ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«Μεθοδολογία Βιοϊατρικής Έρευνας, Βιοστατιστική και Κλινική Βιοπληροφορική»

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

«Assess the reporting quality of studies investigating the diagnostic accuracy of fecal calprotectin in the diagnosis of Inflammatory Bowel Diseases published from inception to July 2019 using the STARD statement»

«Αξιολόγηση της ποιότητας αναφοράς μελετών διερεύνησης της διαγνωστικής ακρίβειας της καλπροτεκτίνης κοπράνων στην διάγνωση των Ιδιοπαθών Φλεγμονωδών Νοσημάτων του Εντέρου οι οποίες δημοσιεύθηκαν μέχρι τον Ιούλιο του 2019 χρησιμοποιώντας το εργαλείο STARD»

Τριμελής Επιτροπή

Επιβλέπων: Στεφανίδης Ιωάννης

Ζιντζαράς Ηλίας

Δοξάνη Χρυσούλα

Τσούνης Ευθύμιος (ΑΜ:190)

ΛΑΡΙΣΑ, 2019

A. ABSTRACT

Background Fecal calprotectin (FC) has an established role as a biomarker to detect inflammatory bowel diseases (IBD). The aim of this study is to assess the reporting quality of relative diagnostic accuracy studies using the STARD 2015 initiative.

Methods PubMed and Cochrane libraries were systematically searched for studies evaluating the diagnostic accuracy of FC in diagnosing IBD published from inception through 30th of July 2019. Quality of reporting was assessed using STARD statement, an evidence-based tool consisting of 30 items. For each item and each study included an overall score was calculated. The relationship between the adherence and the variables: year of publication, number of authors, 5-year journal impact factor and number of participants; was also investigated.

Results The search yielded 26 eligible studies. The mean study STARD score was 67.8% (range 44.1%-88.2%, SD 11.1%). Only one study reported less than the 50% of the items whereas half of the studies reported more than 70%. STARD 2015 update had no significant impact on the score of subsequent studies. Better compliance was noted only when the number of subjects was larger.

Conclusions Quality of reporting is suboptimal with sharp divergence between different sections of the STARD checklist. All stakeholders are invited to promote STARD 2015 implementation in order to improve completeness and transparency of reporting in biomedical research.

Abbreviations

FC, fecal calprotectin
IBD, inflammatory bowel diseases
UC, ulcerative colitis
IBS, irritable bowel syndrome
NSAID, nonsteroidal anti-inflammatory drugs
STARD, standards for reporting of diagnostic accuracy studies

Keywords

Diagnostic accuracy studies, STARD, fecal calprotectin, inflammatory bowel diseases, ulcerative colitis, Crohn’s disease, reporting quality

Α.ΠΕΡΙΛΗΨΗ

Εισαγωγή Η καλπροτεκτίνη κοπράνων έχει εγκατεστημένο ρόλο σαν δείκτης στη διαγνωστική διερεύνηση των ιδιοπαθών φλεγμονωδών νοσημάτων του εντέρου (ΙΦΝΕ). Στόχος αυτής της μελέτης είναι η αξιολόγηση της ποιότητας αναφοράς των σχετικών ερευνών διαγνωστικής ακρίβειας με τη χρήση του STARD 2015 εργαλείου.

Μέθοδοι Πραγματοποιήθηκε αναζήτηση στις ηλεκτρονικές βάσεις PubMed και Cochrane για μελέτες της διαγνωστικής ακρίβειας της καλπροτεκτίνης στα ΙΦΝΕ δημοσιευμένων μέχρι την 30η Ιουλίου 2019. Η αξιολόγηση της ποιότητας αναφοράς έγινε με τη δήλωση STARD, η οποία εμπεριέχει μια λίστα με 30 απαραίτητα σημεία προς έλεγχο. Υπολογίστηκε η αθροιστική βαθμολογία για κάθε σημείο και για κάθε μελέτη που διερευνήθηκε. Επιπλέον, έγινε ανάλυση της συσχέτισης μεταξύ της βαθμολογίας των άρθρων και των κάτωθι μεταβλητών: έτος δημοσίευσης, αριθμός συγγραφέων, 5ετής συντελεστής απήχησης περιοδικού, αριθμός συμμετεχόντων

Αποτελέσματα 26 δημοσιεύσεις που πληρούσαν τις προϋποθέσεις ελέγχθηκαν βάσει των σημείων του STARD. Η μέση βαθμολογία άρθρου ανέρχεται στο 67.8% (εύρος 44.1%-88.2%, SD 11.1%). Μόνο μια μελέτη είχε βαθμολογία κάτω από 50% ενώ οι μισές μελέτες ανέφεραν πάνω από το 70% των σημείων της λίστας. Η ανανέωση του εργαλείου το 2015 δεν φάνηκε να έχει αντίκτυπο στη βαθμολογία των μεταγενέστερων μελετών. Καλύτερη συμμόρφωση παρατηρήθηκε μόνο σε έρευνες με μεγαλύτερο αριθμό συμμετεχόντων.

Συμπεράσματα Η ποιότητα αναφοράς κατά τα πρότυπα STARD κρίνεται μέτρια έως καλή με μεγάλες διακυμάνσεις στα επιμέρους τμήματα της λίστας. Κρίνεται απαραίτητη η καθολικότερη εφαρμογή του STARD 2015 ώστε να εξασφαλίζεται η διαφάνεια και η εγκυρότητα της βιοϊατρικής έρευνας.

