Postprint Version	1.0
Journal website	http://www3.interscience.wiley.com/journal/122491283/abstract
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/19575763
DOI	10.1111/j.1365-2036.2009.04087.x

Systematic review: accuracy of symptom-based criteria for diagnosis of irritable bowel syndrome in primary care

P. JELLEMA*, D. A. W. M. VAN DER WINDT*,†, F. G. SCHELLEVIS*,‡ & H. E. VAN DER HORST*

*Department of General Practice, EMGO Institute for Health and Care research, VU university medical center, Amsterdam, The Netherlands ;
†The Arthritis Research Campaign National Primary Care Centre, Keele University, Keele, Staffordshire, UK ;
‡NIVEL (Netherlands Institute for Health Services Research), Utrecht, The Netherlands

Correspondence to Dr H. E. van der Horst, Department of General Practice, EMGO Institute for Health and Care research, VU university medical center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands. E-mail: he.vanderhorst@vumc.nl

ABSTRACT

Background Despite the trend towards making a positive diagnosis of irritable bowel syndrome (IBS), many health care providers approach IBS as a diagnosis of exclusion.

Aim To summarize available evidence on the diagnostic performance of symptom-based IBS criteria in excluding organic diseases, and of individual signs and symptoms in diagnosing IBS and to additionally assess the influence of sources of heterogeneity on diagnostic performance.

Methods We searched PubMed and EMBASE and screened references. Studies were selected if the design was a primary diagnostic study; the patients were adults consulting because of non-acute abdominal symptoms; the diagnostic test included an externally validated set of IBS criteria, signs, or symptoms. Data extraction and quality assessment were performed by two reviewers independently. The review adhered to the most recent guidelines as described in the Cochrane Diagnostic Reviewers' Handbook.

Results A total of 25 primary diagnostic studies were included in the review. The performance of symptom-based criteria in the exclusion of organic disease was highly variable. Patients fulfilling IBS criteria had, however, a lower risk of organic diseases than those not fulfilling the criteria.

Conclusions With none of the criteria showing sufficiently homogeneous and favourable results, organic disease cannot be accurately excluded by symptom-based IBS criteria alone. However, the low pre-test probability of organic disease especially among patients who meet symptom-based criteria in primary care argues against exhaustive diagnostic evaluation. We advise validation of the new Rome III criteria in primary care populations.

INTRODUCTION

Irritable bowel syndrome (IBS) is a chronic condition characterized by abdominal pain or discomfort associated with disordered bowel habit[1] and is one of the most common gastrointestinal (GI) disorders.[2] Prevalence estimates in Western general populations range from 10% to 26%.[3] Despite the bothersomeness

of its symptoms, only a minority (30%) of patients with IBS seek medical care,[4, 5] with most of them being managed in primary care.[4, 6] Given the high prevalence of IBS, this results in a large number of primary health care visits.

Roughly, two opposing approaches can be distinguished in establishing the diagnosis of IBS: IBS as a 'diagnosis of exclusion' vs. IBS as syndromic condition on its own. The first approach is characterized by more or less extensive diagnostic testing and invasive investigations to exclude possible organic causes,[1] while the second approach is characterized by using symptom-based IBS criteria to establish a positive diagnosis of IBS.[6] The symptom-based approach was initiated in 1978 with the design of the Manning criteria.[7] The most recent criteria are the Rome III criteria.[8]

Despite a growing trend towards making a positive diagnosis of IBS, many health care providers approach IBS as a diagnosis of exclusion.[9] This practice is expensive,[1, 10] may increase patients' anxiety[11] and is in conflict with several guidelines (e.g. AGA position Statement).[12] Review articles conclude that the routine use of diagnostic tests to exclude organic GI disease is not recommended in patients fulfilling IBS symptom criteria without demonstrating alarm features,[1, 13, 14] but may be indicated among those with alarm features.[13–15] However, several of those reviews and other reviews on the diagnosis of IBS lack a systematic approach in collecting the evidence, do not assess its methodological quality, do not differentiate between settings of care or need updating.

In this systematic review, we first studied the diagnostic performance of several symptom-based IBS criteria in excluding organic GI diseases, investigating the probability of uncovering an organic cause of symptoms when people have a positive score on the IBS criteria. Secondly, we studied which individual signs and symptoms (including alarm features) are important in differentiating IBS from organic diseases. Finally, we assessed the influence of sources of heterogeneity like differences in setting, study population and study design on diagnostic performance.

MATERIALS AND METHODS

Data sources and searches

We searched PubMed and Embase for all eligible diagnostic studies (to September 2008). The search strategy used MeSH/EMTREE terms and free text words and included sub-searches related to the index test, target condition, study population and publication type. A methodological filter for the identification of diagnostic studies was added to increase the specificity of the search. The full search strategies can be obtained from the corresponding author on request.

Reference lists of all retrieved primary diagnostic studies were checked for additional relevant diagnostic studies. Additionally, references were checked of relevant reviews, meta-analyses, guidelines and commentaries.

Study selection

Two authors (PJ, DvdW) independently applied the pre-defined selection criteria (see below). PJ checked all citations (titles and abstracts) identified by the search strategy, while DvdW checked eligibility of all citations assessed by PJ as (possibly) relevant, as well as a random selection of citations assessed as not relevant. Next, full publications were retrieved for studies that seemed relevant and for those for which relevance was still unclear. A third reviewer was consulted in cases of persisting disagreement.

Participants, setting and study design. We considered primary diagnostic studies eligible if the study population consisted of adult patients (≥18 years) consulting because of non-acute abdominal symptoms (symptom duration ≥2 weeks). We intended to include only primary care studies, but because of the low number, we decided to include also studies including patients visiting an outpatient GI clinic.

We considered primary diagnostic studies with a cohort design eligible. Of the case-control study designs, only nested case-control designs in which controls formed a representative sample of all patients with abdominal symptoms and case-control designs in which controls were diagnosed with a specific gastro-intestinal disease were eligible. We excluded studies for which no two-by-two table could be extracted or reconstructed (including case series), papers written in other languages than English, Dutch, German or French, and reviews, editorials and case reports.

Target condition and reference test. For the first research question ('performance of symptom-based criteria in excluding organic GI diseases'), IBS is a diagnosis of exclusion. Therefore, only studies including diagnostic work-up or a clinical follow-up of at least 1 year as 'reference test' were considered for inclusion.

For the second research question ('performance of signs and symptoms in identifying IBS'), IBS is a syndromic condition on its own. Therefore, only studies using symptom-based IBS criteria as reference test were eligible. As estimates of performance will always be overly optimistic for the population in which a set of criteria have been developed,[16] only IBS criteria that have been externally validated were included.

Index tests. For the first research question, a study had to include externally validated IBS criteria as index test. To answer the second research question, a study had to include information on (elements of) history taking (symptoms) and/or physical examination (signs). We were interested in signs and symptoms that are regarded as 'alarm features' (e.g. rectal bleeding) and in non-alarm features (e.g. abdominal distension). To prevent overestimation of diagnostic performance, signs and symptoms should not form part of the criteria list studied.

