies in 15 countries. We extracted 94 measures of inappropriateness (39 underuse, 55 overuse) from included studies for 45 different diagnostic tests.

The overall rate of inappropriate diagnostic test ordering varied substantially (0.2% to 100%).

18 tests were underused >50% of the time. Of these, echocardiography was the most frequently studied (n=4 measures) and was consistently underused (between 54% and 89%). There was large variation in the rate of inappropriate underuse of pulmonary function tests (38% to 78%, n = 8) and colonoscopy (8% to 69%, n=2).

Nine tests were inappropriately overused >50% of the time. Echocardiography was consistently overused (78% to 92%), whereas inappropriate overuse of both urinary cultures and upper endoscopy varied widely, from 36% to 77% (n=3) and 10% to 54% (n=10) respectively.

#### Conclusions

There is marked variation in the appropriate use diagnostic tests in primary care. Specifically, the use of echocardiography (both under and overuse) is consistently poor. There is substantial variation in the rate of inappropriate underuse of pulmonary function tests and colonoscopy and overuse of upper endoscopy and urinary cultures.

Registration number: PROSPERO Registration ID: CRD42016048832

Manuscript word count: 3,252

## Strengths and limitations of this study

#### Strengths

- Generates rate of under and overtesting for specific diagnostic tests against national or international guidelines
- Only includes data from real clinical encounters rather than surveys or hypothetical clinical vignettes.
- Quantified inappropriate ordering of all types of diagnostic tests, rather than just laboratory.

#### Limitations

- Systematic reviews are restricted to published literature, thus rates of inappropriate ordering is not available for all tests available to primary care physicians.
- Included studies measure appropriateness of testing in a particular health care setting against a particular guideline, thus reflect test ordering in a specific health care setting.

Topper territory only

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

#### Page 4 of 41

#### Introduction

Reaching a diagnosis in primary care is exceedingly complex. The combination of undifferentiated symptoms, a low prevalence of serious disease, a high degree of symptom overlap between serious and benign conditions, patients with multiple complaints, and psychological or social distress manifesting somatically all complicate reaching a diagnosis [1]. In around 40% of primary care consultations a diagnosis cannot be established from the history and physical examination alone [2], and tests are therefore often needed [1,3].

Primary care consultations make up most of the care provided in healthcare systems (90% of consultations in the UK [4], 55% of consultations in the USA[5]) and inappropriate diagnostic testing in primary care therefore has enormous resource implications. Given the calls for £22 billion in efficiency savings from the UK's National Health Service (NHS) [6] and the \$660 billion US Medicare deficit predicted by 2023 [7], ensuring the appropriateness of primary care diagnostic testing is crucial to the sustainability of healthcare systems [8].

Inappropriate diagnostic tests in primary care can be both inappropriately underused and overused. Underuse of tests, failure to order a test when clearly indicated, can lead to diagnostic errors and delays in diagnosis and the delivery of effective treatment, leading to adverse patient outcomes and further healthcare costs [9,10]. Overuse of tests, the delivery of tests with no clear benefit or when potential harms outweigh potential benefits, subjects patients to direct harms, such as radiation exposure, as well as potential adverse outcomes (e.g. contrast nephropathy) [11], incidental findings [12], and overdiagnosis [13]. Overuse is also a waste of finite healthcare expenditure, diverting resources from beneficial tests and treatments [14–16].

Many drivers encourage inappropriate under and overuse of diagnostic tests in primary care. Greater access to tests [17], the medicolegal consequences of under-testing [18], few if any disincentives to overinvestigate [14], and clinical performance measures [19] may all contribute to overuse. Increasing primary care workload [4], time constraints [19], and difficulty keeping up-to-date with rapidly increasingly evidence [20] may contribute to both inappropriate underuse and overuse.

Guidelines set the standard of care across most health-care settings [21,22]. Furthermore, they provide a medicolegal framework [23], inform health-care policy, and improve both care outcomes and processes of care [24]. Despite some recognised limitations, including varying quality of guidelines [25–27], guidelines are often used as markers of health-care appropriateness [28–31]. Zhi et al, for instance, used guidelines as a measure of appropriateness to estimate under and overuse of laboratory testing [29]. They estimated that 45% (95%CI 34 – 56%) of secondary care laboratory testing is underused and 21% (95%CI 16 – 25%) is overused.

Despite the increasing use of healthcare resources [32], rising healthcare expenditure [6–8], increasing demands placed on primary care [4], and apparent drivers of inappropriate testing [1,4,14,17–20], it is not clear how often diagnostic tests are inappropriately overused or underused in primary care. We therefore conducted a systematic review to quantify the frequency of appropriate ordering of all types of diagnostic tests from primary care in relation to their respective guidelines.

#### Methods

This study was conducted and is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33] and Meta-analysis of Observational Studies in Epidemiology (MOOSE) statements [34].

#### Protocol and Registration

The protocol has been published and is available online (open access) via the International prospective register for systematic reviews (PROSPERO) database (Registration ID: CRD42016048832).

#### Patient involvement

Patient and Public Involvement (PPI) interviews have been conducted, which underpin a series of studies examining test ordering from primary care. However, patients were not explicitly involved in the design, analyses or interpretation of this study.

#### Search Strategy

We searched EMBASE (OvidSP) and MEDLINE (OvidSP) databases from January 1999 to February 2016 for studies of any design measuring how often diagnostic test guidelines were followed in primary care (Supplementary file: Search Strategy). Conference abstracts published after 2015 were also searched for in these databases to capture data not yet published. We also searched the WHO International Clinical Trials Registry Platform (<u>http://apps.who.int/trialsearch/</u>), ClinicalTrials.gov, and the reference lists of included studies.

#### Eligibility Criteria

We included studies of any design if they measured the rate of inappropriate ordering (overuse) or not ordering (underuse) of diagnostic tests ordered from primary care against national or international guidelines. We considered all diagnostic tests ordered in adults. We also included studies that measured diagnostic tests ordered from primary care but performed in secondary care (e.g. upper endoscopy). We included the control arms of RCTs if they offered exclusively usual care, and the pre-intervention periods of studies that used interrupted time series designs (before and after studies).

We excluded studies if they met the following criteria: >20% of participants were children (>20% under 18 years old); diagnostic tests not ordered by General Practitioners; screening or monitoring tests, or publication before 1999 (studies after 1999 were considered to ensure that results would more closely reflect current practice). We defined a screening test as a test on an asymptomatic or symptomatic person without signs or symptoms related to that test [35,36]. We defined monitoring tests as 'a test for a patient with an established diagnosis, for which the test is used to measure progression of the disease' [37]. We excluded studies if they did not give a measure of appropriateness or if appropriateness was measured against local guidelines, such as a guideline specific to a hospital or region, rather than international or national guidelines.

#### Study selection and data extraction

Three reviewers (JS and AA or BN) independently screened titles, abstracts, and full texts for eligibility. The same reviewers assessed risks of bias and extracted the following data from included studies: patient demographics, eligibility criteria, name and type of diagnostic test, duration of study (days), guideline name and recommendation, total number of tests performed, and the number of tests ordered when the specific guideline recommended not ordering (inappropriate overuse) or the number of tests not ordered when the guideline recommended ordering it (inappropriate underuse). The last two data points (overuse and underuse) represent 'measures of inappropriateness'. When studies measured inappropriateness of multiple tests we extracted data on each test and presented them as

individual measures of inappropriateness. When studies measured tests across different periods we extracted measures for each time point and considered each one as an individual measure of inappropriateness.

We assessed the quality of included studies using a modified version of the Hoy risk of bias tool [38]. This tool has been validated to assess the internal and external validity of prevalence studies [38]. Our modified version of this tool kept the same domains, but adjusted the wording of the tool to reflect prevalence of inappropriate testing rather than prevalence of disease. Our tool (and results) is available in Supplementary Table 3.

# Statistical analysis

The primary outcome was the prevalence of inappropriate diagnostic testing. Inappropriate testing was measured in two ways:

1) Overuse: A diagnostic test was ordered when the relevant guideline recommends not ordering it, for instance, imaging for non-red flag low back pain (LBP).

2) Underuse: A diagnostic test was not ordered when the relevant guideline recommended ordering it, for instance, spirometry for suspected COPD.

We expressed measures of inappropriateness as proportions (%), where the numerator represents the total number of times a guideline recommendation was not followed and the denominator the total number of times a guideline recommendation could have been followed. For instance, the number of times imaging was inappropriately ordered for non-red flag headache as a proportion of the total number of patients who presented with non-red flag headache. Given these data are proportions, we calculated Clopper-Pearson 95% confidence intervals for each individual measure of appropriateness. We conducted sensitivity analyses with high risk of bias studies excluded.

Where the same guideline and recommendation were used by multiple studies (e.g. five studies measured inappropriate underuse of spirometry testing in patients with COPD [39–43] using the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guideline) we pooled the measures and assessed heterogeneity. We combined measures of inappropriateness using a random-effects meta-analysis with 95% confidence intervals (Clopper-Pearson), for this reason each measure of appropriateness contributed relatively evenly to pooled estimates. We performed double arcsine transformation on prevalence data to stabilize the variance [44], and pooled the data using the inverse variance method [45]. We assessed heterogeneity using the I<sup>2</sup> statistic [46]. We did not combined measures of overuse and underuse, as they have different denominators: overuse involves the total number of tests ordered, whereas underuse involves the total number of times a test should have been ordered. We performed analyses using R version 3.3.2 (R project).

#### Study selection and characteristics

We included 55 observational studies from 12,824 references identified from independent searches by two authors (JOS and AA or BN) (see Figure 1). These studies were conducted in 15 countries and included 206,601 patients (Supplementary Table 1). Supplementary Table 2 shows the 94 measures of inappropriateness extracted from included studies for 45 different diagnostic tests measured against 69 guideline recommendations (39 measured underuse and 55 measured overuse). Guideline recommendations came from 35 different guideline organisations from 15 countries.

Fourteen studies measured inappropriateness of more than one diagnostic tests for the same condition (e.g. chest x-ray (CXR), electrocardiography (ECG), and transthoracic echocardiography (TTE) to confirm or refute a diagnosis of heart failure). Two studies [47,48] measured inappropriateness across multiple time periods.

Included studies measured inappropriateness in one of three ways:

1. Patients with specific symptoms were recruited and followed up to see if they had received an inappropriate diagnostic test (overuse) or hadn't received the appropriate diagnostic test (underuse) in line with the relevant guideline recommendation (e.g. patients with non-red flag LBP recruited and followed up to see if they received imaging [49]).

2. Patients who had undergone a diagnostic test were identified (via hospital or national databases) and an assessment of whether the test was inappropriate (as per the defined guideline recommendations) via individual patient data was made (overuse). For instance, patients who had an upper endoscopy[50]).

3. Patients with a diagnosis were identified via hospital or national databases and assessed to see whether they had received the appropriate diagnostic test (as per the defined guideline) to confirm or refute the diagnosis via individual patient data (underuse). For instance, assessing if patients with a diagnosis of COPD had spirometry to confirm or refute the diagnosis [39]).

#### Risk of bias

Two thirds of the studies (n=36) were graded as being at low risk of bias, 15 (27%) at moderate risk, and 4 (7%) at high risk (Supplementary Table 3). Moderate or high risk studies were at an increased risk of non-response bias (>20%), non-objective collection of data, and/or unclear intervals between symptom onset and diagnostic test use. Supplementary Table 3 outlines risk of bias scores in detail.

#### Proportions of diagnostic tests ordered in line with specific guideline recommendations

There was large variation in the rate of inappropriate diagnostic test ordering. The 94 diagnostic test guideline recommendations were not followed 0.2 - 100% of the time (Figure 2), wide variation was largely sustained (0.2 - 99.94%) when a further analysis was conducted excluding studies judged high risk of bias. The prevalence of underuse varied 8.2% to 100%, whereas overuse prevalence varied between 0.2% and 94.2%. Similarly, this variation was essentially maintained upon exclusion of high risk studies (under use 9.8% - 99.9%, overuse 0.2 - 99.9%).

Underused tests

#### **BMJ** Open

Table 2 (supplementary) shows that 18 tests were underused more than 50% of the time. Echocardiography was the most frequently studied (n=4 measures in Poland, UK (2), Brazil). In patients with heart failure, echocardiography was underused between 54% and 89% (n=3) of the time and in atrial fibrillation 56-64% (n=2).

For some tests there was large variation in the rate of underuse (Figure 3). Underuse of pulmonary function tests (PFTs) to confirm or refute COPD, measured against the Global Initiative for Chronic Obstructive Lung Disease (GOLD), NICE (UK) and Danish National Board of Health guidelines, varied from 38% to 78% (n=8). Similarly, underuse of colonoscopy for numerous clinical scenarios (such as 'unexplained iron deficiency anaemia') also varied substantially, from 8% to 69% (n=2). None of the studies that studied echocardiography, PFTs or colonoscopy were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

#### Overused tests

Nine tests were overused more than 50% of the time (Supplementary Table 2). Echocardiography was consistently overused, for instance in 'routine perioperative evaluation of ventricular function with no symptoms or signs of cardiovascular disease', whereas other tests (urinary cultures and endoscopy) were overused at varying rates. The over use of echocardiography was studied in the UK [51] and the Netherlands [52]. The rates of overuse varied between the two settings: between 78% (Netherlands) and 92% (UK). Overuse of urinary cultures for uncomplicated urinary tract infections was studied in the USA [53,54] and Spain [55] the rate varied from 57% to 77% in the USA, but was as low as 36% in Spain. Overuse of upper endoscopy was studied widely (n=10); in Australia [56,57], Saudi Arabia [58,59], UK [60], Italy [61–63], USA [50], and Malaysia [64]. The overuse varied markedly, from and 7.5% to 54% (n=10) respectively (figure 4, Supplementary Table 2). None of the above studies were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

Our results also suggest that the inappropriate overuse of CT and MRI scans for non-red flag headache (a headache without symptoms suggesting a malignant underlying pathology) has more than doubled in the last ten years in the USA (2000: 6.7% (95%CI: 5.4 to 8.2%, 2010: 14% (95%CI 12. to 16%) (Supplementary Table 2) [48]. Conversely, the rate of inappropriate overuse of radiology tests for non-red flag low back pain was consistently low, with all (n=11) but one measure showing inappropriate overuse less than 25% of the time (Supplementary Table 2). The one study [65] that showed inappropriate overuse around 50% of the time was conducted in 2001, which may reflect improvements in practice over time. None of these studies were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

#### Variation of inappropriateness against the same guideline recommendation

Eleven different guideline recommendations were studied more than once (Figure 2). There was significant heterogeneity ( $I^2 > 50\%$ ) in nine of these pooled measures. Significant heterogeneity may have occurred for several reasons: 1) vastly different populations (for instance, one studied measured the inappropriateness of upper endoscopy in Saudi Arabia [59] using the American Gastroenterological Association recommendations, whereas another study used the same recommendations in the USA [66]; 2) Contrasting healthcare systems [67,68]; 3) Relevance and applicability of one country's national guideline to another country [69]; 4) A low number of measures for meta-analysis [46] and/or 5) Significant heterogeneity, reflecting significant variation in inappropriate ordering.

#### Overall rates of inappropriate testing

To be consist with previous research [29], we calculated combined rates of over and under test of testing. Overuse of testing occurred on average one-fourth of the time (25%, 95%CI 16 to 35%;

n=55), whereas underuse was more prevalent, occurring in nearly two-thirds of the tests ordered (60% 95%CI 51 to 69%; n=39). The overall rate of inappropriate overuse for laboratory tests was 49% (95%CI: 30% to 68%; n=6), compared with only 15% for radiology tests (95%CI: 8.6% to 25%; n=38) and 34% (95%CI: 26 to 44%) for 'other tests'. These results combine estimates from multiple different health care settings and capture only the studied selection of diagnostic tests available in primary care, thus we feel conclusions from these estimates should be made with caution.

#### Discussion

There is marked variation in the rate of underuse and overuse of diagnostic tests from many primary care settings across the world. This variation suggest improvement can be made in the rate of appropriate diagnostic test ordering.

Primary care use of echocardiography is consistently poor. Echocardiography is inappropriately underused for some clinical situations, e.g. confirming a diagnosis of heart failure, and inappropriately overused in others, e.g. perioperative assessment. This was consistent across the countries where appropriateness of echocardiogram has been studied. This is of concern, given the expertise and resource requirements to perform the test and the increasing availability of direct access ordering for primary care physicians.

For four tests we found marked variation in the rate of inappropriate use. Underuse of pulmonary function tests and colonoscopy varied by 40% and 60% respectively, whereas overuse of urinary cultures and endoscopy both varied by around 40%.

Radiology tests for both non-red flag low back pain and non-red flag headache were frequently *not* overused, but the rate of overuse imaging for non-red flag headache showed concerning trends, more than doubling from 2000 to 2010 (Supplementary Table 2).

#### Implications

Four conclusions can be drawn from our results: 1.Ordering of echocardiograms from primary care appears to require improvement, 2. Markedly varying rates of inappropriate use for pulmonary function tests (underuse), colonoscopy (underuse), upper endoscopy (overuse), and urinary cultures (overuse) suggests that ordering can be improved, 3. Determining reasons for deviation from guidelines is an appropriate next step, and 4. An assessment of the quality of guidelines supporting diagnostic test use would be advantageous.

#### Strengths in relation to other studies

Compared with other studies of inappropriate use of healthcare resources, we used data from real clinical encounters. This allowed a more robust assessment of diagnostic test inappropriateness, where other studies used surveys and hypothetical clinical vignettes [19,75,76]. Furthermore, we quantified the appropriateness of all types of diagnostic tests, rather than focusing on a specific test or specific disease (such as only laboratory tests [29]). Zhi et al [29] quantified the mean rates of overuse and underuse of laboratory tests. Our review is the first systematic pooling of studies that measured inappropriateness of all diagnostic tests ordered from primary care.

Our use of guideline recommendations as the metric of appropriateness allowed a direct measure of diagnostic test appropriateness. Other studies that have assessed temporal and geographical variation in the use of diagnostic tests [77,78] have noted substantial differences in diagnostic practices across different regions, irrespective of disease prevalence and patient characteristics [78]. These studies, however, could not quantify what proportion of the temporal increase in the use of a diagnostic test is inappropriate and what proportion of variation between regions is inappropriate. We have quantified the proportion of inappropriate testing.

#### Limitations

The use of guidelines to quantify appropriateness of diagnostic tests could be considered a limitation of this study. Guidelines are often criticised for varying quality [25–27,79] and panel members' conflicts of interests [80]. However, clinical practice guidelines have been shown to improve both care outcomes and processes of care [24], allow assessment of care on a population level, inform health policy [81,82], set the standard of care across many health care settings [21,22], and provide a medicolegal framework [23]. One major medical insurance company advises that 'doctors must be prepared to explain and justify their decisions and actions, especially if they depart from guidelines produced by a nationally recognised body' [23]. Furthermore, guidelines have been used to measure appropriateness of the use of tests in other published peer-review studies [29]. There will always be times when it is appropriate to depart from guidelines, but dramatic, consistent variation from guidelines requires investigation and is unlikely to be caused entirely by the quality of guidelines.

Furthermore, our study includes only a selection of diagnostic tests and is thus not an allencompassing reflection of clinical practice. The data reflects the use of a specific test, sometimes for a particular clinical situation, in a particular country's health care system. Thus, policy makers and those interested in improving the quality of primary care diagnostic test use, can use our results as a resource to identify tests in their healthcare setting that require improvement and/or investigation to decipher why such deviation from guidelines exists. Our conclusions from this paper, however, are not generalisble to all primary care settings nor all primary care diagnostic tests.

#### Conclusion

There is marked variation in under and overuse of appropriate diagnostic test use in primary care across the world. From the available data, echocardiograms are ordered particularly poorly, while the substantial variation in appropriate ordering of pulmonary function tests, colonoscopy, upper endoscopy, and urinary cultures suggest a need for improvement.

# Funding:

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

All author declare no conflicts of interests.

Ethical approval: Not required

**Data sharing:** Data extracted from the included studies in this review are available on request from the corresponding author.

**Registration:** PROSPERO protocol Registration ID: CRD42016048832 (https://www.crd.york.ac.uk/prospero/display\_record.asp?ID=CRD42016048832)

# Competing interest statement.

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and jointly declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

# **Contribution statement:**

Conception and design: Jack O'Sullivan, Rafael Perera and Carl Heneghan

Screening, extraction and risk of bias: Jack O'Sullivan, Ali Albasri and Brian Nicholson.

Analysis and interpretation of the data: Jack O'Sullivan, Rafael Perera, Jeffrey Aronson and Carl Heneghan.

Drafting of the article: Jack O'Sullivan (all authors critically reviewed and approved manuscript)

Statistical expertise: Rafael Perera

Clinical expertise: Jack O'Sullivan, Brian Nicholson, Jeffrey Aronson and Carl Heneghan

Jack O'Sullivan is the guarantor.

# **Copyright Statement**

The corresponding author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

# References

1 Foot C, Naylor C, Imison C. The quality of GP diagnosis and referral. 2010. http://amapro.isabelhealthcare.com/pdf/Kings\_Fund\_Diagnosis\_and\_Referral\_2010.pdf

- For peer review only http://bmjopen.bmj.com/site/about/guidelines.xhtml

## **BMJ** Open

Koch H, van Bokhoven MA, ter Riet G, et al. Ordering blood tests for patients with unexplained fatigue in general practice: what does it yield? Results of the VAMPIRE trial. Br J Gen Pract 2009;59:e93-100. doi:10.3399/bjgp09X420310 Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. BMJ 2009;338. Hobbs FDR, Bankhead C, Mukhtar T, et al. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. Lancet 2016;387:2323-30. doi:10.1016/S0140-6736(16)00620-6 Centers for Disease Control and Prevention, National Center for Health Statistics. National Ambulatory Medical Care Survey: 2012 Summary Tables. 2012;:5.http://www.cdc.gov/nchs/data/ahcd/namcs summary/2010 namcs web tables.pdf Alderwick H, Robertson R, Appleby J, et al. Better value in the NHS The role of changes in clinical practice. 2015. Fisher ES, Bynum JP, Skinner JS. Slowing the growth of health care costs--lessons from regional variation. N Engl J Med 2009:360:849-52. doi:10.1056/NEJMp0809794 Appleby J, Thompson J, Jabbal J. Quarterly Monitoring Report: How is the NHS performing? King's Fund 2016;:1-42. Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. BMJ Qual Saf 2013;22 Suppl 2:ii6-ii10. doi:10.1136/bmjqs-2012-001621 Gandhi TK, Kachalia A, Thomas EJ, et al. Annals of Internal Medicine Article Missed and Delayed Diagnoses in the Ambulatory Setting : Ann Intern Med 2006;145:488-96. Katzberg RW, Lamba R. Contrast-induced nephropathy after intravenous administration: fact or fiction? Radiol Clin North Am 2009;47:789-800, v. doi:10.1016/j.rcl.2009.06.002\rS0033-8389(09)00094-3 [pii] Lumbreras B, Donat L, Hernández-Aguado I. Incidental findings in imaging diagnostic tests: a systematic review. Br J Radiol 2010;83:276-89. doi:10.1259/bjr/98067945 Welch, H. Gilbert, Schwartz, Lisa, Woloshin S. Overdiagnosed: Making people sick in the pursuit of health. Beacon Press, 2011 2011. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *Bmj* 2012;**344**:e3502–e3502. doi:10.1136/bmj.e3502 Berwick D, Hackbarth AD. Eliminating Waste in US Health Care. JAMA 2012;307:1513. doi:10.1001/jama.2012.362 Cecchini M, Lee S. Tackling Wasteful Spending on Healthcare. 2017. http://networks.sustainablehealthcare.org.uk/sites/default/files/resources/Tackling Wasteful Spending on Health.pdf#page=117 Health D of. NHS 2010–2015: from good to great. preventative, people-centred, productive. London: 2009. Esmail A, Neale G, Elstein M, Firth-Cozens J, Davy C VC. Case Studies in Litigation: Claims reviews in four specialties. Manchester: 2004. Sirovich BE, Woloshin S, Schwartz LM. Too Little? Too Much? Primary care physicians' views on US health care: a brief report. Arch Intern Med 2011;171:1582-5. doi:10.1001/archinternmed.2011.437 For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

#### **BMJ** Open

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.	Erasmushogeschool.	3MJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA
--	--------------------	---

20	Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day:
21	How will we ever keep up? <i>PLoS Med</i> 2010; <b>7</b> . doi:10.1371/journal.pmed.1000326 Garber AM. Evidence-based guidelines as a foundation for performance incentives. <i>Health Aff</i> ( <i>Millwood</i> ) 2005; <b>24</b> :174–9. doi:10.1377/hlthaff.24.1.174
22	Ransohoff DF, Pignone M, Sox HC, <i>et al.</i> How to Decide Whether a Clinical Practice Guideline Is Trustworthy. <i>JAMA</i> 2013; <b>309</b> :139. doi:10.1001/jama.2012.156703
23	Fryar C. Doctors can depart from guidelines in patients' best interests. <i>BMJ</i> 2015; <b>350</b> .
24	Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. <i>Lancet (London, England)</i> 1993; <b>342</b> :1317–22.http://www.ncbi.nlm.nih.gov/pubmed/7901634 (accessed 31 Aug 2016).
25	Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. <i>JAMA</i> 1999; <b>281</b> :1900–5.http://www.ncbi.nlm.nih.gov/pubmed/10349893 (accessed 7 Dec 2016).
26	Grilli R, Magrini N, Penna A, <i>et al.</i> Practice guidelines developed by specialty societies: the need for a critical appraisal. <i>Lancet (London, England)</i> 2000; <b>355</b> :103–6. doi:10.1016/S0140-6736(99)02171-6
27	Lenzer J. Why we can't trust clinical guidelines. BMJ 2013;346.
28	Spyridonidis D, Calnan M. Opening the black box: A study of the process of NICE guidelines implementation. <i>Health Policy (New York)</i> 2011; <b>102</b> :117–25. doi:10.1016/j.healthpol.2011.06.011
29	Zhi M, Ding EL, Theisen-Toupal J, <i>et al.</i> The Landscape of Inappropriate Laboratory Testing: A 15-Year Meta-Analysis. <i>PLoS One</i> 2013; <b>8</b> :e78962. doi:10.1371/journal.pone.0078962
80	McGlynn E, Asch S, Adams J, <i>et al.</i> Quality of health care delivered to adults in the United States. <i>N Engl J Med</i> 2003; <b>349</b> :1866-1868-1868. doi:10.1056/NEJMsa022615
1	Sheldon T a, Cullum N, Dawson D, <i>et al.</i> What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. <i>BMJ</i> 2004; <b>329</b> :999. doi:10.1136/bmj.329.7473.999
32	National Health Service. NHS Imaging and Radiodiagnostic activity in England. 2013;:1–7.http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/KH12-release-2012-13.pdf
33	Moher D, Liberati A, Tetzlaff J, <i>et al.</i> Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. <i>BMJ</i> 2009; <b>339</b> :b2535.http://www.ncbi.nlm.nih.gov/pubmed/19622551 (accessed 22 Aug 2016).
34	Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of Observational Studies in Epidemiology. JAMA 2000;283:2008. doi:10.1001/jama.283.15.2008
35	Wald NJ. Guidance on terminology. J Med Screen 2008;15:50–50. doi:10.1258/jms.2008.008got
36	Raffle A, Gray J. Screening: Evidence and Practice. Oxford University Press 2007.
37	Glasziou P, Irwig L, Aronson J. Evidence-based medical monitoring: from principles to practice. Oxford (UK): Blackwell Publishing, BMJ books 2008.
38	Hoy D, Brooks P, Woolf A, <i>et al.</i> Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. <i>J Clin Epidemiol</i> 2012; <b>65</b> :934–9.
13	
	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**BMJ** Open

doi:10.1016/j.jclinepi.2011.11.014

- 39 Belletti D, Liu J, Zacker C, *et al.* Results of the CAPPS: COPD--assessment of practice in primary care study. *Curr Med Res Opin* 2013;**29**:957–66. doi:10.1185/03007995.2013.803957
- 40 Bertella E, Zadra A, Vitacca M, *et al.* COPD management in primary care: is an educational plan for GPs useful? *Multidiscip Respir Med* 2013;**8**:24. doi:10.1186/2049-6958-8-24
- 41 Chavez PC, Shokar NK. Diagnosis and management of chronic obstructive pulmonary disease (COPD) in a primary care clinic. *COPD* 2009;**6**:446–51. doi:10.3109/15412550903341455
- 42 Lange P, Rasmussen FV, Borgeskov H, *et al.* The quality of COPD care in general practice in Denmark: the KVASIMODO study. *Prim Care Respir J* 2007;**16**:174–81. doi:10.3132/pcrj.2007.00030
- 43 Ulrik CS, Sørensen TB, Højmark TB, *et al.* Adherence to COPD guidelines in general practice: impact of an educational programme delivered on location in Danish general practices. *Prim Care Respir J* 2013;**22**:23–8. doi:10.4104/pcrj.2012.00089
- 44 Barendregt JJ, Doi SA, Lee YY, *et al.* Meta-analysis of prevalence. *J Epidemiol Community Heal* 2013;:974–8. doi:10.1136/jech-2013-203104
- Doi SAR, Barendregt JJ, Khan S, *et al.* Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. *Contemp Clin Trials* 2015;45:130–8. doi:10.1016/j.cct.2015.05.009
- 46 Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60. doi:10.1136/bmj.327.7414.557
- 47 Mafi JN, McCarthy EP, Davis RB, *et al.* Worsening trends in the management and treatment of back pain. *JAMA Intern Med* 2013;**173**:1573–81. doi:10.1001/jamainternmed.2013.8992
- 48 Mafi JN, Edwards ST, Pedersen NP, et al. Trends in the Ambulatory Management of Headache: Analysis of NAMCS and NHAMCS Data 1999-2010. J Gen Intern Med 2015;30:548–55. doi:10.1007/s11606-014-3107-3
- 49 Williams CM, Maher CG, Hancock MJ, et al. Low back pain and best practice care: A survey of general practice physicians. Arch Intern Med 2010;170:271–7. doi:10.1001/archinternmed.2009.507
- 50 Cai JX, Campbell EJ, Richter JM. Concordance of Outpatient Esophagogastroduodenoscopy of the Upper Gastrointestinal Tract With Evidence-Based Guidelines. JAMA Intern Med 2015;175:1563–4. doi:10.1001/jamainternmed.2015.3533
- 51 Gurzun M-M, Ionescu A. Appropriateness of use criteria for transthoracic echocardiography: are they relevant outside the USA? *Eur Hear J - Cardiovasc Imaging* 2014;**15**:450–5. doi:10.1093/ehjci/jet186
- 52 van Gurp N, Boonman-De winter LJM, Meijer Timmerman Thijssen DW, et al. Benefits of an open access echocardiography service: A Dutch prospective cohort study. Netherlands Hear J 2013;21:399–405. doi:10.1007/s12471-013-0416-9
- 53 Johnson JD, O'Mara HM, Durtschi HF, et al. Do Urine Cultures for Urinary Tract Infections Decrease Follow-up Visits? J Am Board Fam Med 2011;24:647–55. doi:10.3122/jabfm.2011.06.100299
- 54 Grover ML, Bracamonte JD, Kanodia AK, *et al.* Assessing Adherence to Evidence-Based Guidelines for the Diagnosis and Management of Uncomplicated Urinary Tract Infection. *Mayo Clin Proc* 2007;**82**:181–5. doi:10.4065/82.2.181
- 55 Llor C, Rabanaque G, Lopez A, et al. The adherence of GPs to guidelines for the diagnosis