Λέξεις Κλειδιά: Μελέτες διαγνωστικής ακρίβειας, STARD, καλπροτεκτίνη κοπράνων, ιδιοπαθή φλεγμονώδη νοσήματα του εντέρου (ΙΦΝΕ), ελκώδης κολίτιδα, κολίτιδα Crohn, ποιότητα αναφοράς

B. INTRODUCTION

The Inflammatory Bowel Diseases challenge

Inflammatory Bowel Diseases (IBS) is an umbrella term used to describe disorders of the intestinal tract marked by chronic inflammation. IBD incidence rates have been increasing affecting 2.5-3 million people in Europe [1][2]. Their course is progressive with consecutive periods of amelioration and relapse, finally leading half of the patients to a surgical operation [3.] In parallel, patients are in increased risk of colorectal carcinoma and other types of neoplasia [4]. IBD encompasses two types of idiopathic intestinal disease: Crohn's disease and ulcerative colitis (UC). Crohn's disease is characterized by transmural inflammation causing fibrosis, granuloma formation and involvement of sporadic lesions throughout the entire tract. As a result common complications include strictures, fissures, obstructive clinical manifestations and fistula formation [5]. In contrast, inflammation in UC is restricted to the mucosa and submucosa of the colon; it typically starts in the rectum expanding proximally in a continuous manner [6]. Both conditions present abdominal pain, diarrhea, bloating, fever and weight loss as clinical features. In addition, rectal bleeding and anemia increase the probability of diagnosis [7]. Lower abdominal symptoms are a common manifestation of Irritable Bowel Syndrome (IBS), a functional gastrointestinal disorder with a higher prevalence (9-23%) but less serious repercussions [8]. Hence the spectrum of differential diagnosis is broadening. Considering that signs and symptoms which clinically confirm IBD do not exist, the diagnosis might be troublesome and time consuming. Consequently, further investigation usually includes endoscopic evaluation with histopathological sampling, an invasive and unpleasant procedure requiring sedation [9]. More than 60% of colonoscopies in younger patients with suspicion of IBD are negative, while a third of adults with rectal bleeding symptoms do not have pathological endoscopic findings [10][11]. Ostensibly a tool, which stratifies the patients and indicates the high risk group necessitating endoscopic examination, would be of major importance.

The role of calprotectin

Fecal calprotectin is a 36.5 kDa nonglycosylated calcium and zinc binding protein found in large amounts in the cytosol of neutrophil granulocytes. It appears to have an immunomodulotary role with anti-proliferative, anti-inflammatory and antimicrobial

capacities [12, 13]. In conditions of bowel inflammation the disruption of the mucus layer permits the invasion of neutrophils in the intestinal tract resulting in an increase of the protein fecal levels [14]. Moreover, FC is resilient against bacterial proteolytic action in the intestinal tract, it demonstrates a high stability in stool samples (more than 7 days in room temperature) and it can be easily measured [15]. All these qualities underline the role of FC as low-cost and reliable biomarker for diagnosing organic bowel diseases (IBD, colorectal cancer, NSAID enteropathy) [16, 17]. There is a plethora of studies examining the diagnostic accuracy of FC in the diagnosis of IBD. Aim of our study is to assess the reporting quality of these studies using the STARD statement.

The STARD statement

Diagnostic accuracy studies focus on evaluating whether a method correctly classifies the subjects as having a target condition. The method under examination is called index test and it is typically compared with a reference standard. If the test results are binary (positive or negative) the cross tabulation can lead to the calculation of specificity, sensitivity, positive prognostic value and negative prognostic value. For no binary outcome other parameters can quantify the overall accuracy of the index test such as the area under the receiver operating curve (ROC). In similarity to other clinical studies, they are at risk of bias. Flaws in methodological design and defective sampling procedure lead to results that systematically deviate from the reality. Furthermore, data collection errors, fallacious analysis and poor interpretation could undermine the validity of the study [18, 19]. Several surveys have shown that diagnostic accuracy studies do not achieve the desired level of trustworthiness and reporting quality [20, 21]. Obviously healthcare algorithms based on such studies are inaccurate and provide misleading recommendations, thus increasing expenditures and sabotaging patients' clinical outcome. The applicability and transparency of these studies can be improved with the utilization of the STARD tool in accordance to Consolidated Standards the Reporting of Trials or CONSORT statement for reporting randomized controlled trials [19]. The STARD was first launched in 2003 and then updated in 2015. There is evidence that the implementation of the STARD 2003 initiative contributed to a modest, yet statistically significant, improvement in the reporting quality of diagnostic accuracy studies [20]. The renewal of the statement had as a purpose to simplify the utilization of the tool and embody modern evidence

about origins of variability and bias. The revised STARD 2015 checklist consists of 30 essential items which facilitate the author, the reviewer and the reader to investigate the completeness and the integrity of a diagnostic accuracy study [22]. 4 items are divided in two parts and in that manner the final list incorporates 34 elements for examination.

C. METHODS

Search strategy

A systematic search was performed on 30 July 2019 in PubMed and Cochrane library with results included from inception until 30 July 2019. The search strategy for Pubmed was “crohn's disease” [MeSH] OR “ulcerative colitis” [MeSH]) OR “IBD” [tw] OR “inflammatory bowel disease” [MeSH] AND “diagnostic accuracy” OR “diagnosis” AND “calprotectin” [tw]. The key words used in the COCHRANE library were: calprotectin, IBD, inflammatory bowel diseases, Crohn’s disease, ulcerative colitis. Eligible studies from auto-alerts were included up to 20 August 2019. Reference lists of included studies were checked for additional sources.

Inclusion and exclusion criteria

Inclusion criteria were: (1) diagnostic accuracy studies estimating the precision of an index test, (2) studies measuring calprotectin in stool samples as an index test. Exclusion criteria were: (1) studies concerning pediatric population, (2) studies using FC as a monitoring tool for IBD, (3) non-English language publications, (4) conference publications and abstracts, (5) unpublished studies, (6) animal studies.

Study selection and assessment of quality

After the removal of duplicates, the title and abstracts of initial search results were screened for relevance. The full texts of the remaining results were assessed for eligibility based on predetermined criteria. For all included studies the following data were collected: year of publication, 5 year journal impact factor, number of authors and reference to “STARD”. The reporting quality of the studies was assessed using the STARD 2015 statement. Every element of the checklist was answered “YES”, “NO”, “NOT APPLICABLE (N/A)” with each “YES” scoring 1 point. Each one of 34 items was weighted equally. An overall reporting quality score percentage was

calculated for each item and for each study by dividing the number of gathered points by the total available (excluding N/A).

