Dysphagia (2016) 31:721–729
DOI 10.1007/s00455-016-9737-3

REVIEW ARTICLE

Diagnostic Accuracy of the Modified Evan's Blue Dye Test in Detecting Aspiration in Patients with Tracheostomy: A Systematic Review of the Evidence

Sibylle Béchet[1] · Fiona Hill[2] · Órla Gilheaney[1] · Margaret Walshe[1]

Received: 28 April 2016 / Accepted: 21 July 2016 / Published online: 16 August 2016

Abstract Oropharyngeal aspiration (OPA) is a common occurrence in patients with tracheostomy. The modified Evan's blue dye test (MEBDT) is an easily administered bedside procedure for the assessment of tracheostomised patients. However, studies evaluating the diagnostic accuracy of the MEBDT reach conflicting results. Therefore, we conducted a systematic review to determine the overall accuracy of the MEBDT in detecting OPA in adults with tracheostomy. The search strategy incorporated searching electronic databases, checking reference lists and citations and retrieving unpublished data. Data of primary studies were extracted and examined by three independent reviewers. The assessment of the methodological quality of included studies was performed using the QUADAS-2 tool. Six studies met the inclusion criteria for this systematic review. The studies presented significant disparities in study design and patient characteristics. Furthermore, high discrepancies in the administration of MEBDT across studies were noted. Therefore, a meta-analysis was not considered appropriate. Sensitivity estimates varied widely across the studies (38–95 %), indicating that the MEBDT is unreliable in detecting OPA. However, the studies emerge with overall high specificity values, ranging from 79 to 100 %. This true negative rate suggests that the MEBDT correctly identifies patients without OPA. This review highlights the need for further research studies assessing the accuracy of the MEBDT in detecting aspiration in patients with tracheostomy, using a standardised and reliable procedure. Outcomes from such studies will update the current level of evidence in relation to the MEBDT and consequently define best clinical practice.

Keywords Deglutition disorders · Oropharyngeal aspiration · Modified Evan's blue dye test · Diagnostic accuracy · Tracheostomy

Introduction

Within the last decade, the number of tracheostomy tube placements has drastically increased and the insertion of a tracheostomy has become a frequently performed procedure in intensive care medicine [1–3]. Reasons include advances in critical care medicine, such as the development of less invasive tracheostomy techniques, and improved management of critically ill patients as well as a growing elderly population [4, 5]. Different manifestations of swallowing difficulties have been reported in combination with the presence of a tracheostomy tube [6, 7]. While it is recognised that oropharyngeal aspiration (OPA) frequently occurs in patients with long-term artificial airways, currently the literature lacks clinical evidence of a causal relationship between tracheostomy and aspiration [8]. An early diagnosis is therefore crucial to enable immediate management as well as accurate intervention. The restricted availability and the need of specialised healthcare practitioners complicate or even prevent the use of instrumental assessment in evaluating deglutition dysfunction of tracheostomised patients [9]. Thus, clinical practice would highly benefit from an easily administered and accurate screening tool for aspiration in tracheostomised patients.

✉ Sibylle Béchet
bechets@tcd.ie

[1] Department of Clinical Speech and Language Studies, Trinity College Dublin, 7-9 South Leinster Street, Dublin 2, Ireland

[2] Department of Speech and Language Therapy, Tallaght Hospital, Dublin 22, Ireland

The modified Evan's blue dye test (MEBDT) is a recommended bedside procedure for the evaluation of deglutition and aspiration in tracheostomised patients [10, 11] and involves mixing blue dye with water or semisolid food. The presence of blue dye in or around the tracheostomy tube indicates a possible aspiration.

However, primary studies assessing the diagnostic accuracy (DA) of the MEBDT reach conflicting results. Current research lacks clear and reliable results, which could be used as a basis for clinical practice. Despite a lack of strong evidence of its accuracy, the MEBDT is used in clinical practice to diagnose deglutition disorders underlying various diseases. A survey on clinical consistency among SLTs in the management of tracheostomies indicates that out of the 64 participants, 45 % of the SLTs reported having used the MEBDT in their clinical practice [12]. To date, no systematic review identifying and summarising the findings of all relevant primary studies and thereby making the available evidence accessible to health care practitioners has been conducted. The presence of conflicting evidence and the lack of studies making a strong impact on clinical practice therefore justify the need for an objective analysis of all available evidence on the accuracy of the MEBDT.

The aims of this systematic review are to investigate the overall DA of the MEBDT in assessing the occurrences of OPA in adult patients with tracheostomy and consequently define the implications of the findings for research and clinical practice.