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

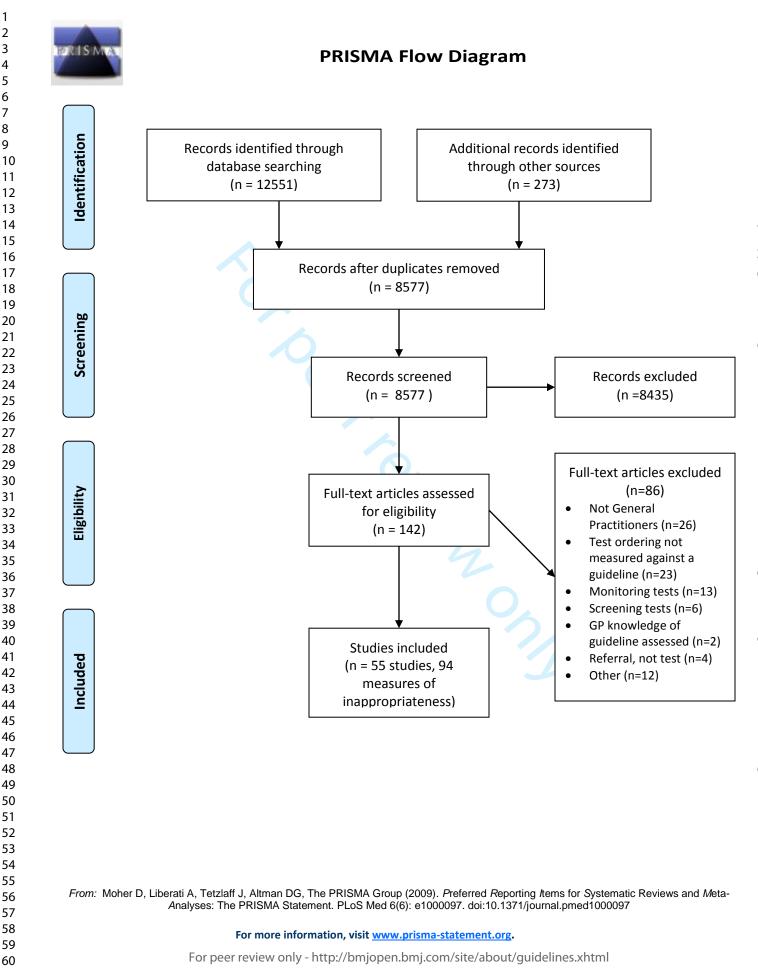
# BMJ Open

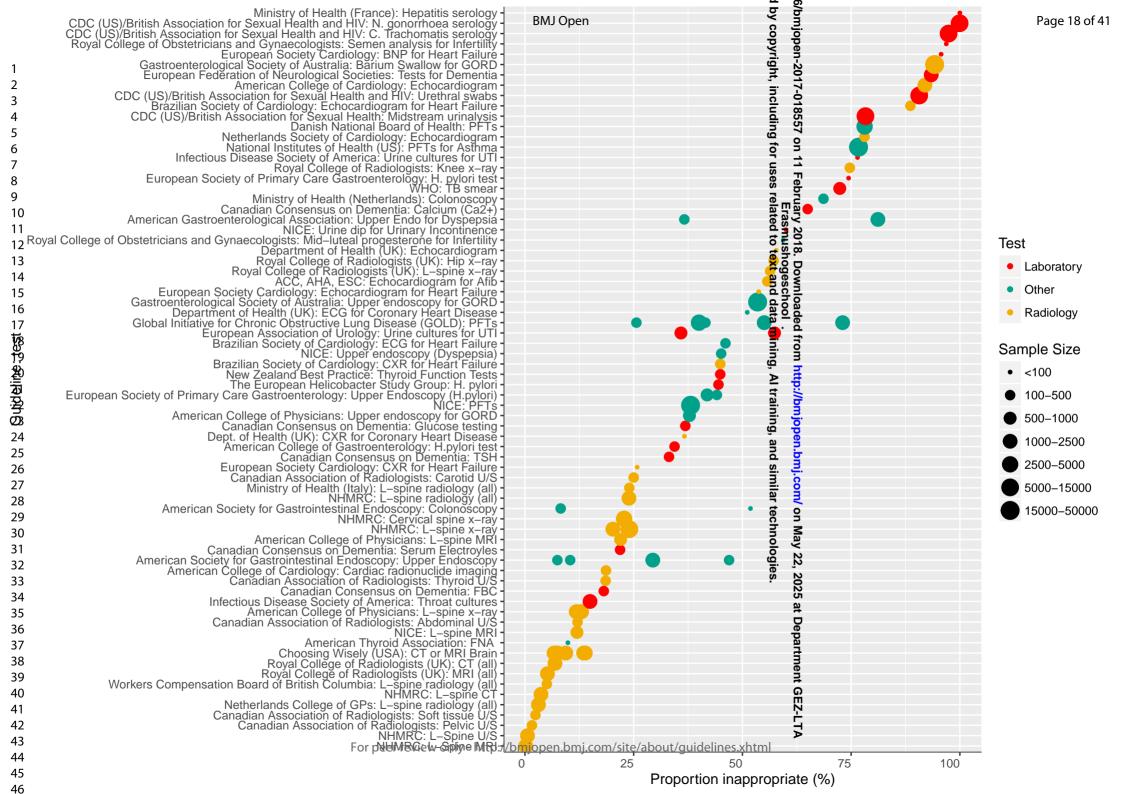
	and treatment of lower urinary tract infections in women is poor. <i>Fam Pract</i> 2011; <b>28</b> :294–9. doi:10.1093/fampra/cmq107
56	Leon P, Catherine K, Mark N, <i>et al.</i> Gastro-oseophageal reflux disease. The impact of guidelines on GP management. 2008.
57	Hughes-Anderson W, Rankin SL, House J, <i>et al.</i> Open access endoscopy in rural and remote Western Australia: does it work? <i>ANZ J Surg</i> 2002; <b>72</b> :699–703.http://www.ncbi.nlm.nih.gov/pubmed/12534377 (accessed 7 Dec 2016).
58	Aljebreen AM, Alswat K, Almadi MA. Appropriateness and diagnostic yield of upper gastrointestinal endoscopy in an open-access endoscopy system. <i>Saudi J Gastroenterol</i> 2013; <b>19</b> :219–22. doi:10.4103/1319-3767.118128
59	Azzam NA, Almadi MA, Alamar HH, <i>et al.</i> Performance of American Society for Gastrointestinal Endoscopy guidelines for dyspepsia in Saudi population: Prospective observational study. <i>World J Gastroenterol</i> 2015; <b>21</b> :637–43. doi:10.3748/wjg.v21.i2.637
60	Elwyn G, Owen D, Roberts L, <i>et al.</i> Influencing referral practice using feedback of adherence to NICE guidelines: a quality improvement report for dyspepsia. <i>Qual Saf Health Care</i> 2007; <b>16</b> :67–70. doi:10.1136/qshc.2006.019992
61	Cardin F, Zorzi M, Bovo E, <i>et al.</i> Effect of Implementation of a Dyspepsia and Helicobacter pylori Eradication Guideline in Primary Care. <i>Digestion</i> 2005; <b>72</b> :1–7. doi:10.1159/000087215
62	Cardin F, Zorzi M, Terranova O. Implementation of a guideline versus use of individual prognostic factors to prioritize waiting lists for upper gastrointestinal endoscopy. <i>Eur J Gastroenterol Hepatol</i> 2007; <b>19</b> :549–53. doi:10.1097/01.meg.0000216942.42306.d5
63	Hassan C, Bersani G, Buri L, <i>et al.</i> Appropriateness of upper-GI endoscopy: an Italian survey on behalf of the Italian Society of Digestive Endoscopy. <i>Gastrointest Endosc</i> 2007; <b>65</b> :767–74. doi:10.1016/j.gie.2006.12.058
64	Chan Y-M, Goh K-L. Appropriateness and diagnostic yield of EGD: a prospective study in a large Asian hospital. <i>Gastrointest Endosc</i> 2004; <b>59</b> :517–24. doi:10.1016/S0016-5107(04)00002-1
65	Eccles M, Steen N, Grimshaw J, <i>et al.</i> Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. <i>Lancet</i> 2001; <b>357</b> :1406–9. doi:10.1016/S0140-6736(00)04564-5
66	Majumdar SR, Soumerai SB, Farraye FA, <i>et al.</i> Chronic acid-related disorders are common and underinvestigated. <i>Am J Gastroenterol</i> 2003; <b>98</b> :2409–14. doi:10.1111/j.1572-0241.2003.07706.x
67	Basu S, Andrews J, Kishore S, <i>et al.</i> Comparative performance of private and public healthcare systems in low- and middle-income countries: A systematic review. <i>PLoS Med</i> 2012; <b>9</b> :19. doi:10.1371/journal.pmed.1001244
68	Ridic G, Gleason S, Ridic O. Comparisons of Health Care Systems in the United States, Germany and Canada. <i>Mat Soc Med</i> 2012; <b>24</b> :112–20. doi:10.5455/msm.2012.24.112-120.Comparisons
69	Gagliardi AR, Brouwers MC. Do guidelines offer implementation advice to target users? A systematic review of guideline applicability. <i>BMJ Open</i> 2015; <b>5</b> :e007047–e007047. doi:10.1136/bmjopen-2014-007047
70	Morgan DJ, Brownlee S, Leppin AL, <i>et al.</i> Setting a research agenda for medical overuse. <i>Bmj</i> 2015; <b>4534</b> :h4534. doi:10.1136/bmj.h4534
71	Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. Health
15	
	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

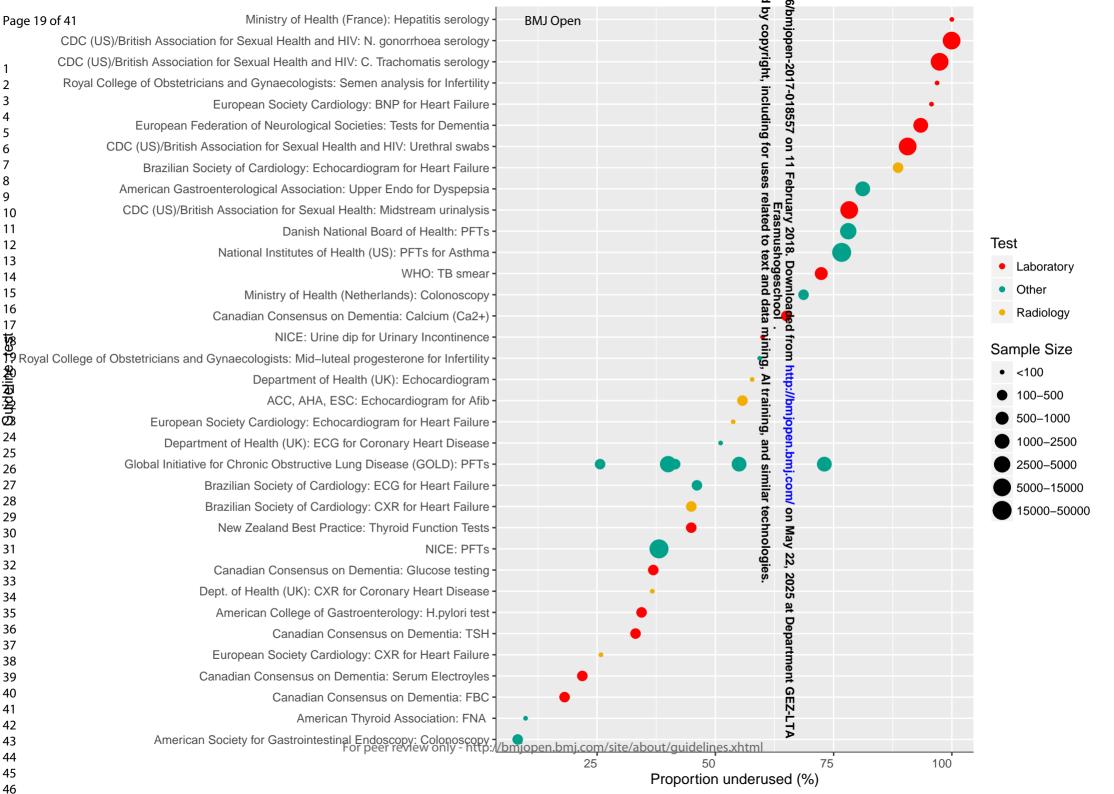
## **BMJ** Open

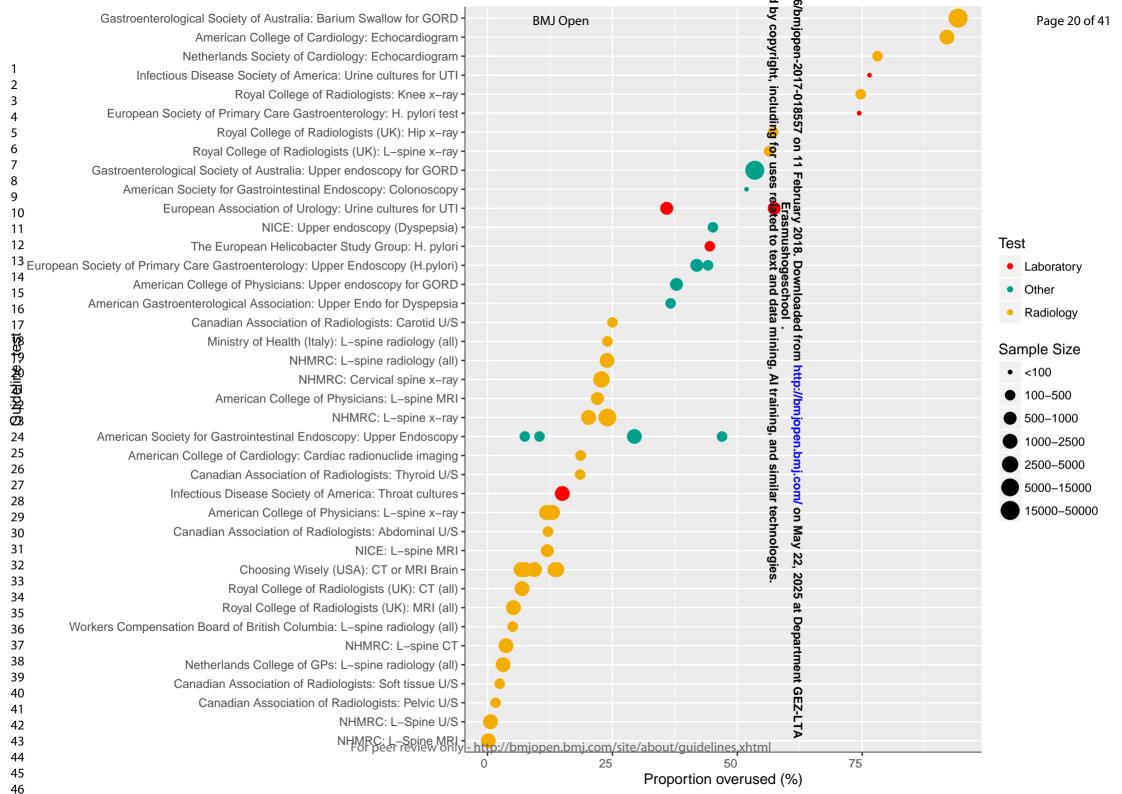
*Aff (Millwood)*;:W96-114.http://www.ncbi.nlm.nih.gov/pubmed/12703563 (accessed 29 Sep 2016).

- 72 Chassin MR. Is health care ready for Six Sigma quality? *Milbank Q* 1998;**76**:565–91, 510. doi:10.1111/1468-0009.00106
- Smith R. Where is the Wisdom...? The Poverty of Medical Evidence. *Br Med J* 1991;303:798–9.http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1671173/pdf/bmj00147-0006.pdf%5Cnfiles/2118/Smith-1991-Where is the Wisdom...\_ The Poverty.pdf
- 74 Schoen C, Osborn R, Doty MM, et al. Toward higher-performance health systems: Adults' health care experiences in seven countries, 2007. *Health Aff* 2007;26. doi:10.1377/hlthaff.26.6.w717
- 75 Kachalia A, Berg A, Fagerlin A, *et al.* Overuse of testing in preoperative evaluation and syncope: a survey of hospitalists. *Ann Intern Med* 2015;**162**:100–8. doi:10.7326/M14-0694
- 76 Swennen MHJ, Rutten FH, Kalkman CJ, et al. Do general practitioners follow treatment recommendations from guidelines in their decisions on heart failure management? A crosssectional study. BMJ Open 2013;3:e002982. doi:10.1136/bmjopen-2013-002982
- Parker L, Levin DC, Frangos A, *et al.* Geographic variation in the utilization of noninvasive diagnostic imaging: national medicare data, 1998-2007. *AJR Am J Roentgenol* 2010;194:1034–9. doi:10.2214/AJR.09.3528
- 78 Song Y, Skinner J, Bynum J, et al. Regional Variations in Diagnostic Practices. N Engl J Med 2010;363:45–53. doi:10.1056/NEJMsa0910881
- 79 Burgers JS, Fervers B, Haugh M, et al. International Assessment of the Quality of Clinical Practice Guidelines in Oncology Using the Appraisal of Guidelines and Research and Evaluation Instrument. J Clin Oncol 2004;22:2000–7. doi:10.1200/JCO.2004.06.157
- 80 Gale EAM. Conflicts of interest in guideline panel members. *BMJ* 2011;**343**.
- 81 IoM C to A the PHS on CPG. Clinical Practice Guidelines: Directions for a New Program. Washington: 1990. doi:10.1097/SPV.0b013e31828a2951
- 82 Browman GP, Snider A, Ellis P. Negotiating for change. The healthcare manager as catalyst for evidence-based practice: changing the healthcare environment and sharing experience. *Healthc Pap* 2003;**3**:10–22.http://www.ncbi.nlm.nih.gov/pubmed/12811083 (accessed 7 Nov 2016).









1 2 3 4 5 6	
7 8 9	Stuc
10 11	Ahmad 20
12 13	Belletti 20
14	Bertella 2
15	Caplan 20
16 17	Chavez 2
18	Droogend 2011
19 20	Gerrits 20
20	Gibbons 2
22	Gnani 200
23 24	Girard 20
25	Hughes-
26	Anderson Lange 20
27	Lipczynsk
28 29	2012
30	Loo 2009
31	Majumdar
32	Moscavito
33 34	2009 Musicco 2
35	
36	Nicholson
37	NICHOISON
38 39	
40	
41	
42 43	
43 44	
45	
46 47	ATJ-Z3Ð friðmt

# Table 1: Study Characteristics

Study	Country	Study length (days)	N (men%)	Population	Test
				Under-use	
Ahmad 2012	Indonesia	181	554 (41%)	Patients registered at health clinics where TB was suspected	Sputum smear microscopy
Belletti 2013	USA	N/S	1517 (46%)	Patients with COPD	Pulmonary function tests (PFT)
Bertella 2013	Italy	1765	437 (286)	Patients with COPD	PFTs
Caplan 2000	USA	365	81	Patients who under went FNA of thyroid	FNA of thyroid
Chavez 2009	USA	2920	200 (48%)	Patients with COPD	PFT
Droogendijk 2011	Netherlands	730	287 (45%)	Women >50yrs and men >18 years with Iron Deficiency Anaemia	Upper endoscopy and colonoscopy
Gerrits 2008	Netherlands	2556	65 (0%)	Women aged 18 – 65 yrs with newly diagnosed urinary incontinence	Urine dipstick
Gibbons 2010	New Zealand	364	265	Patients with subclinical hypothyroidism	Free T4
Gnani 2004	UK	365	90 (53%)	Patients with heart failure	CXR, ECG and Echocardiogram
Girard 2010	France	28	19 (37%)	Patients with acute hepatitis	Hepatitis serology (HBs antigens, anti-HBc anitbodies)
Hughes- Anderson 2002a	Australia	1613	4400 (55%)	Patients who had colonoscopy	Colonoscopy
Lange 2007	Denmark	91	2549 (44%)	Patients with COPD	PFTs
Lipczynska 2012	Poland	61	93	Aged ≥ 55 with Heart Failure (HF) or HF risk factors	Echocardiogram, BNP, CXR
Loo 2009	UK	364	131 (50%)	Patients with Atrial Fibrillation	Echocardiogram
Majumdar 2003	USA	2371	1130 (47%)	Patients with dyspepsia and patients with peptic ulcer disease (PUD)	Upper endoscopy and H.pylori
Moscavitch 2009	Brazil	61	167 (43%)	Patients with Heart Failure	ECG, CXR, Echocardiogram
Musicco 2004	Italy	N/S	1549 (38%)	Patients being assessed for Dementia	Collection of laboratory tests to rule out conditions with similar presenting symptoms to dementia
Nicholson 2010	UK	1827	6943 (100%)	Men with epididymo-orchtitis	C. trachomatis, N. gonorrhoeae, urethral swabs and midstream urinalysis.

Nicopoullos 2003	UK	242	32	Patients with subfertility	Mid-luteal progesterone and semen analysis
Pimlott 2006	Canada	1611	160 (34%)	Patients with Dementia	FBC, TSH, serum electrolytes, serum calcium, glucose
Smith 2008	UK	731	29870 (52%)	Patients with COPD	PFT
Sokol 2015	USA	3652	75902 (23%)	Patients with Asthma	PFT
Ulrik 2010	Denmark	121	1716 (44%)	Patients with COPD	PFT
Ulrik 2013	Denmark	731	4058	Patients with COPD	PFT
	•	•	6	<u>Over-use</u>	•
Aljebreen 2013	Saudi Arabia	365	147 (51%)	Patients who had upper endoscopy	Upper endoscopy
Azzam 2015	Saudi Arabia	121	161 (30%)	Dyspeptic patients who had upper endoscopy	Upper endoscopy
Bhatt 2001	UK	504	437 (65%)	Patients referred for pelvis x-rays	Pelvis x-ray
Bishop 2003	Canada	28	139	Patients with non-red flag LBP	Advanced imaging (CT, MRI or bone scan)
Cai 2015	USA	121	550 (46%)	Patients who under went upper endoscopy	Upper endoscopy
Chan 2004	Malaysia	153	250 (45%)	Patients who under went upper endoscopy	Upper endoscopy
Chan 2006	Malaysia	184	27 (63%)	Patients who underwent 'diagnostic colonoscopies'	Colonoscopy
Cardin 2005	Italy	151	1678	Patients with uninvestigated dyspepsia	Upper endoscopy
Cardin 2007	Italy	182	1/04/2001	Dyspeptic patients who had upper endoscopy	Upper endoscopy
Eccles 2001	UK	182	275	Patients who had knee or lumbar x-ray	Lumbar or knee x-ray
Elwyn 2007	UK	184	215	Patients who under went upper endoscopy	Upper endoscopy
Grover 2007	USA	364	68 (0%)	Patients with uncomplicated UTI	Urine culture and sensitivity analysis
Gurzun 2014	UK	7	1070 (54%)	Echocardiogram	Echocardiogram
Hassan 2007	Italy	30	3769 (46%)	Patients who under went upper endoscopy	Upper endoscopy
Hughes- Anderson 2002b	Australia	1613	4400 (55%)	Patients who had upper endoscopy,	Upper endoscopy
lp 2014	USA	1096	100 (43%)	Patients with non-red flag LBP	MRI lumbar spine

Johnson 2011	USA	510	779 (0%)	Patients with uncomplicated UTI	Urine culture
Kovacs 2013	Spain	183	602 (48%)	Patients with non-red flag LBP	MRI lumbar spine
Lalude 2014	USA	121	102	Patients who had SPECT Myocardial perfusion imaging (MPI) studies	Single Photon Emission CT (SPECT) MPI
Landry 2011	USA	272	124	Patients who had U/S of thyroid, pelvis, abdo, carotid or soft tissue	Thyroid, pelvis, abdomen, caroti or soft tissue ultrasound
Linder 2006	USA	608	1076 (19%)	Patients with pharyngitis	Strep testing (rapid antigen detection test, throat culture)
Llor 2011	Spain	122	658 (0%)	Women with UTI	Urine cultures
Mafi 2013	USA	4377	8066	Patients with non-red flag LBP	X-ray, CT or MRI
Mafi 2015	USA	4018	9362 (25%)	Patients with uncomplicated headache (non-red flag	CT and MRI
Michaleff 2012	Australia	3621	3070 (70%)	Patients reporting first time neck pain or LBP (non-specific, non red flag)	Any radiological test
Noya 2008	Israel	N/S	209 (35%)	Patients who had H.pylori testing	H. pylori test
Piccoliori 2013	Italy	63	475 (43%)	Acute or chronic non-red flag LBP	Any radiological test
Piterman 2008	Australia	550	19219	Patients with GORD	Endoscopy. Barium Swallow
Remedios 2014	UK	N/S	2026	Patients who had CTs and/or MRIs	CT and/or MRI
Schers 2000	Netherlands	214	1096 (50%)	Patients with non-red flag LBP	X-ray
Van Gurp 2013	Netherlands	366	155 (38%)	Patients who had Echocardiogram	Echocardiogram
Williams 2010	Australia	1005	1706 (43%)	Patients with non-red flag LBP	All imaging
					1

AT-LZA Downloaded from http://bmjopen.bmj.com/ on May 22, 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool Protected by copyright, instructory felpted for text and idata, mitring, but training, and similar technologies.

Table 2: Measures of inapp	propriateness
----------------------------	---------------

Study	Test	Guideline authority and recommendation	Measure of inappropriateness (95%Cl)
		<u>Under-use</u>	
Girard 2010	0 Hepatitis B serology Ministry of Health (France): Hepatitis serology for suspected acute hepatitis		100% (82.4 to 100%)
Nicholson 2010	Neisseria Gonorrhoea serology	CDC (US)/British Association for Sexual Health and HIV: Test for N. gonorrhoea for suspected Epididymitis	99.9% (99.85 to 99.98%)
Nicholson 2010	Chlamydia Trachomatis	CDC (US)/British Association for Sexual Health and HIV: Test for C. Trachomatis for suspected Epididymitis	97.4% (97.0 to 97.8%)
Nicopoullos 2003	Semen Analysis	Royal College of Obstetricians and Gynaecologists: Semen analysis for Infertility	96.9% (95%CI: 83.8 to 99.9%)
Lipczynska 2012	Brain Natriuretic Peptide (BNP)	European Society Cardiology: BNP for Heart Failure	95.7% (95%CI: 89.4 to 98.8%)
Musicco 2004	Collection of laboratory tests	European Federation of Neurological Societies: Collection of laboratory tests to rule out conditions with similar presenting symptoms to dementia	93.42% (95%Cl: 92.1 to 94.6%)
Nicholson 2010	Urethral swabs	CDC (US)/British Association for Sexual Health and HIV: Urethral swabs for suspected epididymitis (Urethral swabs)	90.7% (95%CI:89.9 to 91.3%)
Moscavitch 2009	Echocardiogram	Brazilian Society of Cardiology: Échocardiography for Heart Failure	88.6% (95%CI: 82.8 to 93.0%)
Majumdar 2003	Upper Endoscopy	American Gastroenterological Association: Appropriate use of Upper Endoscopy for Dyspepsia	81.2% (78.8 to 83.4%)
Nicholson 2010	Mid stream	CDC (US)/British Association for Sexual Health and HIV: Midstream urinalysis for suspected Epididymitis	78.2 (95%CI: 77.3 to 79.3%)
Ulrik 2013	Pulmonary function tests (PFTs)	Danish National Board of Health: PFTs to diagnosis COPD	78.1 (76.8% to 79.4%)
Sokol 2015	Pulmonary function tests (PFTs)	National Asthma Education and Prevention Program (US): PFTs for asthma	76.5% (64.6 to 85.9%)
Belletti 2013	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	73.0% (10.7 to 75.3%)
Ahmad 2012	Tuberculosis smear	World Health Organisation: Smear for suspected TB	72.4% (68.5 to 76.1%)

BMJ Open

Droogendijk 2011	Colonoscopy	Ministry of Health (Netherlands): Colonoscopy for unexplained Iron Deficiency Anaemia	68.6% (62.9 to 74
Pimlott 2006	Serum Calcium	Canadian Consensus Conference on Dementia: Serum Calcium for Dementia	65.0 (57.1 to 72
Gerrits 2008	Urine dip stick	NICE: Urine dip stick for urinary incontinence	60.0% (47.1 to 7
Nicopoullos 2003	Mid-luteal progesterone	Royal College of Obstetricians and Gynaecologists: Mid-luteal progesterone for Infertility	59.4% (40.6 to 7
Gnani 2004	Echocardiogram	Department of Health (UK): Echocardiogram for Heart Failure	57.8% (46.1 to 68
Loo 2009	Echocardiogram	ACC, AHA, ESC: Echocardiogram to identify causes or complications of atrial fibrillation	55.7% (46.8 to 64
Ulrik 2010	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	55.0% (52.6 to 57
Lipczynska 2012	Echocardiogram	European Society Cardiology: Echocardiogram for Heart Failure	53.8% (43.1 to 64
Gnani 2004	ECG	Department of Health (UK): ECG for Heart Failure	51.1% (40.4% to 6
Moscavitch 2009	ECG	Brazilian Society of Cardiology: ECG for Heart Failure	46.1 (38.4 to 54
Moscavitch 2009	Chest X-ray	Brazilian Society of Cardiology: CXR for Heart Failure	44.9% (37.2 to 52
Gibbons 2010	Thyroid function tests	New Zealand Best Practice: Appropriate use of Thyroid Function tests	44.9% (38.8 to 5 <sup>2</sup>
Chavez 2009	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	41.5% (34.6 to 48
Lange 2007	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	40.0% (38.1 to 4
Smith 2008	Pulmonary Function Tests (PFTs)	NICE: PFTs for COPD	38.1% (37.5 to 38
Pimlott 2006	Glucose testing	Canadian Consensus Conference on Dementia: Glucose testing for Dementia	36.9% (29.4% to 4
Gnani 2004	Chest X-ray	Department of Health (UK): CXR for Heart Failure	36.7% (26.8 to 47
Majumdar 2003	H.pylori	American Gastroenterological Association/American College of Gastroenterology: appropriateness of H.pylori test	34.4% (28.9 to 40
Pimlott 2006	Thyroid Stimulating Hormone (TSH)	Canadian Consensus Conference on Dementia: TSH for dementia	33.1% (25.9 to 4

AT-LZA Downloaded from http://bmjopen.bmj.com/ on May 22, 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool Protected by copyright, instructory felpted for text and idata, mitring, but training, and similar technologies.

Lipczynska 2012	Chest x-ray (CXR)	European Society Cardiology: CXR for Heart Failure	25.8% (17.3 to 35.9%)
Bertella 2013	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	25.6% (21.6 to 30.0%)
Pimlott 2006	Serum electrolytes	Canadian Consensus Conference on Dementia: Serum electrolytes for dementia	21.9% (15.7 to 29.1%)
Pimlott 2006	Full Blood Count (FBC)	Canadian Consensus Conference on Dementia: FBC for dementia	18.1% (12.5 to 25.0%)
Caplan 2000	Fine needle aspiration (FNA) of thyroid	American Thyroid Association/American Association of Clinical Endocrinologists: FNA for thyroid nodules	9.9% (4.4 to 18.5%)
Hughes-Anderson 2002a	Colonoscopy	American Society for Gastrointestinal Endoscopy: Appropriateness of Colonoscopy	8.2% (5.3 to 12.1%)
		<u>Over-use</u>	
Piterman 2008	Barium Swallow	Gastroenterological Society of Australia: Barium Swallow for GORD	94.20% (95%Cl: 93.9 to 94.5%)
Gurzun 2014	Echocardiogram	American College of Cardiology: Appropriate use of Echocardiography	92.0% (95%Cl: 90.2% to 93.5%)
van Gurp 2013	Echocardiogram	Netherlands Society of Cardiology: Appropriate use of Echocardiography	76.7% (76.4 to 77.0%)
Grover 2007	Urine cultures	Infectious Disease Society of America: Urine cultures not required for uncomplicated UTI diagnosis	76.5% (64.6 to 85.9%)
Eccles 2001	Knee x-ray	Royal College of Radiologists: No x-ray for knee pain without restriction of movement	74.7% (69.6 to 79.3%)
Cardin 2005	H. Pylori breath test	European Society of Primary Care Gastroenterology: Appropriate use of H. pylori	74.4% (58.8 to 86.5%)
Johnston 2011	Urine cultures	European Association of Urology: Urinary cultures not required for uncomplicated urinary tract infections	57.4% (53.8 to 60.9%)
Bhatt 2001	Hip x-ray	Royal College of Radiologists (UK): No hip x-ray for hip pain without restriction of movement	57.2% (52.5 to 61.8%)
Eccles 2001	Lumbar spine x-ray	Royal College of Radiologists (UK): no x-ray for non- red flag LBP	56.4% (50.3 to 62.3%)
Piterman 2008	Upper endoscopy	Gastroenterological Society of Australia: Upper endoscopy for GORD	53.5% (52.8 to 54.2%)
Chan 2006	Colonoscopy	American Society for Gastrointestinal Endoscopy: Appropriateness of Colonoscopy	51.9% (32.0 to 71.3%)

АТЈ

 BMJ Open

Aljebreen 2013	Upper endoscopy	The American Society for Gastrointestinal Endoscopy: Appropriateness of Upper Endoscopy	46.9% (38.7 to 55.39
Elwyn 2007	Upper endoscopy	NICE: Appropriate tests for dyspepsia	45.1% (38.3 to 52.09
Noya 2008	H.Pylori	The European Helicobacter Study Group: Appropriate use of H. pylori	44.5 (37.6 to 51.5%
Cardin 2005	Upper endoscopy	European Society of Primary Care Gastroenterology: Upper Endoscopy for H.pylori	44.1% (35.9 to 52.69
Cardin 2007	Upper endoscopy	European Society of Primary Care Gastroenterology: Upper Endoscopy for H.pylori	41.9% (38.3 to 45.59
Cai 2015	Upper endoscopy	American College of Physicians: Upper endoscopy for GORD	37.7% (33.8 to 42.09
Azzam 2015	Upper endoscopy	American Gastroenterological Association: Upper Endoscopy for Dyspepsia	36.7 (29.2 to 44.6%
Llor 2011	Urine cultures	European Association of Urology: Urinary cultures not required for uncomplicated urinary tract infections	35.9% (32.2 to 40.09
Hassan 2007	Upper endoscopy	The American Society for Gastrointestinal Endoscopy: Appropriateness of Upper Endoscopy	29.4% (28.0 to 30.99
Landry 2011	Carotid ultrasound	Canadian Association of Radiologists 2005 guidelines: Carotid U/S	25.0% (17.7 to 33.6°
Piccoliori 2013	Lumbar spine radiology (all)	Ministry of Health (Italy): No imaging for non-red flag LBP	24.0% (20.2 to 28.19
Michaleff 2012	Lumbar spine x-ray	National Health and Medical Research Council (Australia) (NHMRC): No x-ray for non-red flag LBP	24.0% (22.9 to 25.19
Williams 2010	Lumbar spine radiology (all)	National Health and Medical Research Council (Australia) (NHMRC): No imaging for non-red flag LBP	23.9% (21.9 to 26.09
Michaleff 2012	Cervical spine x-ray	Australian National Health and Medical Research Council: No x-ray for neck pain	22.8% (21.3 to 24.39
lp 2014	Lumbar spine MRI	American College of Physicians/American Pain Society: no MRI for non-red flag LBP	22.0% (14.3 to 31.49
Williams 2010	Lumbar spine x-ray	National Health and Medical Research Council (Australia) (NHMRC): No x-ray for non-red flag LBP	20.2% (18.3 to 22.29
Landry 2011	Thyroid ultrasound	Canadian Association of Radiologists 2005 guidelines: Thyroid U/S	19.0% (12.1 to 27.09
Lalude 2014	Single Photon Emission Computed Tomography	American College of Cardiology: SPECT for chest pain	18.6% (11.6 to 27.69

AT-LZA Townloaded from http://bmjopen.bmj.com/ on العادية المعادية المعاد المعادية الم ا

Linder 2006	Streptococcal throat cultures	American College of Physicians/Infectious Disease Society of America: Pharyngitis	15.0 (12.9 to 17.2%)
'Mafi 2013	Lumbar spine x-ray	American College of Physicians/American Pain Society: no x-ray for non-red flag LBP: 2009-2010	13.0% (11.1 to 15.1%)
		2007-2008	12.9% (11.1 to 14.9%)
		2005-2006	12.8% (11.0 to 14.8%)
		2003-2004	12.3% (10.7 to 14.0%)
		2001-2002	12.0% (10.3 to 13.8%)
		1999 - 2000	11.8% (10.2 to 13.6%)
Landry 2011	Abdominal ultrasound	Canadian Association of Radiologists 2005 guidelines: Abdominal U/S	12.1% (6.9 to 19.2%)
Mafi 2015	CT or MRI Brain	The American Headache Society/American Academy of Neurology for Choosing Wisely: No CT or MRI for non-red flag headache 2009 - 2010	13.9% (12.2 to 15.7%)
		2007 – 2008	13.5% (11.8 to 15.3%)
		2005 – 2006	9.4% (8.0 to 11.0%)
		2003 – 2004	7.5% (6.3 to 8.9%)
		2001 – 2002	7.1% (5.9 to 8.4%)
		1999 - 2000	6.7% (5.4 to 8.2%)
Kovacs 2013	Lumbar spine radiology tests (all)	NICE, ACP: No imaging for LBP	12.0% (9.5 to 14.8%)
Chan 2004	Upper endoscopy	The American Society for Gastrointestinal Endoscopy: Appropriateness of Upper Endoscopy	10.4% (6.9 to 14.9%)
Hughes-Anderson 2002b	Upper endoscopy	The American Society for Gastrointestinal Endoscopy: Appropriateness of Upper Endoscopy	7.5% (4.7 to 11.1%)
	CT (any)	Royal College of Radiologists (UK): CT	6.9% (5.8 to 8.1%)
Remedios 2014 —	MRI (any)	Royal College of Radiologists (UK): MRI	5.2% (4.1 to 6.5%)
Bishop 2003	Lumbar spine radiology tests (all)	Workers Compensation Board of British Columbia: No imaging for non-red flag LBP	5.0% (2.1 to 10.1%)
Williams 2010	Lumbar spine CT	National Health and Medical Research Council (Australia) (NHMRC): No CT for non-red flag LBP	3.7% (2.9 to 4.7%)

Schers 2000	Lumbar spine radiology tests (all)	The Netherlands College of General Practitioners: No imaging for non-red flag LBP	3.1% (2.2 to 4.3%)
Landry 2011	Soft tissue ultrasound	Canadian Association of Radiologists 2005 guidelines: Soft tissue U/S	2.4% (0.5 to 6.9%
Landry 2011	Pelvic ultrasound	Canadian Association of Radiologists 2005 guidelines: Pelvic U/S	1.6% (0.2 to 5.7%)
Williams 2010	Lumbar spine Ultrasound	National Health and Medical Research Council (Australia) (NHMRC): No U/S for non-red flag LBP	0.59% (0.28 to 1.1%)
Williams 2010	Lumbar spine MRI	National Health and Medical Research Council (Australia) (NHMRC): No MRI for non-red flag LBP	0.18% (0.04 to 0.5%)
		(Australia) (NHMRC): No MRI for non-red flag LBP	

Protected by copyright, includes telefed to text and lease set and lease within or the training and alminer technologies.