Data extraction and risk of bias

The reviewers (PJ, DvdW, HvdH) extracted data on setting and design, study population, test characteristics and test results using a standardized form. Risk of bias was assessed using a modified version of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool,[17] which is recommended by the Cochrane Diagnostic Reviewers' Handbook.[18] This modified version consists of 11 items on methodological characteristics that have the potential to introduce bias (Table S1). Items were scored as 'positive' (i.e. no bias), 'negative' (i.e. potential bias) or 'unclear'.

Data extraction and risk of bias were first pre-tested using two studies not included in the review. After this, PJ scored all studies, while DvdW and HvdH each scored about half of the studies, independently of each other. Inter-observer agreement was quantified and disagreements were resolved by consensus meetings, and, in cases of persisting disagreement, by consulting a third reviewer.

As recommended by the designers of the QUADAS, we did not apply weights to the QUADAS items, nor used a summary quality score in the analysis. Instead, we decided, a priori, to explore by subgroup or meta-regression analyses whether scores on the following items explained variation in diagnostic performance: item 1 (validity of study sample), item 2 (index test blind for reference standard) and 8 (reference standard blind for index test).

Data synthesis and analysis

We present diagnostic two-by-two tables and diagnostic performance measures per study. Diagnostic performance measures and corresponding 95% confidence intervals (95%CIs) were calculated using MetaDiSc statistical software.[19] When four or more studies on a specific index test showed sufficient clinical and statistical homogeneity, we used bivariate analyses to calculate pooled estimates of sensitivity and specificity and 95%CIs for the summary estimates.[20] The bivariate analyses take into account both within- and between-study variability and perform better than SROC regression models derived with the Moses and Littenberg method,[21] which departs from a fixed effects model. We refrained from pooling when there was considerable clinical or statistical heterogeneity.

Investigations of heterogeneity. Factors that may contribute to variation in diagnostic performance across studies (heterogeneity) include differences in (a) setting: primary vs. secondary care; (b) cut-off scores: Manning 2 of 4 vs. 3 of 6; (c) study population: inclusion vs. exclusion of patients with alarm features or upper GI disorders; (d) QUADAS items 1, 2 or 8 (as described above).

Subgroup analyses were only performed when each subgroup included data of at least two diagnostic studies. In case each subgroup included data of at least four studies with homogenous results per subgroup, we calculated per subgroup, a pooled estimate of sensitivity and specificity using bivariate analyses. In case subgroups included data of less than four studies or data of at least four studies showing, however, heterogeneous results on visual inspection, we presented the range of sensitivity and specificity per subgroup. Studies providing insufficient information on a factor were not included in that specific subgroup-analysis.

RESULTS

Literature search and study selection

The literature search yielded 858 references. A total of 89 full papers were retrieved, of which 26 were finally considered relevant for the review.[7, 22–46] Reference checking yielded two additional relevant papers.[47, 48] With results of two studies being published in a total of five papers,[35, 36, 40–42] a total of 25 primary diagnostic studies were included in the review. A summary of the search results is presented in Figure 1.

[FIGURE 1]

Study characteristics

Table S2 presents details of the primary diagnostic studies included in the review. Four studies[26, 28, 34, 43] were case-control studies; three studies[7, 24, 36] used chart review for data collection. Nine studies included primary care patients, with three[22, 33, 36] including solely primary care patients. Although all studies mention that they included 'patients with IBS', studies implicitly or explicitly defined IBS as (a) non-organic/functional disease without differentiating between upper and lower disease; (b) lower functional GI disease; or (c) a diagnosis using symptom-based IBS criteria. In the first two definitions, IBS is a diagnosis by exclusion, while in the third definition IBS is a positive diagnosis.

The following symptom-based IBS criteria were included as index test: the Manning ($n = 11$), Kruis ($n = 7$), Rome I ($n = 3$), Rome II ($n = 4$), Bellentani ($n = 2$), Mazumdar ($n = 2$), Talley ($n = 2$), and Wasson criteria ($n = 3$) (see Appendix S1 for a description of all sets of criteria). Only two studies[28, 34] used signs and symptoms as index test while using symptom-based criteria (Rome I $n = 1$; Rome II $n = 1$) as reference standard. While most studies did not report on the presence of alarm features in the study population, one study[22] excluded patients presenting with alarm features and two other studies[7, 44] explored the influence of alarm features on diagnostic performance. Three studies[7, 23, 44] explicitly excluded patients with upper GI disease, while six studies[24, 29, 30, 37, 38, 41] explored the influence of including upper functional disease on diagnostic performance.

Risk of bias. On average, the reviewers disagreed in 4 of 11 items (range 1–8). Disagreements mainly concerned incorporation bias (item 3), bias by presence of other clinical information (item 4), and bias by differential verification (item 7). All disagreements were resolved during consensus meetings. Table S3 presents the results of the assessment. Most studies performed well (score '+') on blind assessment of index test (item 2), use of an appropriate reference standard (item 5) and avoidance of partial verification (item 6). Especially, the time between index and reference test (item 9) was poorly described (score '?'). Generally, five studies25, 30, 37, 41, 44 performed well receiving a positive assessment of at least 8 out of 11 QUADAS items.

Can symptom-based IBS criteria exclude organic disease?

Twenty studies investigated the performance of symptom-based criteria to exclude any organic disease and nine studies the performance to exclude a specific disease (e.g. colorectal cancer).

Exclusion of any organic disease. Table 1 presents the diagnostic performance of symptom-based IBS criteria in the exclusion of any organic disease. Kruis and Manning criteria were the most frequently studied criteria. Estimates of diagnostic performance varied widely, both across sets of criteria and between studies evaluating the same set of criteria except for the Wasson criteria that showed consistently low values for sensitivity (34–46%), but high values for specificity (82–91%).[36, 39, 45] The risk of organic disease in test positives ranged from 5% to 24% and in test negatives from 24% to 54%. The diagnostic performance of especially the Rome II criteria appeared to be poor with sensitivities ranging from 31% to 65% and specificities ranging from 30% to 100%.[7, 22, 29, 46] The poorest results for the Rome II criteria were reported by Banerjee et al.,[22] who used strict selection criteria by including only patients with chronic or recurrent lower abdominal pain (>2 years), altered bowel habits and bloating in whom extensive testing had failed to show an organic cause. The diagnostic performance of the Bellentani, Mazumdar, and Talley criteria appeared to be considerably worse in the validation cohorts than in the development cohort (Table 1).