Methods

Study Selection

The index test assessed within the scope of this review was the MEBDT as described by Thompson-Henry and Braddock [13]. The chosen reference standard tests were the videofluoroscopy (VFSS) and fiberoptic endoscopic evaluation of swallowing (FEES), as both assessment methods are widely used in clinical practice and show good concordance regarding laryngeal aspiration [14, 15]. All types of studies assessing the DA of MEBDT against VFSS or FEES were included, regardless of their publication status. No restrictions were made regarding the data-collection design, including both retrospective and prospective studies. Inclusion criteria comprised tracheostomised adults (>18 years) with diagnosed or suspected dysphagia, without restrictions concerning the underlying medical cause of the tracheostomy. Studies were excluded if (1) they did not provide sufficient data for calculating a 2×2 contingency table; (2) it was impossible to retrieve data concerning the relevant participant group in case of a mixed population (adults and paediatrics).

Search Strategy

The search strategy for identification of studies incorporated searching eight different electronic databases (PubMed, CINAHL, Embase, ScienceDirect, Web of Science, PsycINFO, LILACS, ProQuest Nursing and Allied Health Source, ProQuest Dissertation & Theses). Databases were sought from October 2014 to April 2016. A preliminary search in the databases PubMed and CINAHL was conducted to determine the relevant indexing terms. Because of the paucity of studies assessing the DA of the blue dye test, the electronic database search was performed using only the index test as key search term. After this initial search, the reference standards and target condition served as inclusion criteria in the screening process of the identified records.

Furthermore, the search strategy incorporated scanning reference lists and citations of relevant studies as well as retrieving unpublished data for the purposes of minimising publication bias [16]. In addition, a search for ongoing studies assessing the diagnostic accuracy of the MEBDT was carried out. The search process was documented and reported using a flow diagram, as recommended by the PRISMA statement [17].

Data Collection and Quality Assessment

An initial screening on the basis of title and abstract of all articles generated by the literature search allowed to exclude studies that were clearly not meeting the inclusion criteria. Studies that appeared to meet the inclusion criteria were assessed for eligibility, and data extraction and analysis was undertaken by four independent reviewers (SB, FH, OG and MW). Disagreements were discussed and resolved by consensus. Data for each included study were recorded on a standard data extraction form. The following items were extracted: (1) publication details (title, authors, country where the study was conducted; (2) study design (prospective or retrospective); (3) patient characteristics (n included; population type, mean age); (4) TP, FP, TN and FN from the individual studies; (5) index test (protocol and type of blue dye used); (6) reference standard tests and time interval between the latter and the index test; (7) follow-up duration and reported adverse effects of the blue dye. In addition, the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) was used as a checklist for the reporting of DTA studies [18].

The risk of bias was evaluated using the revised quality assessment of diagnostic accuracy studies-2 tool [19, 20]. Each record was evaluated both in terms of the risk of bias and of concerns regarding the applicability of the findings. Results of the methodological quality assessment were managed on the Cochrane Review writing RevMan software© [21], and a methodological quality graph was constructed for all the included studies.

Statistical Analysis and Data Synthesis

The statistical analysis of the data was carried out using the RevMan software© [21]. The type of data was binary, with either positive or negative reporting of the target condition. The following statistical measures provided in the primary study were re-calculated using the contingency table: sensitivity, specificity, predictive values, likelihood ratios and diagnostic odds ratios. The data of each reference standard were subsequently analysed and presented separately. Two forms of graphical display were used to provide an illustrated overview of the estimates of sensitivity and specificity of the primary studies: summary ROC plots and forest plots.

Results

Characteristics and Methodological Quality of Studies

Figure 1 shows the flow of the literature through the search and screening process. Out of 39 different records identified through the searching process, 6 studies met the eligibility criteria. Table 1 lists the characteristics of the included studies. An overview of the methodological quality of the included studies is presented graphically in Figs. 2 and 3. The highest risk of bias involves the patient selection, followed by the domain flow and timing. In general, all studies presented low applicability concerns in the patient selection and their choice of reference standard. However, the applicability concerns were greater with regard to the index test, as the lack of information on the MEBDT protocol in some studies limited the transparency and consequently their applicability.

The analysis of the methodological quality by means of the QUADAS-2 tool shows that the methodology used in the study by Thompson-Henry and Braddock [13] is open to question. The small sample size ($n = 5$), the poorly articulated inclusion and exclusion criteria as well as the lack of information on the subject selection largely reduce the power of the study's findings. In addition, the use of two different reference standard procedures and the lack of consistency in their use limit the trustworthiness of these findings. Inferences made during the study were based on two events, with a