Ioodosegodesumasia

AT-Last instruction on May 22, 2025 at Department GEZ-LTA from http://manopaded from http://manopaded.instruction.com/ on May 22, 2025 at Department GEZ-LTA

				BMJ Open		6/bmjopen-2017-018557 on J by copyright, including fo			Ρας
	Was the study"s target population a close representation of the national population in relation to relevant variables?	Does the inclusion criteria match the target population of guideline?	Were all eligible participants included in the study?	Was the likelihood of non-response bias <2High%?	Was an acceptable disease, test or symptom definition used?	r uses Was das was das was dat was dat was dat was was was was was was was was was was	Was the interval from symptoms to test clinically appropriate for the diagnosis of interest?	Did they report extractable measures?	Othe
Ahmad2012	Low	Unclear	Low	Unclear	Low		Unclear	Low	L
Aljebreen 2013	Low	Low	Low	Low	Unclear	Unclearer S. Down	Low	Low	F
Azzam 2015	Low	Low	Unclear	High	Low	Unclear	Low	Low	L
Belletti2013	Low	Unclear	Unclear	Unclear	Low	Unclear Low an .	Low	Low	L
Bertella 2013	Low	Low	Low	Low	Unclear	Low Low	Unclear	Low	L
Bhatt 2001	Low	Low	Low	High	High	Unclear 📮	Unclear	Low	Н
Bishop 2003	High	Low	Low	Low	Low	Low <b>P</b>	Low	Low	L
Cai 2015	Low	Low	Unclear	Low	Unclear	Unclear 🔒 💦	Unclear	Low	L
Caplan 2000	Low	Low	Low	Low	Unclear	Unclearing	Unclear	Low	L
Cardin 2005	Low	Low	Low	Low	Low	Low a	Unclear	Low	L
Cardin 2007	Low	Low	Low	Unclear	Low	Low nd b	Low	Low	L
Chan 2004	Low	Low	Low	Low	Low	Low 💁 🛃	Unclear	Low	L
Chan 2006	Low	Low	Low	High	Unclear	Low <b><u>B</u>. S</b>	Unclear	Low	L
Chavez 2009	Low	Low	Low	Low	Low	Low Low	High	Low	н
Droogendijk 2011	Unclear	Low	High	Unclear	Low	n Ma Unclearno	Low	Low	L
Eccles 2001	Low	Low	Low	Low	Unclear	Unclear N	Unclear	Low	L
Elwyn 2007	Low	Low	Unclear	Unclear	Unclear	Low gie 2,	Low	High	Н
Gerrits2008	Low	Unclear	High	Low	Low	Low <b>S</b> 202	Unclear	Low	L
Gibbons 2010c	Low	Low	Low	Low	Low	High at	Low	Low	L
Girard 2010	High	Low	Unclear	High	Unclear	High D Low D	Unclear	Low	Н
Gnani 2004	Low	Low	Unclear	Low	Unclear		Unclear	Low	L
Grover 2007	Low	Low	Unclear	Low	Low	High <b>rtme</b>	Unclear	Low	L

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 31	of 41
---------	-------

# BMJ Open

						6/bmjopen-201 J by copyright,			
						6/bmjopen-2017- J by copyright, in			
Gurzun 2014	Low	Low	High	Low	High	Low Low	Unclear	Low	
Hassan 2007	Low	Low	Low	High	Unclear	Low both	Unclear	Low	
Hughes-Anderson 2002	High	Low	Unclear	Low	Unclear	Unclear S	unclear	Low	
lp2014	Low	Low	High	Unclear	Low	Unclear 📩	Unclear	Low	
Johnson 2011	Low	Unclear	Unclear	Low	High	Low Low Low	Low	Low	
Kovacs 2013	Low	Unclear	High	Low	Low	ruary Era <sup>Low</sup>	Unclear	Low	
Lalude 2014	Low	Low	Low	Low	High	2018 Sinus Low	Unclear	Low	
Landry 2011	Low	Unclear	Low	Low	Unclear	Low Low Unclear	Unclear	Low	
Lange 2007	Low	Low	Unclear	Low	Unclear	Unclear	Unclear	Low	
Linder 2006	Low	High	Unclear	Low	Low		Low	Unclear	
Lipczynska 2012	High	High	Unclear	Low	Low	Uncleard data Uncleard data Uncleard data Uncleard a data	Low	Low	
Llor 2011	Low	Low	Low	High	Low	Low <b>in</b>	Low	Low	
Loo 2009	Low	Low	Low	Low	Low	Low	Unclear	Low	
Mafi2013	Low	Low	Unclear	Unclear	Low	Low 📜 🛃	Unclear	Low	
Mafi2015	Low	Low	Unclear	Unclear	Low		Unclear	Low	
Majumdar 2003	Low	Low	Unclear	Low	Low	Low Ini	Unclear	Low	
Michaleff 2012	Low	Low	Low	Unclear	High	ng, a	Unclear	Low	
Moscavitch 2009	Low	Low	Unclear	Low	Low	Unclear <mark>similer</mark> Unclear <b>similer</b>	Low	Low	
Musicco 2004	High	Low	Low	Unclear	Unclear	Unclear milar	Unclear	High	
Nicholson 2010	Low	Low	Low	Low	Unclear	Low Low Low	Low	Low	
Nicopoullos 2003	High	High	Low	High	Low	Unclear <b>olo</b>	Unclear	Low	
Noya 2008	Low	Low	Low	Low	Unclear	Uncleare ,	Unclear	Low	
Piccoliori 2013	Low	Low	Low	Low	Low	Unclear. 20	Low	Low	
Pimlott 2006	Low	Low	Unclear	High	Unclear	Unclear 🕉	Unclear	Low	
Piterman 2008	Low	Unclear	Low	Unclear	Unclear	Low A	Unclear	High	
Remedios 2014	Low	Unclear	Low	High	Unclear	Low D	Unclear	Low	
Schers 2000	Low	Low	Low	Unclear	Unclear	Low par	Unclear	Low	
Smith 2008	Low	Low	Low	Low	Unclear	Low <b>r</b>	Unclear	Low	
Sokol 2015	Low	Low	Low	Low	Low	Low <b>ne</b> Low <b>nt</b>	High	Low	
Ulrik 2010	Low	Low	Low	High	Low	Low It GEZ-LTA	unclear	Low	

Ulrik 2013
van Gurp 2013
Williams2010
Williams2010

1	
2	
3	
4	MEDLINE Search Strategy
5	
6	1. Ambulatory Care/
7	2. exp Ambulatory Care Facilities/
8	3. general practice/ or family practice/
9	4. general practitioners/ or physicians, family/ or physicians, primary care/
10 11	5. Primary Health Care/
12	6. Office visits/
12	7. Academic Medical Centers/
14	8. (ambulatory adj3 (care or setting? or facilit* or ward? or department? or
15	service?)).ti,ab.
16	
17	9. ((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.
18	10. (primary care or primary health care or primary healthcare or family
19	medicine or community medicine or community health).ti,ab.
20	11. (gp or gps).ti,ab.
21	12. (after hour? or afterhour? or "out of hour?" or ooh).ti,ab.
22	13. (clinic? or visit?).ti,ab.
23	14. ((health* or medical) adj2 (center? or centre?)).ti,ab.
24	15. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
25	16. exp Emergency Service, Hospital/
26	17. Emergency Medical Services/
27	18. (emergency adj3 (care or setting? or facilit* or ward? or department? or
28	service? or room?)).ti,ab.
29	19. (emergency medicine or ed or er or a&e).ti,ab.
30	20. 16 or 17 or 18 or 19
31	21. 15 or 20
32	22. guidelines as topic/ or practice guidelines as topic/
33	23. (guideline? or guidance?).ti,ab.
34 35	24. 22 or 23
36	25. (adhere* or non-adhere* or nonadhere* or concord* or non-concord* or
37	nonconcord* or discord* or comply or complian* or non-complian* or
38	
39	noncomplian* or align* or nonalign* or nonalign* or congruen* or incongruen*
40	or consisten* or inconsisten* or contradict*).ti,ab.
41	26. ((does or "does not" or doesn?t or did or "did not" or didn?t or "not" or fail*)
42	adj3 (follow* or met or meet or meeting or match or matching or "in line
43	with")).ti,ab.
44	27. ((follow* or met or meet or meeting or match or matching or "in line with" or
45	keep or kept or keeping or utili?ation or utile?e? or change?) adj5 (criteria or
46	recommend* or guideline? or guidance)).ti,ab.
47	28. Physician's Practice Patterns/
48	29. clinical competence/ or nursing competence/
49	30. 25 or 26 or 27 or 28 or 29
50	31. 24 and 30
51	32. Guideline Adherence/
52	33. 31 or 32
53	34. exp *"diagnostic techniques and procedures"/
54	35. exp "diagnostic techniques and procedures"/ut
55 56	36. (diagnos <sup>*</sup> or detect <sup>*</sup> or test <sup>*</sup> or screen <sup>*</sup> or manag <sup>*</sup> ).ti.
56 57	solution of acteer of test of server of manag Jul
57	
58 59	

**BMJ** Open

1	
2 3	37. (imaging or radiolog* or tomogra* or ct scan* or pet scan* or echocardiogra*
4	or angiogra* or ultrasound* or ultrasonogra*).ti,ab.
5	38. ((medical or clinical or diagnos <sup>*</sup> or screening or routine or laboratory) adj5
6	(test* or investigation?)).ti,ab.
7	39. ((h?ematolog* or blood or urin* or saliva*) adj5 test*).ti,ab.
8	40. ((stress* or physical or function*) adj5 test*).ti,ab.
9	41. 34 or 35 or 36 or 37 or 38 or 39 or 40
10	42. 21 and 33 and 41
11 12	43. ((necessary or unnecessary or appropriate* or inappropriate* or waste* or
13	utili?ation or indicated or excess* or less or more or increas* or decreas*) adj10
14	(test* or imaging or radiolog* or tomogra* or ct scan* or pet scan* or
15	echocardiogra* or angiogra* or ultrasound* or ultrasonogra* or
16	investigation?)).ti,ab.
17	44. ((order* or request*) adj5 (test* or imaging or radiolog* or tomogra* or ct
18	scan* or pet scan* or echocardiogra* or angiogra* or ultrasound* or
19	ultrasonogra* or investigation?)).ti,ab.
20 21	45. Unnecessary Procedures/
22	46. 43 or 44 or 45
23	47. 21 and 24 and 46
24	48. 21 and 41 and 45
25	49. 42 or 47 or 48
26	50. limit 49 to yr="1999 -Current"
27	51. limit 50 to english language
28	52. exp animals/ not humans.sh.
29 30	53. 51 not 52
31	
	EMBASE Search Strategy
32 33	
32	1. Ambulatory Care/
32 33 34 35	1. Ambulatory Care/ 2. general practice/
32 33 34 35 36	1. Ambulatory Care/ 2. general practice/ 3. general practitioners/
32 33 34 35 36 37	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> </ol>
32 33 34 35 36 37 38	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or</li> </ol>
32 33 34 35 36 37 38 39	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine</li> </ol>
32 33 34 35 36 37 38 39 40 41	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 46	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(health* or medical) adj2 (center? or centre?)).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(health* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>((health* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(grimary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(health* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>1 ((health* or so f or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or</li> </ol>
32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 46 47 48 49 50 51 52	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>(general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(health* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(primary care or primary health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(health* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(emergency medicine or ed or er or a&amp;e).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(fealth* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>10 (clinic? or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(emergency medicine or ed or er or a&amp;e).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or primary healthcare or family medicine or community medicine or community health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(fearth or medical) adj2 (center? or centre?)).ti,ab.</li> <li>10 (clinic? or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(emergency medicine or ed or er or a&amp;e).ti,ab.</li> <li>13 or 14 or 15</li> <li>12 or 16</li> </ol>
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	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(fealth* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>10 (clinic? or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(emergency medicine or ed or er or a&amp;e).ti,ab.</li> </ol>
32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or primary healthcare or family medicine or community medicine or community health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(fearth or medical) adj2 (center? or centre?)).ti,ab.</li> <li>10 (clinic? or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(emergency medicine or ed or er or a&amp;e).ti,ab.</li> <li>13 or 14 or 15</li> <li>12 or 16</li> </ol>
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	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(lealth* or medical) adj2 (center? or centre?)).ti,ab.</li> <li>(lealth* or medical) adj2 (center? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(mergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(mergency medicine or ed or er or a&amp;e).ti,ab.</li> <li>(1 3 or 14 or 15</li> <li>12 or 16</li> <li>*practice guideline/</li> </ol>
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	<ol> <li>Ambulatory Care/</li> <li>general practice/</li> <li>general practitioners/</li> <li>Primary Health Care/</li> <li>(ambulatory adj3 (care or setting? or facilit* or ward? or department? or service?)).ti,ab.</li> <li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li> <li>(general or family) adj2 (practi* or primary healthcare or family medicine or community medicine or community health care or primary healthcare or family medicine or community medicine or community health).ti,ab.</li> <li>(gp or gps).ti,ab.</li> <li>(after hour? or afterhour? or "out of hour?" or ooh).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(clinic? or visit?).ti,ab.</li> <li>(fearth or medical) adj2 (center? or centre?)).ti,ab.</li> <li>10 (clinic? or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11</li> <li>Emergency Ward/</li> <li>(emergency adj3 (care or setting? or facilit* or ward? or department? or service? or room?)).ti,ab.</li> <li>(emergency medicine or ed or er or a&amp;e).ti,ab.</li> <li>13 or 14 or 15</li> <li>12 or 16</li> </ol>

#### BMJ Open

1	
2	
3	19. (guideline? or guidance?).ti,ab.
4	20. 18 or 19
5	21. (adhere* or non-adhere* or nonadhere* or concord* or non-concord* or
6	nonconcord* or discord* or comply or complian* or non-complian* or
7	noncomplian* or align* or nonalign* or nonalign* or congruen* or incongruen*
8	or consisten* or inconsisten* or contradict*).ti,ab.
9	22. ((does or "does not" or doesn?t or did or "did not" or didn?t or "not" or fail*)
10	adj3 (follow* or met or meet or meeting or match or matching or "in line
11	with")).ti,ab.
12 13	23. ((follow <sup>*</sup> or met or meet or meeting or match or matching or "in line with" or
14	keep or kept or keeping or utili?ation or utile?e? or change?) adj5 (criteria or
15	recommend* or guideline? or guidance)).ti,ab.
16	
17	24. clinical competence/ or nursing competence/
18	25. 21 or 22 or 23 or 24
19	26. 20 and 25
20	27. diagnostic procedure/ or exp blood examination/ or exp cardiovascular
21	system examination/ or exp digestive system examination/ or exp endocrine
22	system examination/ or exp neurologic examination/ or exp respiratory tract
23	examination/ or exp urogenital system examination/
24	28. (diagnos* or detect* or test* or screen* or manag*).ti.
25	29. (imaging or radiolog* or tomogra* or ct scan* or pet scan* or echocardiogra*
26	or angiogra* or ultrasound* or ultrasonogra*).ti,ab.
27	30. ((medical or clinical or diagnos* or screening or routine or laboratory) adj5
28	(test* or investigation?)).ti,ab.
29	31. ((h?ematolog* or blood or urin* or saliva*) adj5 test*).ti,ab.
30	32. ((stress* or physical or function*) adj5 test*).ti,ab.
31	33. 27 or 28 or 29 or 30 or 31 or 32
32	34. 17 and 26 and 33
33	35. ((necessary or unnecessary or appropriate* or inappropriate* or waste* or
34	utili?ation or indicated or excess* or less or more or increas* or decreas*) adj10
35 36	(test* or imaging or radiolog* or tomogra* or ct scan* or pet scan* or
37	
38	echocardiogra* or angiogra* or ultrasound* or ultrasonogra* or
39	investigation?)).ti,ab.
40	36. ((order* or request*) adj5 (test* or imaging or radiolog* or tomogra* or ct
41	scan* or pet scan* or echocardiogra* or angiogra* or ultrasound* or
42	ultrasonogra* or investigation?)).ti,ab.
43	37. Unnecessary Procedures/
44	38. 35 or 36 or 37
45	39. 17 and 20 and 38
46	40. 17 and 33 and 37
47	41. 34 or 39 or 40
48	42. limit 41 to yr="1999 -Current"
49	43. limit 42 to english language
50	44. (exp animals/ or nonhuman/) not human/
51	45. 43 not 44
52	46. conference*.pt.
53	47. 45 and 46
54	48. 45 not 46
55	49. exp child/ not (exp Child/ and exp Adult/)
56 57	The complement not leve ennot and exp Addit )
57	
58 59	
22	For peer review only - http://bmionen.hmi.com/site/about/quidelines.yhtml

**BMJ** Open

50.48 not 49 51.48 not 49 52. limit 47 to yr="2015 -Current"

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

#### MOOSE Statement - Reporting Checklist for Authors, Editors, and Reviewers of Meta-analyses of Observational Studies

Reporting Criteria	Reported (Yes/No)	Reported on Page
Reporting of Background		
Problem definition	YES	4
Hypothesis statement	YES	4
Description of Study Outcome(s)	YES	4
Type of exposure or intervention used	N/A	N/A
Type of study design used	YES	5, 6
Study population	YES	5
Reporting of Search Strategy		
Qualifications of searchers (eg, librarians and investigators)	YES	5
Search strategy, including time period included in the synthesis and keywords	YES	5, supplementary fil
Effort to include all available studies, including contact with authors	YES	5
Databases and registries searched	YES	5
Search software used, name and version, including special features used (eg, explosion)	YES	5
Use of hand searching (eg, reference lists of obtained articles)	YES	5
List of citations located and those culture content of citation content of the co	NO	
Method for addressing articles	NO	
published in languages other than English	0	
Method of handling abstracts and unpublished studies	YES	5
Description of any contact with authors	N/A	
Reporting of Methods		
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	YES	5,6
Rationale for the selection and coding of data (eg, sound clinical principles or convenience)	YES	6
Documentation of how data were classified and coded (eg, multiple raters, blinding, and interrater reliability)	YES	6
Assessment of confounding (eg, comparability of cases and controls in studies where appropriate	N/A	N/A
Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results	YES	5,6
Assessment of heterogeneity	YES	6

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	
25 26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
42	
44	
45	
46	
47	
48	
49	
50	
50 51	
51	
52	
53	
54	
55	
56	
57	
58	
59 60	
60	

Description of statistical methods (eg, complete description of fixed or random	YES	6
effects models, justification of whether		
the chosen models account for predictors		
of study results, dose-response models,		
or cumulative meta-analysis) in sufficient		
detail to be replicated		
Provision of appropriate tables and	YES	Tables 1,2, Figures
graphics		2,3,4
Reporting of Results		
Table giving descriptive information for	YES	Table 1 and Table 2
each study included		
Results of sensitivity testing (eg,	N/A	N/A
subgroup analysis)		
Indication of statistical uncertainty of	YES	6,8,9
findings		
Reporting of Discussion		
Quantitative assessment of bias (eg, publication bias)	YES	8,9
Justification for exclusion (eg, exclusion	YES	5
of non–English-language citations)		
Assessment of quality of included studies	YES	7, Table 3
Reporting of Conclusions		
Consideration of alternative explanations	YES	9, 10
for observed results		
Generalization of the conclusions (ie,	YES	10
appropriate for the data presented and		
within the domain of the literature review)		
Guidelines for future research	YES	9, 10
Disclosure of funding source	YES	11

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.



## PRISMA 2009 Checklist

Section/topic	_#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 'Search strategy'
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5 & supplementary figure
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5&6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5&6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5, 6, 7 & supplementary figure
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5,6



## **PRISMA 2009 Checklist**

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.	6
	•	Page 1 of 2	
Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7, 8
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	7, 8, 9
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	7
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	7, 8
DISCUSSION	<u> </u>		
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	9
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	10
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	10
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	11

43

46

47

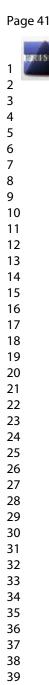
44 From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. 45 doi:10.1371/journal.pmed1000097 Protected by copyright, whellighed for test selected to test and take and the maining the maining and a maining to the connologies.

Erasmushogeschool

ATA-ZEB them in the store of th

Page 41 of 41

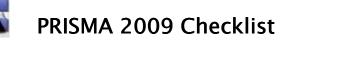
**BMJ** Open



45

46

47



For more information, visit: www.prisma-statement.org

r r

**BMJ** Open

# **BMJ Open**

# Over and undertesting in primary care: a systematic review and meta-analysis.

Journal:	BMJ Open
Manuscript ID	bmjopen-2017-018557.R1
Article Type:	Research
Date Submitted by the Author:	25-Oct-2017
Complete List of Authors:	O'Sullivan, Jack; Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences Albasri, Ali Nicholson, Brian; University of Oxford, Perera, Rafael; University of Oxford, Primary Health Care Aronson, Jeffrey; University of Oxford, Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences Roberts, Nia; University of Oxford, UK, Bodleian Health Care Libraries, Heneghan, Carl; Oxford University, Primary Health Care
<b>Primary Subject Heading</b> :	Epidemiology
Secondary Subject Heading:	Diagnostics, General practice / Family practice
Keywords:	PRIMARY CARE, RADIOLOGY & IMAGING, EPIDEMIOLOGY, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE<sup>™</sup> Manuscripts

2/

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1 2	
3	Over an
4 5	meta-aı
6 7	O'Sullivan
8	
9 10	<sup>1</sup> Centre for
11 12	University of
13	<sup>2</sup> Bodleian I
14 15	
16 17	Jack W O'S
18	Ali Albasri,
19 20	Brian D Nic
21 22	Rafael Pere
23	Jeffrey Aro
24 25	Nia Roberts
26 27	Carl Heneg
28	
29 30	Correspon
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	1
59	-

# **Over and undertesting in primary care: a systematic review and neta-analysis.**

D'Sullivan J<sup>1</sup>, Albasri A<sup>1</sup>, Nicholson B<sup>1</sup>, Perera R<sup>1</sup>, Aronson J<sup>1</sup>, Roberts N<sup>2</sup>, Heneghan C<sup>1</sup>

<sup>1</sup> Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Science, University of Oxford, UK

<sup>2</sup> Bodleian Health Care Libraries, University of Oxford.

Jack W O'Sullivan, Clinical Researcher, jack.osullivan@phc.ox.ac.uk Ali Albasri, Clinical Researcher, <u>ali.albasri@gtc.ox.ac.uk</u> Brian D Nicholson, Clinical Researcher, <u>brian.nicholson@phc.ox.ac.uk</u> Rafael Perera, Professor of Medical Statistics, <u>rafael.perera@phc.ox.ac.uk</u> Jeffrey Aronson, Reader in Evidence-Based Medicine, jeffrey.aronson@phc.ox.ac.uk Nia Roberts, Medical Librarian, <u>nia.roberts@bodleian.ox.ac.uk</u> Carl Heneghan, Professor of Evidence-Based Medicine, <u>carl.heneghan@phc.ox.ac.uk</u>

Correspondence to: Dr Jack O'Sullivan

Centre for Evidence-Based Medicine Nuffield Department of Primary Care Health Sciences Radcliffe Observatory Quarter, Oxford, OX2 6GG

#### Abstract

#### Background

Health systems are currently subject to unprecedented financial strains. Inappropriate test use wastes finite health resources (overuse) and delays diagnoses and treatment (underuse). As most patient care is provided in primary care, it represents an ideal setting to mitigate waste.

#### Objective

To identify over and under use of diagnostic tests in primary care.

#### Design

Systematic review and meta-analysis.

#### Data sources and eligibility criteria

We searched MEDLINE and EMBASE from January 1999 to October 2017 for studies that measured the inappropriateness of any diagnostic test (measured against a national or international guideline) ordered for adult patients in primary care.

#### Results

We included 357,171 patients from 63 studies in 15 countries. We extracted 103 measures of inappropriateness (41 underuse, 62 overuse) from included studies for 47 different diagnostic tests.

The overall rate of inappropriate diagnostic test ordering varied substantially (0.2% to 100%).

17 tests were underused >50% of the time. Of these, echocardiography (n=4 measures) was consistently underused (between 54% and 89%, n=4). There was large variation in the rate of inappropriate underuse of pulmonary function tests (38% to 78%, n = 8).

Ten tests were inappropriately overused >50% of the time. Echocardiography was consistently overused (77% to 92%), whereas inappropriate overuse of urinary cultures, upper endoscopy and colonoscopy varied widely, from 36% to 77% (n=3), 10% to 54% (n=10) and 8% to 52% (n=2) respectively.

#### Conclusions

There is marked variation in the appropriate use of diagnostic tests in primary care. Specifically, the use of echocardiography (both under and overuse) is consistently poor. There is substantial variation in the rate of inappropriate underuse of pulmonary function tests and the overuse of upper endoscopy, urinary cultures and colonoscopy.

Registration number: PROSPERO Registration ID: CRD42016048832

#### Manuscript word count: 3,531

#### Strengths and limitations of this study

#### Strengths

- Generates rate of under and overtesting for specific diagnostic tests against national or international guidelines
- Only includes data from real clinical encounters rather than surveys or hypothetical clinical vignettes.
- Quantified inappropriate ordering of all types of diagnostic tests, rather than just laboratory.

#### Limitations

- Systematic reviews are restricted to published literature, thus rates of inappropriate ordering are not available for all tests available to primary care physicians.
- Included studies measure appropriateness of testing in a particular health care setting against a particular guideline, thus reflect test ordering in a specific health care setting.

Topper territory only

#### Page 4 of 42

#### Introduction

Reaching a diagnosis in primary care is exceedingly complex. The combination of undifferentiated symptoms, a low prevalence of serious disease, a high degree of symptom overlap between serious and benign conditions, patients with multiple complaints, and psychological or social distress manifesting somatically all complicate reaching a diagnosis [1]. In around 40% of primary care consultations a diagnosis cannot be established from the history and physical examination alone [2], and tests are therefore often needed [1,3].

Primary care consultations make up most of the care provided in healthcare systems (90% of consultations in the UK [4], 55% of consultations in the USA[5]) and inappropriate diagnostic testing in primary care therefore has enormous resource implications. Given the calls for £22 billion in efficiency savings from the UK's National Health Service (NHS) [6] and the \$660 billion US Medicare deficit predicted by 2023 [7], ensuring the appropriateness of primary care diagnostic testing is crucial to the sustainability of healthcare systems [8].

Inappropriate diagnostic tests in primary care can be both inappropriately underused and overused. Underuse of tests, failure to order a test when clearly indicated, can lead to diagnostic errors and delays in diagnosis and the delivery of effective treatment, leading to adverse patient outcomes and further healthcare costs [9,10]. Overuse of tests, the delivery of tests with no clear benefit or when potential harms outweigh potential benefits, subjects patients to direct harms, such as radiation exposure, as well as potential adverse outcomes (e.g. contrast nephropathy) [11], incidental findings [12], and overdiagnosis [13]. Overuse is also a waste of finite healthcare expenditure, diverting resources from beneficial tests and treatments [14–16].

Many drivers encourage inappropriate under and overuse of diagnostic tests in primary care. Greater access to tests [17], the medicolegal consequences of under-testing [18], few if any disincentives to overinvestigate [14], and clinical performance measures [19] may all contribute to overuse. Increasing primary care workload [4], time constraints [19], and difficulty keeping up-to-date with rapidly increasingly evidence [20] may contribute to both inappropriate underuse and overuse.

Guidelines set the standard of care across most health-care settings [21,22]. Furthermore, they provide a medicolegal framework [23], inform health-care policy, and improve both care outcomes and processes of care [24]. Despite some recognised limitations, including varying quality of guidelines [25–27], guidelines are often used as markers of health-care appropriateness [28–31]. Zhi et al, for instance, used guidelines as a measure of appropriateness to estimate under and overuse of laboratory testing [29]. They estimated that 45% (95%CI 34 – 56%) of secondary care laboratory testing is underused and 21% (95%CI 16 – 25%) is overused.

Despite the increasing use of healthcare resources [32], rising healthcare expenditure [6–8], increasing demands placed on primary care [4], and apparent drivers of inappropriate testing [1,4,14,17–20], it is not clear how often diagnostic tests are inappropriately overused or underused in primary care. We therefore conducted a systematic review to quantify the frequency of appropriate ordering of all types of diagnostic tests from primary care in relation to their respective guidelines and identify tests that are frequently over and underused.

#### Methods

This study was conducted and is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33] and Meta-analysis of Observational Studies in Epidemiology (MOOSE) statements [34].

#### Protocol and Registration

The protocol has been published and is available online (open access) via the International prospective register for systematic reviews (PROSPERO) database (Registration ID: CRD42016048832).

#### Search Strategy

We searched EMBASE (OvidSP) and MEDLINE (OvidSP) databases from January 1999 to October 2017 for studies of any design measuring how often diagnostic test guidelines were followed in primary care (Supplementary File 1: Search Strategy). Our search strategy can be summarised as: 'Ambulatory Care AND adherence AND guideline AND diagnostic tests AND inappropriate'. Conference abstracts published after 2015 were also searched for in these databases to capture data not yet published. We also searched the WHO International Clinical Trials Registry Platform (http://apps.who.int/trialsearch/), ClinicalTrials.gov, and the reference lists of included studies.

#### Eligibility Criteria

We included studies of any design if they measured the rate of inappropriate ordering (overuse) or not ordering (underuse) of diagnostic tests ordered from primary care against national or international guidelines. We considered all diagnostic tests ordered in adults. We also included studies that measured diagnostic tests ordered from primary care but performed in secondary care (e.g. upper endoscopy). We included the control arms of RCTs if they offered exclusively usual care, and the pre-intervention periods of studies that used interrupted time series designs (before and after studies).

We excluded studies if they met the following criteria: >20% of participants were children (>20% under 18 years old); diagnostic tests not ordered by General Practitioners; screening or monitoring tests, or publication before 1999 (studies after 1999 were considered to ensure that results would more closely reflect current practice). We defined a screening test as a test on an asymptomatic or symptomatic person without signs or symptoms related to that test [35,36]. We defined monitoring tests as 'a test for a patient with an established diagnosis, for which the test is used to measure progression of the disease' [37]. We excluded studies if they did not give a measure of appropriateness was measured against local guidelines, such as a guideline specific to a hospital or region, rather than international or national guidelines.

#### Study selection and data extraction

Three reviewers (JS and AA or BN) independently screened titles, abstracts, and full texts for eligibility. The same reviewers assessed risks of bias and extracted the following data from included studies: patient demographics, eligibility criteria, name and type of diagnostic test, duration of study (days), guideline name and recommendation, total number of tests performed, and the number of tests ordered when the specific guideline recommended not ordering (inappropriate overuse) or the number of tests not ordered when the guideline recommended ordering it (inappropriate underuse). The last two data points (overuse and underuse) represent 'measures of inappropriateness'. When studies measured inappropriateness of multiple tests we extracted data on each test and presented them as individual measures of inappropriateness. When studies measured tests across different periods we

extracted measures for each time point and considered each one as an individual measure of inappropriateness.

We assessed the quality of included studies using a modified version of the Hoy risk of bias tool [38]. This tool has been validated to assess the internal and external validity of prevalence studies [38]. Our modified version of this tool kept the same domains, but adjusted the wording of the tool to reflect prevalence of inappropriate testing rather than prevalence of disease. Our tool (and results) is available in Supplementary File 2: Risk of Bias.

#### Statistical analysis

The primary outcome was the prevalence of inappropriate diagnostic testing. Inappropriate testing was measured in two ways:

1) Overuse: A diagnostic test was ordered when the relevant guideline recommends not ordering it, for instance, imaging for non-red flag low back pain (LBP).

2) Underuse: A diagnostic test was not ordered when the relevant guideline recommended ordering it, for instance, spirometry to confirm or refute the diagnosis of COPD.

We expressed measures of inappropriateness as proportions (%), where the numerator represents the total number of times a guideline recommendation was not followed and the denominator the total number of times a guideline recommendation could have been followed. For instance, the number of times imaging was inappropriately ordered for non-red flag headache as a proportion of the total number of patients who presented with non-red flag headache. Given these data are proportions, we calculated Clopper-Pearson 95% confidence intervals for each individual measure of appropriateness. We conducted sensitivity analyses with high risk of bias studies excluded.

Where the same guideline and recommendation were used by multiple studies (e.g. five studies measured inappropriate underuse of spirometry testing in patients with COPD [39–43] using the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guideline) we pooled the measures and assessed heterogeneity. We combined measures of inappropriateness using a random-effects meta-analysis with 95% confidence intervals (Clopper-Pearson), for this reason each measure of appropriateness contributed relatively evenly to pooled estimates. We performed double arcsine transformation on prevalence data to stabilize the variance [44], and pooled the data using the inverse variance method [45]. We assessed heterogeneity using the I<sup>2</sup> statistic [46]. We did not combine measures of overuse and underuse, as they have different denominators: overuse involves the total number of tests ordered, whereas underuse involves the total number of times a test should have been ordered. We performed analyses using R version 3.3.2 (R project).

#### Results

#### Study selection and characteristics

We included 63 studies from 14,716 references identified from independent searches by two authors (JOS and AA or BN) (see Figure 1). Of the 63 included studies, 55 were observational studies, 6 were before and after studies and 2 were RCTs. These studies were conducted in 15 countries and included 357,171 patients (Supplementary File 3: Table 1). Table 2 (Supplementary File 4: Table 2) shows the 103 measures of inappropriateness extracted from included studies for 47 different diagnostic tests measured against 77 guideline recommendations (41 measured underuse and 62 measured overuse). Guideline recommendations came from 42 different guideline organisations from 15 countries.

Fourteen studies measured inappropriateness of more than one diagnostic tests for the same condition (e.g. chest x-ray (CXR), electrocardiography (ECG), and transthoracic echocardiography (TTE) to confirm or refute a diagnosis of heart failure). Two studies [47,48] measured inappropriateness across multiple time periods. No studies measured both under and overuse of the same test.

Included studies measured inappropriateness in one of three ways:

1. Patients with specific symptoms were assessed (prospectively or retrospectively) to see if they had received an inappropriate diagnostic test (overuse) or hadn't received the appropriate diagnostic test (underuse) in line with the relevant guideline recommendation (e.g. records for patients with non-red flag LBP to see if they received imaging [49]). 18 studies used this method.