[TABLE 1]

Exclusion of specific diseases. Eight studies investigated the performance of IBS criteria to exclude specific diseases (Table 2). The risk of specific disease among those with a positive score on the criteria, [i.e. 1 - positive predictive value (1-PPV)] and among those with a negative score [negative predictive value (NPV)] is presented in Table 2. The 1-PPV ranges from 1% for inflammatory bowel disease to 14% for non-ulcer dyspepsia, while the NPV ranges from 1% for malabsorption to 46% for non-ulcer dyspepsia. These calculations were not possible for the two studies with a case-control design.[26, 43]

We also present a ratio representing the risk of specific disease for patients with a positive score on the IBS criteria, relative to the risk for those with a negative score. For the cohort studies, this ratio is expressed as a relative risk ratio and for case-control studies, as an odds ratio. With the exception of the study of Poynard *et al.*[37] which showed a significantly increased risk for inflammatory bowel disease in test positives, the ratio for other studies varied between 0 and 1 for cancer, inflammatory bowel disease, peptic ulcer and non-ulcer dyspepsia, implying that patients who fulfil the criteria for IBS have a lower risk for these diseases than those not fulfilling these criteria. For most studies, the ratios were significantly different from 1.0 ($P < 0.05$). For both studies on malabsorption diseases, the ratio was higher than 1, but the confidence intervals were wide and the findings not statistically significant ($P > 0.05$).

Which signs and symptoms are important?

Only two studies investigated the diagnostic performance of individual signs and symptoms in the identification of IBS as diagnosed with an IBS symptom list; one study used the Rome I criteria[34] and the other study the Rome II criteria[28] (Table S4). Overall, the specificity of individual signs and symptoms was higher when IBS was compared with a combination of upper and lower organic diseases than with lower organic disease only.

To identify IBS reliably, an alarm feature's specificity must be close to 0, while a non-alarm feature's specificity should be close to 1. When this condition is met, the absence of an alarm feature (e.g. blood in stools) confirms IBS and so does the presence of a non-alarm feature (e.g. abdominal distension). With values varying between 27% and 93%, none of the alarm features appears to have a low specificity. Specificity of non-alarm features varied between 42% and 88% and was higher for 'infrequent bowel movements'. Presence of infrequent bowel movements may therefore be helpful in establishing an IBS diagnosis.

Pre-planned subgroup analyses

For reasons of the minimum number of studies required to enable a subgroup-analysis, pre-planned analyses were only possible for the diagnostic performance of symptom-based IBS criteria in the exclusion of organic disease as a group. Study results were entered only once in a between-study subgroup-analysis (see marks in Table 1).

The subgroup-analysis on setting of care was possible for the Kruis and Rome II criteria. When Kruis criteria were applied in primary care,[23, 33] sensitivity appeared to be higher and specificity lower than in secondary care.[24, 25, 27, 31, 39] For the Rome II criteria, sensitivity and specificity were lower in primary care[22, 46] compared with secondary care.[7, 29] Persisting heterogeneity in subgroups led us to refrain from statistical pooling. Variation in diagnostic estimates for the Manning criteria was not explained by using either a cut-off score of 2 out of 4[32, 39, 47] or 3 out of 6 criteria.[25, 29, 30, 37, 38, 41]

The subgroup-analysis on the influence of upper functional GI disease was only possible for the Manning criteria. Five studies presented within-study comparisons[29, 30, 37, 38, 41] and all showed that sensitivity of the Manning criteria was lower when the functional disease group also included patients with upper functional disease (pooled estimate 0.51, 95% CI 0.48–0.54 vs. 0.64, 95% CI 0.59–0.68). Results in the between-study subgroups were too heterogeneous to confirm this finding.[25, 29, 30, 32, 37–39, 41, 47] Two studies[7, 44] on the Rome criteria explored the influence of alarm features on diagnostic performance. While the exclusion of patients with alarm features had no influence on the diagnostic performance of the Rome I criteria,[44] it resulted in somewhat better performance measures for the Rome II criteria.[7] However, in this last study, confidence intervals were wide and this study had few positive scores on the QUADAS tool. Results in the between-study subgroups were too heterogeneous to confirm this finding.[7, 22, 29, 46]

For reasons of many scores remaining 'unclear' on the QUADAS tool and the low number of studies per index test, only the potential influence of a selective study sample (item 1) as source of heterogeneity could

be studied and only for the Manning criteria.[29, 32, 37–39, 41] Results in subgroups were, however, too heterogeneous to detect a pattern.

DISCUSSION

The performance of symptom-based IBS criteria in the exclusion of organic diseases was highly variable, both across criteria sets and between studies evaluating the same IBS criteria. With regard to the exclusion of specific GI diseases, our results showed that patients who fulfilled IBS criteria have a lower risk for specific GI diseases than those not fulfilling the criteria. The absence of an alarm symptom does not seem to be helpful in confirming an IBS diagnosis, while presence of infrequent bowel movements does.

How to reliably diagnose IBS?

With none of the criteria showing sufficiently homogeneous and favourable results, organic disease cannot be accurately excluded by symptom-based IBS criteria alone. Our conclusion is in agreement with the conclusions of previous reviews.[49] Considering this, one may be tempted to conclude that a diagnostic work-up (including invasive tests) is the best approach to establish a diagnosis of IBS in all patients with abdominal symptoms. The literature remains divided on this subject[50] and views of gastroenterologists may conflict with those of primary care physicians. However, when deciding if an invasive diagnostic test is needed, clinicians should also consider the pretest probability (prevalence) of organic disease in their population.[13] Cash *et al.*[14] found that the prevalence rate of inflammatory bowel disease, colon cancer, lactose malabsorption and thyroid dysfunction in patients who fulfilled IBS criteria was similar to or even lower than the prevalence rates found in the general population, with the exception of coeliac disease. Studies in their review, however, included solely patients with an IBS diagnosis according to a symptom-based criteria list. As such a diagnosis is often made after a diagnostic work-up is performed, their prevalence rates may underestimate the prevalence of organic disease in primary care populations. The prevalence rates presented in Table 2, on the other hand, overestimate the prevalence of organic disease in primary care as our analyses were primarily performed on (often referred) out-patients. This, combined with the finding that fulfilment of IBS criteria leads to a significant reduction in risk of a specific GI disease (Table 2), and considering the costs and patient burden involved in invasive diagnostic invasive testing, leads us to agree with the recommendation that in primary care patients meeting symptom-based IBS criteria, no exhaustive diagnostic testing should be performed.[14, 50]

However, as IBS is so common, it may coexist with organic GI disease.[49] How should a GP identify these patients without exhaustive diagnostic testing? One strategy is to pay attention to individual alarm features as their presence is felt to identify a subgroup of patients with a greater pretest probability of organic disease.[51] Results of our review do not unambiguously support this view. Individual alarm features appeared to be frequently absent in patients with organic disease, while they were often present in patients diagnosed with IBS [Table S4]. Within-study subgroup-analyses showed that inclusion of patients reporting alarm features hardly affected diagnostic performance. Furthermore, as alarm features are very common in primary care, a majority of patients would be referred for diagnostic work-up. In the study of Whitehead *et al.*,[46] 84% of the study sample reported alarm features, while only 6% were eventually diagnosed with an organic GI disease. Additionally, evidence is accumulating for poor predictive value of individual alarm features.[52, 53] Instead of individual symptoms, one may use combinations of symptoms to detect patients with a high probability of organic disease. In the study of Hammer *et al.*[28] the best discrimination between IBS and organic disease was provided by a multivariate model including alarm features, non-alarm features and three or more Manning criteria. In the United Kingdom, referral criteria for bowel cancer have been developed in which rectal bleeding is only an alarm feature when present in combination with a persistent change in bowel habit.[54]