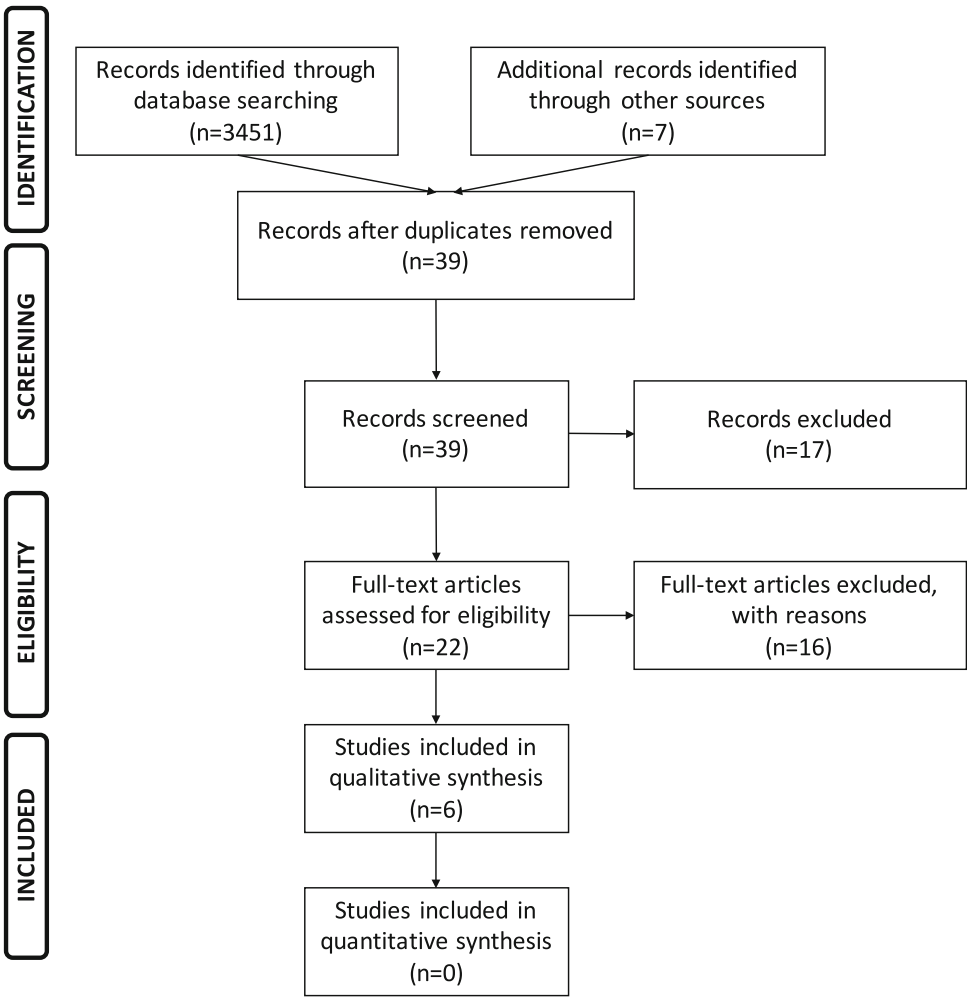

Fig. 1 Flow diagram showing the process of selection of records and studies for this review

Table 1 Study characteristics

Study	*n*	Data collection	Sampling method	Reference standard	Consistencies trialled	Time interval between RST and IT
Brady et al. [23]	21	Prospective	Convenience	FEES	Pureed and solid consistencies, ice chips	Simultaneous procedures
Donzelli et al. [24]	15	Prospective	Consecutive	FEES	Not specified	Simultaneous procedures
O'Neill-Pirozzi et al. [22]	37	Prospective	Unreported	MBS	• Liquid consistencies: thin liquid; thick nectar; saliva • Pureed solid consistencies	Simultaneous procedures
Peruzzi et al. [8]	20	Prospective	Consecutive	VFSS	Liquid consistencies (water)	Consecutive procedures (immediate)
Thompson-Henry and Braddock [13]	5	Retrospective	Case study	VFSS & FEES	Liquid and semisolid consistencies	Consecutive procedures (4–22 days)
Winkl-maier et al. [25]	30	Prospective	Consecutive	FEES	Liquid consistencies (water, artificial saliva)	Consecutive procedures (immediate)

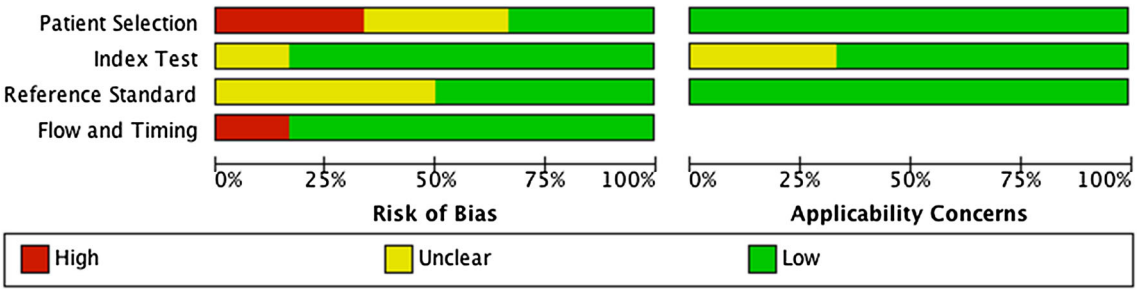

Fig. 2 Risk of bias and applicability concerns across included studies as per review authors' judgement for each domain presented as percentages