Statistical Analysis

Statistical analysis was carried out using IBM SPSS Statistics V.25 and Microsoft Excel 2011. Pearson's product-moment correlation was used to estimate the correlation between STARD and pre-specified variables (5-year journal impact factor, year of publication, author number). Studies were divided in two categories based on year of publication: pre- and post-STARD 2015 update. T-test for independent data was used to test for binary parametric data. In addition, comparison of adherence was conducted between studies with greater and shorter number of participants. As a cut-off n=140 was selected, since it represented the 50^{th} quartile of the number of participants of all the included studies. Mann-Whitney U test was used to test for binary non-parametric data. The normality check and the equality of variances check was performed using the Shapiro-Wilk and the Levene's test, respectively. A p-value<0.05 was considered as statistical significant. Journal metrics were obtained from Resourchify search home. Risk of bias for each study was not analyzed.

D. RESULTS

Study search results

Initial search identified 883 potential studies (PubMed n= 804, Cochrane n=79). After the removal of duplicates (n=14) and non-relevant articles (n=668), 201 articles were full text assessed in accordance to predetermined criteria. After the eligibility evaluation 21 studies were included to the study. Another 5 studies from the reference list of previews systematic reviews and meta-analyses [10, 23] were added (figure 1). In total, 26 studies were included for analysis [24-49].

Reporting quality - Main Results

Table 1 contains the STARD 2015 checklist with the total score of each item after the assessment of the studies while Figure 2 displays the percentage of studies adequately reporting each item where applicable.

Figure 1 Flow chart of study search, selection, inclusion and exclusion of articles

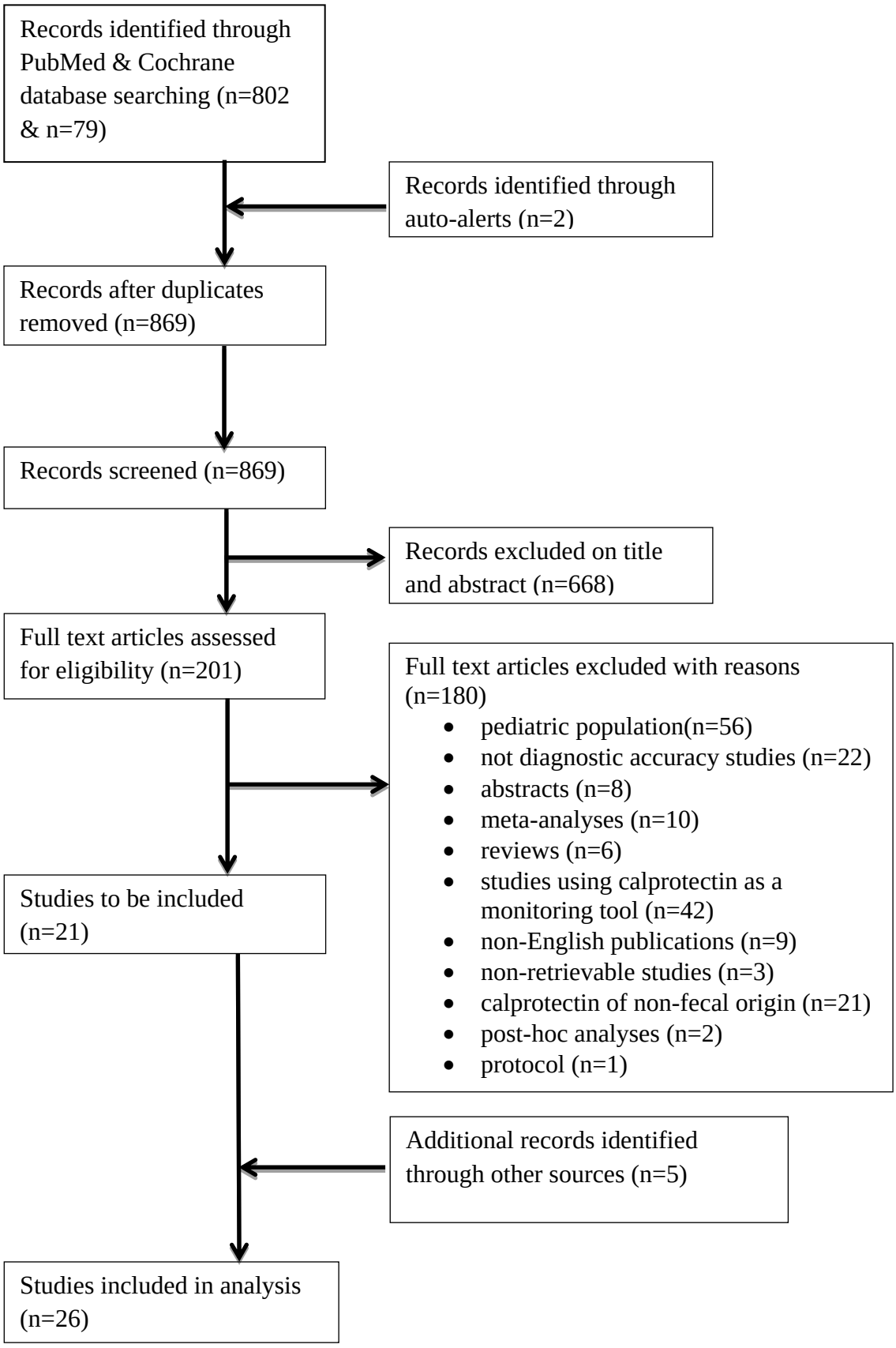


Table 1 The STARD 2015 checklist and the score of each item

Section & Topic	*No*	*Item*	*Score*
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	**26/26 (100%)**
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	**25/26 (96.2%)**
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	**26/26 (100%)**
	4	Study objectives and hypotheses	**26/26 (100%)**
METHODS			
Study Design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	**25/26 (96.2%)**
Participants	6	Eligibility criteria	**23/26 (88.5%)**
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	**23/26 (88.5%)**
	8	Where and when potentially eligible participants were identified (setting, location, and dates)	**22/26 (84.6%)**
	9	Whether participants formed a consecutive, random, or convenience series	**16/26 (61.5%)**
	10a	Index test, in sufficient detail to allow replication	**26/26 (100%)**
	10b	Reference standard, in sufficient detail to allow replication	**26/26 (100%)**
	11	Rationale for choosing the reference standard (if alternatives exist)	**26/26 (100%)**
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	**26/26 (100%)**
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	**24/26 (92.3%)**