Strength and weaknesses

In this review, where possible, we adhered to the most recent guidelines for conducting a diagnostic review as described in the Cochrane Diagnostic Reviewers' Handbook. We used an extensive search strategy, but included a methodological filter to limit number of hits. By reference checking, we tried to track down those publications our search strategy had failed to identify. The inclusion criteria of our review reflect our priority for gathering diagnostic performance data that are relevant to clinical practice. We solely extracted or reconstructed diagnostic data collected from symptomatic patients, excluding information from healthy controls. This decision resulted in lower specificity rates than sometimes presented in the original

publications, or even led to exclusion of studies. Furthermore, we decided to exclude novel technological developments such as rectal distension tests, as their role in clinical practice has not yet been defined and as they are not available in primary health care.[55] Lastly, symptom-based IBS criteria were only eligible for inclusion when they had been externally validated. Our review, as well as Starmans et al.,[39] shows the importance of this criterion: for most criteria lists, the diagnostic performance in validation studies was much lower than in the cohorts in which the system was developed. If we had selected all IBS criteria lists including those that have not been tested in other populations, this would have resulted in overestimation of the diagnostic performance of symptom-based IBS criteria. A consequence of our strict selection criteria is that we included less studies than might have been expected, which is certainly true with regard to our second research question ('performance of signs and symptoms in identifying IBS'). Considering the importance of homogeneous patient groups and our interest in IBS as a syndromic condition on its own, we had a priori decided to include only studies using symptom-based IBS criteria as reference test. Unfortunately, only two studies fulfilled this criterion. In a future systematic review evaluating the diagnostic value of individual signs and symptoms, it may be relevant to include also studies using a diagnostic work-up or clinical follow-up as a reference test.

Other implications for research and clinical practice

Few diagnostic studies were conducted in primary care, which is remarkable as patients with IBS are mainly diagnosed and managed by GPs. We therefore strongly advise a careful validation of the recently developed Rome III criteria in a primary care setting. The study population should consist of all patients consulting the GP because of non-acute abdominal pain or discomfort, including those demonstrating alarm features. By assessing the presence/absence of these features, researchers may analyse which combination of Rome III and alarm features discriminates best between IBS and organic disease. The study population should not include patients with clear upper GI symptoms as these patients are not representative of those who will receive the test (i.e. Rome III criteria) in daily clinical practice. If the validity of the Rome III criteria is confirmed, we believe that the Rome III criteria need a proper introduction in primary care to facilitate implementation. A recent survey showed that nearly 80% of GPs had no knowledge of symptom-based IBS criteria and only 4% used them in clinical practice.[10]

The fact that none of the sets of criteria showed sufficiently homogeneous and favourable results may be particularly disappointing for those physicians regularly facing diagnostic uncertainty in patients with non-acute abdominal signs or symptoms. However, considering the findings of our review in the light of recent advances in the field of GI disease based on multivariable diagnostic modelling,[28, 54] and considering the available evidence on individual alarm features, we suggest a diagnostic aid which may help (somewhat) reduce diagnostic uncertainty: GPs can accurately diagnose IBS in patients who meet symptom-based IBS criteria and lack any alarm features, but should consider (immediate) referral to secondary care in those who demonstrate multiple alarm features.[54] For patients who not only meet symptom-based IBS criteria but also demonstrate a single alarm feature, clinical expertise and judgment will be needed to guide decision-making.

REFERENCES

1 Malagelada JR. A symptom-based approach to making a positive diagnosis of irritable bowel syndrome with constipation. Int J Clin Pract 2006: 60; 57–63.

2 Holten KB, Wetherington A, Bankston L. Diagnosing the patient with abdominal pain and altered bowel habits: is it irritable bowel syndrome? Am Fam Physician 2003: 67; 2157–62.

3 Hillila MT, Farkkila MA. Prevalence of irritable bowel syndrome according to different diagnostic criteria in a non-selected adult population. Aliment Pharmacol Ther 2004: 20; 339–45.

4 Hatlebakk JG, Hatlebakk MV. Diagnostic approach to suspected irritable bowel syndrome. Best Pract Res Clin Gastroenterol 2004: 18; 735–46.

5 Olden KW. Diagnosis of irritable bowel syndrome. Gastroenterology 2002: 122; 1701–14.

6 Corsetti M, Tack J. Are symptom-based diagnostic criteria for irritable bowel syndrome useful in clinical practice? Digestion 2004: 70; 207–9.

7 Vanner SJ, Depew WT, Paterson WG, et al. Predictive value of the Rome criteria for diagnosing the irritable bowel syndrome. Am J Gastroenterol 1999: 94; 2912–7.

8 Drossman DA. The functional gastrointestinal disorders and the Rome III process. Gastroenterology 2006: 130; 1377–90.

9 Lin OS. Colonoscopy in irritable bowel syndrome: whom are we reassuring? Gastrointest Endosc 2005: 62; 900–2.

10 Lea R, Hopkins V, Hastleton J, Houghton LA, Whorwell PJ. Diagnostic criteria for irritable bowel syndrome: utility and applicability in clinical practice. Digestion 2004: 70; 210–3.

11 Fitzpatrick R. Telling patients there is nothing wrong. BMJ 1996: 313; 311–2.

12 Drossman DA, Camilleri M, Mayer EA, Whitehead WE. AGA technical review on irritable bowel syndrome. Gastroenterology 2002: 123; 2108–31.

13 Brandt LJ, Bjorkman D, Fennerty MB, et al. Systematic review on the management of irritable bowel syndrome in North America. Am J Gastroenterol 2002: 97; S7–26.

14 Cash BD, Schoenfeld P, Chey WD. The utility of diagnostic tests in irritable bowel syndrome patients: a systematic review. Am J Gastroenterol 2002: 97; 2812–9.

15 De Giorgio R, Barbara G, Stanghellini V, et al. Diagnosis and therapy of irritable bowel syndrome. Aliment Pharmacol Ther 2004: 20 Suppl 2; 10–22.

16 Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001: 54; 774–81.

17 Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004: 8; iii1-iii234. 18 DTA Editorial Team. Cochrane Diagnostic Reviewer's Handbook. Draft version 0.3 [updated July 2005]. The Cochrane Collaboration, 2005: In future avaiable from http://www.cochrane.org.