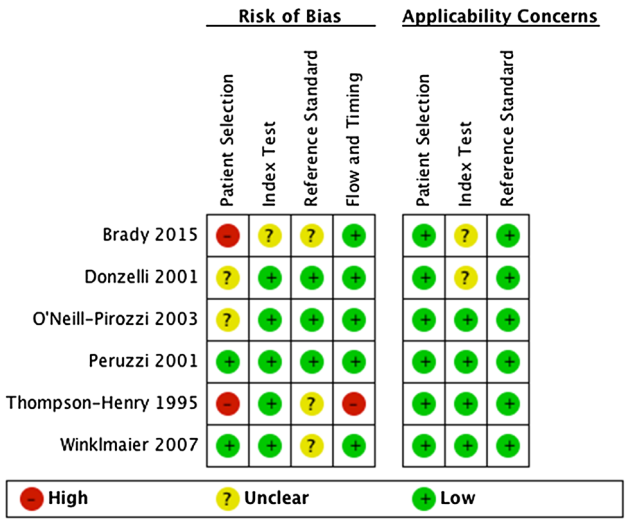

Fig. 3 Summary of the risks of bias and applicability concerns for included studies as judged by the review authors for each domain

delay reaching up to 22 days. This methodological flaw introduces a high risk for misclassification due to recovery or deterioration of the condition of the patients. Furthermore, the lack of information on the analysis and interpretation of the reference standard raises potential for bias and concerns regarding the investigator objectivity. Owing to these methodological flaws, the study by Thompson-Henry and Braddock [13] was excluded from the analysis, graphs and tables, as its data would distort the results of this review.

Estimates of Diagnostic Accuracy

Pooling of the studies in a meta-analysis was felt to be inadvisable due to the high diversity of the studies, participant samples, and index test protocols of the included studies.

Both studies comparing the MEBDT to VFSS reported estimates of sensitivity and specificity [8, 22]. However, the sensitivity and specificity values in O'Neill-Pirozzi et al. [22] are incorrectly inversed. Table 2 provides an

Table 2 Estimates of accuracy of the MEBDT compared with VFSS

Study	*n*	Sens (95 % CI)	Spec (95 % CI)	Prevalence (95 % CI)	PPV (95 % CI)	NPV (95 % CI)	LR^+ (95 % CI)	LR^- (95 % CI)
O'Neill-Pirozzi et al. [22]	50	0.619048 (0.3869–0.8105)	0.793103 (0.5974–0.9129)	0.42 (0.28489–0.5673)	0.684211 (0.43498–0.8644)	0.741935 (0.5507–0.8746)	2.992059 (1.3612–6.5770)	0.480331 (0.2736–0.8433)
Peruzzi et al. [8]	20	0.384615 (0.1513–0.6772)	1.00 (0.5609–1.00)	0.65 (0.4094–0.8369)	1.00 (0.4629–1.00)	0.466667 (0.2228–0.7258)	Infinity	0.615385 (0.4004–0.9458)

overview of the statistical measures of the MEBDT compared with VFSS. The prevalence of OPA in the study samples ranged from 42 to 65 %. However, the high variability patient characteristics and the overall small sample sizes of the primary studies reduce the representativity of the calculated prevalence. Sensitivity and specificity estimates of the primary studies were highly variable and ranged from 38 to 62 % and from 79 to 100 %, respectively. No procedure emerged with both high sensitivity and specificity estimates, although high levels of specificity were more commonly reported than high levels of sensitivity. Comparative analyses of sensitivity and specificity estimates of the MEBDT versus VFSS are displayed in a coupled forest plot in Fig. 4 and in a SROC plot in Fig. 5.

The above three studies investigated and compared the estimates of accuracy of the MEBDT versus FEES. [13, 23–25]. The estimates of the accuracy of the MEBDT compared with FEES are presented in Table 3. The prevalence of OPA in the study samples ranged from 53 to 71 %. Estimates of sensitivity varied across studies and ranged from 40 to 95 %. All the three studies agree on the specificity of the MEBDT compared with FEES, reporting an estimate of 100 %. These results suggest that the MEBDT is highly accurate in diagnosing aspiration when it is present, rather than excluding it. Comparative analyses of sensitivity and specificity estimates of the MEBDT versus FEES are displayed in a coupled forest plot in Fig. 6 and in a SROC plot in Fig. 7.

Overall, the reported low sensitivity of the MEBDT indicates its inaccuracy in unfailingly identifying patient with OPA. The observed high true negative rate, on the other hand, seems to suggest that MEBDT is reliable in identifying patients without OPA. Extrapolation of these findings into clinical settings adverts that a clinician can rely on the positive results of the MEBDT, indicating the patient most certainly has OPA.

Discussion

No definite conclusions could be drawn regarding the accuracy of the MEBDT in detecting OPA in adult patients with tracheostomy. The completeness and applicability of evidence was limited due to the paucity of primary studies investigating the DA of the MEBDT. Moreover, the strength of results yielded in this review is limited by the discrepancies in MEBDT protocols used across the included studies.