2. Patients who had undergone a diagnostic test were identified (via hospital or national databases) and an assessment of whether the test was inappropriate (as per the defined guideline recommendations) via individual patient data was made (overuse). For instance, patients who had an upper endoscopy[50]). 22 studies used this method.

3. Patients with a diagnosis were identified via hospital or national databases and assessed to see whether they had received the appropriate diagnostic test (as per the defined guideline) to confirm or refute the diagnosis via individual patient data (underuse). For instance, assessing if patients with a diagnosis of COPD had spirometry to confirm or refute the diagnosis [39]). 23 studies used this method.

#### Risk of bias

Two thirds of the studies (n=44) were graded as being at low risk of bias, 15 (24%) at moderate risk, and 4 (6%) at high risk (Supplementary File 2 Risk of Bias). Moderate or high risk studies were at an increased risk of non-response bias (>20%), non-objective collection of data, and/or unclear intervals between symptom onset and diagnostic test use. Supplementary File 2 Risk of Bias outlines risk of bias scores in detail.

#### Proportions of diagnostic tests ordered in line with specific guideline recommendations

There was large variation in the rate of inappropriate diagnostic test ordering. The 103 diagnostic test guideline recommendations were not followed 0.2 - 100% of the time (Supplementary File 4 Table 2), wide variation was largely sustained (0.2 - 99.94%) when a further analysis was conducted excluding studies judged to be of high risk of bias. The prevalence of underuse varied 8.2% to 100%, whereas overuse varied between 0.2% and 94.2%. Similarly, this variation was essentially maintained upon exclusion of high risk studies (under use 9.8% - 99.9%, overuse 0.2 - 94.2%).

#### Underused tests

Table 2 (Supplementary File 4) shows that 17 tests were underused more than 50% of the time. Echocardiography was the most frequently studied (n=4 measures in Poland, UK (2), Brazil). In

patients with heart failure, echocardiography was underused between 54% and 89% (n=3) of the time and in atrial fibrillation 56% (n=1).

For some tests there was large variation in the rate of underuse (Figure 2). Underuse of pulmonary function tests (PFTs) to confirm or refute COPD, measured against the Global Initiative for Chronic Obstructive Lung Disease (GOLD), NICE (UK) and Danish National Board of Health guidelines, varied from 26% to 78% (n=8). None of the studies that studied echocardiography, or PFTs were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

#### Overused tests

Ten tests were overused more than 50% of the time (Figure 3). Echocardiography was consistently overused, for instance in 'routine perioperative evaluation of ventricular function with no symptoms or signs of cardiovascular disease', whereas other tests (urinary cultures, upper endoscopy and colonoscopy) were overused at varying rates. The over use of echocardiography was studied in the UK [51] and the Netherlands [52]. The rates of overuse varied between the two settings: between 77% (Netherlands) and 92% (UK). Overuse of urinary cultures for uncomplicated urinary tract infections was studied in the USA [53,54], Spain [55] and Sweden [56] the rate varied from 57% to 77% in the USA, was around 50% in Sweden and was as low as 36% in Spain. Overuse of upper endoscopy was studied widely (n=11); in Australia [57,58], Saudi Arabia [59,60], UK [61], Italy [62–64], USA [50,65], and Malaysia [66]. The overuse varied markedly, from 7.5% to 54% (n=11) respectively (Figure 3, Supplementary File 4 Table 2). Similarly, the inappropriate over-use of colonoscopy varied substantially; from 8% in Australia [58] to 52% in Malaysia [67]. None of the above studies were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

Our results also suggest that the inappropriate overuse of CT and MRI scans for non-red flag headache (a headache without symptoms suggesting a malignant underlying pathology) has more than doubled in the last ten years in the USA (2000: 6.7% (95%CI: 5.4 to 8.2%, 2010: 14% (95%CI 12. to 16%) (Supplementary File 4 Table 2) [48]. Conversely, the rate of inappropriate overuse of radiology tests for non-red flag low back pain was consistently low, with all (n=18 measures) but two measure showing inappropriate overuse less than 25% of the time (Supplementary File 4 Table 2). One of these studies [68] estimated overuse to be about 50%, but was conducted in 2001 and thus may reflect improvements over time. The other study is current, but used a small sample size [69]. None of these studies were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

#### Variation of inappropriateness against the same guideline recommendation

Eleven different guideline recommendations were studied more than once. There was significant heterogeneity ( $I^2 > 50\%$ ) in nine of these pooled measures. Significant heterogeneity may have occurred for several reasons: 1) vastly different populations (for instance, one study measured the inappropriateness of upper endoscopy in Saudi Arabia [60] using the American Gastroenterological Association recommendations, whereas another study used the same recommendations in the USA [70]; 2) Contrasting healthcare systems [71,72]; 3) Relevance and applicability of one country's national guideline to another country [73]; 4) A low number of measures for meta-analysis [46] and/or 5) Significant heterogeneity, reflecting significant variation in inappropriate ordering.

There is marked variation in the rate of underuse and overuse of diagnostic tests from many primary care settings across the world. This variation suggests improvement can be made in the rate of appropriate diagnostic test ordering.

Primary care use of echocardiography is consistently poor. Echocardiography is inappropriately underused for some clinical situations, e.g. confirming a diagnosis of heart failure, and inappropriately overused in others, e.g. perioperative assessment. This was consistent across the countries where appropriateness of echocardiogram has been studied. This is of concern, given the expertise and resource requirements to perform the test and the increasing availability of direct access ordering for primary care physicians.

For four tests we found marked variation in the rate of inappropriate use. Underuse of pulmonary function tests varied by >50%, whereas overuse of urinary cultures, upper endoscopy and colonoscopy all varied by around 40%.

Radiology tests for both non-red flag low back pain and non-red flag headache were frequently *not* overused, but the rate of overuse of imaging for non-red flag headache showed concerning trends, more than doubling from 2000 to 2010 (Supplementary File 4 Table 2).

#### Implications and future research

Two principle conclusions can be drawn from our results: 1.Ordering of echocardiograms from primary care appears to require improvement, 2. Markedly varying rates of inappropriate use for pulmonary function tests (underuse), colonoscopy (overuse), upper endoscopy (overuse), and urinary cultures (overuse) suggests that ordering can be improved.

Future research should focus on: Determining the reasons for deviation from guidelines, assessing the quality of guidelines supporting diagnostic test use and systematic reviews quantifying inappropriate screening and monitoring tests. Further, investigators wishing to undertake primary studies measuring inappropriate use should focus on developing objective data extraction methods for assessing patient notes and define clearly the interval they (investigators) will consider a test ordered for a particular symptom or disease.

#### Strengths in relation to other studies

Compared with other studies of inappropriate use of healthcare resources, we used data from real clinical encounters. This allowed a more robust assessment of diagnostic test inappropriateness, where other studies used surveys and hypothetical clinical vignettes [19,74,75]. Furthermore, we quantified the appropriateness of all types of diagnostic tests, rather than focusing on a specific test or specific disease (such as only laboratory tests [29]). Our paper is the first systematic review of studies that measured inappropriateness of all diagnostic tests ordered from primary care. Zhi et al [29] quantified the mean rates of overuse and underuse of laboratory tests in secondary care and focused on quantifying an overall rate of over and under use. They estimated that over and underuse of laboratory tests was around 21% and 45% respectively [29]. We choose not to quantify an overall rate of over and under use because we feel the results would not be representative; we would be combining data from multiple different health care settings and data captured only the studied selection of diagnostic tests available in primary care.

Our use of guideline recommendations as the metric of appropriateness allowed a direct measure of diagnostic test appropriateness. Other studies that have assessed temporal and geographical variation in the use of diagnostic tests [76,77] have noted substantial differences in diagnostic practices across different regions, irrespective of disease prevalence and patient characteristics [77]. These studies, however, could not quantify what proportion of the temporal increase in the use of a diagnostic test is

inappropriate and what proportion of variation between regions is inappropriate. We have quantified the proportion of inappropriate testing.

Although beyond the scope of our review, ultimately, interventions should be implemented to improve test use. A 2015 systematic review [78] concluded that 'Interventions such as educational strategies, feedback and changing test order forms may improve the efficient use of laboratory tests in primary care'. Thus, doctors, academics and policy makers can use our results to identify diagnostic tests in their particular health care settings which may benefit from intervention.

#### Limitations

The use of guidelines to quantify appropriateness of diagnostic tests could be considered a limitation of this study. Guidelines are often criticised for varying quality [25–27,79] and panel members' conflicts of interests [80]. However, clinical practice guidelines have been shown to improve both care outcomes and processes of care [24], allow assessment of care on a population level, inform health policy [81,82], set the standard of care across many health care settings [21,22], and provide a medicolegal framework [23]. One major medical insurance company advises that 'doctors must be prepared to explain and justify their decisions and actions, especially if they depart from guidelines produced by a nationally recognised body' [23]. Furthermore, guidelines have been used to measure appropriateness of the use of tests in other published peer-review studies [29]. There will always be times when it is appropriate to depart from guidelines, but dramatic, consistent variation from guidelines requires investigation and is unlikely to be caused entirely by the quality of guidelines.

Furthermore, our study includes only a selection of diagnostic tests and is thus not an allencompassing reflection of clinical practice. The data reflects the use of a specific test, sometimes for a particular clinical situation, in a particular country's health care system. Thus, policy makers and those interested in improving the quality of primary care diagnostic test use, can use our results as a resource to identify tests in their healthcare setting that require improvement and/or investigation to decipher why such deviation from guidelines exists. Our conclusions from this paper, however, are not generalisable to all primary care settings nor all primary care diagnostic tests.

Lastly, caution must be taken when comparing results that measured inappropriateness using different denominators. The results from studies that measured inappropriateness using patients who had undergone a diagnostic test as a denominator should be interpreted differently to studies that used patients with a diagnosis or symptoms as a denominator (and vice versa).

#### Conclusion

There is marked variation in under and overuse of appropriate diagnostic test use in primary care across the world. From the available data, echocardiograms are ordered particularly poorly, while the substantial variation in appropriate ordering of pulmonary function tests, colonoscopy, upper endoscopy, and urinary cultures suggest a need for improvement.

#### Funding:

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

All author declare no conflicts of interests.

Ethical approval: Not required

**Data sharing:** Data extracted from the included studies in this review are available on request from the corresponding author.

**Registration:** PROSPERO protocol Registration ID: CRD42016048832 (https://www.crd.york.ac.uk/prospero/display\_record.asp?ID=CRD42016048832)

#### Competing interest statement.

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and jointly declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

#### **Contribution statement:**

Conception and design: Jack O'Sullivan, Rafael Perera and Carl Heneghan

Search Strategy: Nia Roberts and Jack O'Sullivan

Screening, extraction and risk of bias: Jack O'Sullivan, Ali Albasri and Brian Nicholson.

Analysis and interpretation of the data: Jack O'Sullivan, Rafael Perera, Jeffrey Aronson and Carl Heneghan.

Drafting of the article: Jack O'Sullivan (all authors critically reviewed and approved manuscript)

Statistical expertise: Rafael Perera

Clinical expertise: Jack O'Sullivan, Brian Nicholson, Jeffrey Aronson and Carl Heneghan

Jack O'Sullivan is the guarantor.

#### **Copyright Statement**

The corresponding author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

#### References

1 Foot C, Naylor C, Imison C. The quality of GP diagnosis and referral. 2010.

	http://amapro.isabelhealthcare.com/pdf/Kings_Fund_Diagnosis_and_Referral_2010.pdf
2	Koch H, van Bokhoven MA, ter Riet G, <i>et al.</i> Ordering blood tests for patients with unexplained fatigue in general practice: what does it yield? Results of the VAMPIRE trial. <i>Br J Gen Pract</i> 2009; <b>59</b> :e93-100. doi:10.3399/bjgp09X420310
3	Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. BMJ 2009; <b>338</b> .
4	Hobbs FDR, Bankhead C, Mukhtar T, <i>et al.</i> Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. <i>Lancet</i> 2016; <b>387</b> :2323–30. doi:10.1016/S0140-6736(16)00620-6
5	Centers for Disease Control and Prevention, National Center for Health Statistics. National Ambulatory Medical Care Survey: 2012 Summary Tables. 2012;:5.http://www.cdc.gov/nchs/data/ahcd/namcs_summary/2010_namcs_web_tables.pdf
6	Alderwick H, Robertson R, Appleby J, <i>et al.</i> Better value in the NHS The role of changes in clinical practice. 2015.
7	Fisher ES, Bynum JP, Skinner JS. Slowing the growth of health care costslessons from regional variation. <i>N Engl J Med</i> 2009; <b>360</b> :849–52. doi:10.1056/NEJMp0809794
8	Appleby J, Thompson J, Jabbal J. Quarterly Monitoring Report: How is the NHS performing? <i>King's Fund</i> 2016;:1–42.
9	Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. <i>BMJ Qual Saf</i> 2013; <b>22 Suppl 2</b> :ii6-ii10. doi:10.1136/bmjqs-2012-001621
10	Gandhi TK, Kachalia A, Thomas EJ, et al. Annals of Internal Medicine Article Missed and Delayed Diagnoses in the Ambulatory Setting : Ann Intern Med 2006;145:488–96.
11	Katzberg RW, Lamba R. Contrast-induced nephropathy after intravenous administration: fact or fiction? <i>Radiol Clin North Am</i> 2009; <b>47</b> :789–800, v. doi:10.1016/j.rcl.2009.06.002\rS0033-8389(09)00094-3 [pii]
12	Lumbreras B, Donat L, Hernández-Aguado I. Incidental findings in imaging diagnostic tests: a systematic review. <i>Br J Radiol</i> 2010; <b>83</b> :276–89. doi:10.1259/bjr/98067945
13	Welch, H. Gilbert, Schwartz, Lisa, Woloshin S. Overdiagnosed: Making people sick in the pursuit of health. Beacon Press, 2011 2011.
14	Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. <i>Bmj</i> 2012; <b>344</b> :e3502–e3502. doi:10.1136/bmj.e3502
15	Berwick D, Hackbarth AD. Eliminating Waste in US Health Care. <i>JAMA</i> 2012; <b>307</b> :1513. doi:10.1001/jama.2012.362
16	Cecchini M, Lee S. <i>Tackling Wasteful Spending on Healthcare</i> . 2017. http://networks.sustainablehealthcare.org.uk/sites/default/files/resources/Tackling Wasteful Spending on Health.pdf#page=117
17	Health D of. NHS 2010–2015: from good to great. preventative, people-centred, productive. London: 2009.
18	Esmail A, Neale G, Elstein M, Firth-Cozens J, Davy C VC. Case Studies in Litigation: Claims reviews in four specialties. Manchester: 2004.
19	Sirovich BE, Woloshin S, Schwartz LM. Too Little? Too Much? Primary care physicians' views on US health care: a brief report. <i>Arch Intern Med</i> 2011; <b>171</b> :1582–5.
12	

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

#### BMJ Open

2		
3		doi:10.1001/archinternmed.2011.437
4 5 6	20	Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? <i>PLoS Med</i> 2010; <b>7</b> . doi:10.1371/journal.pmed.1000326
7 8	21	Garber AM. Evidence-based guidelines as a foundation for performance incentives. <i>Health Aff (Millwood)</i> 2005; <b>24</b> :174–9. doi:10.1377/hlthaff:24.1.174
9 10 11	22	Ransohoff DF, Pignone M, Sox HC, <i>et al</i> . How to Decide Whether a Clinical Practice Guideline Is Trustworthy. <i>JAMA</i> 2013; <b>309</b> :139. doi:10.1001/jama.2012.156703
12	23	Fryar C. Doctors can depart from guidelines in patients' best interests. BMJ 2015;350.
13 14 15 16	24	Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. <i>Lancet (London, England)</i> 1993; <b>342</b> :1317–22.http://www.ncbi.nlm.nih.gov/pubmed/7901634 (accessed 31 Aug 2016).
17 18 19 20 21	25	Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. <i>JAMA</i> 1999; <b>281</b> :1900–5.http://www.ncbi.nlm.nih.gov/pubmed/10349893 (accessed 7 Dec 2016).
22 23 24	26	Grilli R, Magrini N, Penna A, <i>et al.</i> Practice guidelines developed by specialty societies: the need for a critical appraisal. <i>Lancet (London, England)</i> 2000; <b>355</b> :103–6. doi:10.1016/S0140-6736(99)02171-6
25 26	27	Lenzer J. Why we can't trust clinical guidelines. BMJ 2013;346.
27 28 29 30	28	Spyridonidis D, Calnan M. Opening the black box: A study of the process of NICE guidelines implementation. <i>Health Policy (New York)</i> 2011; <b>102</b> :117–25. doi:10.1016/j.healthpol.2011.06.011
31 32	29	Zhi M, Ding EL, Theisen-Toupal J, <i>et al</i> . The Landscape of Inappropriate Laboratory Testing: A 15-Year Meta-Analysis. <i>PLoS One</i> 2013; <b>8</b> :e78962. doi:10.1371/journal.pone.0078962
33 34 35	30	McGlynn E, Asch S, Adams J, <i>et al.</i> Quality of health care delivered to adults in the United States. <i>N Engl J Med</i> 2003; <b>349</b> :1866-1868-1868. doi:10.1056/NEJMsa022615
36 37 38	31	Sheldon T a, Cullum N, Dawson D, <i>et al.</i> What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. <i>BMJ</i> 2004; <b>329</b> :999. doi:10.1136/bmj.329.7473.999
<ul><li>39</li><li>40</li><li>41</li><li>42</li></ul>	32	National Health Service. NHS Imaging and Radiodiagnostic activity in England. 2013;:1–7.http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/KH12-release-2012-13.pdf
43 44 45 46	33	Moher D, Liberati A, Tetzlaff J, <i>et al.</i> Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. <i>BMJ</i> 2009; <b>339</b> :b2535.http://www.ncbi.nlm.nih.gov/pubmed/19622551 (accessed 22 Aug 2016).
47 48	34	Stroup DF, Berlin JA, Morton SC, <i>et al.</i> Meta-analysis of Observational Studies in Epidemiology. <i>JAMA</i> 2000; <b>283</b> :2008. doi:10.1001/jama.283.15.2008
49 50 51	35	Wald NJ. Guidance on terminology. <i>J Med Screen</i> 2008; <b>15</b> :50–50. doi:10.1258/jms.2008.008got
52	36	Raffle A, Gray J. Screening: Evidence and Practice. Oxford University Press 2007.
53 54 55	37	Glasziou P, Irwig L, Aronson J. Evidence-based medical monitoring: from principles to practice. Oxford (UK): Blackwell Publishing, BMJ books 2008.
56 57	38	Hoy D, Brooks P, Woolf A, et al. Assessing risk of bias in prevalence studies: modification of
58 59	13	
60		For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

**BMJ** Open

	an existing tool and evidence of interrater agreement. <i>J Clin Epidemiol</i> 2012; <b>65</b> :934–9. doi:10.1016/j.jclinepi.2011.11.014
39	Belletti D, Liu J, Zacker C, <i>et al.</i> Results of the CAPPS: COPDassessment of practice in primary care study. <i>Curr Med Res Opin</i> 2013; <b>29</b> :957–66. doi:10.1185/03007995.2013.803957
40	Bertella E, Zadra A, Vitacca M, <i>et al.</i> COPD management in primary care: is an educational plan for GPs useful? <i>Multidiscip Respir Med</i> 2013; <b>8</b> :24. doi:10.1186/2049-6958-8-24
41	Chavez PC, Shokar NK. Diagnosis and management of chronic obstructive pulmonary disease (COPD) in a primary care clinic. <i>COPD</i> 2009; <b>6</b> :446–51. doi:10.3109/15412550903341455
42	Lange P, Rasmussen FV, Borgeskov H, <i>et al.</i> The quality of COPD care in general practice in Denmark: the KVASIMODO study. <i>Prim Care Respir J</i> 2007; <b>16</b> :174–81. doi:10.3132/pcrj.2007.00030
43	Ulrik CS, Sørensen TB, Højmark TB, <i>et al.</i> Adherence to COPD guidelines in general practice: impact of an educational programme delivered on location in Danish general practices. <i>Prim Care Respir J</i> 2013; <b>22</b> :23–8. doi:10.4104/pcrj.2012.00089
44	Barendregt JJ, Doi SA, Lee YY, et al. Meta-analysis of prevalence. J Epidemiol Community Heal 2013;:974–8. doi:10.1136/jech-2013-203104
45	Doi SAR, Barendregt JJ, Khan S, <i>et al.</i> Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. <i>Contemp Clin Trials</i> 2015; <b>45</b> :130–8. doi:10.1016/j.cct.2015.05.009
46	Higgins JPT, Thompson SG, Deeks JJ, <i>et al.</i> Measuring inconsistency in meta-analyses. <i>BMJ</i> 2003; <b>327</b> :557–60. doi:10.1136/bmj.327.7414.557
47	Mafi JN, McCarthy EP, Davis RB, <i>et al.</i> Worsening trends in the management and treatment of back pain. <i>JAMA Intern Med</i> 2013; <b>173</b> :1573–81. doi:10.1001/jamainternmed.2013.8992
48	Mafi JN, Edwards ST, Pedersen NP, <i>et al.</i> Trends in the Ambulatory Management of Headache: Analysis of NAMCS and NHAMCS Data 1999-2010. <i>J Gen Intern Med</i> 2015; <b>30</b> :548–55. doi:10.1007/s11606-014-3107-3
49	Williams CM, Maher CG, Hancock MJ, <i>et al.</i> Low back pain and best practice care: A survey of general practice physicians. <i>Arch Intern Med</i> 2010; <b>170</b> :271–7. doi:10.1001/archinternmed.2009.507
50	Cai JX, Campbell EJ, Richter JM. Concordance of Outpatient Esophagogastroduodenoscopy of the Upper Gastrointestinal Tract With Evidence-Based Guidelines. <i>JAMA Intern Med</i> 2015; <b>175</b> :1563–4. doi:10.1001/jamainternmed.2015.3533
51	Gurzun M-M, Ionescu A. Appropriateness of use criteria for transthoracic echocardiography: are they relevant outside the USA? <i>Eur Hear J - Cardiovasc Imaging</i> 2014; <b>15</b> :450–5. doi:10.1093/ehjci/jet186
52	van Gurp N, Boonman-De winter LJM, Meijer Timmerman Thijssen DW, <i>et al.</i> Benefits of an open access echocardiography service: A Dutch prospective cohort study. <i>Netherlands Hear J</i> 2013; <b>21</b> :399–405. doi:10.1007/s12471-013-0416-9
53	Johnson JD, O'Mara HM, Durtschi HF, <i>et al.</i> Do Urine Cultures for Urinary Tract Infections Decrease Follow-up Visits? <i>J Am Board Fam Med</i> 2011; <b>24</b> :647–55. doi:10.3122/jabfm.2011.06.100299
54	Grover ML, Bracamonte JD, Kanodia AK, <i>et al.</i> Assessing Adherence to Evidence-Based Guidelines for the Diagnosis and Management of Uncomplicated Urinary Tract Infection. <i>Mayo Clin Proc</i> 2007; <b>82</b> :181–5. doi:10.4065/82.2.181
14	

#### **BMJ** Open

Llor C, Rabanaque G, Lopez A, <i>et al.</i> The adherence of GPs to guidelines for the diagnosis and treatment of lower urinary tract infections in women is poor. <i>Fam Pract</i> 2011; <b>28</b> :294–9. doi:10.1093/fampra/cmq107
Lindbäck H, Lindbäck J, Melhus Å. Inadequate adherence to Swedish guidelines for uncomplicated lower urinary tract infections among adults in general practice. <i>Apmis</i> 2017; <b>125</b> :816–21. doi:10.1111/apm.12718
Leon P, Catherine K, Mark N, et al. Gastro-oseophageal reflux disease. The impact of guidelines on GP management. 2008.
Hughes-Anderson W, Rankin SL, House J, <i>et al.</i> Open access endoscopy in rural and remote Western Australia: does it work? <i>ANZ J Surg</i> 2002; <b>72</b> :699–703.http://www.ncbi.nlm.nih.gov/pubmed/12534377 (accessed 7 Dec 2016).
Aljebreen AM, Alswat K, Almadi MA. Appropriateness and diagnostic yield of upper gastrointestinal endoscopy in an open-access endoscopy system. <i>Saudi J Gastroenterol</i> 2013; <b>19</b> :219–22. doi:10.4103/1319-3767.118128
Azzam NA, Almadi MA, Alamar HH, <i>et al.</i> Performance of American Society for Gastrointestinal Endoscopy guidelines for dyspepsia in Saudi population: Prospective observational study. <i>World J Gastroenterol</i> 2015; <b>21</b> :637–43. doi:10.3748/wjg.v21.i2.637
Elwyn G, Owen D, Roberts L, <i>et al.</i> Influencing referral practice using feedback of adherence to NICE guidelines: a quality improvement report for dyspepsia. <i>Qual Saf Health Care</i> 2007; <b>16</b> :67–70. doi:10.1136/qshc.2006.019992
Cardin F, Zorzi M, Bovo E, <i>et al.</i> Effect of Implementation of a Dyspepsia and Helicobacter pylori Eradication Guideline in Primary Care. <i>Digestion</i> 2005; <b>72</b> :1–7. doi:10.1159/000087215
Cardin F, Zorzi M, Terranova O. Implementation of a guideline versus use of individual prognostic factors to prioritize waiting lists for upper gastrointestinal endoscopy. <i>Eur J Gastroenterol Hepatol</i> 2007; <b>19</b> :549–53. doi:10.1097/01.meg.0000216942.42306.d5
Hassan C, Bersani G, Buri L, <i>et al.</i> Appropriateness of upper-GI endoscopy: an Italian survey on behalf of the Italian Society of Digestive Endoscopy. <i>Gastrointest Endosc</i> 2007; <b>65</b> :767–74. doi:10.1016/j.gie.2006.12.058
Fiorenza JP, Tinianow AM, Chan WW. The Initial Management and Endoscopic Outcomes of Dyspepsia in a Low-Risk Patient Population. <i>Dig Dis Sci</i> 2016; <b>61</b> :2942–8. doi:10.1007/s10620-016-4051-3
Chan Y-M, Goh K-L. Appropriateness and diagnostic yield of EGD: a prospective study in a large Asian hospital. <i>Gastrointest Endosc</i> 2004; <b>59</b> :517–24. doi:10.1016/S0016-5107(04)00002-1
CHAN T, GOH K. Appropriateness of colonoscopy using the ASGE guidelines: experience in a large Asian hospital. <i>Chin J Dig Dis</i> 2006; <b>7</b> :24–32. doi:10.1111/j.1443-9573.2006.00240.x
Eccles M, Steen N, Grimshaw J, <i>et al.</i> Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. <i>Lancet</i> 2001; <b>357</b> :1406–9. doi:10.1016/S0140-6736(00)04564-5
Tahvonen P, Oikarinen H, Niinimäki J, <i>et al.</i> Justification and active guideline implementation for spine radiography referrals in primary care. <i>Acta radiol</i> 2016; <b>58</b> :586–92. doi:10.1177/0284185116661879
Majumdar SR, Soumerai SB, Farraye FA, <i>et al.</i> Chronic acid-related disorders are common and underinvestigated. <i>Am J Gastroenterol</i> 2003; <b>98</b> :2409–14. doi:10.1111/j.1572-0241.2003.07706.x
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool .

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

For peer review only - http://b

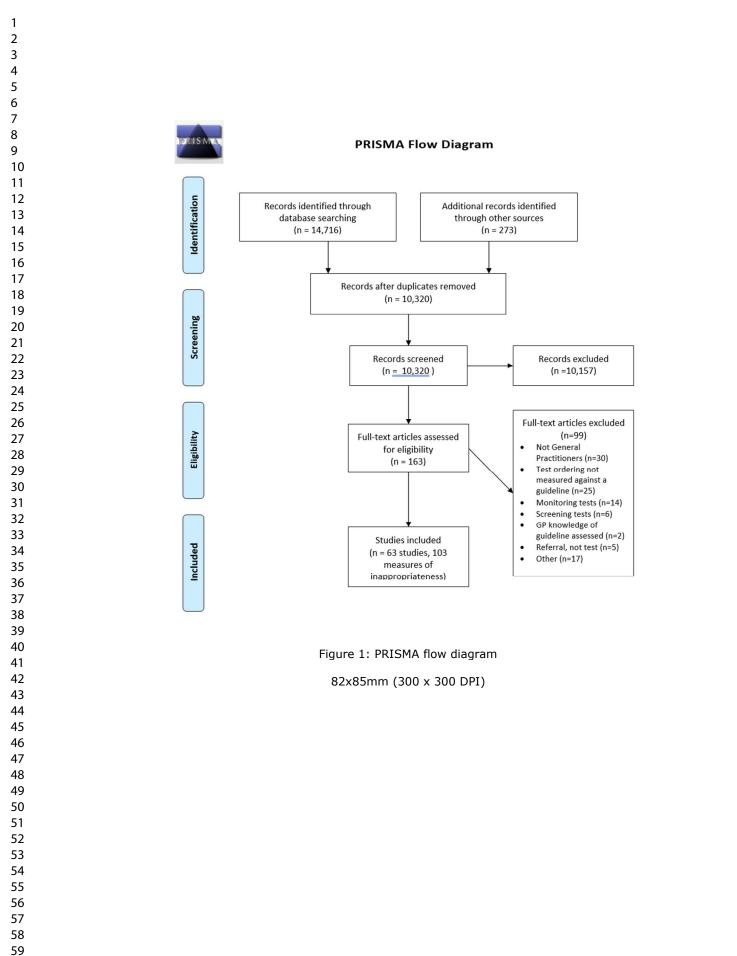
- 71 Basu S, Andrews J, Kishore S, *et al.* Comparative performance of private and public healthcare systems in low- and middle-income countries: A systematic review. *PLoS Med* 2012;**9**:19. doi:10.1371/journal.pmed.1001244
- 72 Ridic G, Gleason S, Ridic O. Comparisons of Health Care Systems in the United States, Germany and Canada. *Mat Soc Med* 2012;**24**:112–20. doi:10.5455/msm.2012.24.112-120.Comparisons
- 73 Gagliardi AR, Brouwers MC. Do guidelines offer implementation advice to target users? A systematic review of guideline applicability. *BMJ Open* 2015;5:e007047–e007047. doi:10.1136/bmjopen-2014-007047
- 74 Kachalia A, Berg A, Fagerlin A, *et al.* Overuse of testing in preoperative evaluation and syncope: a survey of hospitalists. *Ann Intern Med* 2015;**162**:100–8. doi:10.7326/M14-0694
- 75 Swennen MHJ, Rutten FH, Kalkman CJ, *et al.* Do general practitioners follow treatment recommendations from guidelines in their decisions on heart failure management? A cross-sectional study. *BMJ Open* 2013;**3**:e002982. doi:10.1136/bmjopen-2013-002982
- Parker L, Levin DC, Frangos A, *et al.* Geographic variation in the utilization of noninvasive diagnostic imaging: national medicare data, 1998-2007. *AJR Am J Roentgenol* 2010;**194**:1034–9. doi:10.2214/AJR.09.3528
- 77 Song Y, Skinner J, Bynum J, et al. Regional Variations in Diagnostic Practices. N Engl J Med 2010;363:45–53. doi:10.1056/NEJMsa0910881
- 78 Cadogan SL, Browne JP, Bradley CP, et al. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement* Sci 2015;10:167. doi:10.1186/s13012-015-0356-4
- 79 Burgers JS, Fervers B, Haugh M, *et al.* International Assessment of the Quality of Clinical Practice Guidelines in Oncology Using the Appraisal of Guidelines and Research and Evaluation Instrument. *J Clin Oncol* 2004;22:2000–7. doi:10.1200/JCO.2004.06.157
- Gale EAM. Conflicts of interest in guideline panel members. *BMJ* 2011;**343**.
- 81 IoM C to A the PHS on CPG. Clinical Practice Guidelines: Directions for a New Program. Washington: 1990. doi:10.1097/SPV.0b013e31828a2951
- 82 Browman GP, Snider A, Ellis P. Negotiating for change. The healthcare manager as catalyst for evidence-based practice: changing the healthcare environment and sharing experience. *Healthc Pap* 2003;**3**:10–22.http://www.ncbi.nlm.nih.gov/pubmed/12811083 (accessed 7 Nov 2016).

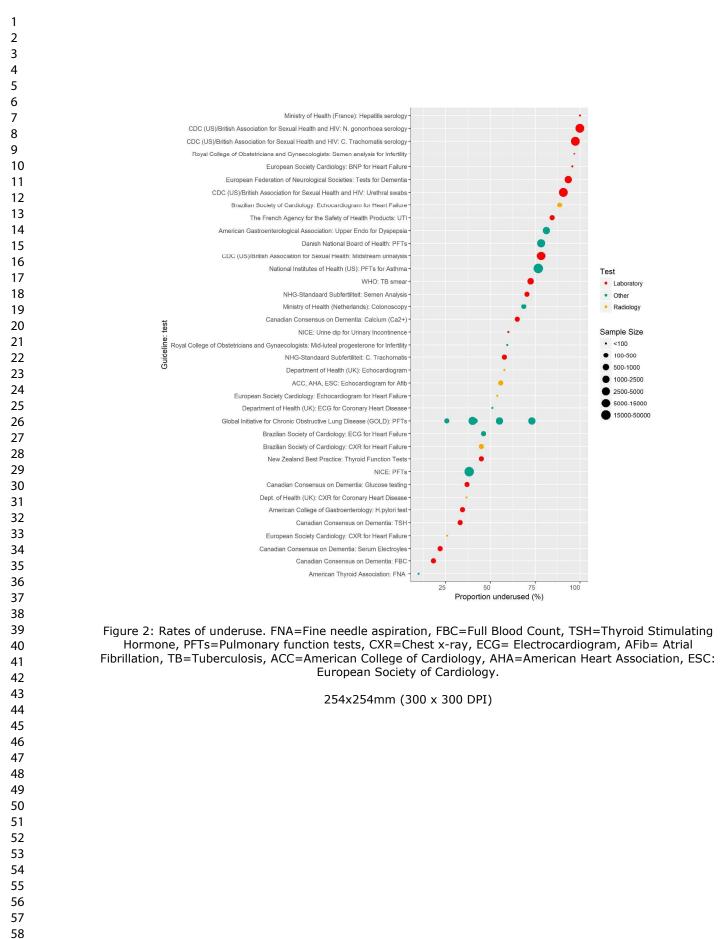
Figure legends

Figure 1: PRISMA Flow Diagram

Figure 2: Rates of underuse. FNA=Fine needle aspiration, FBC=Full Blood Count, TSH=Thyroid Stimulating Hormone, PFTs=Pulmonary function tests, CXR=Chest x-ray, ECG= Electrocardiogram, AFib= Atrial Fibrillation, TB=Tuberculosis, ACC=American College of Cardiology, AHA=American Heart Association, ESC: European Society of Cardiology.