19 Zamora J, Muriel A, Abrair V. Statistical methods. MetaDiSc version 1.4. Clinical BioStatistics Unit - Hospital Ramón y Cajal 2004.

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

21 Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. AJR Am J Roentgenol 2006: 187; 271–81.

22 Banerjee R, Choung OW, Gupta R, et al. Rome I criteria are more sensitive than Rome II for diagnosis of irritable bowel syndrome in Indian patients. Indian J Gastroenterol 2005: 24; 164–6.

23 Bellentani S, Baldoni P, Petrella S, et al. A simple score for the identification of patients at high risk of organic diseases of the colon in the family doctor consulting room. The Local IBS Study Group. Fam Pract 1990: 7; 307–12.

24 Chalubinski K, Brunner H. [Positive diagnosis of irritable colon: a scored chart or standardized anamnesis?]. Wien Klin Wochenschr 1987: 99; 819–24.

25 Dogan UB, Unal S. Kruis scoring system and Manning's criteria in diagnosis of irritable bowel syndrome: is it better to use combined? Acta Gastroenterol Belg 1996: 59; 225–8.

26 Enck P, Whitehead WE, Schuster MM, Wienbeck M. [Psychosomatic aspects of irritable bowel syndrome. Specificity of clinical symptoms, psychopathological features and motor activity of the rectosigmoid]. Dtsch Med Wochenschr 1988: 113; 459–62.

27 Frigerio G, Beretta A, Orsenigo G, Tadeo G, Imperiali G, Minoli G. Irritable bowel syndrome. Still far from a positive diagnosis. Dig Dis Sci 1992: 37; 164–7.

28 Hammer J, Eslick GD, Howell SC, Altiparmak E, Talley NJ. Diagnostic yield of alarm features in irritable bowel syndrome and functional dyspepsia. Gut 2004: 53; 666–72.

29 Hammer J, Talley NJ. Value of different diagnostic criteria for the irritable bowel syndrome among men and women. J Clin Gastroenterol 2008: 42; 160–6.

30 Jeong H, Lee HR, Yoo BC, Park SM. Manning criteria in irritable bowel syndrome: its diagnostic significance. Korean J Intern Med 1993: 8; 34–9.

31 Kruis W, Thieme C, Weinzierl M, Schussler P, Holl J, Paulus W. A diagnostic score for the irritable bowel syndrome. Its value in the exclusion of organic disease. Gastroenterology 1984: 87; 1–7.

32 Manning AP, Thompson WG, Heaton KW, Morris AF. Towards positive diagnosis of the irritable bowel. Br Med J 1978: 2; 653–4.

33 Mohamed AS, Khan BA. A more positive diagnosis of irritable bowel syndrome in Saudi patients. Annals of Saudi Medicine 1999: 19; 459–61.

34 Mouzas IA, Fragkiadakis N, Moschandreas J, et al. Validation and results of a questionnaire for functional bowel disease in out-patients. BMC Public Health 2002: 2; 8.

35 Orient JO, Kettel LJ, Lim J. A test of a linear discriminant for identifying low-risk abdominal pain. Med Decis Making 1985: 5; 77–87.

36 Orient JO. Evaluation of abdominal pain: clinicians' performance compared with three protocols. South Med J 1986: 79; 793–9.

37 Poynard T, Couturier D, Frexinos J, et al. French experience of Manning's criteria in the irritable bowel syndrome. Eur J Gastroenterol Hepatol 1992: 4; 747–52.

38 Rao KP, Gupta S, Jain AK, Agrawal AK, Gupta JP. Evaluation of Manning's criteria in the diagnosis of irritable bowel syndrome. J Assoc Physicians India 1993: 41; 357–8, 363.

39 Starmans R, Muris JW, Fijten GH, Schouten HJ, Pop P, Knottnerus JA. The diagnostic value of scoring models for organic and non-organic gastrointestinal disease, including the irritable-bowel syndrome. Med Decis Making 1994: 14; 208–16.

40 Talley NJ, Phillips SF, Melton J III, Wiltgen C, Zinsmeister AR. A patient questionnaire to identify bowel disease. Ann Intern Med 1989: 111; 671–4.

41 Talley NJ, Phillips SF, Melton LJ, Mulvihill C, Wiltgen C, Zinsmeister AR. Diagnostic value of the Manning criteria in irritable bowel syndrome. Gut 1990: 31; 77–81.

42 Talley NJ, Phillips SF, Wiltgen CM, Zinsmeister AR, Melton LJ III. Assessment of functional gastrointestinal disease: the bowel disease questionnaire. Mayo Clin Proc 1990: 65; 1456–79.

43 Thompson WG. Gastrointestinal symptoms in the irritable bowel compared with peptic ulcer and inflammatory bowel disease. Gut 1984: 25; 1089–92.

44 Tibble JA, Sigthorsson G, Foster R, Forgacs I, Bjarnason I. Use of surrogate markers of inflammation and Rome criteria to distinguish organic from nonorganic intestinal disease. Gastroenterology 2002: 123; 450–60.

45 Wasson JH, Sox HC Jr, Sox CH. The diagnosis of abdominal pain in ambulatory male patients. Med Decis Making 1981: 1; 215–24.

46 Whitehead WE, Palsson OS, Feld AD, et al. Utility of red flag symptom exclusions in the diagnosis of irritable bowel syndrome. Aliment Pharmacol Ther 2006: 24; 137–46.

47 Kapoor KK, Nigam P, Rastogi CK, Kumar A, Gupta AK. Clinical profile of irritable bowel syndrome. Indian J Gastroenterol 1985: 4; 15–6.

48 Mazumdar TN, Prasad KV, Bhat PV. Formulation of a scoring chart for irritable bowel syndrome (IBS): a prospective study. Indian J Gastroenterol 1988: 7; 101–2.

49 Longstreth GF. Definition and classification of irritable bowel syndrome: current consensus and controversies. Gastroenterol Clin North Am 2005: 34; 173–87.

50 Talley NJ. When to conduct testing in patients with suspected irritable bowel syndrome. Rev Gastroenterol Disord 2003: 3 Suppl 3; S18–24.

51 Cash BD, Chey WD. Diagnosis of irritable bowel syndrome. Gastroenterol Clin North Am 2005: 34; 205–20 vi.

52 Hamilton W, Sharp D. Diagnosis of colorectal cancer in primary care: the evidence base for guidelines. Fam Pract 2004: 21; 99–106.

53 Vakil N, Moayyedi P, Fennerty MB, Talley NJ. Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: systematic review and meta-analysis. Gastroenterology 2006: 131; 390–401.

54 Thompson MR, Heath I, Ellis BG, Swarbrick ET, Wood LF, Atkin WS. Identifying and managing patients at low risk of bowel cancer in general practice. BMJ 2003: 327; 263–5.

55 Delvaux M. Do we need to perform rectal distention tests to diagnose IBS in clinical practice? Gastroenterology 2002: 122; 2075–8.