The lack of a universally accepted standard MEBDT protocol led to high variations across the studies' techniques in the administration of the index test. This review shows significant differences in the amount and

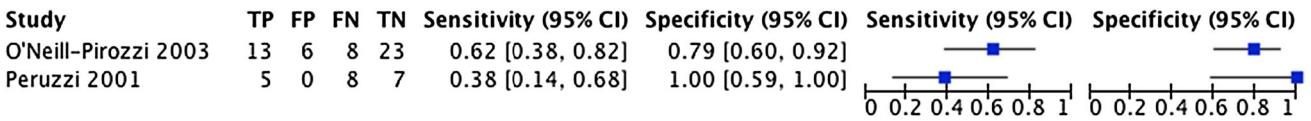

Fig. 4 Coupled forest plot of the estimates of sensitivity and specificity in VFSS studies. *TP* true positive, *FP* false positive, *FN* false negative, *TN* true negative

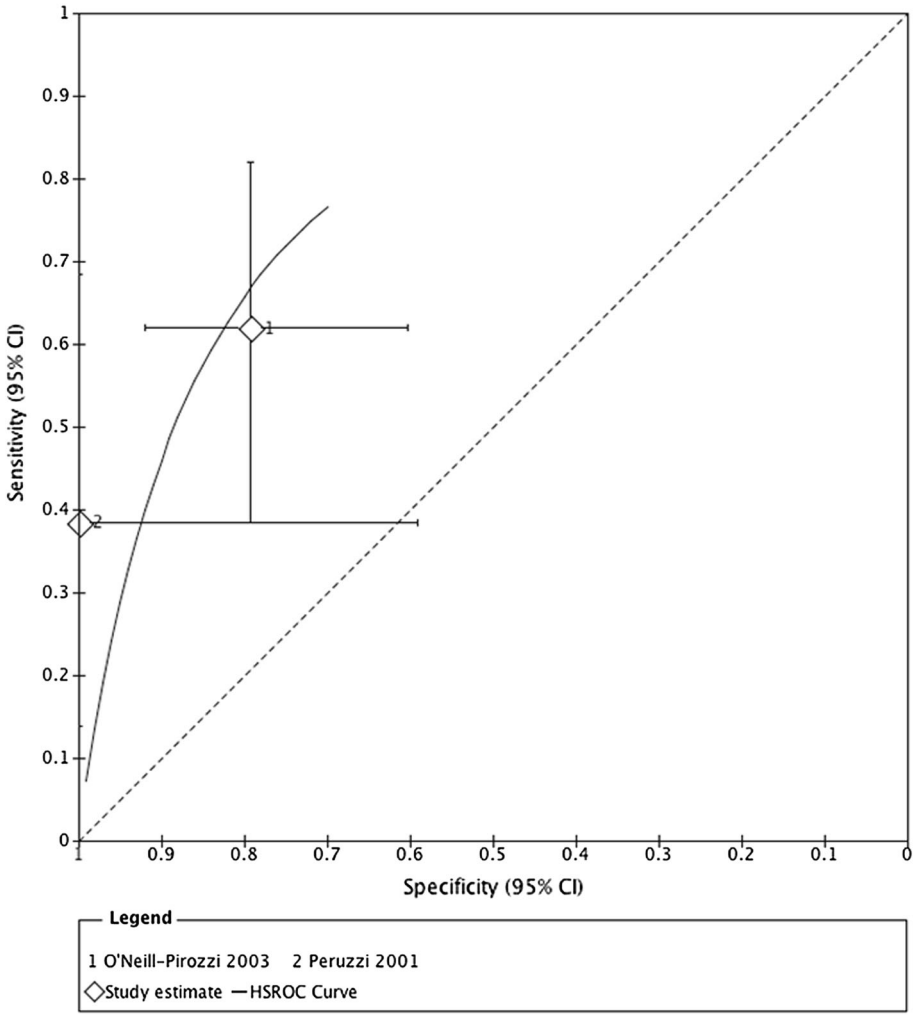

Fig. 5 Summary ROC plot of estimates of the accuracy of MEBDT compared with VFSS

consistency of the administered boluses as well as in the nature and amount of blue dye used to tint the test material. These discrepancies may account for the high variability in the reported accuracy measures, which consequently makes a comparison of the studies' findings difficult. Considering the paucity of primary studies assessing the MEBDT and the lack of consensus regarding their sensitivity and specificity values, there is a definite need for further research in that area. It is therefore essential for further research to provide large and well-designed primary diagnostic accuracy studies using a standardised MEBDT protocol in order to yield generalizable DA measures.