Figure 3: Rates of overuse. NHMRC= National Health and Medical Research Council, U/S=Ultrasound





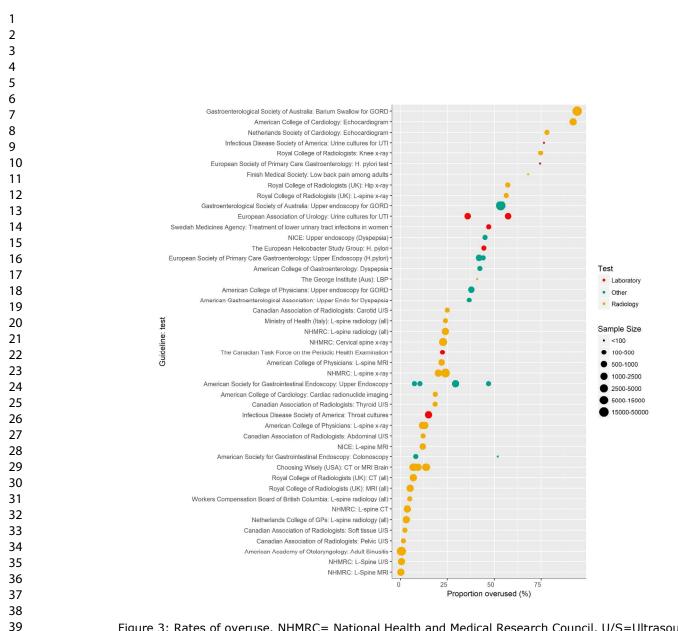


Figure 3: Rates of overuse. NHMRC= National Health and Medical Research Council, U/S=Ultrasound

254x254mm (300 x 300 DPI)

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1	
2 3	
4 5	
6	
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	

60

### MEDLINE Search Strategy

1. Ambulatory Care/
2. exp Ambulatory Care Facilities/
3. general practice/ or family practice/
4. general practitioners/ or physicians, family/ or physicians, primary care/
5. Primary Health Care/
6. Office visits/
7. Academic Medical Centers/
8. (ambulatory adj3 (care or setting? or facilit* or ward? or department? or
service?)).ti,ab.
9. ((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.
10. (primary care or primary health care or primary healthcare or family
medicine or community medicine or community health).ti,ab.
11. (gp or gps).ti,ab.
12. (after hour? or afterhour? or "out of hour?" or ooh).ti,ab.
13. (clinic? or visit?).ti,ab.
14. ((health* or medical) adj2 (center? or centre?)).ti,ab.
15. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16. exp Emergency Service, Hospital/
17. Emergency Medical Services/
18. (emergency adj3 (care or setting? or facilit* or ward? or department? or
service? or room?)).ti,ab.
19. (emergency medicine or ed or er or a&e).ti,ab.
20. 16 or 17 or 18 or 19
21. 15 or 20
22. guidelines as topic/ or practice guidelines as topic/
23. (guideline? or guidance?).ti,ab.
24. 22 or 23
25. (adhere* or non-adhere* or nonadhere* or concord* or non-concord* or
nonconcord* or discord* or comply or complian* or non-complian* or
noncomplian* or align* or nonalign* or nonalign* or congruen* or incongruen*
or consisten* or inconsisten* or contradict*).ti,ab.
26. ((does or "does not" or doesn?t or did or "did not" or didn?t or "not" or fail*)
adj3 (follow* or met or meet or meeting or match or matching or "in line
with")).ti,ab.
27. ((follow* or met or meet or meeting or match or matching or "in line with" or
keep or kept or keeping or utili?ation or utile?e? or change?) adj5 (criteria or
recommend* or guideline? or guidance)).ti,ab.
28. Physician's Practice Patterns/
29. clinical competence/ or nursing competence/
30. 25 or 26 or 27 or 28 or 29
31. 24 and 30
32. Guideline Adherence/
33. 31 or 32
34. exp *"diagnostic techniques and procedures"/
35. exp "diagnostic techniques and procedures"/ut
36. (diagnos* or detect* or test* or screen* or manag*).ti.

2	
2 3	37. (imaging or radiolog* or tomogra* or ct scan* or pet scan* or echocardiogra*
4	or angiogra* or ultrasound* or ultrasonogra*).ti,ab.
5	38. ((medical or clinical or diagnos <sup>*</sup> or screening or routine or laboratory) adj5
6	(test* or investigation?)).ti,ab.
7	
8 9	39. ((h?ematolog* or blood or urin* or saliva*) adj5 test*).ti,ab.
9 10	40. ((stress* or physical or function*) adj5 test*).ti,ab.
11	41. 34 or 35 or 36 or 37 or 38 or 39 or 40
12	42. 21 and 33 and 41
13	43. ((necessary or unnecessary or appropriate* or inappropriate* or waste* or
14	utili?ation or indicated or excess* or less or more or increas* or decreas*) adj10
15	(test* or imaging or radiolog* or tomogra* or ct scan* or pet scan* or
16	echocardiogra* or angiogra* or ultrasound* or ultrasonogra* or
17	investigation?)).ti,ab.
18	44. ((order* or request*) adj5 (test* or imaging or radiolog* or tomogra* or ct
19	scan* or pet scan* or echocardiogra* or angiogra* or ultrasound* or
20	
21 22	ultrasonogra* or investigation?)).ti,ab.
22	45. Unnecessary Procedures/
24	46. 43 or 44 or 45
25	47. 21 and 24 and 46
26	48. 21 and 41 and 45
27	49. 42 or 47 or 48
28	50. limit 49 to yr="1999 -Current"
29	51. limit 50 to english language
30	52. exp animals/ not humans.sh.
31	53. 51 not 52
32	
33 34	EMBASE Search Strategy
34 35	ENDIDE Bear en berategy
36	1. Ambulatory Care/
37	
38	2. general practice/
39	3. general practitioners/
40	4. Primary Health Care/
41	5. (ambulatory adj3 (care or setting? or facilit* or ward? or department? or
42	service?)).ti,ab.
43	<ol><li>((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.</li></ol>
44 45	7. (primary care or primary health care or primary healthcare or family medicine
45 46	or community medicine or community health).ti,ab.
47	8. (gp or gps).ti,ab.
48	9. (after hour? or afterhour? or "out of hour?" or ooh).ti,ab.
49	10. (clinic? or visit?).ti,ab.
50	11. ((health* or medical) adj2 (center? or centre?)).ti,ab.
51	12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
52	
53	13. Emergency Ward/
54	14. (emergency adj3 (care or setting? or facilit* or ward? or department? or
55	service? or room?)).ti,ab.
56 57	15. (emergency medicine or ed or er or a&e).ti,ab.
57 58	16. 13 or 14 or 15
50 59	17. 12 or 16
60	18. *practice guideline/

2	
3	10 (mideline) en miden ee) tiek
4	19. (guideline? or guidance?).ti,ab.
5	20. 18 or 19
6	21. (adhere* or non-adhere* or nonadhere* or concord* or non-concord* or
7	nonconcord* or discord* or comply or complian* or non-complian* or
8	noncomplian* or align* or nonalign* or nonalign* or congruen* or incongruen*
9	or consisten* or inconsisten* or contradict*).ti,ab.
10 11	22. ((does or "does not" or doesn?t or did or "did not" or didn?t or "not" or fail*)
12	adj3 (follow* or met or meet or meeting or match or matching or "in line
13	with")).ti,ab.
14	23. ((follow* or met or meet or meeting or match or matching or "in line with" or
15	keep or kept or keeping or utili?ation or utile?e? or change?) adj5 (criteria or
16	recommend* or guideline? or guidance)).ti,ab.
17	24. clinical competence/ or nursing competence/
18	25. 21 or 22 or 23 or 24
19	26. 20 and 25
20 21	27. diagnostic procedure/ or exp blood examination/ or exp cardiovascular
22	system examination/ or exp digestive system examination/ or exp endocrine
23	system examination/ or exp neurologic examination/ or exp respiratory tract
24	
25	examination/ or exp urogenital system examination/
26	28. (diagnos* or detect* or test* or screen* or manag*).ti.
27	29. (imaging or radiolog* or tomogra* or ct scan* or pet scan* or echocardiogra*
28	or angiogra* or ultrasound* or ultrasonogra*).ti,ab.
29 30	30. ((medical or clinical or diagnos* or screening or routine or laboratory) adj5
31	(test* or investigation?)).ti,ab.
32	31. ((h?ematolog* or blood or urin* or saliva*) adj5 test*).ti,ab.
33	32. ((stress* or physical or function*) adj5 test*).ti,ab.
34	33. 27 or 28 or 29 or 30 or 31 or 32
35	34. 17 and 26 and 33
36	35. ((necessary or unnecessary or appropriate* or inappropriate* or waste* or
37 38	utili?ation or indicated or excess* or less or more or increas* or decreas*) adj10
39	(test* or imaging or radiolog* or tomogra* or ct scan* or pet scan* or
40	echocardiogra* or angiogra* or ultrasound* or ultrasonogra* or
41	investigation?)).ti,ab.
42	36. ((order* or request*) adj5 (test* or imaging or radiolog* or tomogra* or ct
43	scan* or pet scan* or echocardiogra* or angiogra* or ultrasound* or
44	ultrasonogra* or investigation?)).ti,ab.
45 46	37. Unnecessary Procedures/
46 47	38. 35 or 36 or 37
48	39. 17 and 20 and 38
49	40. 17 and 33 and 37
50	41. 34 or 39 or 40
51	42. limit 41 to yr="1999 -Current"
52	43. limit 42 to english language
53	44. (exp animals/ or nonhuman/) not human/
54 55	45. 43 not 44
55 56	46. conference*.pt.
57	•
58	47. 45 and 46
59	48.45 not 46
60	49. exp child/ not (exp Child/ and exp Adult/)

1	
2	
3	50. 48 not 49
4	51. 48 not 49
5	
6	52. limit 47 to yr="2015 -Current"
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	
57	
58	
59 60	
UO	

Page 24 of 4	42
--------------	----

	Was the study"s target population a close representation of the national population in relation to relevant variables?	Does the inclusion criteria match the target population of guideline?	Were all eligible participants included in the study?	Was the likelihood of non- response bias <20?	Was an acceptable disease, test or symptom definition used?	Was data extracted/collected in an objective way?	Was the interval from symptoms to test clinically appropriate for the diagnosis of interest?	Did they report extractable measures?	Other bias?
Ahmad2012	Low	Unclear	Low	Unclear	Low	Unclear	Unclear	Low	Low
Aljebreen 2013	Low	Low	Low	Low	Unclear	Unclear	Low	Low	High
Azzam 2015	Low	Low	Unclear	High	Low	Unclear	Low	Low	Low
Belletti2013	Low	Unclear	Unclear	Unclear	Low	Low	Low	Low	Low
Bertella 2013	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Low
Bhatt 2001	Low	Low	Low	High	High	Unclear	Unclear	Low	High
Birk-Urovitz 2017	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Bishop 2003	High	Low	Low	Low	Low	Low	Low	Low	Low
Cai 2015	Low	Low	Unclear	Low	Unclear	Unclear	Unclear	Low	Low
Caplan 2000	Low	Low	Low	Low	Unclear	Unclear	Unclear	Low	Low
Cardin 2005	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Cardin 2007	Low	Low	Low	Unclear	Low	Low	Low	Low	Low
Chan 2004	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Chan 2006	Low	Low	Low	High	Unclear	Low	Unclear	Low	Low
Chavez 2009	Low	Low	Low	Low	Low	Low	High	Low	High
Droogendijk 2011	Unclear	Low	High	Unclear	Low	Unclear	Low	Low	Low
Eccles 2001	Low	Low	Low	Low	Unclear	Unclear	Unclear	Low	Low
Elwyn 2007	Low	Low	Unclear	Unclear	Unclear	Low	Low	High	High

+00	puc	+70	+ 2+	P040
loou	เวรอ	6ou	Isnu	usei

2002								1
lp2014	Low	Low	High	Unclear	Low	Unclear	Unclear	Low
Johnson 2011	Low	Unclear	Unclear	Low	High	Low	Low	Low
Kinouani 2017	Low	Low	low	Low	Unclear	Yes	Unclear	Low
Kovacs 2013	Low	Unclear	High	Low	Low	Low	Unclear	Low
Lalude 2014	Low	Low	Low	Low	High	Low	Unclear	Low
Landry 2011	Low	Unclear	Low	Low	Unclear	Unclear	Unclear	Low
Lange 2007	Low	Low	Unclear	Low	Unclear	Unclear	Unclear	Low
Lin 2016	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low
Linder 2006	Low	High	Unclear	Low	Low	Low	Low	Unclear
Lipczynska 2012	High	High	Unclear	Low	Low	Unclear	Low	Low
Llor 2011	Low	Low	Low	High	Low	Low	Low	Low

Fiorenza 2017	Low	Low	Low	Unclear	Low	Low	Unclear	Low	Low
Gerrits2008	Low	Unclear	High	Low	Low	Low	Unclear	Low	Low
Gibbons 2010c	Low	Low	Low	Low	Low	High	Low	Low	Low
Girard 2010	High	Low	Unclear	High	Unclear	High	Unclear	Low	High
Gnani 2004	Low	Low	Unclear	Low	Unclear	Low	Unclear	Low	Low
Grover 2007	Low	Low	Unclear	Low	Low	High	Unclear	Low	Low
Gurzun 2014	Low	Low	High	Low	High	Low	Unclear	Low	High
Hassan 2007	Low	Low	Low	High	Unclear	Low	Unclear	Low	Low
Heidi Lindbäck 2017	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Hughes- Anderson 2002	High	Low	Unclear	Low	Unclear	Unclear	unclear	Low	Low
lp2014	Low	Low	High	Unclear	Low	Unclear	Unclear	Low	High
Johnson 2011	Low	Unclear	Unclear	Low	High	Low	Low	Low	Low
Kinouani 2017	Low	Low	low	Low	Unclear	Yes	Unclear	Low	Low
Kovacs 2013	Low	Unclear	High	Low	Low	Low	Unclear	Low	High
Lalude 2014	Low	Low	Low	Low	High	Low	Unclear	Low	High
Landry 2011	Low	Unclear	Low	Low	Unclear	Unclear	Unclear	Low	Low
Lange 2007	Low	Low	Unclear	Low	Unclear	Unclear	Unclear	Low	Low
Lin 2016	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low	Low
Linder 2006	Low	High	Unclear	Low	Low	Low	Low	Unclear	Low
Lipczynska 2012	High	High	Unclear	Low	Low	Unclear	Low	Low	High
Llor 2011	Low	Low	Low	High	Low	Low	Low	Low	Low

BMJ Open: first published as 10.525, 2025 at Department GEZ-LTA. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA. Et action of the standard of t

Protected by copyright

7-1-2	P		<b>T</b> ~	71		
looy	iosə	6ou	Isn	u	isi	sı

diainiasi di	8.,eninim,etab)	brejxstotbetek	jf/sðsni jój (bu	ibulati <sub>xu</sub>
	1 1001	aca6aucaucau		

Loo 2009	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Mafi2013	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low	Low
Mafi2015	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low	Low
Majumdar 2003	Low	Low	Unclear	Low	Low	Low	Unclear	Low	Low
Michaleff 2012	Low	Low	Low	Unclear	High	Low	Unclear	Low	Low
Moscavitch 2009	Low	Low	Unclear	Low	Low	Unclear	Low	Low	Low
Musicco 2004	High	Low	Low	Unclear	Unclear	Unclear	Unclear	High	High
Nicholson 2010	Low	Low	Low	Low	Unclear	Low	Low	Low	Low
Nicopoullos 2003	High	High	Low	High	Low	Unclear	Unclear	Low	High
Noya 2008	Low	Low	Low	Low	Unclear	Unclear	Unclear	Low	High
Piccoliori 2013	Low	Low	Low	Low	Low	Unclear	Low	Low	High
Pimlott 2006	Low	Low	Unclear	High	Unclear	Unclear	Unclear	Low	Low
Piterman 2008	Low	Unclear	Low	Unclear	Unclear	Low	Unclear	High	High
Remedios 2014	Low	Unclear	Low	High	Unclear	Low	Unclear	Low	High
Schers 2000	Low	Low	Low	Unclear	Unclear	Low	Unclear	Low	Low
Smith 2008	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Low
Sokol 2015	Low	Low	Low	Low	Low	Low	High	Low	High
Tahvonen 2017	Low	Low	High	Low	Low	Low	Unclear	Low	Low
Ulrik 2010	Low	Low	Low	High	Low	Low	unclear	Low	High
Ulrik 2013	Low	Low	Low	High	Low	Low	unclear	Low	Low
van der Pluijm- Schouten 2017	Low	Low	Low	Unclear	Low	Unclear	Unclear	Low	Low

BMJ Open

van Gurp 2013	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Low
Williams 2010	Low	Low	Unclear	Low	Low	Unclear	Unclear	Low	Low
1-ZƏĐ tuəmtısc	.səipo	londət technol	ngeriquist da <sub>i</sub> x	2nințim,≼țsbiene 	ixaiofbajelar/si Soysnussia	Low Unclear	«ງໍຢູ່ອີຢູ່ນີ້	Protecte	luq fərif :nəqO LMI

	6/bmjopen-2017-018557 9 by copyright, including	Table 1: Study Characteristics				
Test	8557 on 11 uding for u	Population	N (men%)	Study length (days)	Country	Study
	Ses e	Under-use		(uujo)		
Sputum smear microscopy	s suspected	Patients registered at health clinics where TB wa	554 (41%)	181	Indonesia	Ahmad 2012
Pulmonary function tests (PFT)	ed t	Patients with COPD	1517 (46%)	N/S	USA	Belletti 2013
PFTs	te h	Patients with COPD	437 (286)	1765	Italy	Bertella 2013
FNA of thyroid	S SUSArted to text and d	Patients with a thyroid nodule	81	365	USA	Caplan 2000
PFT	vnlo sch	Patients with COPD	200 (48%)	2920	USA	Chavez 2009
Upper endoscopy and colonoscopy	iciener 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Women >50yrs and men >18 years with Iron Def	287 (45%)	730	Netherlands	Droogendijk 2011
Urine dipstick	urinar zin dintinence	Women aged 18 – 65 yrs with newly diagnosed u	65 (0%)	2556	Netherlands	Gerrits 2008
Free T4	19, /	Patients with subclinical hypothyroidism	265	364	New Zealand	Gibbons 2010
CXR, ECG and Echocardiogram	L tra	Patients with heart failure	90 (53%)	365	UK	Gnani 2004
Hepatitis serology (HBs antigen anti-HBc anitbodies)	aining	Patients with acute hepatitis	19 (37%)	28	France	Girard 2010
Urine Dipstick	, an	Patients with urinary tract infections	61 (18%)	150	France	Kinuoani 2017
PFTs	d si	Patients with COPD	2549 (44%)	91	Denmark	Lange 2007
Echocardiogram, BNP, CXR	ors <b>milar</b>	Aged ≥ 55 with Heart Failure (HF) or HF risk factor	93	61	Poland	Lipczynska 2012
Echocardiogram	on	Patients with Atrial Fibrillation	131 (50%)	364	UK	Loo 2009
Upper endoscopy		Patients >50 years, on Proton Pump Inhibitors wi dyspepsia	531 (47%)	2371	USA	Majumdar 2003a
H.pylori	22, 202 29 jies.	Patients with peptic ulcer disease (PUD)	132 (47%)	2371	USA	Majumdar 2003b
ECG, CXR, Echocardiogram	25 at [	Patients with Heart Failure	167 (43%)	61	Brazil	Moscavitch 2009
Collection of laboratory tests to rule out conditions with similar	Departm	Patients being assessed for Dementia	1549 (38%)	NR	Italy	Musicco 2004

 Page 28 of 42

Page	29	of	42
------	----	----	----

#### BMJ Open

				BMJ Open BMJ Open-201	
Nicholson 2010	UK	1827	6943 (100%)	Men with epididymo-orchtitis	C. trachomatis, N. gonorrho urethral swabs and midstre- urinalysis.
Nicopoullos 2003	UK	242	32	Patients with subfertility	Mid-luteal progesterone and semen analysis
Pimlott 2006	Canada	1611	160 (34%)	Patients with Dementia	FBC, TSH, serum electroly serum calcium, glucose
Smith 2008	UK	731	29870 (52%)	Patients with COPD	PFT
Sokol 2015	USA	3652	75902 (23%)	Patients with Dementia     estimate       Patients with COPD     attant 2018       Patients with Asthma     text of using estimate       Patients with COPD     and store       Patients with COPD     and store	PFT
Ulrik 2010	Denmark	121	1716 (44%)	Patients with COPD	PFT
Ulrik 2013	Denmark	731	4058	Patients with COPD	PFT
van der Pluijm- Schouten 2017	Netherlands	840	100%*	Patients (couples) referred to IVF clinics	Chlamydia Antibody Titre a Semen Analysis
				<u>Over-use</u>	
Aljebreen 2013	Saudi Arabia	365	147 (51%)	Patients who had upper endoscopy	Upper endoscopy
Azzam 2015	Saudi Arabia	121	161 (30%)	Dyspeptic patients who had upper endoscopy	Upper endoscopy
Bhatt 2001	UK	504	437 (65%)	Patients referred for pelvis x-rays	Pelvis x-ray
Birk-Urovitz 2017	Canada	1538	77 (38%)	Patients that had a Thyroid Stimulating Hormone (TSH) test	TSH
Bishop 2003	Canada	28	139	Patients with non-red flag LBP	Advanced imaging (CT, MF bone scan)
Cai 2015	USA	121	550 (46%)	Patients who under went upper endoscopy	Upper endoscopy
Chan 2004	Malaysia	153	250 (45%)	Patients who under went upper endoscopy	Upper endoscopy
Chan 2006	Malaysia	184	27 (63%)	Patients who underwent 'diagnostic colonoscopies'	Colonoscopy
Cardin 2005	Italy	151	1678	Patients who had upper endoscopy	Upper endoscopy
Cardin 2007	Italy	182	NR	Dyspeptic patients who had upper endoscopy	Upper endoscopy
Eccles 2001	UK	182	275	Patients who had knee or lumbar x-ray	Lumbar or knee x-ray
Elwyn 2007	UK	184	215	Patients who under went upper endoscopy	Upper endoscopy
Fiorenza 2017	USA	456	45 (34%)	Patients who under went upper endoscopy	Upper Endoscopy
				GEZ-LTA	

				BMJ Open BMJ Open BMJ Open BMJ Open BMJ Open	Pag
				in Z	
Grover 2007	USA	364	68 (0%)	Patients with uncomplicated UTI	Urine culture and sensitivity analysis
Gurzun 2014	UK	7	1070 (54%)	Patients who underwent an echocardiogram	Echocardiogram
Hassan 2007	Italy	30	3769 (46%)	Patients who underwent upper endoscopy	Upper endoscopy
Hughes- Anderson 2002a	Australia	1613	154 (55%)	Patients who had colonoscopy     Ferrary       Patients who had upper endoscopy,     Ferrary       Patients with non-red flag LBP     Ferrary       Patients with uncomplicated UTI     Ferrary       Patients with non-red flag LBP     Ferrary       Patients with non-red flag LBP     Ferrary       Patients with non-red flag LBP     Ferrary	Colonoscopy
Hughes- Anderson 2002b	Australia	1613	162 (55%)	Patients who had upper endoscopy,	Upper endoscopy
lp 2014	USA	1096	100 (43%)	Patients with non-red flag LBP	MRI lumbar spine
Johnson 2011	USA	510	779 (0%)	Patients with uncomplicated UTI	Urine culture
Kovacs 2013	Spain	183	602 (48%)	Patients with non-red flag LBP	MRI lumbar spine
Lalude 2014	USA	121	102	Patients who had SPECT Myocardial perfusion imaging (MPI) studies	Single Photon Emission CT (SPECT) MPI
Landry 2011	USA	272	124	Patients who had U/S of thyroid, pelvis, abdo, carotid gr soft tissue	Thyroid, pelvis, abdomen, card or soft tissue ultrasound
Lin 2016	Australia	NR	NR	Patients with non-red flag LBP	Lumbar Spine X-ray
Lindbäck 2017	Sweden	59	0	Patients that had urinary cultures	Urinary Culture
Linder 2006	USA	608	1076 (19%)	Patients with pharyngitis	Strep testing (rapid antigen detection test, throat culture)
Llor 2011	Spain	122	658 (0%)	Women with UTI	Urine cultures
Mafi 2013	USA	4377	8066	Patients with non-red flag LBP	X-ray, CT or MRI
Mafi 2015	USA	4018	9362 (25%)	Patients with uncomplicated headache (non-red flag	CT and MRI
Michaleff 2012	Australia	3621	3070 (70%)	Patients reporting first time neck pain or LBP (non-specific non red flag)	Any radiological test
Noya 2008	Israel	N/S	209 (35%)	Patients who had H.pylori testing	H. pylori test
Piccoliori 2013	Italy	63	475 (43%)	Acute or chronic non-red flag LBP	Any radiological test
Piterman 2008	Australia	550	19219	Patients with GORD	Endoscopy. Barium Swallow
Remedios 2014	UK	NR	2026	Patients who had CTs and/or MRIs	CT and/or MRI
Sharp 2015	USA	730	37,464	Patient with Acute Sinusitis	CT Sinuses
				GEZ-LTA	

Page 31 of 42	BMJ Open BMJ Open-201						
1 2					yright, in	-201	
3	Schers 2000	Netherlands	214	1096 (50%)	Patients with non-red flag LBP	X-ray	
4 5	Tahvonen 2017	Finland	180	18 (35%)	Patients with non-red flag LBP	ر Lumbar Spine X-ray	
6 7	Van Gurp 2013	Netherlands	366	155 (38%)		Echocardiogram	
8	Williams 2010	Australia	1005	1706 (43%)	Patients who had Echocardiogram Patients with non-red flag LBP	All imaging	
9 10	*Both men and wo	men			່ ຕີ ຊ		
11					ted	asm	
12 13					o te	18.1	
14					X ar	oges	
15					a a		
16 17							
18					mining, Al training, and similar technologies.	Tro	
19					ing,	3	
20 21					A t	₩.	
22					in in in it is a set of the set o	//bm	
23					ing	- Jop	
24 25					and a second sec	http://bmjopen.bmj.com/ on May 22.	
26						- <u>ă</u>	
27						- Sector - S	
28 29					r teo	<b>v</b>	
30							
31						1Y 2	
32					yi. es		
33 34						2025	
35						at	
36						Dep	
37						ba nt	
38						Te	
39						<u>1</u>	
40 41						ЭЕ Z	
41						GEZ-LTA	
43			E	or neer review	only - http://bmjopen.bmj.com/site/about/guidelines.xhtml		
44			Г	or heer review (	shiy - http://bhijopen.bhij.com/site/about/guidelines.xhtml		
45							
46							

		BMJ Open BMJ Open Measures of inappropriateness Guideline authority and recommendation of	
	Table 2: N	leasures of inappropriateness	
Study	Test	Guideline authority and recommendation ວິ ອີ	Measure of inappropriateness (959
		Under-use	
Girard 2010	Hepatitis B serology	Ministry of Health (France): Hepatitis serol	100% (82.4 to 100%
Nicholson 2010	Neisseria Gonorrhoea serology	CDC (US)/British Association for Sexual Heater and HIV: Test for N. gonorrhoea for suspected of Epididymitis	99.9% (99.85 to 99.98
Nicholson 2010	Chlamydia Trachomatis	CDC (US)/British Association for Sexual Health and HIV: Test for C. Trachomatis for suspected a Epididymitis	97.4% (97.0 to 97.8%
Nicopoullos 2003	Semen Analysis	Royal College of Obstetricians and Government Gynaecologists: Semen analysis for Infedility	96.9% (83.8 to 99.9%
Lipczynska 2012	Brain Natriuretic Peptide (BNP)	European Society Cardiology: BNP for Heater Failure	95.7% (89.4 to 98.8%
Musicco 2004	Collection of laboratory tests	European Federation of Neurological Societies Collection of laboratory tests to rule out conditions with similar presenting symptoms to demonstration	93.42% (92.1 to 94.6%
Nicholson 2010	Urethral swabs	CDC (US)/British Association for Sexual Health and HIV: Urethral swabs for suspected epidid mits (Urethral swabs)	90.7% (89.9 to 91.3%
Moscavitch 2009	Echocardiogram	Brazilian Society of Cardiology: Echocardio and for Heart Failure	88.6% (82.8 to 93.0%
Kinouani 2017	Urine Dipstick	The French Agency for the Safety of Health Products: Urine Dipstick for UTI	84.4% (80.1 to 88.1%
Majumdar 2003a	Upper Endoscopy	American Gastroenterological Associaten: Appropriate use of Upper Endoscopy for Dyspettia	81.2% (78.8 to 83.4%
Ulrik 2013	Pulmonary function tests (PFTs)	Danish National Board of Health: PFTs to diagnosis COPD	78.3 (77.3% to 79.3%
Nicholson 2010	Mid stream	CDC (US)/British Association for Sexual Health and HIV: Midstream urinalysis for suspected Epididy ritis	78.2 (77.3 to 79.3%)
Sokol 2015	Pulmonary function tests (PFTs)	National Asthma Education and Prevention Program (US): PFTs for asthma	76.5% (64.6 to 85.9%
Belletti 2013	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung	73.0% (70.7 to 75.3%

Page	33	of	42
------	----	----	----

		6/bmjopen-201 4 by copyright, BMJ Open	
Ahmad 2012	Tuberculosis smear	World Health Organisation: Smear for suspected TB	72.4% (68.5 to 76.19
van der Pluijm- Schouten 2017	Semen Analysis	NHG-Standaard Subfertiliteit: Semen Anaysis	70.4% (61.9 to 77.99
Droogendijk 2011	Colonoscopy	Ministry of Health (Netherlands): Colonoscopy for unexplained Iron Deficiency Anaemi	68.6% (62.9 to 74.09
Pimlott 2006	Serum Calcium	Canadian Consensus Conference on Dementa Serum Calcium for Dementia និច្ចពុច្	65.0 (57.1 to 72.4%
Gerrits 2008	Urine dip stick	NICE: Urine dip stick for urinary incontin	60.0% (47.1 to 72.09
Nicopoullos 2003	Mid-luteal progesterone	Royal College of Obstetricians and Gynaecoogenets: Mid-luteal progesterone for Infertility	59.4% (40.6 to 76.39
Gnani 2004	Echocardiogram	Heart Failure 우울등	57.8% (46.1 to 68.19
Loo 2009	Echocardiogram	ACC, AHA, ESC: Echocardiogram to identify a ses or complications of atrial fibrillation	55.7% (46.8 to 64.39
Ulrik 2010	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Eung Disease (GOLD): PFTs for COPD	55.0% (52.6 to 57.49
Lipczynska 2012	Echocardiogram	European Society Cardiology: Echocardiogeam	53.8% (43.1 to 64.29
van der Pluijm- Schouten 2017	Chlamydia Trachomatis	NHG-Standaard Subfertiliteit: Chlamy	57.8% (49.0 to 66.29
Gnani 2004	ECG	Department of Health (UK): ECG for Heart all are	51.1% (40.4% to 61.8
Moscavitch 2009	ECG	Brazilian Society of Cardiology: ECG for Fea	46.1 (38.4 to 54.0)
Moscavitch 2009	Chest X-ray	Brazilian Society of Cardiology: CXR for 🛱 eag Failure	44.9% (37.2 to 52.89
Gibbons 2010	Thyroid function tests	New Zealand Best Practice: Appropriate Thyroid Function tests	44.9% (38.8 to 51.19
Chavez 2009	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Bung Disease (GOLD): PFTs for COPD	41.5% (34.6 to 48.79
Lange 2007	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	40.0% (38.1 to 42.0
Smith 2008	Pulmonary Function Tests (PFTs)	NICE: PFTs for COPD	38.1% (37.5 to 38.6%
Pimlott 2006	Glucose testing	Canadian Consensus Conference on Demente: Glucose testing for Dementia	36.9% (29.4% to 44.9
Gnani 2004	Chest X-ray	Department of Health (UK): CXR for Heart Failure	36.7% (26.8 to 47.59

Page	34	of	42
. uge	<b>.</b>	<u> </u>	

		6/bmjopen-2017 J by copyright, i BMJ Open	
Majumdar 2003b	H.pylori	American Gastroenterological Association/American College of Gastroenterology: appropriateness of H.pylori test	34.4% (28.9 to 40.3%
Pimlott 2006	Thyroid Stimulating Hormone (TSH)	Canadian Consensus Conference on Demeraia: TSH for dementia	33.1% (25.9 to 41.0%
Lipczynska 2012	Chest x-ray (CXR)	European Society Cardiology: CXR for Hear	25.8% (17.3 to 35.9%
Bertella 2013	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive ដំហាច Disease (GOLD): PFTs for COPD គ្គី ខ្លុំ ្	25.6% (21.6 to 30.0%
Pimlott 2006	Serum electrolytes	Canadian Consensus Conference on Denæitia: Serum electrolytes for dementia ਛ	21.9% (15.7 to 29.1%
Pimlott 2006	Full Blood Count (FBC)	Canadian Consensus Conference on Demention BC	18.1% (12.5 to 25.0%
Caplan 2000	Fine needle aspiration (FNA) of thyroid	American Thyroid Association/American Asଲ୍ଲେଟ୍ଟିୟ୍ପion of Clinical Endocrinologists: FNA for thyroid ଅୟାନ୍ଥାରେ	9.9% (4.4 to 18.5%)
		<u>Over-use</u>	
Piterman 2008	Barium Swallow	Gastroenterological Society of Australia:	94.20% (93.9 to 94.5%
Gurzun 2014	Echocardiogram	American College of Cardiology: Appropriate use of Echocardiography	92.0% (90.2% to 93.5
van Gurp 2013	Echocardiogram	Netherlands Society of Cardiology: Appropriate se of Echocardiography	76.7% (76.4 to 77.0%
Grover 2007	Urine cultures	Infectious Disease Society of America: Urine cultures not required for uncomplicated UTI diagrosise.	76.5% (64.6 to 85.9%
Eccles 2001	Knee x-ray	Royal College of Radiologists: No x-ray for kaee an without restriction of movement	74.7% (69.6 to 79.3%
Cardin 2005	H. Pylori breath test	European Society of Primary Care Gastroen broggy: Appropriate use of H. pylori 3	74.4% (58.8 to 86.5%
Tahvonen 2017	L spine x-ray	Finish Medical Society: LBP among addits	68.0% (53.3 to 80.489
Johnston 2011	Urine cultures	European Association of Urology: Urinary of Itures not required for uncomplicated urinary tract of infections	57.4% (53.8 to 60.9%
Bhatt 2001	Hip x-ray	Royal College of Radiologists (UK): No hip x-raufor hip pain without restriction of movement	57.2% (52.5 to 61.8%
Eccles 2001	Lumbar spine x-ray	Royal College of Radiologists (UK): no x-ray for	56.4% (50.3 to 62.3%
		nt GEZ-LTA	