TABLES AND FIGURES

Figure 1. Flow chart depicting search and selection processes.

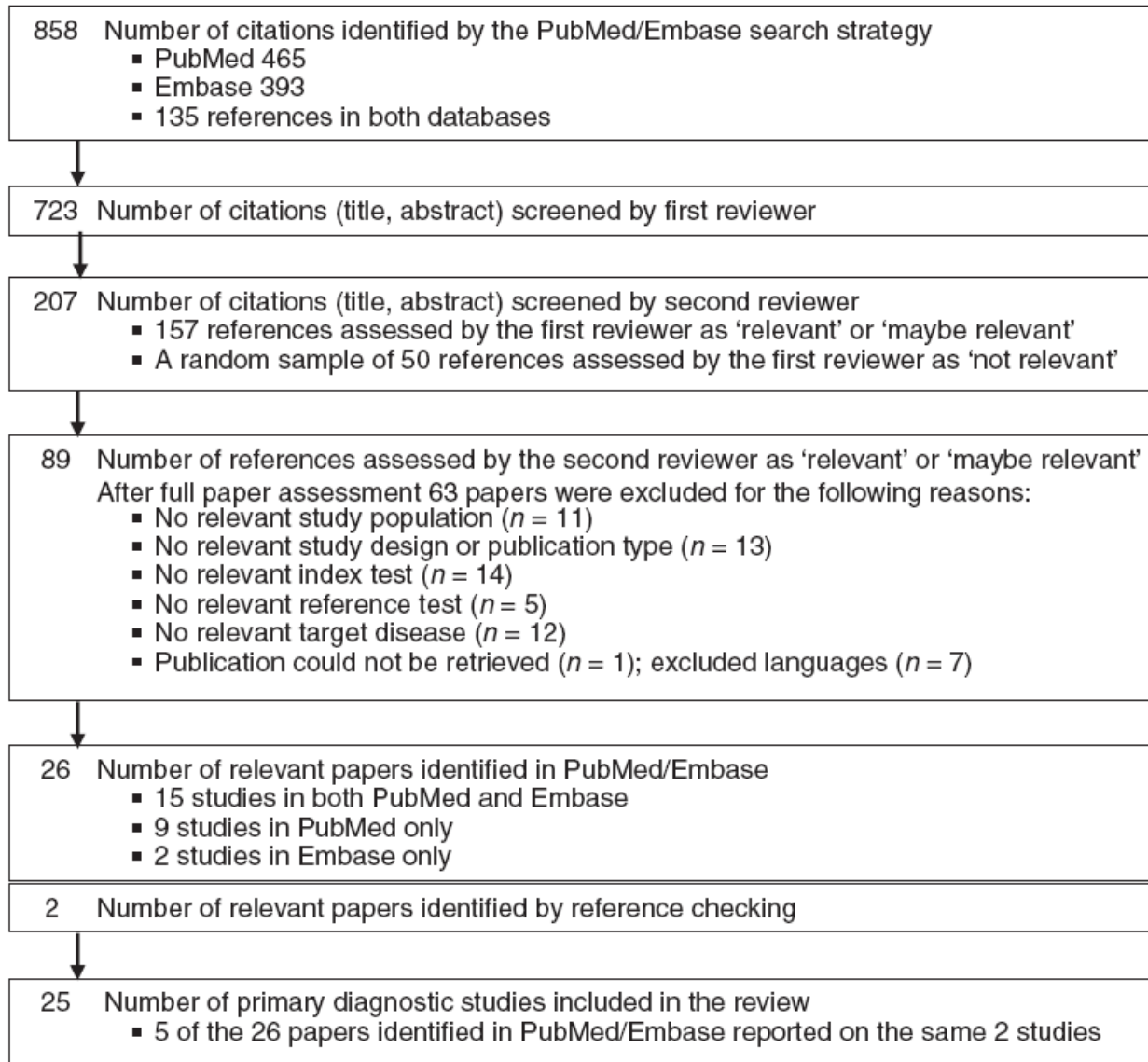|	
858	Number of citations identified by the PubMed/Embase search strategy ■ PubMed 465 ■ Embase 393 ■ 135 references in both databases
723	Number of citations (title, abstract) screened by first reviewer
207	Number of citations (title, abstract) screened by second reviewer ■ 157 references assessed by the first reviewer as 'relevant' or 'maybe relevant' ■ A random sample of 50 references assessed by the first reviewer as 'not relevant'
89	Number of references assessed by the second reviewer as 'relevant' or 'maybe relevant' After full paper assessment 63 papers were excluded for the following reasons: ■ No relevant study population ($n = 11$) ■ No relevant study design or publication type ($n = 13$) ■ No relevant index test ($n = 14$) ■ No relevant reference test ($n = 5$) ■ No relevant target disease ($n = 12$) ■ Publication could not be retrieved ($n = 1$); excluded languages ($n = 7$)
26	Number of relevant papers identified in PubMed/Embase ■ 15 studies in both PubMed and Embase ■ 9 studies in PubMed only ■ 2 studies in Embase only
2	Number of relevant papers identified by reference checking
25	Number of primary diagnostic studies included in the review ■ 5 of the 26 papers identified in PubMed/Embase reported on the same 2 studies

Table 1. Diagnostic performance of symptom-based IBS criteria sets in the exclusion of organic disease