	13a	Whether clinical information and reference standard results were available to the performers or readers of the index test	**12/26 (46.2%)**
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	**18/26 (69.2%)**
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	**26/26 (100%)**
	15	How indeterminate index test or reference standard results were handled	**5/26 (19.2%)**
	16	How missing data on the index test and reference standard were handled	**16/26 (61.5%)**
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	**15/15 (100%)**
	18	Intended sample size and how it was determined	**5/26 (19.2%)**
RESULTS			
Participants	19	Flow of participants, using a diagram	**10/26 (38.5%)**
	20	Baseline demographic and clinical characteristics of participants	**15/26 (57.7%)**
	21a	Distribution of severity of disease in those with the target condition	**8/26 (30.8%)**
	21b	Distribution of alternative diagnoses in those without the target condition	**21/26 (80.8%)**
	22	Time interval and any clinical interventions between index test and reference standard	**19/26 (73.1%)**
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	**13/26 (50%)**
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	**20/26 (76.9%)**
	25	Any adverse events from performing the index test or the reference standard	**0/26 (0%)**
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalizability	**12/26 (46.2%)**
	27	Implications for practice, including the intended use and clinical role of the index test	**26/26 (100%)**
OTHER INFORMATION			
	28	Registration number and name of registry	**1/26 (3.8%)**
	29	Where the full study protocol can be accessed	**0/26 (0%)**
	30	Sources of funding and other support; role of funders	**17/26 (65.4%)**

10/34 (29.4%) items were reported in all studies whereas 9/34 (26.5%) items were reported in less than 50% of the studies.

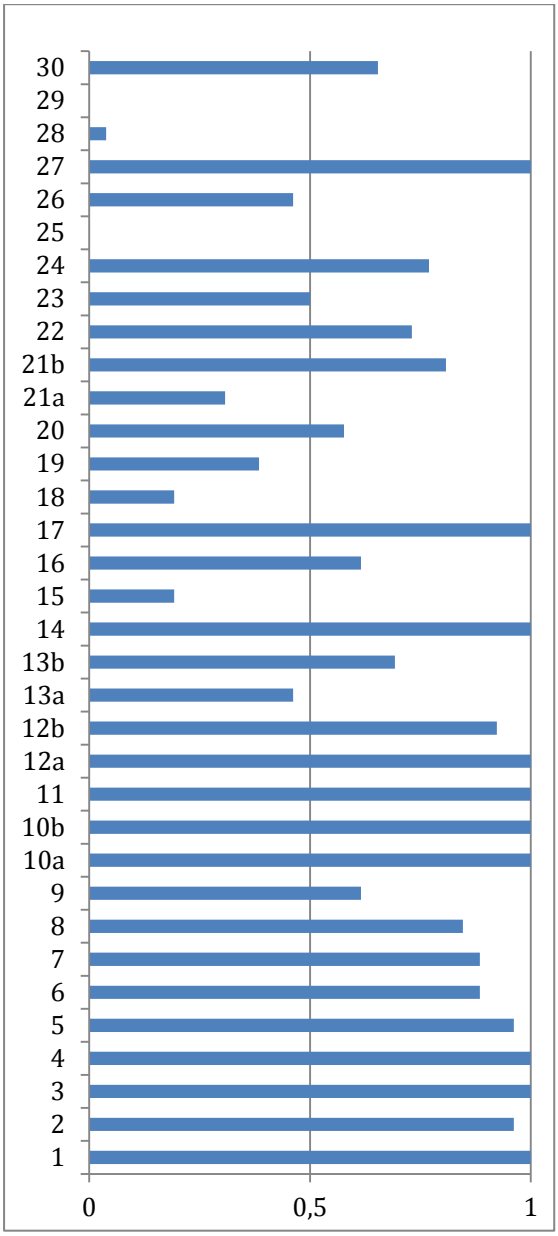


Figure 2 Percentage of studies adequately reporting each STARD item

The best reported elements were:

- Item 1: identification as a study of diagnostic accuracy in the title or abstract (26/26)
- Item 3, 4: a comprehensive introduction clarifying the scientific background of the index test and demonstrating the study's objectives and hypotheses (26/26)
- Item 10a, 10b: index and reference test in detail to allow replication (26/26)
- Item 11: most studies used as reference standard the gold standard method to diagnose IBD (medical history and endoscopic evaluation with biopsies for histological evaluation) (26/26)
- Item 12a: most studies used for test positivity cut-offs of the index test the margin values proposed by the manufacturer of the correspondent FC measurement kit (26/26)
- Item 14: analysis and statistical methodology (26/26)
- Item 17: assessment of differences in accuracy across subgroups of participants as per protocol (15/15)
- Item 27: a thorough discussion section including implication for practice and the clinical role of the index test (26/26)