		6/bmjopen-2017- 9 by copyright, ir BMJ Open	
Piterman 2008	Upper endoscopy	Gastroenterological Society of Australia: Epper endoscopy for GORD American Society for Gastrointestinal Endecopy:	53.5% (52.8 to 54.
Chan 2006	Colonoscopy	American Society for Gastrointestinal Endecopy: Appropriateness of Colonoscopy	51.9% (32.0 to 71.
Heidi Lindbäck 2017	Urine Culture	Swedish Medicines Agency: Treatment of sower urinary tract infections in women 🕱 📮	47.0% (40.8 to 53.3
Aljebreen 2013	Upper endoscopy	The American Society for Gastrointesting	46.9% (38.7 to 55.
Elwyn 2007	Upper endoscopy	NICE: Appropriate tests for dyspepsie	45.1% (38.3 to 52.
Noya 2008	H.Pylori	The European Helicobacter Study Gro 65 여 여 여 여 여 여 여 여 여 여 여 여 여 여 여 여 여 여	44.5 (37.6 to 51.5
Cardin 2005	Upper endoscopy	European Society of Primary Care Gastroen and Society of Primary Care Gastroen	44.1% (35.9 to 52.
Lin 2016	L spine x-ray	Upper Endoscopy for H.pylori	40.9% (26.3 to 56.
Cardin 2007	Upper endoscopy	European Society of Primary Care Gastroen Ero gg: Upper Endoscopy for H.pylori	41.9% (38.3 to 45.
Fiorenza 2017	Upper Endoscopy	American College of Gastroenterology: Dyspepaia	42.4% (36.8 to 48.
Cai 2015	Upper endoscopy	American College of Physicians: Upper end scorpy for GORD	37.7% (33.8 to 42.
Azzam 2015	Upper endoscopy	American Gastroenterological Association Upper	36.7 (29.2 to 44.6
Llor 2011	Urine cultures	European Association of Urology: Urinary cultures not required for uncomplicated urinary tract of infections	35.9% (32.2 to 40.
Hassan 2007	Upper endoscopy	The American Society for Gastrointestinal Endoscopy: Appropriateness of Upper Endescopy	29.4% (28.0 to 30.
Landry 2011	Carotid ultrasound	Canadian Association of Radiologists 2005	25.0% (17.7 to 33.
Piccoliori 2013	Lumbar spine radiology (all)	Ministry of Health (Italy): No imaging for nongred flag LBP	24.0% (20.2 to 28.
Michaleff 2012	Lumbar spine x-ray	National Health and Medical Research Council (Australia) (NHMRC): No x-ray for non-red flag	24.0% (22.9 to 25.
Williams 2010	Lumbar spine radiology (all)	National Health and Medical Research Cound (Australia) (NHMRC): No imaging for non-red flag LBP	23.9% (21.9 to 26.
		nt GEI	
		EZ-LTA	

Page	36	of 42	
rage	50	01 42	

		6/bmjopen-2017 4 by copyright, i BMJ Open	
Michaleff 2012	Cervical spine x-ray	بة ع ت 7 Australian National Health and Medical Research	22.8% (21.3 to 24.3%)
		Council: No x-ray for neck pain	
Birk-Urovitz 2017	Thyroid Stimulating Hormone	The Canadian Task Force on Preventative Heath	22.4% (16.9 to 28.8%)
lp 2014	Lumbar spine MRI	American College of Physicians/American Pain Society: no MRI for non-red flag LBR	22.0% (14.3 to 31.4%)
Williams 2010	Lumbar spine x-ray	National Health and Medical Research Council (Australia) (NHMRC): No x-ray for non-red ដ្រើឮឆ្និBP	20.2% (18.3 to 22.2%)
Landry 2011	Thyroid ultrasound	Canadian Association of Radiologists 2	19.0% (12.1 to 27.0%)
Lalude 2014	Single Photon Emission Computed Tomography	American College of Cardiology: SPECT for chest pain	18.6% (11.6 to 27.6%)
Linder 2006	Streptococcal throat cultures	American College of Physicians/Infectious E	15.0 (12.9 to 17.2%)
ʻMafi 2013	Lumbar spine x-ray	American College of Physicians/Americarத் Paத Society: no x-ray for non-red flag LBP: 2009-2010	13.0% (11.1 to 15.1%)
		2007-2008 ni rom	12.9% (11.1 to 14.9%)
		2005-2006 ≥	12.8% (11.0 to 14.8%)
		2003-2004 Tain b	12.3% (10.7 to 14.0%)
		ي 2001-2002 ي و	12.0% (10.3 to 13.8%)
		1999 - 2000 and b	11.8% (10.2 to 13.6%)
Landry 2011	Abdominal ultrasound	Canadian Association of Radiologists 2005 guidelines: Abdominal U/S	12.1% (6.9 to 19.2%)
Mafi 2015	CT or MRI Brain	The American Headache Society/American Academy of Neurology for Choosing WiselyoNo-CT or MRI for non-red flag headache 2009 - 2015	13.9% (12.2 to 15.7%)
		2007 – 2008 <b>5</b> <u>5</u> <u>5</u> <u>5</u> <u>5</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u> <u>7</u>	13.5% (11.8 to 15.3%)
		2005 – 2006 <sup>8</sup> 8	9.4% (8.0 to 11.0%)
		2003 – 2004 ar	7.5% (6.3 to 8.9%)
		2001 – 2002 pg 1999 - 2000 m	7.1% (5.9 to 8.4%)
			6.7% (5.4 to 8.2%)
Kovacs 2013	Lumbar spine radiology tests (all)	NICE, ACP: No imaging for LBP	12.0% (9.5 to 14.8%)
Kovacs 2013	Lumbar spine radiology tests (all)	NICE, ACP: No imaging for LBP	12.0% (9.5 to 14.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Chan 2004	Upper endoscopy	The American Society for Gastrointestigal	10.4% (6.9 to 14.9
Hughes-Anderson	Colonoscopy	Endoscopy: Appropriateness of Upper Endescopy: American Society for Gastrointestinal Endescopy:	8.2% (5.3 to 12.1
2002a Hughes-Anderson 2002b	Upper endoscopy	Appropriateness of Colonoscopy 🤤 🚬 The American Society for Gastrointestigal 🗎 Endoscopy: Appropriateness of Upper End	7.5% (4.7 to 11.1
	CT (any)	Royal College of Radiologists (UK): Ca	6.9% (5.8 to 8.19
Remedios 2014 —	MRI (any)	Royal College of Radiologists (UK): M	5.2% (4.1 to 6.5%
Bishop 2003	Lumbar spine radiology tests (all)	Workers Compensation Board of British Coណ៍គ្នាឆ្លាំឆ្ល	5.0% (2.1 to 10.1
Williams 2010	Lumbar spine CT	No imaging for non-red flag LBP 폭호 National Health and Medical Research C알였여 (Australia) (NHMRC): No CT for non-red flag을 타	3.7% (2.9 to 4.79
Schers 2000	Lumbar spine radiology tests (all)	The Netherlands College of General Practing Seas: No imaging for non-red flag LBP	3.1% (2.2 to 4.3)
Landry 2011	Soft tissue ultrasound	Canadian Association of Radiologists 2005 guidelines: Soft tissue U/S	2.4% (0.5 to 6.9
Landry 2011	Pelvic ultrasound	Canadian Association of Radiologists 2005	1.6% (0.2 to 5.7
Sharp 2015	CT Sinuses	American Academy of Otolaryngology: Adulg Sinusitis	0.60% (0.56 to 0.6
Williams 2010	Lumbar spine Ultrasound	National Health and Medical Research Council (Australia) (NHMRC): No U/S for non-red flag LaP	0.59% (0.28 to 1.
Williams 2010	Lumbar spine MRI	National Health and Medical Research Council (Australia) (NHMRC): No MRI for non-red flag LBP	0.18% (0.04 to 0.
		m/ on May 22, 2025 at Department ar technologies.	

n/ on May 22, 2025 at Department GEZ-LTA r technologies.

Reporting Criteria	Reported (Yes/No)	Reported on Page
Reporting of Background		
Problem definition	YES	4
Hypothesis statement	YES	4
Description of Study Outcome(s)	YES	4
Type of exposure or intervention used	N/A	N/A
Type of study design used	YES	5, 6
Study population	YES	5
Reporting of Search Strategy		
Qualifications of searchers (eg, librarians	YES	5
and investigators)	TES	5
Search strategy, including time period	YES	5 gunnlementer (fil
	TES	5, supplementary fil
included in the synthesis and keywords Effort to include all available studies,	VEQ	F
	YES	5
including contact with authors		F
Databases and registries searched	YES	5
Search software used, name and	YES	5
version, including special features used		
(eg, explosion)		
Use of hand searching (eg, reference	YES	5
lists of obtained articles)	110	
List of citations located and those	NO	
excluded, including justification		
Method for addressing articles	NO	
published in languages other than		
English	1/50	
Method of handling abstracts and	YES	5
unpublished studies		
Description of any contact with authors	N/A	
Reporting of Methods	) (F.O.	
Description of relevance or	YES	5,6
appropriateness of studies assembled for		
assessing the hypothesis to be tested		
Rationale for the selection and coding of	YES 🥥	6
data (eg, sound clinical principles or		
convenience)		
Documentation of how data were	YES	6
classified and coded (eg, multiple raters,		
blinding, and interrater reliability)		
Assessment of confounding (eg,	N/A	N/A
comparability of cases and controls in		
studies where appropriate		
Assessment of study quality, including	YES	5,6
blinding of quality assessors;		
stratification or regression on possible		
predictors of study results		
Assessment of heterogeneity	YES	6

# MOOSE Statement - Reporting Checklist for Authors Editors and

1 2 3 4 5 6		
7 8 9 1 1 1	1 2	
1. 1. 1. 1. 1. 1. 1. 1. 1.	4 5 6 7 8	
2 2 2 2 2 2	0 1 2 3 4 5	
2 2 2 2 3 3 3 3 3	7 8 9 0 1	
3 3 3 3 3 3 3 3	3 4 5 6 7 8	
3 4 4 4 4 4 4	0 1 2 3 4	
4 4 4 5 5	6 7 8 9	
5 5 5 5 5 5 5 5	2 3 4 5 6 7	

Description of statistical methods (eg,	YES	6
complete description of fixed or random		
effects models, justification of whether		
the chosen models account for predictors		
of study results, dose-response models,		
or cumulative meta-analysis) in sufficient		
detail to be replicated	VEC	Tablas 1.2 Figure
Provision of appropriate tables and	YES	Tables 1,2, Figure
graphics Reporting of Results		2,3,4
	YES	Table 1 and Table
Table giving descriptive information for each study included	TES	
Results of sensitivity testing (eg,	N/A	7, 8
subgroup analysis)	IN/ <i>P</i>	7,0
Indication of statistical uncertainty of	YES	6,7, 8,9
findings	TEO	0,7, 0,0
Reporting of Discussion		
Quantitative assessment of bias (eg,	YES	8,9
publication bias)		
Justification for exclusion (eg, exclusion	YES	5
of non–English-language citations)		
Assessment of quality of included studies	YES	7, Table 3
Reporting of Conclusions		
Consideration of alternative explanations	YES	9, 10
for observed results		
Generalization of the conclusions (ie,	YES	10
appropriate for the data presented and		
within the domain of the literature review)		
Guidelines for future research	YES	9, 10
Disclosure of funding source	YES	11



## PRISMA 2009 Checklist

ection/topic # Checklist item			
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 'Search strategy'
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5 & supplementary figure
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5&6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5&6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5, 6, 7 & supplementary figure
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5,6

Page 41 of 42

**BMJ** Open



3 2

### **PRISMA 2009 Checklist**

Synthesis of results	14 Describe the methods of handling data and combining results of studies, if done, including measures of consistence (e.g., I <sup>2</sup> ) for each meta-analysis.				
		Page 1 of 2			
Section/topic	#	Checklist item	Reported on page #		
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	6		
Additional analyses	16	16 Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.			
RESULTS					
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7		
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7		
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7		
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7, 8		
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	7, 8, 9		
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	7		
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).			
DISCUSSION	•				
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	9		
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	10		
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	10		
FUNDING	1				
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	11		

44 From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. 45 doi:10.1371/journal.pmed1000097 Protected by copyright, instriction for test and test tand data minus, by the maining and similar technologies.

ATA-ZEB them in the store of th

46 47



For more information, visit: www.prisma-statement.org

Page 2 of 2

For beer review only

AT-LZB trammarge of the room o

Protected by copyright, insuling to received selected so text and take within on the maining and almost rechnologies.

# **BMJ Open**

## Over and undertesting in primary care: a systematic review and meta-analysis.

Journal:	BMJ Open
Manuscript ID	bmjopen-2017-018557.R2
Article Type:	Research
Date Submitted by the Author:	12-Dec-2017
Complete List of Authors:	O'Sullivan, Jack; Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences Albasri, Ali Nicholson, Brian; University of Oxford, Perera, Rafael; University of Oxford, Primary Health Care Aronson, Jeffrey; University of Oxford, Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences Roberts, Nia; University of Oxford, UK, Bodleian Health Care Libraries, Heneghan, Carl; Oxford University, Primary Health Care
<b>Primary Subject Heading</b> :	Epidemiology
Secondary Subject Heading:	Diagnostics, General practice / Family practice
Keywords:	PRIMARY CARE, RADIOLOGY & IMAGING, EPIDEMIOLOGY, Protocols & guidelines < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE<sup>™</sup> Manuscripts

2/

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1	
2 3	(
4	n
5 6	0
7 8	0
9	1
10 11	U
12 13	2
14	
15 16	Ja
17 18	А
19	В
20 21	R
22 23	Je
24 25	N
26	С
27 28	
29 30	С
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	
58 59	1

### **Over and undertesting in primary care: a systematic review and** neta-analysis.

<sup>2</sup>Sullivan JW<sup>1</sup>, Albasri A<sup>1</sup>, Nicholson B<sup>1</sup>, Perera R<sup>1</sup>, Aronson J<sup>1</sup>, Roberts N<sup>2</sup>, Heneghan C<sup>1</sup>

Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Science, Iniversity of Oxford, UK

Bodleian Health Care Libraries, University of Oxford.

ack W O'Sullivan, Clinical Researcher and Doctoral Candidate, jack.osullivan@phc.ox.ac.uk li Albasri, Clinical Researcher and Doctoral Candidate, ali.albasri@gtc.ox.ac.uk rian D Nicholson, Clinical Researcher and Doctoral Candidate, brian.nicholson@phc.ox.ac.uk afael Perera, Professor of Medical Statistics, rafael.perera@phc.ox.ac.uk effrey Aronson, Reader in Evidence-Based Medicine, jeffrey.aronson@phc.ox.ac.uk ia Roberts, Medical Librarian, nia.roberts@bodleian.ox.ac.uk arl Heneghan, Professor of Evidence-Based Medicine, carl.heneghan@phc.ox.ac.uk

Dr Jack O'Sullivan orrespondence to:

> Centre for Evidence-Based Medicine Nuffield Department of Primary Care Health Sciences Radcliffe Observatory Quarter, Oxford, OX2 6GG

#### Abstract

#### Background

Health systems are currently subject to unprecedented financial strains. Inappropriate test use wastes finite health resources (overuse) and delays diagnoses and treatment (underuse). As most patient care is provided in primary care, it represents an ideal setting to mitigate waste.

#### Objective

To identify over and under use of diagnostic tests in primary care.

#### Design

Systematic review and meta-analysis.

#### Data sources and eligibility criteria

We searched MEDLINE and EMBASE from January 1999 to October 2017 for studies that measured the inappropriateness of any diagnostic test (measured against a national or international guideline) ordered for adult patients in primary care.

#### Results

We included 357,171 patients from 63 studies in 15 countries. We extracted 103 measures of inappropriateness (41 underuse, 62 overuse) from included studies for 47 different diagnostic tests.

The overall rate of inappropriate diagnostic test ordering varied substantially (0.2% to 100%).

17 tests were underused >50% of the time. Of these, echocardiography (n=4 measures) was consistently underused (between 54% and 89%, n=4). There was large variation in the rate of inappropriate underuse of pulmonary function tests (38% to 78%, n = 8).

Eleven tests were inappropriately overused >50% of the time. Echocardiography was consistently overused (77% to 92%), whereas inappropriate overuse of urinary cultures, upper endoscopy and colonoscopy varied widely, from 36% to 77% (n=3), 10% to 54% (n=10) and 8% to 52% (n=2) respectively.

#### Conclusions

There is marked variation in the appropriate use of diagnostic tests in primary care. Specifically, the use of echocardiography (both under and overuse) is consistently poor. There is substantial variation in the rate of inappropriate underuse of pulmonary function tests and the overuse of upper endoscopy, urinary cultures and colonoscopy.

Registration number: PROSPERO Registration ID: CRD42016048832

#### Manuscript word count: 3,531

#### Strengths and limitations of this study

#### Strengths

- Generates rate of under and overtesting for specific diagnostic tests against national or international guidelines
- Only includes data from real clinical encounters rather than surveys or hypothetical clinical vignettes.
- Quantified inappropriate ordering of all types of diagnostic tests, rather than just laboratory.

#### Limitations

- Systematic reviews are restricted to published literature, thus rates of inappropriate ordering are not available for all tests available to primary care physicians.
- Included studies measure appropriateness of testing in a particular health care setting against a particular guideline, thus reflect test ordering in a specific health care setting.

Topper territory only

#### Page 4 of 43

#### Introduction

Reaching a diagnosis in primary care is exceedingly complex. The combination of undifferentiated symptoms, a low prevalence of serious disease, a high degree of symptom overlap between serious and benign conditions, patients with multiple complaints, and psychological or social distress manifesting somatically all complicate reaching a diagnosis [1]. In around 40% of primary care consultations a diagnosis cannot be established from the history and physical examination alone [2], and tests are therefore often needed [1,3].

Primary care consultations make up most of the care provided in healthcare systems (90% of consultations in the UK [4], 55% of consultations in the USA[5]) and inappropriate diagnostic testing in primary care therefore has enormous resource implications. Given the calls for £22 billion in efficiency savings from the UK's National Health Service (NHS) [6] and the \$660 billion US Medicare deficit predicted by 2023 [7], ensuring the appropriateness of primary care diagnostic testing is crucial to the sustainability of healthcare systems [8].

Inappropriate diagnostic tests in primary care can be both inappropriately underused and overused. Underuse of tests, failure to order a test when indicated, can lead to diagnostic errors and delays in diagnosis and the delivery of effective treatment, leading to adverse patient outcomes and further healthcare costs [9,10]. Overuse of tests, the delivery of tests with no clear benefit or when potential harms outweigh potential benefits, subjects patients to direct harms, such as radiation exposure, as well as potential adverse outcomes (e.g. contrast nephropathy) [11], incidental findings [12], and overdiagnosis [13]. Overuse is also a waste of finite healthcare expenditure, diverting resources from beneficial tests and treatments [14–16].

Many drivers encourage inappropriate under and overuse of diagnostic tests in primary care. Greater access to tests [17], the medicolegal consequences of under-testing [18], few if any disincentives to overinvestigate [14], and clinical performance measures [19] may all contribute to overuse. Increasing primary care workload [4], time constraints [19], and difficulty keeping up-to-date with rapidly increasingly evidence [20] may contribute to both inappropriate underuse and overuse.

Guidelines set the standard of care across most health-care settings [21,22]. Furthermore, they provide a medicolegal framework [23], inform health-care policy, and improve both care outcomes and processes of care [24]. Despite some recognised limitations, including varying quality of guidelines [25–27], guidelines are often used as markers of health-care appropriateness [28–31]. Zhi et al, for instance, used guidelines as a measure of appropriateness to estimate under and overuse of laboratory testing [29]. They estimated that 45% (95%CI 34 – 56%) of secondary care laboratory testing is underused and 21% (95%CI 16 – 25%) is overused.

Despite the increasing use of healthcare resources [32], rising healthcare expenditure [6–8], increasing demands placed on primary care [4], and the apparent drivers of inappropriate testing [1,4,14,17–20], it is not clear how often diagnostic tests are inappropriately overused or underused in primary care. We therefore conducted a systematic review to quantify the frequency of inappropriate ordering of all types of diagnostic tests from primary care in relation to their respective guidelines and identify tests that are frequently over and underused.

#### Methods

This study was conducted and is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [33] and Meta-analysis of Observational Studies in Epidemiology (MOOSE) statements [34].

#### Protocol and Registration

The protocol has been published and is available online (open access) via the International prospective register for systematic reviews (PROSPERO) database (Registration ID: CRD42016048832).

#### Search Strategy

We searched EMBASE (OvidSP) and MEDLINE (OvidSP) databases from January 1999 to October 2017 for studies of any design measuring how often diagnostic test guidelines were followed in primary care (Supplementary File 1: Search Strategy). Our search strategy can be summarised as: 'Ambulatory Care AND adherence AND guideline AND diagnostic tests AND inappropriate'. Conference abstracts published after 2015 were also searched for in these databases to capture data not yet published. We also searched the WHO International Clinical Trials Registry Platform (http://apps.who.int/trialsearch/), ClinicalTrials.gov, and the reference lists of included studies.

#### Eligibility Criteria

We included studies of any design if they measured the rate of inappropriate ordering (overuse) or not ordering (underuse) of diagnostic tests ordered from primary care against national or international guidelines. We considered all diagnostic tests ordered in adults. We also included studies that measured diagnostic tests ordered from primary care but performed in secondary care (e.g. upper endoscopy). We included the control arms of RCTs if they offered exclusively usual care, and the pre-intervention periods of studies that used interrupted time series designs (before and after studies).

We excluded studies if they met the following criteria: >20% of participants were children (>20% under 18 years old); diagnostic tests not ordered by General Practitioners; screening or monitoring tests, or publication before 1999 (studies after 1999 were considered to ensure that results would more closely reflect current practice). We defined a screening test as a test on an asymptomatic or symptomatic person without signs or symptoms related to that test [35,36]. We defined monitoring tests as 'a test for a patient with an established diagnosis, for which the test is used to measure progression of the disease' [37]. We excluded studies if they did not give a measure of appropriateness was measured against local guidelines, such as a guideline specific to a hospital or region, rather than international or national guidelines.

#### Study selection and data extraction

Three reviewers (JS and AA or BN) independently screened titles, abstracts, and full texts for eligibility. The same reviewers assessed risks of bias and extracted the following data from included studies: patient demographics, eligibility criteria, name and type of diagnostic test, duration of study (days), guideline name and recommendation, total number of tests performed, and the number of tests ordered when the specific guideline recommended not ordering (inappropriate overuse) or the number of tests not ordered when the guideline recommended ordering it (inappropriate underuse). The last two data points (overuse and underuse) represent 'measures of inappropriateness'. When studies measured inappropriateness of multiple tests we extracted data on each test and presented them as individual measures of inappropriateness. When studies measured tests across different periods we

extracted measures for each time point and considered each one as an individual measure of inappropriateness.

We assessed the quality of included studies using a modified version of the Hoy risk of bias tool [38]. This tool has been validated to assess the internal and external validity of prevalence studies [38]. Our modified version of this tool kept the same domains, but adjusted the wording of the tool to reflect prevalence of inappropriate testing rather than prevalence of disease. Our tool (and results) is available in Supplementary File 2: Risk of Bias.

#### Statistical analysis

The primary outcome was the prevalence of inappropriate diagnostic testing. Inappropriate testing was measured in two ways:

1) Overuse: A diagnostic test was ordered when the relevant guideline recommends not ordering it, for instance, imaging for non-red flag low back pain (LBP).

2) Underuse: A diagnostic test was not ordered when the relevant guideline recommended ordering it, for instance, spirometry to confirm or refute the diagnosis of COPD.

We expressed measures of inappropriateness as percentages (%), where the numerator represents the total number of times a guideline recommendation was not followed and the denominator the total number of times a guideline recommendation could have been followed. For instance, the number of times imaging was inappropriately ordered for non-red flag headache as a percentage of the total number of patients who presented with non-red flag headache. Given these data are percentages, we calculated Clopper-Pearson 95% confidence intervals for each individual measure of appropriateness. We conducted sensitivity analyses with high risk of bias studies excluded.

Where the same guideline and recommendation were used by multiple studies (e.g. five studies measured inappropriate underuse of spirometry testing in patients with COPD [39–43] using the Global Initiative for Chronic Obstructive Lung Disease (GOLD) guideline) we pooled the measures and assessed heterogeneity. We combined measures of inappropriateness using a random-effects meta-analysis with 95% confidence intervals (Clopper-Pearson), for this reason each measure of appropriateness contributed relatively evenly to pooled estimates. We performed double arcsine transformation on prevalence data to stabilize the variance [44], and pooled the data using the inverse variance method [45]. We assessed heterogeneity using the I<sup>2</sup> statistic [46]. We did not combine measures of overuse and underuse, as they have different denominators: overuse involves the total number of tests ordered, whereas underuse involves the total number of times a test should have been ordered. We performed analyses using R version 3.3.2 (R project).

#### Results

#### Study selection and characteristics

We included 63 studies from 14,716 references identified from independent searches by two authors (JOS and AA or BN) (see Figure 1). Of the 63 included studies, 55 were observational studies, 6 were before and after studies and 2 were RCTs. The two RCTs investigated the effect of implementing an intervention to reduce inappropriate testing. These studies were conducted in 15 countries and included 357,171 patients (Supplementary File 3: Table 1). Table 1 (Supplementary File 4: Table 1) shows the 103 measures of inappropriateness extracted from included studies for 47 different diagnostic tests measured against 77 guideline recommendations (41 measured underuse and 62 measured overuse). Guideline recommendations came from 42 different guideline organisations from 15 countries.

Fourteen studies measured inappropriateness of more than one diagnostic tests for the same condition (e.g. chest x-ray (CXR), electrocardiography (ECG), and transthoracic echocardiography (TTE) to confirm or refute a diagnosis of heart failure). Two studies [47,48] measured inappropriateness across multiple time periods. No studies measured both under and overuse of the same test.

Included studies measured inappropriateness in one of three ways:

1. Patients with specific symptoms were assessed (prospectively or retrospectively) to see if they had received an inappropriate diagnostic test (overuse) or hadn't received the appropriate diagnostic test (underuse) in line with the relevant guideline recommendation (e.g. records for patients with non-red flag LBP to see if they received imaging [49]). 18 studies used this method.

2. Patients who had undergone a diagnostic test were identified (via hospital or national databases) and an assessment of whether the test was inappropriate (as per the defined guideline recommendations) via individual patient data was made (overuse). For instance, patients who had an upper endoscopy[50]). 22 studies used this method.

3. Patients with a diagnosis were identified via hospital or national databases and assessed to see whether they had received the appropriate diagnostic test (as per the defined guideline) to confirm or refute the diagnosis via individual patient data (underuse). For instance, assessing if patients with a diagnosis of COPD had spirometry to confirm or refute the diagnosis [39]). 23 studies used this method.

#### Risk of bias

Two thirds of the studies (n=44) were graded as being at low risk of bias, 15 (24%) at moderate risk, and 4 (6%) at high risk (Supplementary File 2 Risk of Bias). Moderate or high risk studies were at an increased risk of non-response bias (>20%), non-objective collection of data, and/or unclear intervals between symptom onset and diagnostic test use. Supplementary File 2 Risk of Bias outlines risk of bias scores in detail.

#### Percentage of diagnostic tests ordered in line with specific guideline recommendations

There was large variation in the rate of inappropriate diagnostic test ordering. The 103 diagnostic test guideline recommendations were not followed 0.2 - 100% of the time (Supplementary File 4 Table 1), wide variation was largely sustained (0.2 - 99.94%) when a further analysis was conducted excluding studies judged to be of high risk of bias. The prevalence of underuse varied 8.2% to 100%, whereas overuse varied between 0.2% and 94.2%. Similarly, this variation was essentially maintained upon exclusion of high risk studies (under use 9.8% - 99.9%, overuse 0.2 - 94.2%).

Underused tests

Table 1 (Supplementary File 4) shows that 17 tests were underused more than 50% of the time. Echocardiography was the most frequently studied (n=4 measures in Poland, UK (2), Brazil). In patients with heart failure, echocardiography was underused between 54% and 89% (n=3) of the time and in atrial fibrillation 56% (n=1).

For some tests there was large variation in the rate of underuse (Figure 2). Underuse of pulmonary function tests (PFTs) to confirm or refute COPD, measured against the Global Initiative for Chronic Obstructive Lung Disease (GOLD), NICE (UK) and Danish National Board of Health guidelines, varied from 26% to 78% (n=8). None of the studies that studied echocardiography, or PFTs were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

#### Overused tests

Eleven tests were overused more than 50% of the time (Figure 3). Echocardiography was consistently overused, for instance in 'routine perioperative evaluation of ventricular function with no symptoms or signs of cardiovascular disease', whereas other tests (urinary cultures, upper endoscopy and colonoscopy) were overused at varying rates. The over use of echocardiography was studied in the UK [51] and the Netherlands [52]. The rates of overuse varied between the two settings: between 77% (Netherlands) and 92% (UK). Overuse of urinary cultures for uncomplicated urinary tract infections was studied in the USA [53,54], Spain [55] and Sweden [56] the rate varied from 57% to 77% in the USA, was around 50% in Sweden and was as low as 36% in Spain. Overuse of upper endoscopy was studied widely (n=11); in Australia [57,58], Saudi Arabia [59,60], UK [61], Italy [62–64], USA [50,65], and Malaysia [66]. The overuse varied markedly, from 7.5% to 54% (n=11) respectively (Figure 3, Supplementary File 4 Table 1). Similarly, the inappropriate over-use of colonoscopy varied substantially; from 8% in Australia [58] to 52% in Malaysia [67]. None of the above studies were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

Our results also suggest that the inappropriate overuse of CT and MRI scans for non-red flag headache (a headache without symptoms suggesting a malignant underlying pathology) has more than doubled in the last ten years in the USA (2000: 6.7% (95%CI: 5.4 to 8.2%, 2010: 14% (95%CI 12. to 16%) (Supplementary File 4 Table 1) [48]. Conversely, the rate of inappropriate overuse of radiology tests for non-red flag low back pain was consistently low, with all (n=18 measures) but two measure showing inappropriate overuse less than 25% of the time (Supplementary File 4 Table 1). One of these studies [68] estimated overuse to be about 50%, but was conducted in 2001 and thus may reflect improvements over time. The other study is current, but used a small sample size [69]. None of these studies were considered high risk of bias and thus results didn't change upon further analysis excluding high risk studies.

#### Variation of inappropriateness against the same guideline recommendation

Eleven different guideline recommendations were studied more than once. There was significant heterogeneity ( $I^2 > 50\%$ ) in nine of these pooled measures. Significant heterogeneity may have occurred for several reasons: 1) vastly different populations (for instance, one study measured the inappropriateness of upper endoscopy in Saudi Arabia [60] using the American Gastroenterological Association recommendations, whereas another study used the same recommendations in the USA [70]; 2) Contrasting healthcare systems [71,72]; 3) Relevance and applicability of one country's national guideline to another country [73]; 4) A low number of measures for meta-analysis [46] and/or 5) Significant heterogeneity, reflecting significant variation in inappropriate ordering.

#### Discussion

There is marked variation in the rate of underuse and overuse of diagnostic tests from many primary care settings across the world. This variation suggests improvement can be made in the rate of appropriate diagnostic test ordering.

Primary care use of echocardiography is consistently poor. Echocardiography is inappropriately underused for some clinical situations, e.g. confirming a diagnosis of heart failure, and inappropriately overused in others, e.g. perioperative assessment. This was consistent across the countries where appropriateness of echocardiogram has been studied. This is of concern, given the expertise and resource requirements to perform the test and the increasing availability of direct access ordering for primary care physicians.

For four tests we found marked variation in the rate of inappropriate use. Underuse of pulmonary function tests varied by >50%, whereas overuse of urinary cultures, upper endoscopy and colonoscopy all varied by around 40%.

Radiology tests for both non-red flag low back pain and non-red flag headache were frequently *not* overused, but the rate of overuse of imaging for non-red flag headache showed concerning trends, more than doubling from 2000 to 2010 (Supplementary File 4 Table 1).

#### Implications and future research

Two principle conclusions can be drawn from our results: 1.Ordering of echocardiograms from primary care appears to require improvement, 2. Markedly varying rates of inappropriate use for pulmonary function tests (underuse), colonoscopy (overuse), upper endoscopy (overuse), and urinary cultures (overuse) suggests that ordering can be improved.

Future research should focus on: Determining the reasons for deviation from guidelines, assessing the quality of guidelines supporting diagnostic test use and systematic reviews quantifying inappropriate screening and monitoring tests. Further, investigators wishing to undertake primary studies measuring inappropriate use should focus on developing objective data extraction methods for assessing patient notes and define clearly the interval they (investigators) will consider a test ordered for a particular symptom or disease.

#### Strengths in relation to other studies

Compared with other studies of inappropriate use of healthcare resources, we used data from real clinical encounters. This allowed a more robust assessment of diagnostic test inappropriateness, where other studies used surveys and hypothetical clinical vignettes [19,74,75]. Furthermore, we quantified the appropriateness of all types of diagnostic tests, rather than focusing on a specific test or specific disease (such as only laboratory tests [29]). Our paper is the first systematic review of studies that measured inappropriateness of all diagnostic tests ordered from primary care. Zhi et al [29] quantified the mean rates of overuse and underuse of laboratory tests in secondary care and focused on quantifying an overall rate of over and under use. They estimated that over and underuse of laboratory tests was around 21% and 45% respectively [29]. We choose not to quantify an overall rate of over and under use because we feel the results would not be representative; we would be combining data from multiple different health care settings and data captured only the studied selection of diagnostic tests available in primary care.

Our use of guideline recommendations as the metric of appropriateness allowed a direct measure of diagnostic test appropriateness. Other studies that have assessed temporal and geographical variation in the use of diagnostic tests [76,77] have noted substantial differences in diagnostic practices across

different regions, irrespective of disease prevalence and patient characteristics [77]. These studies, however, could not quantify what percentage of the temporal increase in the use of a diagnostic test is inappropriate and what percentage of variation between regions is inappropriate. We have quantified the percentage of inappropriate testing.

Although beyond the scope of our review, ultimately, interventions should be implemented to improve test use. A 2015 systematic review [78] concluded that 'Interventions such as educational strategies, feedback and changing test order forms may improve the efficient use of laboratory tests in primary care'. Thus, doctors, academics and policy makers can use our results to identify diagnostic tests in their particular health care settings which may benefit from intervention.

#### Limitations

The use of guidelines to quantify appropriateness of diagnostic tests could be considered a limitation of this study. Guidelines are often criticised for varying quality [25–27,79] and panel members' conflicts of interests [80]. However, clinical practice guidelines have been shown to improve both care outcomes and processes of care [24], allow assessment of care on a population level, inform health policy [81,82], set the standard of care across many health care settings [21,22], and provide a medicolegal framework [23]. One major medical insurance company advises that 'doctors must be prepared to explain and justify their decisions and actions, especially if they depart from guidelines produced by a nationally recognised body' [23]. Furthermore, guidelines have been used to measure appropriateness of the use of tests in other published peer-review studies [29]. There will always be times when it is appropriate to depart from guidelines, but dramatic, consistent variation from guidelines requires investigation and is unlikely to be caused entirely by the quality of guidelines.