Study	Comparison	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Risk for organic disease in test positives (95% CI)	Risk for organic disease in test negatives (95% CI)
Manning criteria	Cut-off ≥ 2 vs. < 2 out of 4† or ≥ 3 vs. < 3 out of 6								
Manning et al.[32]	F vs. O(tot)*,†	29	10	3	23	0.91 (0.75–0.98)	0.70 (0.51–0.84)	0.26 (0.13–0.42)	0.89 (0.70–0.98)
	F vs. O(tot)	27	8	5	25	0.84 (0.67–0.95)	0.76 (0.58–0.89)	0.23 (0.10–0.40)	0.83 (0.65–0.94)
Kapoor et al.[47]	F vs. O(tot)*,†	50	27	5	100	0.91 (0.80–0.97)	0.79 (0.71–0.86)	0.35 (0.25–0.47)	0.95 (0.89–0.98)
Talley et al.[41]	F(low) vs. O(tot)*	43	26	39	75	0.52 (0.41–0.64)	0.74 (0.65–0.82)	0.38 (0.26–0.50)	0.66 (0.56–0.74)
	F(low) + F(up) vs. O(tot)	52	26	63	75	0.45 (0.36–0.55)	0.74 (0.65–0.82)	0.33 (0.23–0.45)	0.54 (0.46–0.63)
Poynard et al.[37]	F(low) vs. O(tot)*	76	45	44	49	0.63 (0.54–0.72)	0.52 (0.42–0.63)	0.38 (0.29–0.46)	0.53 (0.42–0.63)
	F(low) + F(up) vs. O(tot)	86	45	80	49	0.52 (0.44–0.60)	0.52 (0.42–0.63)	0.34 (0.26–0.43)	0.38 (0.30–0.47)
Jeong et al.[30]	F(low) vs. O(tot)*	39	30	19	57	0.67 (0.54–0.79)	0.66 (0.55–0.75)	0.44 (0.32–0.56)	0.75 (0.64–0.84)
	F(low) + F(up) vs. O(tot)	43	30	42	57	0.51 (0.40–0.62)	0.66 (0.55–0.75)	0.41 (0.30–0.53)	0.58 (0.47–0.68)
	F(low) vs. O(low)	39	7	19	9	0.67 (0.54–0.79)	0.56 (0.30–0.80)	0.15 (0.06–0.29)	0.32 (0.16–0.52)
Rao et al.[38]	F(low) vs. O(low)*	43	9	22	14	0.66 (0.53–0.77)	0.61 (0.39–0.80)	0.17 (0.08–0.30)	0.39 (0.23–0.57)
	F(low) + F(up) vs. O(low)	46	9	54	14	0.46 (0.36–0.56)	0.61 (0.39–0.80)	0.16 (0.08–0.29)	0.21 (0.12–0.32)
Dogan and Unal[25]	F(low) vs. O(tot)*	149	24	16	158	0.90 (0.85–0.94)	0.87 (0.81–0.91)	0.14 (0.09–0.20)	0.91 (0.86–0.95)
Hammer and Talley[29]	F(low) vs. O	155	95	78	210	0.67 (0.60–0.73)	0.69 (0.63–0.74)	0.38 (0.32–0.44)	0.73 (0.67–0.78)
	F(low) + F(up) vs. O	220	95	189	210	0.54 (0.49–0.59)	0.69 (0.63–0.74)	0.30 (0.25–0.36)	0.53 (0.48–0.58)
Starmans et al.[39]	F vs. O(tot)*,†,‡	72	18	86	24	0.46 (0.38–0.54)	0.57 (0.41–0.72)	0.20 (0.12–0.30)	0.22 (0.15–0.31)
Kruis criteria	Cut-off ≥ 44 vs. < 44								
Kruis et al.[31]	F(low) vs. O(tot)*	69	2	39	207	0.64 (0.54–0.73)	0.99 (0.97–1.00)	0.03 (0.00–0.10)	0.84 (0.79–0.89)
Chalubinski and Brunner[24]	F(low) vs. O(tot)*	27	18	26	252	0.51 (0.37–0.65)	0.93 (0.90–0.96)	0.40 (0.26–0.56)	0.91 (0.87–0.94)
	F(low) + F(up) vs. O(tot)	34	18	69	252	0.33 (0.24–0.43)	0.93 (0.90–0.96)	0.35 (0.22–0.49)	0.79 (0.74–0.83)
Bellentani et al.[23]	F(low) vs. O(low)*	125	36	27	66	0.82 (0.75–0.88)	0.65 (0.55–0.74)	0.22 (0.16–0.30)	0.71 (0.61–0.80)
Frigerio et al.[27]	F(low) vs. O(tot)*	29	11	23	190	0.56 (0.41–0.70)	0.95 (0.90–0.97)	0.28 (0.15–0.44)	0.89 (0.84–0.93)
Starmans et al.[39]	F(low) vs. O(tot)*	199	36	90	41	0.69 (0.63–0.74)	0.53 (0.42–0.65)	0.15 (0.11–0.21)	0.31 (0.24–0.40)
Dogan and Unal[25]	F(low) vs. O(tot)*	134	16	31	166	0.81 (0.74–0.87)	0.91 (0.86–0.95)	0.11 (0.06–0.17)	0.84 (0.78–0.89)
Mohamed and Khan[33]	F vs. O*	23	6	6	9	0.79 (0.60–0.92)	0.60 (0.32–0.84)	0.21 (0.08–0.40)	0.60 (0.32–0.84)
Rome I criteria	Cut-off Rome I+ vs. Rome I−								
Hammer and Talley[29]	F(low) vs. O / incl. red flags	156	92	77	213	0.67 (0.61–0.73)	0.70 (0.64–0.75)	0.37 (0.31–0.43)	0.73 (0.68–0.78)
	F(low) + F(up) vs. O / incl. red flags	216	92	193	213	0.53 (0.48–0.58)	0.70 (0.64–0.75)	0.30 (0.25–0.35)	0.53 (0.48–0.57)
Tibble et al.[44]	F(low) vs. O / incl. red flags	240	39	99	224	0.71 (0.66–0.76)	0.85 (0.80–0.89)	0.14 (0.10–0.19)	0.69 (0.64–0.74)
	F(low) vs. O / excl. red flags*	241	40	95	168	0.72 (0.67–0.77)	0.81 (0.75–0.86)	0.14 (0.10–0.19)	0.64 (0.58–0.70)
Banerjee et al.[22]	F(low) vs. O excl. red flags*	110	4	22	2	0.83 (0.76–0.89)	0.33 (0.04–0.78)	0.04 (0.01–0.09)	0.08 (0.01–0.27)
Rome II criteria	Cut-off Rome II+ vs. Rome II−								
Hammer and Talley[29]	F(low) vs. O / incl. red flags	161	104	72	201	0.69 (0.63–0.75)	0.66 (0.60–0.71)	0.39 (0.33–0.45)	0.74 (0.68–0.79)

Table 1. (Continued)