Preliminary evidence emerging from this review suggests that the MEBDT was generally better at excluding OPA, with all studies reporting higher specificity values. OPA-positive findings as assessed by the MEBDT-screening procedure thus provide an accurate basis for diagnosis, as the patient passed the 'exclusion test'. Nevertheless, it is important to stress that all studies reported low sensitivity values, indicating that the MEBDT is associated with potential false negative results. As a result, clinicians are not advised to use the MEBDT in order to rule-in OPA. However, the paucity of primary studies investigating the MEBDT's accuracy and their conflicting

Table 3 Estimates of accuracy of the MEBDT compared with FEES

Study	*n*	Sens (95 % CI)	Spec (95 % CI)	Prevalence (95 % CI)	PPV (95 % CI)	NPV (95 % CI)	LR^+ (95 % CI)	LR^- (95 % CI)
Brady et al. [23]	21	0.40 (0.1746–0.6711)	1.00 (0.5168–1.00)	0.714286 (0.4769–0.8781)	1.00 (0.5168–1.00)	0.40 (0.1746–0.6711)	Infinity	0.60 (0.39692–0.90699)
Donzelli et al. [24]	15	0.5 (0.1745–0.8255)	1.00 (0.5609–1.00)	0.533333 (0.27423–0.7772)	1.00 (0.3958–1.00)	0.636364 (0.3161–0.8763)	Infinity	0.50 (0.250049–0.9998)
Winkl-maier et al. [25]	30	0.95238095 (0.7413–0.9975)	1.00 (0.6288–1.00)	0.70 (0.504421–0.845872)	1.00 (0.7995–1.00)	0.90 (0.5412–0.9948)	Infinity	0.047619 (0.0070–0.32245)

findings fail to make a definite statement about the reliability of the MEBDT as a screening tool. Nevertheless, the use of a colouring agent might be useful as part of a comprehensive bedside swallowing assessment of patients with artificial airways, or in conjunction with other instrumental assessment tools. If a patient demonstrates a positive MEBDT, the timing of further instrumental swallowing evaluations may be postponed until the patient is able to pass the MEBDT [23]. This can provide SLTs with baseline information, and give early indications on the patient's deglutition ability. This procedure may be of value in clinical settings, where instrumental assessment is not readily available. However, the MEBDT should be used with caution, as there is insufficient evidence to support or reject the MEBDT as a sole screening tool.

Furthermore, it is important to acknowledge that the use of FD&C blue no.1. has become highly controversial. The FDA alerted healthcare professionals of several clinical cases of toxicity, including death, associated with the use of FD&C blue no.1 in enteral feeding solutions [26]. None of the included studies address the possible adverse effects of the blue dye. The different protocols of the studies do not include a long-term follow-up after the MEDBT procedure, which precluded an analysis of possible complications due to blue dye absorption within this review. It would therefore be of interest for future research studies to monitor the subjects undergoing the MEBDT with FD&C blue no.1 on a long-term basis, in order to provide information on the possible adverse effect of the blue dye. Outcomes from such studies will update the current level of evidence in relation to the MEBDT and consequently define best clinical practice.

Limitations of the Review

Because of the foreseen paucity of primary studies investigating the diagnostic accuracy of the MEBDT in tracheostomised adults, wide inclusion criteria were set for this review. As a consequence, the included studies were subject to heterogeneity and low overall methodology [27]. Therefore, a major limitation of this review process constitutes the impossibility of pooling the studies together in a meta-analysis in order to yield one overall effect estimate.

In addition, the use of multiple reference standards to determine the diagnostic accuracy of the MEBDT constitutes a limitation of the review, as it may result in biased estimates of the accuracy results if the findings are treated as interchangeable [28]. Practical or ethical constraints may lead researchers to use an alternative and less accurate reference standard instead of the preferred reference standard to assess the disease status of a specific population. This may consequently lead to differential verification [29, 30].

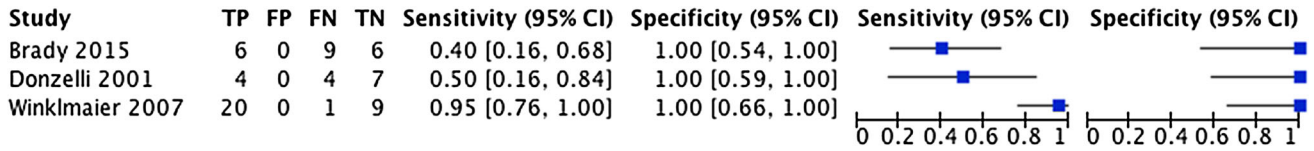

Fig. 6 Coupled forest plot of the estimates of sensitivity and specificity in FEES studies. *TP* true positive, *FP* false positive, *FN* false negative, *TN* true negative