Furthermore, our study includes only a selection of diagnostic tests and is thus not an allencompassing reflection of clinical practice. The data reflects the use of a specific test, sometimes for a particular clinical situation, in a particular country's health care system. Thus, policy makers and those interested in improving the quality of primary care diagnostic test use, can use our results as a resource to identify tests in their healthcare setting that require improvement and/or investigation to decipher why such deviation from guidelines exists. Our conclusions from this paper, however, are not generalisable to all primary care settings nor all primary care diagnostic tests.

Lastly, caution must be taken when comparing results that measured inappropriateness using different denominators. The results from studies that measured inappropriateness using patients who had undergone a diagnostic test as a denominator should be interpreted differently to studies that used patients with a diagnosis or symptoms as a denominator (and vice versa).

#### Conclusion

There is marked variation in under and overuse of appropriate diagnostic test use in primary care across the world. From the available data, echocardiograms are ordered particularly poorly, while the substantial variation in appropriate ordering of pulmonary function tests, colonoscopy, upper endoscopy, and urinary cultures suggest a need for improvement.

#### Acknowledgements

We thank Kate Roche and Jason Hendry for comments on the draft and figures. We also thank the peer reviewers for their constructive feedback.

#### Funding:

This research received no specific grant from any funding agency in the public, commercial or notfor-profit sectors.

All author declare no conflicts of interests.

#### Ethical approval: Not required

**Data sharing:** Data extracted from the included studies in this review are available on request from the corresponding author.

**Registration:** PROSPERO protocol Registration ID: CRD42016048832 (https://www.crd.york.ac.uk/prospero/display\_record.asp?ID=CRD42016048832)

#### Competing interest statement.

We have read and understood BMJ policy on declaration of interests and declare that we have no competing interests.

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and jointly declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

#### **Contribution statement:**

Conception and design: Jack O'Sullivan, Rafael Perera and Carl Heneghan

Search Strategy: Nia Roberts and Jack O'Sullivan

Screening, extraction and risk of bias: Jack O'Sullivan, Ali Albasri and Brian Nicholson.

Analysis and interpretation of the data: Jack O'Sullivan, Rafael Perera, Jeffrey Aronson and Carl Heneghan.

Drafting of the article: Jack O'Sullivan (all authors critically reviewed and approved manuscript)

Statistical expertise: Rafael Perera

Clinical expertise: Jack O'Sullivan, Brian Nicholson, Jeffrey Aronson and Carl Heneghan

Jack O'Sullivan is the guarantor.

#### **Copyright Statement**

The corresponding author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

#### References

- 1 Foot C, Naylor C, Imison C. The quality of GP diagnosis and referral. 2010. http://amapro.isabelhealthcare.com/pdf/Kings\_Fund\_Diagnosis\_and\_Referral\_2010.pdf
- 2 Koch H, van Bokhoven MA, ter Riet G, *et al.* Ordering blood tests for patients with unexplained fatigue in general practice: what does it yield? Results of the VAMPIRE trial. *Br J Gen Pract* 2009;**59**:e93-100. doi:10.3399/bjgp09X420310
- 3 Heneghan C, Glasziou P, Thompson M, *et al.* Diagnostic strategies used in primary care. *BMJ* 2009;**338**.
- 4 Hobbs FDR, Bankhead C, Mukhtar T, *et al.* Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *Lancet* 2016;**387**:2323–30. doi:10.1016/S0140-6736(16)00620-6
- 5 Centers for Disease Control and Prevention, National Center for Health Statistics. National Ambulatory Medical Care Survey: 2012 Summary Tables. 2012;:5.http://www.cdc.gov/nchs/data/ahcd/namcs\_summary/2010\_namcs\_web\_tables.pdf
- 6 Alderwick H, Robertson R, Appleby J, *et al.* Better value in the NHS The role of changes in clinical practice. 2015.
- 7 Fisher ES, Bynum JP, Skinner JS. Slowing the growth of health care costs--lessons from regional variation. *N Engl J Med* 2009;**360**:849–52. doi:10.1056/NEJMp0809794
- 8 Appleby J, Thompson J, Jabbal J. Quarterly Monitoring Report: How is the NHS performing? *King's Fund* 2016;:1–42.
- 9 Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ Qual Saf* 2013;22 Suppl 2:ii6-ii10. doi:10.1136/bmjqs-2012-001621
- 10 Gandhi TK, Kachalia A, Thomas EJ, *et al.* Annals of Internal Medicine Article Missed and Delayed Diagnoses in the Ambulatory Setting : *Ann Intern Med* 2006;**145**:488–96.
- 11 Katzberg RW, Lamba R. Contrast-induced nephropathy after intravenous administration: fact or fiction? *Radiol Clin North Am* 2009;47:789–800, v. doi:10.1016/j.rcl.2009.06.002\rS0033-8389(09)00094-3 [pii]
- 12 Lumbreras B, Donat L, Hernández-Aguado I. Incidental findings in imaging diagnostic tests: a systematic review. *Br J Radiol* 2010;**83**:276–89. doi:10.1259/bjr/98067945
- 13 Welch, H. Gilbert, Schwartz, Lisa, Woloshin S. *Overdiagnosed: Making people sick in the pursuit of health.* Beacon Press, 2011 2011.
- 14 Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *Bmj* 2012;**344**:e3502–e3502. doi:10.1136/bmj.e3502
- 15 Berwick D, Hackbarth AD. Eliminating Waste in US Health Care. *JAMA* 2012;**307**:1513. doi:10.1001/jama.2012.362
- 16 Cecchini M, Lee S. Tackling Wasteful Spending on Healthcare. 2017. http://networks.sustainablehealthcare.org.uk/sites/default/files/resources/Tackling Wasteful Spending on Health.pdf#page=117
- 17 Health D of. NHS 2010–2015: from good to great. preventative, people-centred, productive.

#### BMJ Open

	London: 2009.
18	Esmail A, Neale G, Elstein M, Firth-Cozens J, Davy C VC. Case Studies in Litigation: Claims reviews in four specialties. Manchester: 2004.
19	Sirovich BE, Woloshin S, Schwartz LM. Too Little? Too Much? Primary care physicians' views on US health care: a brief report. <i>Arch Intern Med</i> 2011; <b>171</b> :1582–5. doi:10.1001/archinternmed.2011.437
20	Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? <i>PLoS Med</i> 2010; <b>7</b> . doi:10.1371/journal.pmed.1000326
21	Garber AM. Evidence-based guidelines as a foundation for performance incentives. <i>Health Aff (Millwood)</i> 2005; <b>24</b> :174–9. doi:10.1377/hlthaff.24.1.174
22	Ransohoff DF, Pignone M, Sox HC, <i>et al.</i> How to Decide Whether a Clinical Practice Guideline Is Trustworthy. <i>JAMA</i> 2013; <b>309</b> :139. doi:10.1001/jama.2012.156703
23	Fryar C. Doctors can depart from guidelines in patients' best interests. BMJ 2015;350.
24	Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. <i>Lancet (London, England)</i> 1993; <b>342</b> :1317–22.http://www.ncbi.nlm.nih.gov/pubmed/7901634 (accessed 31 Aug 2016).
25	Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. <i>JAMA</i> 1999; <b>281</b> :1900–5.http://www.ncbi.nlm.nih.gov/pubmed/10349893 (accessed 7 Dec 2016).
26	Grilli R, Magrini N, Penna A, <i>et al.</i> Practice guidelines developed by specialty societies: the need for a critical appraisal. <i>Lancet (London, England)</i> 2000; <b>355</b> :103–6. doi:10.1016/S0140-6736(99)02171-6
27	Lenzer J. Why we can't trust clinical guidelines. BMJ 2013;346.
28	Spyridonidis D, Calnan M. Opening the black box: A study of the process of NICE guidelines implementation. <i>Health Policy (New York)</i> 2011; <b>102</b> :117–25. doi:10.1016/j.healthpol.2011.06.011
29	Zhi M, Ding EL, Theisen-Toupal J, <i>et al.</i> The Landscape of Inappropriate Laboratory Testing: A 15-Year Meta-Analysis. <i>PLoS One</i> 2013; <b>8</b> :e78962. doi:10.1371/journal.pone.0078962
30	McGlynn E, Asch S, Adams J, <i>et al.</i> Quality of health care delivered to adults in the United States. <i>N Engl J Med</i> 2003; <b>349</b> :1866-1868-1868. doi:10.1056/NEJMsa022615
31	Sheldon T a, Cullum N, Dawson D, <i>et al.</i> What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. <i>BMJ</i> 2004; <b>329</b> :999. doi:10.1136/bmj.329.7473.999
32	National Health Service. NHS Imaging and Radiodiagnostic activity in England. 2013;:1–7.http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/04/KH12-release-2012-13.pdf
33	Moher D, Liberati A, Tetzlaff J, <i>et al.</i> Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. <i>BMJ</i> 2009; <b>339</b> :b2535.http://www.ncbi.nlm.nih.gov/pubmed/19622551 (accessed 22 Aug 2016).
34	Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of Observational Studies in Epidemiology. JAMA 2000; <b>283</b> :2008. doi:10.1001/jama.283.15.2008
35	Wald NJ. Guidance on terminology. J Med Screen 2008;15:50-50.
12	
13	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

BMJ Open: first published as 10.1136/bmjopen-2017-018557 on 11 February 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open

doi:10.1258/jms.2008.008got

- 36 Raffle A, Gray J. *Screening: Evidence and Practice*. Oxford University Press 2007.
- 37 Glasziou P, Irwig L, Aronson J. *Evidence-based medical monitoring: from principles to practice*. Oxford (UK): Blackwell Publishing, BMJ books 2008.
- 38 Hoy D, Brooks P, Woolf A, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012;65:934–9. doi:10.1016/j.jclinepi.2011.11.014
- 39 Belletti D, Liu J, Zacker C, *et al.* Results of the CAPPS: COPD--assessment of practice in primary care study. *Curr Med Res Opin* 2013;**29**:957–66. doi:10.1185/03007995.2013.803957
- 40 Bertella E, Zadra A, Vitacca M, *et al.* COPD management in primary care: is an educational plan for GPs useful? *Multidiscip Respir Med* 2013;**8**:24. doi:10.1186/2049-6958-8-24
- 41 Chavez PC, Shokar NK. Diagnosis and management of chronic obstructive pulmonary disease (COPD) in a primary care clinic. *COPD* 2009;**6**:446–51. doi:10.3109/15412550903341455
- 42 Lange P, Rasmussen FV, Borgeskov H, *et al.* The quality of COPD care in general practice in Denmark: the KVASIMODO study. *Prim Care Respir J* 2007;**16**:174–81. doi:10.3132/pcrj.2007.00030
- 43 Ulrik CS, Sørensen TB, Højmark TB, *et al.* Adherence to COPD guidelines in general practice: impact of an educational programme delivered on location in Danish general practices. *Prim Care Respir J* 2013;**22**:23–8. doi:10.4104/pcrj.2012.00089
- 44 Barendregt JJ, Doi SA, Lee YY, *et al.* Meta-analysis of prevalence. *J Epidemiol Community Heal* 2013;:974–8. doi:10.1136/jech-2013-203104
- Doi SAR, Barendregt JJ, Khan S, *et al.* Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. *Contemp Clin Trials* 2015;45:130–8. doi:10.1016/j.cct.2015.05.009
- 46 Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;**327**:557–60. doi:10.1136/bmj.327.7414.557
- 47 Mafi JN, McCarthy EP, Davis RB, *et al.* Worsening trends in the management and treatment of back pain. *JAMA Intern Med* 2013;**173**:1573–81. doi:10.1001/jamainternmed.2013.8992
- 48 Mafi JN, Edwards ST, Pedersen NP, et al. Trends in the Ambulatory Management of Headache: Analysis of NAMCS and NHAMCS Data 1999-2010. J Gen Intern Med 2015;30:548–55. doi:10.1007/s11606-014-3107-3
- 49 Williams CM, Maher CG, Hancock MJ, et al. Low back pain and best practice care: A survey of general practice physicians. Arch Intern Med 2010;170:271–7. doi:10.1001/archinternmed.2009.507
- 50 Cai JX, Campbell EJ, Richter JM. Concordance of Outpatient Esophagogastroduodenoscopy of the Upper Gastrointestinal Tract With Evidence-Based Guidelines. *JAMA Intern Med* 2015;**175**:1563–4. doi:10.1001/jamainternmed.2015.3533
- 51 Gurzun M-M, Ionescu A. Appropriateness of use criteria for transthoracic echocardiography: are they relevant outside the USA? *Eur Hear J - Cardiovasc Imaging* 2014;**15**:450–5. doi:10.1093/ehjci/jet186
- 52 van Gurp N, Boonman-De winter LJM, Meijer Timmerman Thijssen DW, *et al.* Benefits of an open access echocardiography service: A Dutch prospective cohort study. *Netherlands Hear J* 2013;**21**:399–405. doi:10.1007/s12471-013-0416-9

53	Johnson JD, O'Mara HM, Durtschi HF, <i>et al.</i> Do Urine Cultures for Urinary Tract Infections Decrease Follow-up Visits? <i>J Am Board Fam Med</i> 2011; <b>24</b> :647–55. doi:10.3122/jabfm.2011.06.100299
54	Grover ML, Bracamonte JD, Kanodia AK, <i>et al.</i> Assessing Adherence to Evidence-Based Guidelines for the Diagnosis and Management of Uncomplicated Urinary Tract Infection. <i>Mayo Clin Proc</i> 2007; <b>82</b> :181–5. doi:10.4065/82.2.181
55	Llor C, Rabanaque G, Lopez A, <i>et al.</i> The adherence of GPs to guidelines for the diagnosis and treatment of lower urinary tract infections in women is poor. <i>Fam Pract</i> 2011; <b>28</b> :294–9. doi:10.1093/fampra/cmq107
56	Lindbäck H, Lindbäck J, Melhus Å. Inadequate adherence to Swedish guidelines for uncomplicated lower urinary tract infections among adults in general practice. <i>Apmis</i> 2017; <b>125</b> :816–21. doi:10.1111/apm.12718
57	Leon P, Catherine K, Mark N, et al. Gastro-oseophageal reflux disease. The impact of guidelines on GP management. 2008.
58	Hughes-Anderson W, Rankin SL, House J, <i>et al.</i> Open access endoscopy in rural and remote Western Australia: does it work? <i>ANZ J Surg</i> 2002; <b>72</b> :699–703.http://www.ncbi.nlm.nih.gov/pubmed/12534377 (accessed 7 Dec 2016).
59	Aljebreen AM, Alswat K, Almadi MA. Appropriateness and diagnostic yield of upper gastrointestinal endoscopy in an open-access endoscopy system. <i>Saudi J Gastroenterol</i> 2013; <b>19</b> :219–22. doi:10.4103/1319-3767.118128
60	Azzam NA, Almadi MA, Alamar HH, <i>et al.</i> Performance of American Society for Gastrointestinal Endoscopy guidelines for dyspepsia in Saudi population: Prospective observational study. <i>World J Gastroenterol</i> 2015; <b>21</b> :637–43. doi:10.3748/wjg.v21.i2.637
61	Elwyn G, Owen D, Roberts L, <i>et al.</i> Influencing referral practice using feedback of adherence to NICE guidelines: a quality improvement report for dyspepsia. <i>Qual Saf Health Care</i> 2007; <b>16</b> :67–70. doi:10.1136/qshc.2006.019992
62	Cardin F, Zorzi M, Bovo E, <i>et al.</i> Effect of Implementation of a Dyspepsia and Helicobacter pylori Eradication Guideline in Primary Care. <i>Digestion</i> 2005; <b>72</b> :1–7. doi:10.1159/000087215
63	Cardin F, Zorzi M, Terranova O. Implementation of a guideline versus use of individual prognostic factors to prioritize waiting lists for upper gastrointestinal endoscopy. <i>Eur J Gastroenterol Hepatol</i> 2007; <b>19</b> :549–53. doi:10.1097/01.meg.0000216942.42306.d5
64	Hassan C, Bersani G, Buri L, <i>et al.</i> Appropriateness of upper-GI endoscopy: an Italian survey on behalf of the Italian Society of Digestive Endoscopy. <i>Gastrointest Endosc</i> 2007; <b>65</b> :767–74. doi:10.1016/j.gie.2006.12.058
65	Fiorenza JP, Tinianow AM, Chan WW. The Initial Management and Endoscopic Outcomes of Dyspepsia in a Low-Risk Patient Population. <i>Dig Dis Sci</i> 2016; <b>61</b> :2942–8. doi:10.1007/s10620-016-4051-3
66	Chan Y-M, Goh K-L. Appropriateness and diagnostic yield of EGD: a prospective study in a large Asian hospital. <i>Gastrointest Endosc</i> 2004; <b>59</b> :517–24. doi:10.1016/S0016-5107(04)00002-1
67	CHAN T, GOH K. Appropriateness of colonoscopy using the ASGE guidelines: experience in a large Asian hospital. <i>Chin J Dig Dis</i> 2006;7:24–32. doi:10.1111/j.1443-9573.2006.00240.x
68	Eccles M, Steen N, Grimshaw J, <i>et al.</i> Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. <i>Lancet</i> 2001; <b>357</b> :1406–9. doi:10.1016/S0140-6736(00)04564-5
15	
	For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

- 69 Tahvonen P, Oikarinen H, Niinimäki J, *et al.* Justification and active guideline implementation for spine radiography referrals in primary care. *Acta radiol* 2016;**58**:586–92. doi:10.1177/0284185116661879
  - 70 Majumdar SR, Soumerai SB, Farraye FA, *et al.* Chronic acid-related disorders are common and underinvestigated. *Am J Gastroenterol* 2003;**98**:2409–14. doi:10.1111/j.1572-0241.2003.07706.x
  - 71 Basu S, Andrews J, Kishore S, *et al.* Comparative performance of private and public healthcare systems in low- and middle-income countries: A systematic review. *PLoS Med* 2012;**9**:19. doi:10.1371/journal.pmed.1001244
  - 72 Ridic G, Gleason S, Ridic O. Comparisons of Health Care Systems in the United States, Germany and Canada. *Mat Soc Med* 2012;**24**:112–20. doi:10.5455/msm.2012.24.112-120.Comparisons
  - 73 Gagliardi AR, Brouwers MC. Do guidelines offer implementation advice to target users? A systematic review of guideline applicability. *BMJ Open* 2015;5:e007047–e007047. doi:10.1136/bmjopen-2014-007047
  - 74 Kachalia A, Berg A, Fagerlin A, *et al.* Overuse of testing in preoperative evaluation and syncope: a survey of hospitalists. *Ann Intern Med* 2015;**162**:100–8. doi:10.7326/M14-0694
  - 75 Swennen MHJ, Rutten FH, Kalkman CJ, *et al.* Do general practitioners follow treatment recommendations from guidelines in their decisions on heart failure management? A cross-sectional study. *BMJ Open* 2013;3:e002982. doi:10.1136/bmjopen-2013-002982
  - Parker L, Levin DC, Frangos A, *et al.* Geographic variation in the utilization of noninvasive diagnostic imaging: national medicare data, 1998-2007. *AJR Am J Roentgenol* 2010;**194**:1034–9. doi:10.2214/AJR.09.3528
  - Song Y, Skinner J, Bynum J, *et al.* Regional Variations in Diagnostic Practices. *N Engl J Med* 2010;363:45–53. doi:10.1056/NEJMsa0910881
  - 78 Cadogan SL, Browne JP, Bradley CP, et al. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement* Sci 2015;10:167. doi:10.1186/s13012-015-0356-4
  - 79 Burgers JS, Fervers B, Haugh M, et al. International Assessment of the Quality of Clinical Practice Guidelines in Oncology Using the Appraisal of Guidelines and Research and Evaluation Instrument. J Clin Oncol 2004;22:2000–7. doi:10.1200/JCO.2004.06.157
  - 80 Gale EAM. Conflicts of interest in guideline panel members. *BMJ* 2011;343.
  - 81 IoM C to A the PHS on CPG. Clinical Practice Guidelines: Directions for a New Program. Washington: 1990. doi:10.1097/SPV.0b013e31828a2951
  - 82 Browman GP, Snider A, Ellis P. Negotiating for change. The healthcare manager as catalyst for evidence-based practice: changing the healthcare environment and sharing experience. *Healthc Pap* 2003;**3**:10–22.http://www.ncbi.nlm.nih.gov/pubmed/12811083 (accessed 7 Nov 2016).

#### Figure legends

Figure 1: PRISMA Flow Diagram

Figure 2: Rates of underuse. FNA=Fine needle aspiration, FBC=Full Blood Count, TSH=Thyroid Stimulating Hormone, PFTs=Pulmonary function tests, CXR=Chest x-ray, ECG= Electrocardiogram,

#### 

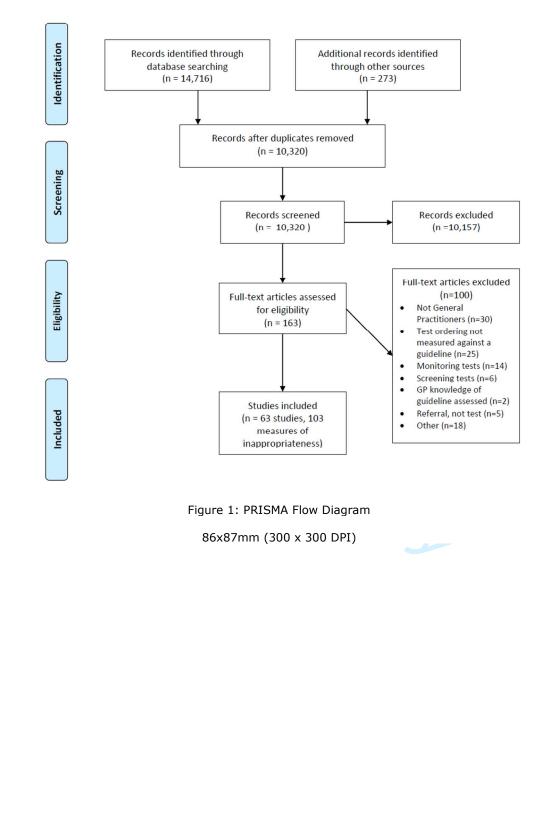
#### **BMJ** Open

Figure 3: Rates of overuse. NHMRC= National Health and Medical Research Council, U/S=Ultrasound

to beet terien only

Page 18 of 43

**BMJ** Open



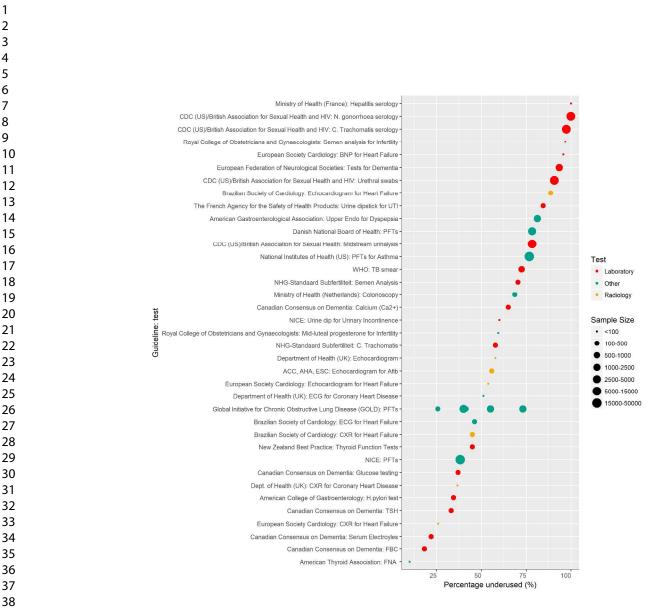


Figure 2: Rates of underuse. FNA: Fine needle aspiration, FBC: Full Blood Count, TSH: Thyroid Stimulating Hormone, PFTs: Pulmonary function tests, CXR: Chest x-ray, ECG:Electrocardiogram, AFib: AtrialFibrillation, TB: Tuberculosis, ACC: American College of Cardiology, AHA: American Heart Association, ESC: European Society of Cardiology, UTI: Urinary Tract Infection.

254x254mm (300 x 300 DPI)

	Gastroenterological Society of Australia: Barium Swallow for GORD -				
	American College of Cardiology: Echocardiogram				
	Infectious Disease Society of America: Throat cultures -			· · · · ·	
	Netherlands Society of Cardiology: Echocardiogram				
	Infectious Disease Society of America: Urine cultures for UTI-				
	Royal College of Radiologists: Knee x-ray				
	European Society of Primary Care Gastroenterology: H. pylori test				
	Finish Medical Society: Low back pain among adults -			•	
	Royal College of Radiologists (UK): Hip x-ray-				
	Royal College of Radiologists (UK): L-spine x-ray -				
Ga	astroenterological Society of Australia: Upper endoscopy for GORD -				
_	European Association of Urology: Urine cultures for UTI -				
Swedish Med	icines Agency: Treatment of lower urinary tract infections in women -				
	NICE: Upper endoscopy (Dyspepsia) -				
	The European Helicobacter Study Group: H. pylori				T
European Soc	iety of Primary Care Gastroenterology: Upper Endoscopy (H.pylori)				Test
	American College of Gastroenterology: Dyspepsia -				<ul> <li>Laboratory</li> </ul>
	The George Institute (Aus): LBP -				Other
A	American College of Physicians: Upper endoscopy for GORD -				<ul> <li>Radiology</li> </ul>
	nerican Gastroenterological Association: Upper Endo for Dyspepsia - Canadian Association of Radiologists: Carotid U/S -				
Guideline: test	Ministry of Health (Italy): L-spine radiology (all)				Sample Size
Je:	NHMRC: L-spine radiology (all)				• <100
ieli	NHMRC. Cervical spine x-ray				• 100-500
Suic	The Canadian Task Force on the Periodic Health Examination				
0	American College of Physicians: L-spine MRI-				• 500-1000
	NHMRC: L-spine viru				• 1000-2500
	merican Society for Gastrointestinal Endoscopy: Upper Endoscopy				2500-5000
~	American College of Cardiology: Cardiac radionuclide imaging -		•		5000-15000
	Canadian Association of Radiologists: Thyroid U/S -				15000-50000
	American College of Physicians: L-spine x-ray	-			-
	Canadian Association of Radiologists: Abdominal U/S				
	NICE: L-spine MRI -	-			
	American Society for Gastrointestinal Endoscopy: Colonoscopy				
	Choosing Wisely (USA): CT or MRI Brain -				
	Royal College of Radiologists (UK): CT (all) -				
	Royal College of Radiologists (UK): MRI (all)				
Work	ers Compensation Board of British Columbia: L-spine radiology (all)				
	NHMRC: L-spine CT -				
	Netherlands College of GPs: L-spine radiology (all)				
	Canadian Association of Radiologists: Soft tissue U/S -				
	Canadian Association of Radiologists: Pelvic U/S				
	American Academy of Otolaryngology: Adult Sinusitis -	<u> </u>			
	NHMRC: L-Spine U/S -	ē			
		ě –			
		0 25	50	75	
		C) = 20	ercentage overused (		

254x254mm (300 x 300 DPI)

1	
2	
3	
4	MEDLINE Search Strategy
5	
6	1 Ambulatowy Caro/
7	1. Ambulatory Care/
8	2. exp Ambulatory Care Facilities/
9	3. general practice/ or family practice/
10	4. general practitioners/ or physicians, family/ or physicians, primary care/
11	5. Primary Health Care/
12	6. Office visits/
13	,
14	7. Academic Medical Centers/
15	8. (ambulatory adj3 (care or setting? or facilit* or ward? or department? or
16	service?)).ti,ab.
17	9. ((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.
18	10. (primary care or primary health care or primary healthcare or family
19	medicine or community medicine or community health).ti,ab.
20	
21	11. (gp or gps).ti,ab.
22 23	12. (after hour? or afterhour? or "out of hour?" or ooh).ti,ab.
23	13. (clinic? or visit?).ti,ab.
25	14. ((health* or medical) adj2 (center? or centre?)).ti,ab.
26	15. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
27	16. exp Emergency Service, Hospital/
28	17. Emergency Medical Services/
29	18. (emergency adj3 (care or setting? or facilit* or ward? or department? or
30	service? or room?)).ti,ab.
31	
32	19. (emergency medicine or ed or er or a&e).ti,ab.
33	20. 16 or 17 or 18 or 19
34	21. 15 or 20
35	22. guidelines as topic/ or practice guidelines as topic/
36	23. (guideline? or guidance?).ti,ab.
37	24. 22 or 23
38	25. (adhere* or non-adhere* or nonadhere* or concord* or non-concord* or
39	nonconcord* or discord* or comply or complian* or non-complian* or
40	
41	noncomplian* or align* or nonalign* or nonalign* or congruen* or incongruen*
42	or consisten* or inconsisten* or contradict*).ti,ab.
43	26. ((does or "does not" or doesn?t or did or "did not" or didn?t or "not" or fail*)
44	adj3 (follow* or met or meet or meeting or match or matching or "in line
45	with")).ti,ab.
46 47	27. ((follow* or met or meet or meeting or match or matching or "in line with" or
47 48	keep or kept or keeping or utili?ation or utile?e? or change?) adj5 (criteria or
48	recommend* or guideline? or guidance)).ti,ab.
50	
51	28. Physician's Practice Patterns/
52	29. clinical competence/ or nursing competence/
53	30. 25 or 26 or 27 or 28 or 29
54	31. 24 and 30
55	32. Guideline Adherence/
56	33. 31 or 32
57	34. exp *"diagnostic techniques and procedures"/
58	35. exp "diagnostic techniques and procedures"/ut
59	
60	36. (diagnos* or detect* or test* or screen* or manag*).ti.

37. (imaging or radiolog* or tomogra* or ct scan* or pet scan* or echocardiogra*
or angiogra* or ultrasound* or ultrasonogra*).ti,ab.
38. ((medical or clinical or diagnos* or screening or routine or laboratory) adj5
(test* or investigation?)).ti,ab.
39. ((h?ematolog* or blood or urin* or saliva*) adj5 test*).ti,ab.
40. ((stress* or physical or function*) adj5 test*).ti,ab.
41. 34 or 35 or 36 or 37 or 38 or 39 or 40
42. 21 and 33 and 41
43. ((necessary or unnecessary or appropriate* or inappropriate* or waste* or
utili?ation or indicated or excess* or less or more or increas* or decreas*) adj10
(test* or imaging or radiolog* or tomogra* or ct scan* or pet scan* or
echocardiogra* or angiogra* or ultrasound* or ultrasonogra* or
investigation?)).ti,ab.
44. ((order* or request*) adj5 (test* or imaging or radiolog* or tomogra* or ct
scan* or pet scan* or echocardiogra* or angiogra* or ultrasound* or
ultrasonogra* or investigation?)).ti,ab.
45. Unnecessary Procedures/
46. 43 or 44 or 45
47. 21 and 24 and 46
48. 21 and 41 and 45
49. 42 or 47 or 48
50. limit 49 to yr="1999 -Current"
51. limit 50 to english language
52. exp animals/ not humans.sh.
53. 51 not 52
EMBASE Search Strategy
$\sim$
1. Ambulatory Care/
2. general practice/
3. general practitioners/
4. Primary Health Care/
5. (ambulatory adj3 (care or setting? or facilit* or ward? or department? or
service?)).ti,ab.
6. ((general or family) adj2 (practi* or physician? or doctor?)).ti,ab.
7. (primary care or primary health care or primary healthcare or family medicine
or community medicine or community health).ti,ab.
8. (gp or gps).ti,ab.
9. (after hour? or afterhour? or "out of hour?" or ooh).ti,ab.
10. (clinic? or visit?).ti,ab.
11. ((health* or medical) adj2 (center? or centre?)).ti,ab.
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. Emergency Ward/
14. (emergency adj3 (care or setting? or facilit* or ward? or department? or
service? or room?)).ti,ab.
15. (emergency medicine or ed or er or a&e).ti,ab.
16. 13 or 14 or 15
17 12 or 16
17. 12 or 16
17. 12 or 16 18. *practice guideline/

1	
2	
3	19. (guideline? or guidance?).ti,ab.
4	20. 18 or 19
5 6	21. (adhere* or non-adhere* or nonadhere* or concord* or non-concord* or
7	nonconcord* or discord* or comply or complian* or non-complian* or
8	noncomplian* or align* or nonalign* or nonalign* or congruen* or incongruen*
9	
10	or consisten* or inconsisten* or contradict*).ti,ab.
11	22. ((does or "does not" or doesn?t or did or "did not" or didn?t or "not" or fail*)
12	adj3 (follow* or met or meet or meeting or match or matching or "in line
13	with")).ti,ab.
14	23. ((follow* or met or meet or meeting or match or matching or "in line with" or
15	keep or kept or keeping or utili?ation or utile?e? or change?) adj5 (criteria or
16	recommend* or guideline? or guidance)).ti,ab.
17	24. clinical competence/ or nursing competence/
18	25. 21 or 22 or 23 or 24
19	26. 20 and 25
20 21	27. diagnostic procedure/ or exp blood examination/ or exp cardiovascular
21	
22	system examination/ or exp digestive system examination/ or exp endocrine
24	system examination/ or exp neurologic examination/ or exp respiratory tract
25	examination/ or exp urogenital system examination/
26	<ol> <li>(diagnos* or detect* or test* or screen* or manag*).ti.</li> </ol>
27	29. (imaging or radiolog* or tomogra* or ct scan* or pet scan* or echocardiogra*
28	or angiogra* or ultrasound* or ultrasonogra*).ti,ab.
29	30. ((medical or clinical or diagnos* or screening or routine or laboratory) adj5
30	(test* or investigation?)).ti,ab.
31	31. ((h?ematolog* or blood or urin* or saliva*) adj5 test*).ti,ab.
32	32. ((stress* or physical or function*) adj5 test*).ti,ab.
33 34	33. 27 or 28 or 29 or 30 or 31 or 32
35	34. 17 and 26 and 33
36	35. ((necessary or unnecessary or appropriate* or inappropriate* or waste* or
37	
38	utili?ation or indicated or excess* or less or more or increas* or decreas*) adj10
39	(test* or imaging or radiolog* or tomogra* or ct scan* or pet scan* or
40	echocardiogra* or angiogra* or ultrasound* or ultrasonogra* or
41	investigation?)).ti,ab.
42	36. ((order* or request*) adj5 (test* or imaging or radiolog* or tomogra* or ct
43	scan* or pet scan* or echocardiogra* or angiogra* or ultrasound* or
44	ultrasonogra* or investigation?)).ti,ab.
45 46	37. Unnecessary Procedures/
46 47	38. 35 or 36 or 37
47 48	39. 17 and 20 and 38
49	40. 17 and 33 and 37
50	41. 34 or 39 or 40
51	
52	42. limit 41 to yr="1999 -Current"
53	43. limit 42 to english language
54	44. (exp animals/ or nonhuman/) not human/
55	45. 43 not 44
56	46. conference*.pt.
57	47. 45 and 46
58 50	48. 45 not 46
59 60	49. exp child/ not (exp Child/ and exp Adult/)

50. 48 not 49 51. 48 not 49 52. limit 47 to yr="2015 -Current"

for occreation work

Erasmushogeschool . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

	Was the study"s target population a close representation of the national population in relation to relevant variables?	Does the inclusion criteria match the target population of guideline?	Were all eligible participants included in the study?	Was the likelihood of non- response bias <20?	Was an acceptable disease, test or symptom definition used?	Was data extracted/collected in an objective way?	Was the interval from symptoms to test clinically appropriate for the diagnosis of interest?	Did they report extractable measures?	Other bias
Ahmad2012	Low	Unclear	Low	Unclear	Low	Unclear	Unclear	Low	Low
Aljebreen 2013	Low	Low	Low	Low	Unclear	Unclear	Low	Low	High
Azzam 2015	Low	Low	Unclear	High	Low	Unclear	Low	Low	Low
Belletti2013	Low	Unclear	Unclear	Unclear	Low	Low	Low	Low	Low
Bertella 2013	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Low
Bhatt 2001	Low	Low	Low	High	High	Unclear	Unclear	Low	High
Birk-Urovitz 2017	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Bishop 2003	High	Low	Low	Low	Low	Low	Low	Low	Low
Cai 2015	Low	Low	Unclear	Low	Unclear	Unclear	Unclear	Low	Low
Caplan 2000	Low	Low	Low	Low	Unclear	Unclear	Unclear	Low	Low
Cardin 2005	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Cardin 2007	Low	Low	Low	Unclear	Low	Low	Low	Low	Low
Chan 2004	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Chan 2006	Low	Low	Low	High	Unclear	Low	Unclear	Low	Low
Chavez 2009	Low	Low	Low	Low	Low	Low	High	Low	High
Droogendijk 2011	Unclear	Low	High	Unclear	Low	Unclear	Low	Low	Low
Eccles 2001	Low	Low	Low	Low	Unclear	Unclear	Unclear	Low	Low
Elwyn 2007	Low	Low	Unclear	Unclear	Unclear	Low	Low	High	High

AT-LZA Downloaded from http://bmjopen.bmj.com/ on May 22, 2018. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA Erasmushogeschool Protected by copyright, instructory felpted for text and idata, mitring, but training, and similar technologies.