Study	Comparison	TP	FP	FN	TN	Sensitivity (95%CI)	Specificity (95%CI)	Risk for organic disease in test positives (95% CI)	Risk for organic disease in test negatives (95% CI)
	F(low) + F(up) vs. O / incl. red flags	235	104	174	201	0.58 (0.53–0.62)	0.66 (0.60–0.71)	0.31 (0.26–0.36)	0.54 (0.48–0.59)
Vanner et al.[7]	F(low) vs. O / incl. red flags	30	6	29	74	0.51 (0.38–0.64)	0.93 (0.84–0.97)	0.17 (0.06–0.33)	0.72 (0.62–0.80)
Vanner et al.[7]	F(low) vs. O / excl. red flags*	30	0	16	52	0.65 (0.50–0.79)	1.00 (0.93–1.00)	0.00 (0.00–0.12)	0.77 (0.65–0.86)
Banerjee et al.[22]	F(low) vs. O / excl. red flags*	41	4	91	2	0.31 (0.23–0.40)	0.33 (0.04–0.78)	0.09 (0.03–0.21)	0.02 (0.00–0.08)
Whitehead et al.[46]	F vs. O / incl. red flags*	682	32	668	52	0.51 (0.48–0.53)	0.62 (0.51–0.72)	0.05 (0.03–0.06)	0.07 (0.05–0.09)
Bellentani criteria	Cut-off < 0 vs. > 0								
Bellentani et al.[23]	F(low) vs. O(low)*	115	18	37	84	0.76 (0.68–0.82)	0.82 (0.74–0.89)	0.14 (0.08–0.21)	0.69 (0.60–0.78)
Starmans et al.[39]	F(low) vs. O(low)*,‡	60	2	112	26	0.35 (0.28–0.43)	0.93 (0.77–0.99)	0.03 (0.00–0.11)	0.19 (0.13–0.26)
Mazumdar criteria	Cut-off > 0 vs. < 0								
Mazumdar et al.[48]	F vs. O(low)*	50	0	5	20	0.91 (0.80–0.97)	1.00 (0.83–1.00)	0.00 (0.00–0.07)	0.80 (0.59–0.93)
Starmans et al.[39]	F vs. O(low)*,‡	110	20	56	14	0.66 (0.59–0.73)	0.41 (0.25–0.59)	0.15 (0.10–0.23)	0.20 (0.11–0.31)
Talley criteria	Cut-off > 0 (?) vs. < 0 (?)								
Talley et al.[40]	F(up) + F(low)vs. O(tot)*	98	40	17	61	0.85 (0.77–0.91)	0.60 (0.50–0.70)	0.29 (0.22–0.37)	0.78 (0.67–0.87)
Talle et al.y[42]	F(up) + F(low) vs. O(tot)	81	32	34	69	0.70 (0.61–0.79)	0.68 (0.58–0.77)	0.28 (0.20–0.38)	0.67 (0.57–0.76)
Starmans et al.[39]	F(up) + F(low) vs. O(tot)*,‡	62	20	80	38	0.44 (0.35–0.52)	0.66 (0.52–0.78)	0.24 (0.16–0.35)	0.32 (0.24–0.41)
Wasson criteria	Cut-off < 0 vs. > 0								
Wasson et al.[45]	F vs. O(tot)*	186	10	252	104	0.43 (0.38–0.47)	0.91 (0.85–0.96)	0.05 (0.03–0.09)	0.29 (0.25–0.34)
Orient[36]	F vs. O(tot)*	54	17	63	75	0.46 (0.37–0.56)	0.82 (0.72–0.89)	0.24 (0.15–0.36)	0.54 (0.46–0.63)
Starmans et al.[39]	F vs. O(tot)*,‡	54	6	106	34	0.34 (0.27–0.42)	0.85 (0.70–0.94)	0.10 (0.04–0.21)	0.24 (0.17–0.32)

* Eligible for beween-study subgroup-analyses per IBS criteria set.
† Cut-off 2 out of 4.
‡ 2 × 2 tables have been reconstructed (details are available from the corresponding author).
F, functional disease; F(low), lower functional disease (including clinical diagnosis of IBS); F(up), upper functional disease; O, organic disease; O(low), lower organic disease; O(tot), upper and lower organic disease; TP, true positives (positive score on IBS criteria set, no organic disease); FP, false positives (positive score on IBS criteria set, organic disease); FN, false negatives (negative score on IBS criteria set, no organic disease); TN, true negatives (negative score on IBS criteria set, organic disease); CI, confidence interval.

Table 2. Diagnostic performance of symptom-based IBS criteria sets in the exclusion of specific diseases

Specific disease	Index test (T+ vs. T−)	Outcome	TP	FP	FN	TN	Risk for specific disease in test positives (95% CI)	Risk for specific disease in test negatives (95% CI)	RR, OR* for specific disease (95%CI)
Cancer (Ca)									
Chalubinski and Brunner[24]	Kruis ≥44 vs. <44	Ca GI− vs. Ca GI+	51	1	279	42	0.02 (0.00–0.10)	0.13 (0.10–0.17)	0.15 (0.02–1.04)
Whitehead et al.[46]	Rome II+ vs. Rome II−	Ca GI− vs. Ca GI+	701	13	697	23	0.02 (0.01–0.03)	0.03 (0.02–0.05)	0.57 (0.29–1.12)
Inflammatory Bowel Disease (IBD)									
Chalubinski snd Brunner[24]	Kruis ≥44 vs. <44	IBD− vs. IBD+	49	3	244	77	0.06 (0.01–0.16)	0.24 (0.19–0.29)	0.24 (0.08–0.73)
Whitehead et al.[46]	Rome II+ vs. Rome II−	IBD− vs. IBD+	707	7	698	22	0.01 (0.00–0.02)	0.03 (0.02–0.05)	0.32 (0.14–0.75)
Poynard et al.[37]	Manning ≥3 vs. <3 of 6	IBD− vs. IBD+	117	14	125	4	0.11 (0.06–0.17)	0.03 (0.01–0.08)	3.45 (1.17–10.20)
Thompson[43]	Manning ≥2 vs. <2 of 4	F(low) vs. IBD+	35	17	15	32	Not possible	Not possible	0.23 (0.10–0.53)*
Malabsorption									
Enck et al.[26]	Manning ≥ 3? vs. <3? of 6	F vs. LI+	6	8	12	11	Not possible	Not possible	1.46 (0.38–5.54)*
Whitehead et al.[46]	Rome II+ vs. Rome II−	malabsorption− vs. malabsorption+	702	12	713	7	0.02 (0.01–0.03)	0.01 (0.00–0.02)	1.73 (0.68–4.37)
Peptic ulcer									
Thompson[43]	Manning ≥2 vs. <2 of 4	F(low) vs. Peptic Ulcer	35	13	15	36	Not possible	Not possible	0.16 (0.06–0.37)*
Non Ulcer Dyspepsia (NUD)									
Chalubinski and Brunner[24]	Kruis ≥44 vs. < 44	NUD− vs. NUD+	45	7	278	43	0.14 (0.06–0.26)	0.13 (0.10–0.18)	1.00 (0.48–2.11)
Poynard et al.[37]	Manning ≥3 vs. <3 of 6	NUD− vs. NUD+	121	10	93	36	0.08 (0.04–0.14)	0.28 (0.20–0.37)	0.27 (0.14–0.53)
Jeong et al.[30]	Manning ≥3 vs. <3 of 6	NUD− vs. NUD+	69	4	76	23	0.06 (0.02–0.13)	0.23 (0.15–0.33)	0.24 (0.09–0.65)
Rao et al.[38]	Manning ≥3 vs. <3 of 6	NUD− vs. NUD+	52	3	36	32	0.06 (0.01–0.15)	0.47 (0.35–0.60)	0.12 (0.04–0.36)
Talley et al.[41]	Manning ≥3 vs. <3 of 6	NUD− vs. NUD+	69	9	114	24	0.12 (0.05–0.21)	0.17 (0.12–0.25)	0.66 (0.32–1.35)

TP, true positives (positive score on IBS criteria set, no specific disease); FP, false positives (positive score on IBS criteria set, specific disease); FN, false negatives (negative score on IBS criteria set, no specific disease); TN, true negatives (negative score on IBS criteria set, specific disease); CI, confidence interval; GI, gastrointestinal; F, functional disease; LI, lactose intolerance; T+, positive test result; T-, negative test result.