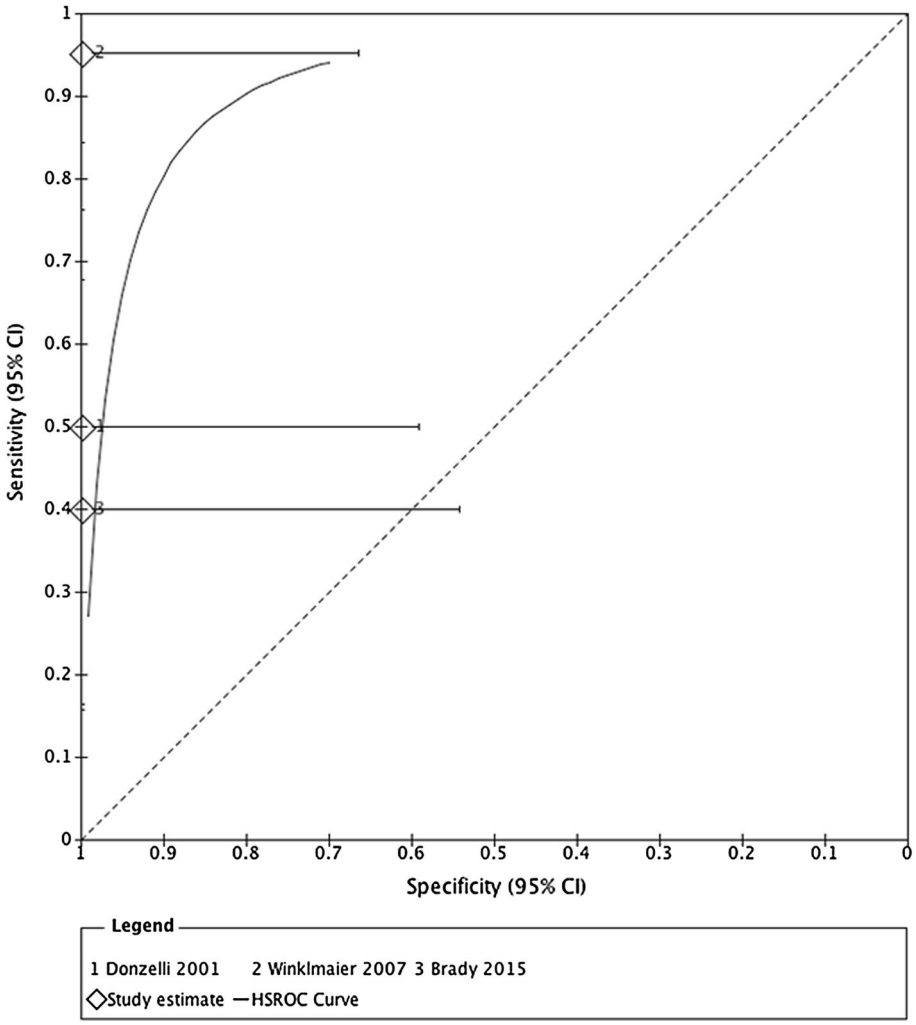

Fig. 7 Summary ROC plot of the estimates of accuracy of MEBDT compared with FEES

Conclusions

This review critically examined evidence pertaining to the effectiveness of the MEBDT in detecting OPA in tracheostomised patients. The results yielded in this review, however, indicate that the MEBDT generally presents with higher specificity values, and thus seems better at excluding OPA. Furthermore, it is noteworthy to stress the low sensitivity of the MEBDT. Clinicians should therefore use the MEBDT screening procedure with caution, as it is associated with high risk of false negative results. However, the strength of the findings reported in this review is limited by the discrepancies in MEBDT protocols used across studies. As a consequence, empirical evidence for or against the use of the MEBDT is not strong enough to direct clinical practice. Therefore, there is a strong necessity for further well-designed diagnostic accuracy studies using a standardised protocol to bridge the gap between clinical practice and documented evidence.

Compliance with Ethical Standards

Conflict of Interest There are no relevant disclosures or conflict of interest to declare for the content of this review.