ATJ-ZED inemineda is 2202, 202, 200, imd.neqojmd//:qiif moni bebeolnwor. 518. Downloaded from http://pinace.com/ on May 22, 2025 at Department GEZ-LTB. Downloaded from http://pinace.com/ on May 22, 2025 at Department GEZ-LTB. Com/ on the second comparison of the second c

Protected by copyright, and hological selected for the selection of the main of the selection of the selecti

steh)	puc	. 170	Ŧ₩Ŧ	hot
IOOU	isse	6ou	Isnu	use

Gurzun 2014	Low	Low	High	Low	High	Low	Unclear	Low
Hassan 2007	Low	Low	Low	High	Unclear	Low	Unclear	Low
Heidi Lindbäck 2017	Low	Low	Low	Low	Low	Low	Unclear	Low
Hughes- Anderson 2002	High	Low	Unclear	Low	Unclear	Unclear	unclear	Low
lp2014	Low	Low	High	Unclear	Low	Unclear	Unclear	Low
Johnson 2011	Low	Unclear	Unclear	Low	High	Low	Low	Low
Kinouani 2017	Low	Low	low	Low	Unclear	Yes	Unclear	Low
Kovacs 2013	Low	Unclear	High	Low	Low	Low	Unclear	Low
Lalude 2014	Low	Low	Low	Low	High	Low	Unclear	Low
Landry 2011	Low	Unclear	Low	Low	Unclear	Unclear	Unclear	Low
Lange 2007	Low	Low	Unclear	Low	Unclear	Unclear	Unclear	Low
Lin 2016	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low
Linder 2006	Low	High	Unclear	Low	Low	Low	Low	Unclear
Lipczynska 2012	High	High	Unclear	Low	Low	Unclear	Low	Low
Llor 2011	Low	Low	Low	High	Low	Low	Low	Low

39

45

46 47

1 2

Fiorenza

2017

Gerrits2008

Gibbons

2010c

Girard 2010

Gnani 2004

Grover

2007

Gurzun

Low

Low

Low

High

Low

Low

Low

Unclear

Low

Low

Low

Low

Low

High

Low

Unclear

Unclear

Unclear

Low

Low

Low

High

Low

Low

High

Low

Low

Low

High

Low

Low

High

High

Low

Low

Low

Low

High

Low

Unclear

Unclear

Low

Unclear

Unclear

Unclear

Low

Low

Low

Low

Low

Low

Low

Low

Low

Unclear

Unclear

Low

Low

Low

High

High

Low

High

Unclear

Low

Low

High

Low

Low

BMJ Open: first published as 10.525, 2025 at Department GEZ-LTA. Downloaded from http://bmjopen.bmj.com/ on May 22, 2025 at Department GEZ-LTA. Etasmushogeschool .

.səipolor Protect

70 0-2	P = 0	7.0		7 19 97
loou	เวรอ	6ou	sn	wse.

ontast neimia ເປັນຜູ້ແມ່ນເອົາກໍ່ເອົາກໍາມີການ	etepipuenxanos parejar/sa	cobAri@th/waenidibio	cred by
	-,-,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,		

7-1-		,		Ŧ ~	7 17	
00	109	ະລຄ	iou	sn	1119	spi
	700					

w	is þaga brinnign	t ¶ƏlbGuliùi	&i/eteblenet	xaj ot bają	ið:F/SðSh	hiqididhide	Ϟϟϟ <b>ϴ</b> ϳͱΛdo
			1 100110006	ou opuiopi-	-		

									_
Loo 2009	Low	Low	Low	Low	Low	Low	Unclear	Low	Low
Mafi2013	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low	Low
Mafi2015	Low	Low	Unclear	Unclear	Low	Low	Unclear	Low	Low
Majumdar 2003	Low	Low	Unclear	Low	Low	Low	Unclear	Low	Low
Michaleff 2012	Low	Low	Low	Unclear	High	Low	Unclear	Low	Low
Moscavitch 2009	Low	Low	Unclear	Low	Low	Unclear	Low	Low	Low
Musicco 2004	High	Low	Low	Unclear	Unclear	Unclear	Unclear	High	High
Nicholson 2010	Low	Low	Low	Low	Unclear	Low	Low	Low	Low
Nicopoullos 2003	High	High	Low	High	Low	Unclear	Unclear	Low	High
Noya 2008	Low	Low	Low	Low	Unclear	Unclear	Unclear	Low	High
Piccoliori 2013	Low	Low	Low	Low	Low	Unclear	Low	Low	High
Pimlott 2006	Low	Low	Unclear	High	Unclear	Unclear	Unclear	Low	Low
Piterman 2008	Low	Unclear	Low	Unclear	Unclear	Low	Unclear	High	High
Remedios 2014	Low	Unclear	Low	High	Unclear	Low	Unclear	Low	High
Schers 2000	Low	Low	Low	Unclear	Unclear	Low	Unclear	Low	Low
Smith 2008	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Low
Sokol 2015	Low	Low	Low	Low	Low	Low	High	Low	Hig
Tahvonen 2017	Low	Low	High	Low	Low	Low	Unclear	Low	Low
Ulrik 2010	Low	Low	Low	High	Low	Low	unclear	Low	High
Ulrik 2013	Low	Low	Low	High	Low	Low	unclear	Low	Low
van der Pluijm- Schouten 2017	Low	Low	Low	Unclear	Low	Unclear	Unclear	Low	Lov

Page	28	of	43
------	----	----	----

	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Low
Williams 2010	Low	Low	Unclear	Low	Low	Unclear	Unclear	Low	Low

				Table 1: Study Characteristics	
Study	Country	Study length (days)	N (men%)	Population 1g 7 on 11	Test
		<u>.</u>	• •	Under-use	-
Ahmad 2012	Indonesia	181	554 (41%)	Patients registered at health clinics where TB was suspected	Sputum smear microscopy
Belletti 2013	USA	N/S	1517 (46%)	Patients with COPD	Pulmonary function tests
Bertella 2013	Italy	1765	437 (286)	Patients with COPD	PFTs
Caplan 2000	USA	365	81	Patients with a thyroid nodule	FNA of thyroid
Chavez 2009	USA	2920	200 (48%)	Patients with COPD	PFT
Droogendijk 2011	Netherlands	730	287 (45%)	Women >50yrs and men >18 years with Iron Deficien	Upper endoscopy and colonoscopy
Gerrits 2008	Netherlands	2556	65 (0%)	Women aged 18 – 65 yrs with newly diagnosed urinar inapintinence	Urine dipstick
Gibbons 2010	New Zealand	364	265	Patients with subclinical hypothyroidism	Free T4
Gnani 2004	UK	365	90 (53%)	Patients with heart failure	CXR, ECG and Echocardi
Girard 2010	France	28	19 (37%)	Patients with acute hepatitis	Hepatitis serology (HBs ar anti-HBc anitbodies)
Kinuoani 2017	France	150	61 (18%)	Patients with urinary tract infections	Urine Dipstick
Lange 2007	Denmark	91	2549 (44%)	Patients with COPD	PFTs
Lipczynska 2012	Poland	61	93	Aged ≥ 55 with Heart Failure (HF) or HF risk factors	Echocardiogram, BNP, C>
Loo 2009	UK	364	131 (50%)	Patients with Atrial Fibrillation	Echocardiogram
Majumdar 2003a	USA	2371	531 (47%)	Patients >50 years, on Proton Pump Inhibitors with persistent dyspepsia	Upper endoscopy
Majumdar 2003b	USA	2371	132 (47%)	Patients with peptic ulcer disease (PUD)	H.pylori
Moscavitch 2009	Brazil	61	167 (43%)	Patients with Heart Failure	ECG, CXR, Echocardiogra
Musicco 2004	Italy	NR	1549 (38%)	Patients being assessed for Dementia	Collection of laboratory tes rule out conditions with sir presenting symptoms to d

Pag	6/bmjopen-2017- 1 by copyright, ir	BMJ Open				
C. trachomatis, N. gonorrhoea urethral swabs and midstream urinalysis.	17-018557 . , including	Men with epididymo-orchtitis	6943 (100%)	1827	UK	Nicholson 2010
Mid-luteal progesterone and semen analysis	on 11 y for us	Patients with subfertility	32	242	UK	Nicopoullos 2003
FBC, TSH, serum electrolytes serum calcium, glucose	Febru Jes re	Patients with Dementia	160 (34%)	1611	Canada	Pimlott 2006
PFT	lary 2 Erasn lated	Patients with COPD	29870 (52%)	731	UK	Smith 2008
PFT	018. ⊑ nusho to tex	Patients with Asthma	75902 (23%)	3652	USA	Sokol 2015
PFT	yov ges an	Patients with COPD	1716 (44%)	121	Denmark	Ulrik 2010
PFT	nloa d da	Patients with COPD	4058	731	Denmark	Ulrik 2013
Chlamydia Antibody Titre and Semen Analysis	ary 2018. Downloaded fr Erasmushogeschool . ated to text and data mi	Patients (couples) referred to IVF clinics	100%*	840	Netherlands	van der Pluijm- Schouten 2017
	ning	<u>Over-use</u>				
Upper endoscopy	http J, Al	Patients who had upper endoscopy	147 (51%)	365	Saudi Arabia	Aljebreen 2013
Upper endoscopy	://b	Dyspeptic patients who had upper endoscopy	161 (30%)	121	Saudi Arabia	Azzam 2015
Pelvis x-ray	nin,	Patients referred for pelvis x-rays	437 (65%)	504	UK	Bhatt 2001
TSH	e (TSH) tegt	Patients that had a Thyroid Stimulating Hormone (	77 (38%)	1538	Canada	Birk-Urovitz 2017
Advanced imaging (CT, MRI o bone scan)	mj.cc simi	Patients with non-red flag LBP	139	28	Canada	Bishop 2003
Upper endoscopy	m/ ol lar te	Patients who under went upper endoscopy	550 (46%)	121	USA	Cai 2015
Upper endoscopy	n M	Patients who under went upper endoscopy	250 (45%)	153	Malaysia	Chan 2004
Colonoscopy	on May 22 ies'	Patients who underwent 'diagnostic colonoscopies	27 (63%)	184	Malaysia	Chan 2006
Upper endoscopy	2, 20) gies.	Patients who had upper endoscopy	1678	151	Italy	Cardin 2005
Upper endoscopy	125 at	Dyspeptic patients who had upper endoscopy	NR	182	Italy	Cardin 2007
Lumbar or knee x-ray	t Dej	Patients who had knee or lumbar x-ray	275	182	UK	Eccles 2001
Upper endoscopy	partr	Patients who under went upper endoscopy	215	184	UK	Elwyn 2007
Upper Endoscopy	ment	Patients who under went upper endoscopy	45 (34%)	456	USA	Fiorenza 2017

Page 31 of 43

## BMJ Open

				BMJ Open BMJ open-201	
Grover 2007	USA	364	68 (0%)	Patients with uncomplicated UTI	Urine culture and sensitivity
Gurzun 2014	UK	7	1070 (54%)	Patients who underwent an echocardiogram	analysis Echocardiogram
Hassan 2007	Italy	30	3769 (46%)		Upper endoscopy
Hughes- Anderson 2002a	Australia	1613	154 (55%)	<u> </u>	Colonoscopy
Hughes- Anderson 2002b	Australia	1613	162 (55%)	Patients who had colonoscopy     s     e       Patients who had upper endoscopy,     range of the short of t	Upper endoscopy
lp 2014	USA	1096	100 (43%)	Patients with non-red flag LBP	MRI lumbar spine
Johnson 2011	USA	510	779 (0%)	Patients with uncomplicated UTI	Urine culture
Kovacs 2013	Spain	183	602 (48%)	Patients with non-red flag LBP	MRI lumbar spine
Lalude 2014	USA	121	102	Patients who had SPECT Myocardial perfusion imaging (MPI) studies	Single Photon Emission CT (SPECT) MPI
Landry 2011	USA	272	124	Patients who had U/S of thyroid, pelvis, abdo, carotid gr soft tissue	Thyroid, pelvis, abdomen, caroti or soft tissue ultrasound
Lin 2016	Australia	NR	NR	Patients with non-red flag LBP	Lumbar Spine X-ray
Lindbäck 2017	Sweden	59	0	Patients that had urinary cultures	Urinary Culture
Linder 2006	USA	608	1076 (19%)	Patients with pharyngitis	Strep testing (rapid antigen detection test, throat culture)
Llor 2011	Spain	122	658 (0%)	Women with UTI	Urine cultures
Mafi 2013	USA	4377	8066	Patients with non-red flag LBP	X-ray, CT or MRI
Mafi 2015	USA	4018	9362 (25%)	Patients with uncomplicated headache (non-red flag g	CT and MRI
Michaleff 2012	Australia	3621	3070 (70%)	Patients reporting first time neck pain or LBP (non-specific non red flag)	Any radiological test
Noya 2008	Israel	N/S	209 (35%)	Patients who had H pylori testing	H. pylori test
Piccoliori 2013	Italy	63	475 (43%)	Acute or chronic non-red flag LBP	Any radiological test
Piterman 2008	Australia	550	19219	Patients with GORD	Endoscopy. Barium Swallow
Remedios 2014	UK	NR	2026	Patients who had CTs and/or MRIs	CT and/or MRI
Sharp 2015	USA	730	37,464	Patient with Acute Sinusitis	CT Sinuses

				BMJ Open	6/bmjop 1 by cop	Page 32
					6/bmjopen-2017-018557 9 by copyright, including	
Schers 2000	Netherlands	214	1096 (50%)	Patients with non-red flag LBP	nclud	X-ray
Tahvonen 2017	Finland	180	18 (35%)	Patients with non-red flag LBP	557 c ding	Lumbar Spine X-ray
Van Gurp 2013	Netherlands	366	155 (38%)	Patients who had Echocardiogram	for 1	Echocardiogram
Williams 2010	Australia	1005	1706 (43%)	Patients with non-red flag LBP	1 Fel	All imaging
				Patients With non-red flag LBP	/ 22, 2025 at Department GEZ-LTA logies.	
		Г	or heer review (	my - mcp.// pmjopen.pmj.com/site/about/guidelines.xm	.1111	

	Table 1:	BMJ Open BMJ Open Measures of inappropriateness Guideline authority and recommendation on	
Study	Test	ن بن ختر Guideline authority and recommendation ق ع ب 11	Measure of inappropriateness
		Under-use	
Girard 2010	Hepatitis B serology	Ministry of Health (France): Hepatitis serol	100% (82.4 to 10
Nicholson 2010	Neisseria Gonorrhoea serology	CDC (US)/British Association for Sexual Heater and HIV: Test for N. gonorrhoea for suspected of Epididymitis	99.9% (99.85 to 99
Nicholson 2010	Chlamydia Trachomatis	CDC (US)/British Association for Sexual Health and HIV: Test for C. Trachomatis for suspected of Epididymitis	97.4% (97.0 to 97
Nicopoullos 2003	Semen Analysis	Royal College of Obstetricians and Grant Gynaecologists: Semen analysis for Infedility	96.9% (83.8 to 99
Lipczynska 2012	Brain Natriuretic Peptide (BNP)	European Society Cardiology: BNP for Hear	95.7% (89.4 to 98
Musicco 2004	Collection of laboratory tests	European Federation of Neurological Societies Collection of laboratory tests to rule out conditions with similar presenting symptoms to demention	93.42% (92.1 to 9
Nicholson 2010	Urethral swabs	CDC (US)/British Association for Sexual Health and HIV: Urethral swabs for suspected epidid mits (Urethral swabs)	90.7% (89.9 to 9 <sup>4</sup>
Moscavitch 2009	Echocardiogram	Brazilian Society of Cardiology: Echocardioara hy for Heart Failure	88.6% (82.8 to 93
Kinouani 2017	Urine Dipstick	The French Agency for the Safety of Heorith Products: Urine Dipstick for UTI 로 드	84.4% (80.1 to 88
Majumdar 2003a	Upper Endoscopy	American Gastroenterological Associaten: Appropriate use of Upper Endoscopy for Dyspe	81.2% (78.8 to 83
Ulrik 2013	Pulmonary function tests (PFTs)	Danish National Board of Health: PFTs to diagnessis COPD	78.3 (77.3% to 79
Nicholson 2010	Mid stream	CDC (US)/British Association for Sexual Health and HIV: Midstream urinalysis for suspected Epididy ditis	78.2 (77.3 to 79.
Sokol 2015	Pulmonary function tests (PFTs)	National Asthma Education and Prevention Program (US): PFTs for asthma	76.5% (64.6 to 85
Belletti 2013	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	73.0% (70.7 to 75

		6/bmjopen-201 J by copyright, BMJ Open	
		5 7	
Ahmad 2012	Tuberculosis smear	World Health Organisation: Smear for suspected TB	72.4% (68.5 to 76.1%
van der Pluijm- Schouten 2017	Semen Analysis	NHG-Standaard Subfertiliteit: Semen Analysis	70.4% (61.9 to 77.9%
Droogendijk 2011	Colonoscopy	Ministry of Health (Netherlands): Colonoscepy for unexplained Iron Deficiency Anaemi	68.6% (62.9 to 74.0%
Pimlott 2006	Serum Calcium	Canadian Consensus Conference on Dementa Serum Calcium for Dementia និញទី	65.0 (57.1 to 72.4%)
Gerrits 2008	Urine dip stick	NICE: Urine dip stick for urinary incontine	60.0% (47.1 to 72.0%
Nicopoullos 2003	Mid-luteal progesterone	Royal College of Obstetricians and Gynaeco be	59.4% (40.6 to 76.3%
Gnani 2004	Echocardiogram	Royal College of Obstetricians and Gynaecologists: Mid-luteal progesterone for Infertility 2 0 Department of Health (UK): Echocardiograms for Heart Failure	57.8% (46.1 to 68.1%
Loo 2009	Echocardiogram	ACC, AHA, ESC: Echocardiogram to identif梁函證es or complications of atrial fibrillation	55.7% (46.8 to 64.39%
Ulrik 2010	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Eung Disease (GOLD): PFTs for COPD	55.0% (52.6 to 57.4%
Lipczynska 2012	Echocardiogram	European Society Cardiology: Echocardiogram	53.8% (43.1 to 64.2%
van der Pluijm- Schouten 2017	Chlamydia Trachomatis	NHG-Standaard Subfertiliteit: Chlamy	57.8% (49.0 to 66.2%
Gnani 2004	ECG	Department of Health (UK): ECG for Heart all are	51.1% (40.4% to 61.8%
Moscavitch 2009	ECG	Brazilian Society of Cardiology: ECG for	46.1 (38.4 to 54.0)
Moscavitch 2009	Chest X-ray	Brazilian Society of Cardiology: CXR for Fieag Failure	44.9% (37.2 to 52.8%
Gibbons 2010	Thyroid function tests	New Zealand Best Practice: Appropriate See of Thyroid Function tests	44.9% (38.8 to 51.1%
Chavez 2009	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Bung	41.5% (34.6 to 48.7%
Lange 2007	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive Lung Disease (GOLD): PFTs for COPD	40.0% (38.1 to 42.0)
Smith 2008	Pulmonary Function Tests (PFTs)	NICE: PFTs for COPD	38.1% (37.5 to 38.6%
Pimlott 2006	Glucose testing	Canadian Consensus Conference on Demente	36.9% (29.4% to 44.9%
Gnani 2004	Chest X-ray	Department of Health (UK): CXR for Heart Fail	36.7% (26.8 to 47.5%

		BMJ Open BMJ Open	
Majumdar 2003b	H.pylori	American Gastroenterological Association/American College of Gastroenterology: appropriateness of H.pylori test	34.4% (28.9 to 40.3
Pimlott 2006	Thyroid Stimulating Hormone (TSH)	Canadian Consensus Conference on Demeraia: TSH for dementia	33.1% (25.9 to 41.0
Lipczynska 2012	Chest x-ray (CXR)	European Society Cardiology: CXR for Hear	25.8% (17.3 to 35.9
Bertella 2013	Pulmonary function tests (PFTs)	Global Initiative for Chronic Obstructive ដំរាញ Disease (GOLD): PFTs for COPD ថ្មី ខ្ល	25.6% (21.6 to 30.0
Pimlott 2006	Serum electrolytes	Canadian Consensus Conference on Denze Ita: Serum electrolytes for dementia	21.9% (15.7 to 29.1
Pimlott 2006	Full Blood Count (FBC)	Canadian Consensus Conference on Demertie BC for dementia	18.1% (12.5 to 25.0
Caplan 2000	Fine needle aspiration (FNA) of thyroid	American Thyroid Association/American Association of Clinical Endocrinologists: FNA for thyroid	9.9% (4.4 to 18.5%
		<u>Over-use</u>	
Piterman 2008	Barium Swallow	Gastroenterological Society of Australia:	94.20% (93.9 to 94.5
Gurzun 2014	Echocardiogram	American College of Cardiology: Appropriate use of Echocardiography	92.0% (90.2% to 93.
Linder 2006	Streptococcal throat cultures	American College of Physicians/Infectious Biserse Society of America: Do not order strept test for centor criteria 0 or 1 in pharyngitis 2	91.5% (89.7 to 93.1
van Gurp 2013	Echocardiogram	Netherlands Society of Cardiology: Appropriate ise of Echocardiography	76.7% (76.4 to 77.0
Grover 2007	Urine cultures	Infectious Disease Society of America: Urine cultures not required for uncomplicated UTI diagrapsis	76.5% (64.6 to 85.9
Eccles 2001	Knee x-ray	Royal College of Radiologists: No x-ray for kalee ani without restriction of movement	74.7% (69.6 to 79.3
Cardin 2005	H. Pylori breath test	European Society of Primary Care Gastroentero	74.4% (58.8 to 86.5
Tahvonen 2017	L spine x-ray	Appropriate use of H. pylori 🦉 😕 Finish Medical Society: LBP among adults 🕃	68.0% (53.3 to 80.48
Johnston 2011	Urine cultures	European Association of Urology: Urinary cultures not required for uncomplicated urinary tracted infections	57.4% (53.8 to 60.9
Bhatt 2001	Hip x-ray	Royal College of Radiologists (UK): No hip x-rag for hip pain without restriction of movement	57.2% (52.5 to 61.8
		GEZ-LTA	

Page	36	of	43

Eccles 2001	Lumbar spine x-ray	BMJ Open BMJ Open Copy right, in Copy righ	56.4% (50.3 to 62.3%
Piterman 2008	Upper endoscopy	Gastroenterological Society of Australia: Oppor endoscopy for GORD	53.5% (52.8 to 54.2%
Chan 2006	Colonoscopy	American Society for Gastrointestinal Endoscopy: Appropriateness of Colonoscopy	51.9% (32.0 to 71.3%
Heidi Lindbäck 2017	Urine Culture	Swedish Medicines Agency: Treatment of dower urinary tract infections in women	47.0% (40.8 to 53.389
Aljebreen 2013	Upper endoscopy	The American Society for Gastrointesting	46.9% (38.7 to 55.3%
Elwyn 2007	Upper endoscopy	NICE: Appropriate tests for dyspepsi	45.1% (38.3 to 52.0%
Noya 2008	H.Pylori	The European Helicobacter Study Gro	44.5 (37.6 to 51.5%)
Cardin 2005	Upper endoscopy	European Society of Primary Care Gastroen Beegy: Upper Endoscopy for H.pylori	44.1% (35.9 to 52.6%
Lin 2016	L spine x-ray	Upper Endoscopy for H.pylori         n.         <	40.9% (26.3 to 56.8%
Cardin 2007	Upper endoscopy	European Society of Primary Care Gastroenterongy: Upper Endoscopy for H.pylori	41.9% (38.3 to 45.5%
Fiorenza 2017	Upper Endoscopy	American College of Gastroenterology: Dy	42.4% (36.8 to 48.1%
Cai 2015	Upper endoscopy	American College of Physicians: Upper endescopy for GORD	37.7% (33.8 to 42.0%
Azzam 2015	Upper endoscopy	American Gastroenterological Association: Uper Endoscopy for Dyspepsia	36.7 (29.2 to 44.6%)
Llor 2011	Urine cultures	European Association of Urology: Urinary altures not required for uncomplicated urinary tractor infections	35.9% (32.2 to 40.0%
Hassan 2007	Upper endoscopy	The American Society for Gastrointestical	29.4% (28.0 to 30.9%
Landry 2011	Carotid ultrasound	Canadian Association of Radiologists 2005 guidelines: Carotid U/S	25.0% (17.7 to 33.6%
Piccoliori 2013	Lumbar spine radiology (all)	Ministry of Health (Italy): No imaging for non-red	24.0% (20.2 to 28.1%
Michaleff 2012	Lumbar spine x-ray	National Health and Medical Research Cound (Australia) (NHMRC): No x-ray for non-red flag	24.0% (22.9 to 25.1%

		6/bmjopen-2017 4 by copyright, i	
		<b>3</b> L	
Williams 2010	Lumbar spine radiology (all)	National Health and Medical Research Council (Australia) (NHMRC): No imaging for non-red figg LBP	23.9% (21.9 to 26.09
Michaleff 2012	Cervical spine x-ray	Australian National Health and Medical Regearch Council: No x-ray for neck pain 5	22.8% (21.3 to 24.3%
Birk-Urovitz 2017	Thyroid Stimulating Hormone	The Canadian Task Force on Preventative	22.4% (16.9 to 28.8%
lp 2014	Lumbar spine MRI	American College of Physicians/American Society: no MRI for non-red flag LBF	22.0% (14.3 to 31.4%
Williams 2010	Lumbar spine x-ray	National Health and Medical Research Codid (Australia) (NHMRC): No x-ray for non-red 協答的	20.2% (18.3 to 22.2%
Landry 2011	Thyroid ultrasound	Canadian Association of Radiologists 2000 guidelines: Thyroid U/S	19.0% (12.1 to 27.0%
Lalude 2014	Single Photon Emission Computed Tomography	American College of Cardiology: SPECT for the st pain	18.6% (11.6 to 27.6%
'Mafi 2013	Lumbar spine x-ray	American College of Physicians/American Path Society: no x-ray for non-red flag LBP: 2009-20010	13.0% (11.1 to 15.1%
		2007-2008 g n ht	12.9% (11.1 to 14.99
		2005-2006 trail b	12.8% (11.0 to 14.8%
		2003-2004	12.3% (10.7 to 14.0%
		2001-2002 ag	12.0% (10.3 to 13.8%
		1999 - 2000 si j	11.8% (10.2 to 13.6%
Landry 2011	Abdominal ultrasound	Canadian Association of Radiologists 2005 guidelines: Abdominal U/S	12.1% (6.9 to 19.2%
Mafi 2015	CT or MRI Brain	The American Headache Society/American Academy of Neurology for Choosing Wisely Noa CT or MRI for non-red flag headache 2009 - 2010	13.9% (12.2 to 15.7%
			13.5% (11.8 to 15.3%
		2005 – 2006 <sup>35</sup>	9.4% (8.0 to 11.0%
		2003 – 2004	7.5% (6.3 to 8.9%)
		2001 – 2002 art	7.1% (5.9 to 8.4%)
		1999 - 2000 <b>Per</b>	6.7% (5.4 to 8.2%)
		GEZ-LTA	

Page	38	of 4	3
		• • •	-

Kovacs 2013	Lumbar spine radiology tests (all)	BMJ Open BMJ Open NICE, ACP: No imaging for LBP	12.0% (9.5 to 14.8%)
Chan 2004	Upper endoscopy	The American Society for Gastrointestigal 긱 Endoscopy: Appropriateness of Upper Endascopy	10.4% (6.9 to 14.9%)
Hughes-Anderson 2002a	Colonoscopy	American Society for Gastrointestinal Endoscopy: Appropriateness of Colonoscopy	8.2% (5.3 to 12.1%)
Hughes-Anderson 2002b	Upper endoscopy	The American Society for Gastrointestinal	7.5% (4.7 to 11.1%)
Remedios 2014	CT (any)	Royal College of Radiologists (UK): C별 방국 Royal College of Radiologists (UK): MR 년 호	6.9% (5.8 to 8.1%)
Remedios 2014 —	MRI (any)	Royal College of Radiologists (UK): M	5.2% (4.1 to 6.5%)
Bishop 2003	Lumbar spine radiology tests (all)	Workers Compensation Board of British Collingia: No imaging for non-red flag LBP	5.0% (2.1 to 10.1%)
Williams 2010	Lumbar spine CT	National Health and Medical Research Congo	3.7% (2.9 to 4.7%)
Schers 2000	Lumbar spine radiology tests (all)	The Netherlands College of General Practitioners:	3.1% (2.2 to 4.3%)
Landry 2011	Soft tissue ultrasound	Canadian Association of Radiologists 2005 guidelines: Soft tissue U/S	2.4% (0.5 to 6.9%
Landry 2011	Pelvic ultrasound	Canadian Association of Radiologists 2005	1.6% (0.2 to 5.7%)
Sharp 2015	CT Sinuses	American Academy of Otolaryngology: 셔odulo	0.60% (0.56 to 0.65%)
Williams 2010	Lumbar spine Ultrasound	National Health and Medical Research Council (Australia) (NHMRC): No U/S for non-red figg LBP	0.59% (0.28 to 1.1%)
Williams 2010	Lumbar spine MRI	National Health and Medical Research Council (Australia) (NHMRC): No MRI for non-red flag L&P	0.18% (0.04 to 0.5%)
		n May 22, 2025 at Department GEZ-LTA chnologies.	

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

## MOOSE Statement - Reporting Checklist for Authors, Editors, and Reviewers of Meta-analyses of Observational Studies

Reporting Criteria	Reported (Yes/No)	Reported on Page
Reporting of Background		
Problem definition	YES	4
Hypothesis statement	YES	4
Description of Study Outcome(s)	YES	4
Type of exposure or intervention used	N/A	N/A
Type of study design used	YES	5, 6
Study population	YES	5
Reporting of Search Strategy		
Qualifications of searchers (eg, librarians and investigators)	YES	5
Search strategy, including time period included in the synthesis and keywords	YES	5, supplementary fil
Effort to include all available studies, including contact with authors	YES	5
Databases and registries searched	YES	5
Search software used, name and version, including special features used (eg, explosion)	YES	5
Use of hand searching (eg, reference lists of obtained articles)	YES	5
List of citations located and those culture control co	NO	
Method for addressing articles	NO	
published in languages other than English	R.	
Method of handling abstracts and unpublished studies	YES	5
Description of any contact with authors	N/A	
Reporting of Methods		
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	YES	5,6
Rationale for the selection and coding of data (eg, sound clinical principles or convenience)	YES 🥏	6
Documentation of how data were classified and coded (eg, multiple raters, blinding, and interrater reliability)	YES	6
Assessment of confounding (eg, comparability of cases and controls in studies where appropriate	N/A	N/A
Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results	YES	5,6
Assessment of heterogeneity	YES	6

1	
2	
3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
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	

Description of statistical methods (eg,	YES	6
complete description of fixed or random		
effects models, justification of whether		
the chosen models account for predictors		
of study results, dose-response models,		
or cumulative meta-analysis) in sufficient		
detail to be replicated	N/50	
Provision of appropriate tables and	YES	Tables 1,2, Figures
graphics		2,3,4
Reporting of Results	N/50	<b>T 11 4 17 11 0</b>
Table giving descriptive information for	YES	Table 1 and Table 2
each study included		7.0
Results of sensitivity testing (eg,	N/A	7, 8
subgroup analysis)		07.00
Indication of statistical uncertainty of	YES	6,7, 8,9
findings		
Reporting of Discussion	YES	
Quantitative assessment of bias (eg, publication bias)	TES	8,9
Justification for exclusion (eg, exclusion	YES	5
of non–English-language citations)	TES	5
Assessment of quality of included studies	YES	7, Table 3
Reporting of Conclusions	120	
Consideration of alternative explanations	YES	9, 10
for observed results	120	
Generalization of the conclusions (ie,	YES	10
appropriate for the data presented and		
within the domain of the literature review)		
Guidelines for future research	YES	9, 10
Disclosure of funding source	YES	11
<b></b>		1

2021

Page 41 of 43



3

47

## PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 'Search strategy'
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5 & supplementary figure
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5&6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5&6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5, 6, 7 & supplementary figure
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5,6

BMJ Open



## **PRISMA 2009 Checklist**

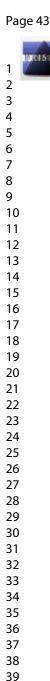
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.	6
		Page 1 of 2	
Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7, 8
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	7, 8, 9
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	7
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	7, 8
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	9
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	10
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	10
FUNDING	<u> </u>		
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	11

44 From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. 45 doi:10.1371/journal.pmed1000097 Protected by copyright, instriction for test and test tand data minus, by the maining and similar technologies.

ATA-ZEB them in the store of th

Page 43 of 43

**BMJ** Open



45

46

47



For more information, visit: www.prisma-statement.org

, "