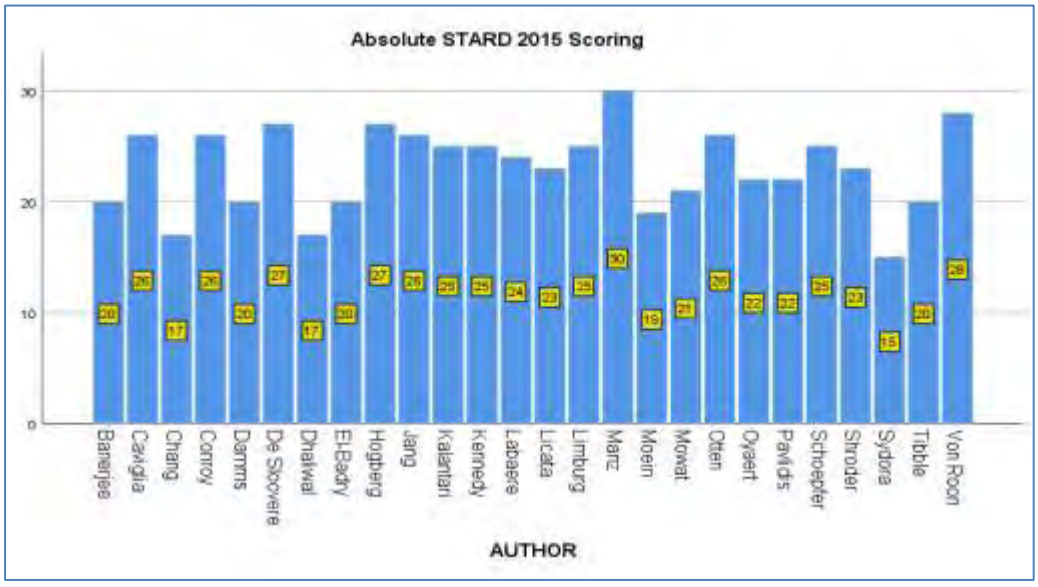
On the contrary, the worst reported elements were:

- Item 25: reporting adverse events during the study (0/26)
- Item 29: access to full study protocol (0/26)
- Item 28: registration number and number of registry (1/26)
- Item 15: indeterminate index test or reference standard results management strategy (5/26)
- Item 18: statistical methodology for sample size estimation (5/26)
- Item 21a: distribution of severity of disease in IBD patients (8/26)
- Item 19: diagram displaying the flow of participants (10/26)

The variability of the adherence between the different sections of STARD list is notable. The title, the abstract and the introduction parts were reported in an excellent level while the adherence of "other information" section was disappointing.

The mean study STARD compliance was 67.8% (range 44.1%-88.2%, SD 11.1%). No article scored 100%. *Manz et al, 2012* [38] was the article of our analysis that attained the highest reporting score 30/34 (88.2%), whereas *Sydora et al, 2012* [46] marked an anemic 15/34 (44.1%)(Figure 2). 13 out of 26 included studies scored ≥70% and 8 of them scored ≥ 75%. Only one study reported < 50% of the items.

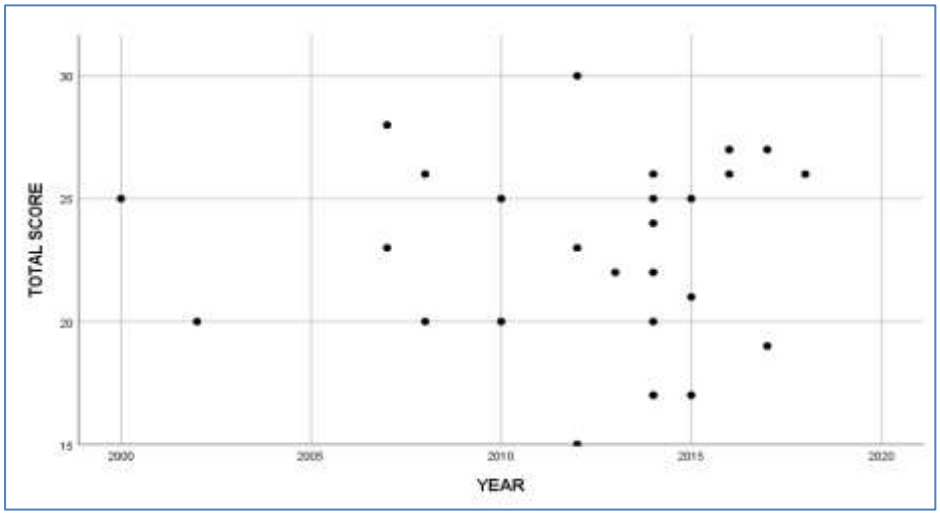
Figure 2 Absolute STARD 2015 score of studies included in the analysis (maximum achievable score: 34)



Time trend and STARD statement update

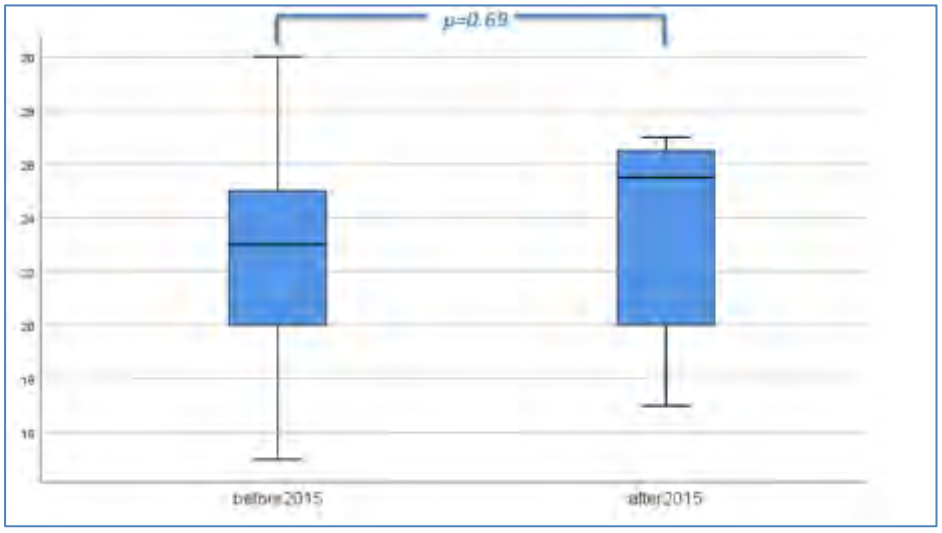
The STARD score did not increase with time and had no statistical significant correlation with the year of publication (r=0.005, p=0.98; figure 3)

Figure 3 Studies score and year of publication



The 2015 update was used to dichotomize the included trials in two categories; pre-STARD 2015 (18 studies) and post-STARD 2015 (8 studies). After assumption testing, a t-test for independent data was executed to compare the groups. Mean study score before 2015 was 22.8 with SD=3.8 whereas mean score after 2015 was 23.5 with SD=3.9. It becomes apparent that the STARD statement update had feeble if any effect in the score of studies published after 2015 (p=0.69; figure 4).