References

1. Duann CW, Hsieh MS, Chen PT, Chou HP, Huang CS. Successful percutaneous tracheostomy via puncture through the thyroid isthmus. Respirol Case Rep. 2004;2:57–60. doi:10.1002/rcr2.48.
2. Kluge S, Baumann HJ, Maier C, et al. Tracheostomy in the intensive care unit: a nationwide survey. Anesth Analg. 2008;107:1639–43. doi:10.1213/ane.0b013e318188b818.
3. Young D, Harrison DA, Cuthbertson BH, Rowan K, Collaborators TracMan. Effect of early vs late tracheostomy placement on survival in patients receiving mechanical ventilation: the tracman randomized trial. JAMA. 2013;309:2121–9. doi:10.1001/jama.2013.5154.
4. Baskin JZ, Panagopoulos G, Parks C, Rothstein S, Komisar A. Clinical outcomes for the elderly patient receiving a tracheotomy. Head Neck. 2004;26:71–5.
5. Durbin CG. Tracheostomy: why, when, and how? Respir Care. 2010;55:1056–68.
6. Garuti G, Reverberi C, Briganti A, Massobrio M, Lombardi F, Lusuardi M. Swallowing disorders in tracheostomised patients: a multidisciplinary/multiprofessional approach in decannulation protocols. Multidiscip Respir Med. 2014;9:1–10. doi:10.1186/2049-6958-9-36.
7. Jung SJ, Kim DY, Kim YW, Koh YW, Joo SY, Kim ES. Effect of decannulation on pharyngeal and laryngeal movement in post-stroke tracheostomized patients. Ann Rehabil Med. 2012;36:356–64. doi:10.5535/arm.2012.36.3.356.
8. Peruzzi WT, Logemann JA, Currie D, Moen SG. Assessment of aspiration in patients with tracheostomies: comparison of the bedside colored dye assessment with videofluoroscopic examination. Respir Care. 2001;46:243–7.
9. Singh S, Hamdy S. Dysphagia in stroke patients. Postgrad Med J. 2006;82:383–91. doi:10.1136/pgmj.2005.043281.
10. Deutsche Gesellschaft für Neurologie. Neurogene Dysphagien. In: Diener H-C, Weimar C, editors. Leitlinien für diagnostik und therapie in der neurologie. Stuttgart: Thieme Verlag; 2012.
11. NHS Health Scotland. Caring for the patient with a tracheostomy: best practice statement. Scotland: NHS quality improvement. 2007. http://www.nhshealthquality.org/nhsqis/files/TRACHEOREV_BPS_MAR07.pdf.
12. Downes LS. Clinical consistency among speech and language therapists in the management of tracheostomies in the Republic of Ireland. Dublin: Department of Clinical Speech and Language Studies, Trinity College Dublin; 2012.
13. Thompson-Henry S, Braddock B. The modified Evan's blue dye procedure fails to detect aspiration in the tracheostomized patient: five case reports. Dysphagia. 1005;10:172–4. doi:10.1007/BF00260973.
14. Brady S, Donzelli J. The modified barium swallow and the functional endoscopic evaluation of swallowing. Otolaryngol Clin North Am. 2013;46:1009–22. doi:10.1016/j.otc.2013.08.001.
15. Lee W-S, Lee S-H, Park YH, et al. Effectiveness of fiberoptic endoscopic evaluation of swallowing (FEES) in patients with swallowing disorders: a systematic review. Brain Neurorehabil. 2008;6:9–16.
16. de Vet HCW, Eisinga A, Riphagen I, Aertgeertes B, Pewsner D. Chapter 7: Searching for studies. Cochrane handbook for systematic reviews of diagnostic test accuracy version 0.4. Oxford: The Cochrane Collaboration; 2008.
17. Moher D, Liberati A, Tetzlaff J, Altman GD, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151:264–9. doi:10.7326/0003-4819-151-4-200908180-00135.
18. Bossuyt MMP, Reitsma BJ, Bruns ED, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clin Chem. 2003;49:1–6. doi:10.1373/49.1.1.
19. Whiting P, Rutjes WSA, Reitsma BJ, Bossuyt MMP, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;10(3):25. doi:10.1186/1471-2288-3-25.
20. Whiting P, Rutjes WSA, Westwood EM, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529–36. doi:10.7326/0003-4819-155-8-201110180-00009.
21. Review Manager (RevMan) (computer program) Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
22. O'Neill-Pirozzi MT, Lisiecki JD, Momose KJ, Connors JJ, Milliner PM. Simultaneous modified barium swallow and blue dye tests: a determination of the accuracy of blue dye test aspiration findings. Dysphagia. 2003;18:32–8. doi:10.1007/s00455-002-0081-4.
23. Brady LS, Krieger R, Wesling M, Scott K, Donzelli J, Pietrantoni M. Sensitivity of the blue dye food test for detecting aspiration in patients with a tracheotomy. BJAST. 2015;7:429–35. doi:10.9734/BJAST/2015/11652.
24. Donzelli J, Brady LS, Wesling M, Craney M. Simultaneous modified Evans blue dye procedure and video nasal endoscopic evaluation of the swallow. Laryngoscope. 2001;111:1746–50.
25. Winklmaier U, Wüst K, Plinkert PK, Wallner F. The accuracy of the modified Evans blue dye test in detecting aspiration in head and neck cancer patients. Eur Arch Otorhinolaryngol. 2007;264:1059–64. doi:10.1007/s00405-007-0299-8.
26. Acheson D (2003) Reports of blue doscoloration and death in patients receiving enteral feedings tinted with the dye FD&C blue no. 1. FDA Public Health Advisory. http://www.fda.gov/ForIndustry/ColorAdditives/ColorAdditivesinSpecificProducts/InMedicalDevices/ucm142395.htm. Accessed 27 May 2015.
27. Kleijnen J, Westwood M, Whiting P. Applicability of diagnostic studies: statistics, bias and estimates of diagnostic accuracy. Z Evid Fortbild Qual Gesundheitswes. 2011;105:498–503. doi:10.1016/j.zefq.2011.07.025.
28. Naaktgeboren AC, de Groot J, van Smeden M, Moons K, Reitsma BJ. Evaluating diagnostic accuracy in the face of multiple reference standards. Ann Intern Med. 2013;159:195–202. doi:10.7326/0003-4819-159-3-201308060-00009.
29. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282:1061–6. doi:10.1001/jama.2015.7649.
30. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006;174:469–76. doi:10.1503/cmaj.050090.

Sibylle Béchet MSc

Fiona Hill MSc

Órla Gilheaney MSc

Margaret Walshe PhD