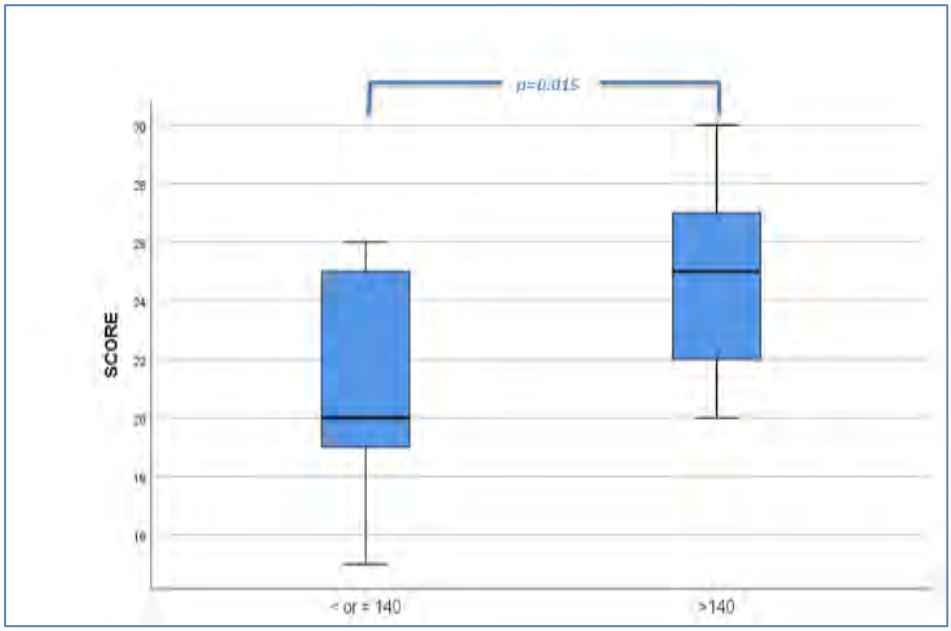
Figure 4 The STARD 2015 update as a landmark



STARD score and other variables

The STARD score had no statistically significant correlation neither with 5-year journal impact factor (r=0.06, p=0.79) nor with number of authors (r=0.04, p=0.87). However, trials that included >140 patients had a statistical significant higher score (mean score 24.8, SD=3) than trials with 140 or less subjects (mean score 21.3, SD=3.75). The mean difference was 3.46 points (95% CI: 0.72-6.2; p=0.015) (figure 5).

Figure 5 The number of subjects impact on trials' STARD score



Out of all trials only 6 reported that their study design was in accordance with STARD guidelines. Nevertheless, there was no statistical significant difference between these groups of studies (p=0.26).

E. DISCUSSION

In this survey, diagnostic accuracy studies examining the role of FC in diagnosing IBD, indexed in PubMed and Cochrane libraries during a period of 18 years, have been identified and have been, for the first time to our knowledge, comprehensively assessed using the STARD 2015 statement.

Our evidence reveals a moderate to satisfactory reporting quality (44.1-88.2%) with 25/26 (96.2%) trials reporting ≥ 50% and 8/26 (30.8%) trials reporting ≥75% of the STARD items. The mean adherence score (67.8%) was comparable to that of previous

publications in other fields of medicine [50, 51]. The variance of the reporting between different sections of the checklist was remarkable. The title, the abstract and the introduction parts approximated 100% adherence but, notwithstanding this, the section "other information" (latterly included in the 2015 update) displayed discouraging results. Explicitly, none of the studies gave access to the full protocol and only one mentioned registration data. Resembling findings were documented by *Zarei et al, 2015,* where items 18, 25, 28 and 29 also marked the lowest score [52]. Although in Zarei's article the low adherence to the unit "other information" was attributed to the nature of radiology studies, we suggest that the author, peer reviewers and journal editors should be further familiarized with the novel items of STARD 2015. The methods section was generally adequately reported (eligibility criteria, trial design, location, duration, index and reference test description) but the sample size estimation, the indeterminate data handling and the blinding of the performers of the FC test were the Achilles' heel with an adherence score below 50%. Finally, the result unit was characterized by a modest level of adherence (no referral to adverse events, only 38.5% of publications use a participant flowchart). The core element of the result section, the cross tabulation of index results against reference standard data, allows recalculating measures of diagnostic accuracy and performing alternative or additional analyses such as meta-analyses. [19] This substantial item scored no more than a tolerable 50% hindering the appraisal and the future analysis of the trials' results.

We did not identify any significant improvement in the reporting quality over time and the 2015 update of STARD checklist had trivial effect. Analogously, *Coppus et al 2006* supported that the introduction of STARD 2003 initiative had no impact in the reporting of test accuracy studies in reproductive medicine [53]. In addition in our study, the number of author's and the journal 5-year impact factor had inconsequential association with the STARD score. These findings are in consistency with previous survey which noted a suboptimal reporting quality even in articles published in journals with impact factor above 4 [54]. We concluded, nonetheless, that only a larger number of participants formed a pattern of ameliorated adherence.

Diagnostic accuracy studies are paramount to the promotion of medical science and the establishment of priorities and recommendations. Scarcity in reporting afflicts the quality of systematic reviews and therefore downgrades the development of clinical

guidelines that aim to detect patients with IBD in an early stage without imposing low risk patients to the inconvenient and expensive endoscopic process. Only 6 studies out of 26 included used the STARD tool proving that the scientific community is not yet adequately informed. Ergo reporting can be substantially improved by disseminating the utilization of the STARD statement; proper education of authors, training researchers, reviewers, funders and journal editors has a key role to prevent against incomplete adherence, one of the largest sources of avoidable waste in biomedical research [55].

Study limitations

Several limitations of the study merit consideration. Firstly, the search strategy was restricted only in PubMed and Cochrane libraries. Subsequently, articles indexed in other databases were omitted. Secondly, non-English literature was excluded increasing the potential risk of selection bias. Thirdly, the outcome measure, STARD score, is a subjective evaluation. Especially in our study the presence of one sole assessor inhibits the measurement of intra-observer agreement as an index of systematic bias. Positively, the use of "N/A" as a supplementary dependent protects included studies against falsely low scores, by only scoring articles out of a relevant total. Finally, it must be accentuated that the quality of the science of an article and its STARD score do not necessarily concur, albeit they do overlap. Besides, high quality reporting allows the readership to accept the authenticity of a trial and the medical society to conduct effective systematic reviews and meta-analyses.

Conclusion

To summarize, it is of high priority to spread and upgrade the role of STARD 2015 in order to ensure a comprehensive reporting status of diagnostic accuracy studies. Some vital sections of the checklist such as the analysis methodology and the results performed below the satisfactory level. Hence, meticulous assessment is required to guarantee the critical appraisal and the credibility of a study. Undoubtedly a basic precondition for the right management decisions by doctors and policymakers is the STARD tool implementation with the greatest of assiduity.

Conflict of interests: None declared

F. REFERENCES

[1] Gismera CS, Aladren BS. Inflammatory bowel diseases: a disease(s) of modern times? Is incidence still increasing? *World Journal of Gastroenteroly* 2008;14:5491-8

[2] Burisch J, Jess T, Martinato M, et al. The burden of inflammatory bowel disease in Europe *Journal of Crohn's Colitis* 2013;7:322–37

[3]Frolkis AD, Dykeman J, Negrón ME, et al. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology* 2013;145:996–1006

[4] Jess T, Loftus EV, Velayos FS, et al. Risk of intestinal cancer in inflammatory bowel disease: a population-based study from olmsted county, Minnesota. *Gastroenterology* 2006;130:1039–1046.

[5] Ha F, Khalil H. Crohn's disease: a clinical update. *Therap Adv Gastroenterol*. 2015 Nov;8(6):352-9.

[6] Lynch WD, Hsu R. Colitis Ulcerative *Stat. Pearls* 2019

[7] Stange EF, Travis SP, Vermeire S, et al. European evidence based consensus on the diagnosis and management of Crohn's disease: definitions and diagnosis. *Gut.* 2006;55(suppl 1):i1-15S

[8] World Gastroenterology Organization. Irritable bowel syndrome: a global perspective. *World Gastroenterology Organisation Global Guideline* 2009.

[9] Ristvedt SL, McFarland EG, Weinstock LB. Patient preferences for CT colonography, conventional colonoscopy, and bowel preparation. *Am J Gastroenterol* 2003;98:578-85.

[10] Waugh N, Cummins E, Royle P, et al. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technol Assess* 2013;17:1–211

[11] Lasson A, Kilander A, Stotzer PO. Diagnostic yield of colonoscopy based on symptoms. *Scand J Gastroenterol* 2008;43:356-62

[12] Dale I, Brandtzaeg P, Fagerhol MK, et al. Distribution of a new myelomonocytic antigen (L1) in human peripheral blood leukocytes.. *Am J Clin Pathol* 1985;84(1):24–34.

[13] Andersson KB, Sletten K, Berntzen HB, et al. The leucocyte L1 protein: identity with the cystic fibrosis antigen and the calcium-binding MRP-8 and MRP-14 macrophage components. *Scand J Immunol* 1988;28(2):241–5.

[14] Summerton CB, Longlands MG, Wiener K, et al. Faecal calprotectin: a marker of inflammation throughout the intestinal tract. *Eur J Gastroenterol Hepatol*. 2002;14(8):841–5

[15] Roseth AG, Fagerhol MK, Aadland E, et al. Assessment of the neutrophil dominating protein calprotectin in feces. A methodologic study. *Scand J Gastroenterol* 1992;27:793-8

[16] Røseth AG, Kristinsson J, Fagerhol MK,et al. Faecal calprotectin: a novel test for the diagnosis of colorectal cancer? *Scand J Gastroenterol.*1993 Dec; 28(12):1073-6.

[17] Tibble JA, Sigthorsson G, Foster R, et al. High prevalence of NSAID enteropathy as shown by a simple faecal test. *Gut*. 1999 Sep; 45(3):362-6.

[18] Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013;66:1093–104

[19] Cohen JF, Korevaar DA, Altman DG et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ* 2016

[20] Korevaar DA, van Enst WA, Spijker R, et al. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med* 2014;19:47–54.

[21] Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6

[22]Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015

[23] Freeman K, Willis BH, Fraser H, et al. Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy. *BMJ* 2018

[24] Banerjee A, Srinivas M, Eyre R, et al. Faecal calprotectin for differentiating between irritable bowel syndrome and inflammatory bowel disease: a useful screen in daily gastroenterology practice. *Frontline Gastroenterol*. 2015;6(1):20–26.
[25] Caviglia G, Pantaleoni S, Touscoz GA, et al. Fecal calprotectin is an effective diagnostic tool that differentiates inflammatory from functional intestinal disorders, *Scandinavian Journal of Gastroenterology* 2014
[26] Chang, M., Chou, J., Chen, et al. Faecal calprotectin as a novel biomarker for differentiating between inflammatory bowel disease and irritable bowel syndrome. *Molecular Medicine Reports* 2014
[27] Conroy S, Hale MF, Cross SS, et al Unrestricted faecal calprotectin testing performs poorly in the diagnosis of inflammatory bowel disease in patients in primary care *Journal of Clinical Pathology* 2018;71:316-322.
[28] Damms A, Bischoff S. Validation and clinical significance of a new calprotectin rapid test for the diagnosis of gastrointestinal diseases. *Int J Colorectal Dis* 2008
[29] De Sloovere, Maxime MW, Dieter De Smet, et al. 2017. Analytical and diagnostic performance of two automated fecal calprotectin immunoassays for detection of inflammatory bowel disease. *Clinical Chemistry and Laboratory Medicine* 2017
[30] Dhaliwal A, Zeino Z, Tomkins C,et al. Utility of faecal calprotectin in inflammatory bowel disease (IBD): what cut-offs should we apply? *Frontline Gastroenterology* 2015; 6:14-19
[31] El-Badry A, Sedrak H, Rashed L. Faecal calprotectin in differentiating between functional and organic bowel diseases. *Arab J Gastroenterol*. 2010;11:70–3
[32] Högberg C, Karling C, Rutegård J, et al. Diagnosing colorectal cancer and inflammatory bowel disease in primary care: The usefulness of tests for faecal haemoglobin, faecal calprotectin, anaemia and iron deficiency. *Scandinavian Journal of Gastroenterology* 2017
[33] Jang HW, Kim HS, Park SJ, et al. Accuracy of three different fecal calprotectin tests in the diagnosis of inflammatory bowel disease. *Intest Res.* 2016;14(4):305–313.
[34] Kalantari H, Taheri A, Yaran M. Fecal calprotectin is a useful marker to diagnose ulcerative colitis from irritable bowel syndrome. *Adv Biomed Res* 2015;4:85.
[35] Kennedy NA, Clark A, Walkden A, et al. Clinical utility and diagnostic accuracy of faecal calprotectin for IBD at first presentation to gastroenterology services in adults aged 16-50 years. *J Crohns Colitis*. 2015;9(1):41–49.
[36] Labaere D, Smismans A, Van Olmen A, et al. Comparison of six different calprotectin assays for the assessment of inflammatory bowel disease. *United European Gastroenterol J*. 2014
[37] Licata A, Randazzo C, Cappello MF, et al. (2012). Fecal calprotectin in clinical practice: a noninvasive screening tool for patients with chronic diarrhea. *Journal of clinical gastroenterology* 2012
[38] Manz M, Burri E, Rothen C, et al. Value of fecal calprotectin in the evaluation of patients with abdominal discomfort. *BMC Gastroenterol*. 2012.
[39] Moein S, Qujeq D, Vaghari T, et al. Diagnostic accuracy of fecal calprotectin in assessing the severity of inflammatory bowel disease: From laboratory to clinic. *Caspian J Intern Med* 2017
[40] Mowat C, Digby J, Strachan JA, et al. Faecal haemoglobin and faecal calprotectin as indicators of bowel disease in patients presenting to primary care with bowel symptoms. *Gut.* 2016;65(9):1463–1469.
[41] Otten C Kok L, Witteman B, et al. Diagnostic performance of rapid tests for detection of fecal calprotectin and lactoferrin and their ability to discriminate inflammatory from irritable bowel syndrome. *Clinical Chemistry and Laboratory Medicine* 2008
[42] Oyaert M, Trouvé C, Baert F, et al.Comparison of two immunoassays for measurement of faecal calprotectin in detection of inflammatory bowel disease: (pre)-analytical and diagnostic performance characteristics. *Clinical Chemistry and Laboratory Medicine* 2013
[43] Pavlidis P, Fergus J, Jeremy A, et al. Diagnostic accuracy and clinical application of faecal calprotectin in adult patients presenting with gastrointestinal symptoms in primary care. *Scandinavian Journal of Gastroenterology* 2013

[44] Schoepfer A, Beglinger C, Straumann A et al. Fecal Calprotectin Correlates More Closely With the Simple Endoscopic Score for Crohn's Disease (SES-CD) than CRP, Blood Leukocytes, and the CDAI. *American Journal of Gastroenterology* 2010
[45] Schroder O, Naumann M, Shastri Y, et al. Prospective evaluation of faecal neutrophil-derived proteins in identifying intestinal inflammation: combination of parameters does not improve diagnostic accuracy of calprotectin. *Alimentary Pharmacology & Therapeutics* 2007
[46] Sydora M, Sydora B, Fedorak R. Validation of a point-of-care desk top device to quantitate fecal calprotectin and distinguish inflammatory bowel disease from irritable bowel syndrome. *Journal of Crohn's and Colitis* 2012
[47] Tibble J, et al. Use of surrogate markers of inflammation and Rome criteria to distinguish organic from nonorganic intestinal disease. *Gastroenterology* 2002
[48] Von Roon A, Karamountzos L, Purkayastha S, et al. Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy. *Am J Gastroenterol.* 2007;102:803–813
[49] Limburg PJ, Ahlquist DA, Sandborn WJ, et al. Fecal calprotectin levels predict colorectal inflammation among patients with chronic diarrhea referred for colonoscopy. *Am J Gastroenterol.* 2000;95:2831–7.
[50] Choi YJ, Chung MS, Koo HJ, et al. Does the Reporting Quality of Diagnostic Test Accuracy Studies, as Defined by STARD 2015, Affect Citation?. *Korean J Radiol.* 2016;17(5):706–714.
[51] Areia M, Soares M, Dinis-Ribeiro M. Quality reporting of endoscopic diagnostic studies in gastrointestinal journals: where do we stand on the use of the STARD and CONSORT statements? *Endoscopy* 2010;42(2):138–147.
[52] Zarei F, Zeinali-Rafsanjani B. Assessment of Adherence of Diagnostic Accuracy Studies Published in Radiology Journals to STARD Statement Indexed in Web of Science, PubMed & Scopus in 2015. *J Biomed Phys Eng*. 2018;8(3):311–324.
[53] Coppus SF, van der Veen F, Bossuyt PM, et al. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertility and Sterility* 2006
[54] Smidt N, Rutjes AW, van der Windt DA, et al. Reproducibility of the STARD checklist: an instrument to assess the quality of reporting of diagnostic accuracy studies. *BMC Med Res Methodol.* 2006;6:12.
[55] Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet.* 2014;383(9913):267–76