Journal of Pediatric Gastroenterology and Nutrition
42:496–505 © May 2006 Lippincott Williams & Wilkins, Philadelphia

Diagnostic Tests in Hirschsprung Disease: A Systematic Review

*F. de Lorijn, †L.C.M. Kremer, ‡J.B. Reitsma, and *M.A. Benninga

*From the *Departments of Paediatric Gastroenterology and Nutrition, †Paediatrics and ‡Clinical Epidemiology and Biostatistics, Emma Children's Hospital/Academic Medical Centre, Amsterdam, The Netherlands.*

ABSTRACT

Objective: We conducted a systematic review to determine and compare the diagnostic accuracy of contrast enema (CE), anorectal manometry (ARM) and rectal suction biopsy (RSB) in infants suspected of Hirschsprung disease.

Design: This is a systematic review.

Data Sources: Articles were identified through electronic searches in Medline, EMBASE.com and Cochrane Controlled Trials Register. Searches were limited to articles published after 1966 in PubMed and after 1980 in EMBASE.com.

Study Selection: Studies were included if infants underwent at least one of the following tests: CE, ARM or RSB, followed by full-thickness biopsy and/or clinical follow-up as the reference standard.

Data Extraction: Two reviewers independently assessed the methods of data collection, patient selection, blinding and prevention of verification bias and description of the test protocol and reference standard. Data to construct 2 × 2 tables were abstracted for each test.

Results: Twenty-four studies met our inclusion criteria, but 2 studies were subsequently excluded for statistical analysis because data was missing to construct the 2 × 2 table. RSB (14 studies for a total of 993 patients) was the most accurate test, having both the highest mean sensitivity (93%; 95% confidence interval [CI], 88%–95%) and mean specificity (98%; 95% CI, 95%–99%). Sensitivity and specificity of ARM (9 studies for a total of 400 patients) were similar to those of RSB (91% vs 93%, $P = 0.73$ and 94% vs 98%, $P = 0.08$, respectively). Sensitivity and specificity of CE (12 studies for a total of 425 patients) were significantly lower than those of RSB and ARM, with mean sensitivity and mean specificity of 70% and 83%, respectively.

Conclusions: RSB and ARM are the most accurate tests in the diagnostic workup of Hirschsprung disease. ***JPGN 42: 496–505, 2006.*** **Abbreviations:** HD, Hirschsprung disease-ARM, norectal manometry—RSB, rectal suction biopsy—CE, contrast enema—FTB, full thickness biopsy—RAIR, recto-anal inhibitory reflex—AChE, acetylcholinesterase—H&E, hematoxylin and eosin. **Key Words:** paediatric constipation—Hirschsprung disease—anorectal manometry—rectal suction biopsy—contrast enema—sensitivity and specificity—diagnosis—accuracy—systematic review.

INTRODUCTION

Constipation is a common problem in children and has often been regarded as a trivial symptom, but it can result in substantial discomfort and disability, impairing social and physical well-being. In only 10% of all children with defecation disorders, constipation is part of an organic disorder (1). About once per 5000 live births, constipation is caused by Hirschsprung disease (HD) (2).

HD is a developmental disorder of the enteric nervous system characterized by absence of ganglion cells in the myenteric and submucosal plexuses along a variable portion of the distal intestine. Literature showed that the aganglionosis is confined to rectosigmoid in 75% of patients; sigmoid, splenic flexure or transverse colon in 17%, total colon in 17% and total colon along with a short segment of the terminal ileum in 8% (3,4). Clinical symptoms are produced in 80% to 90% of all cases of HD and are diagnosed during the neonatal period. The usual presentation of HD in the neonatal period are lack of passage or delayed passage of meconium and signs and symptoms of large-bowel obstruction, which can lead to enterocolitis.

The diagnosis of HD is not always easy to establish. Three tests are available in the diagnostic workup of HD. The presence of a transitional zone is the critical feature to suspect HD in contrast enema (CE) test. Anorectal manometry (ARM) assesses the rectoanal inhibition reflex, and failure to elicit this reflex indicates HD. The third option consists of rectal suction biopsy (RSB), which shows an elevated acetylcholinesterase

Received June 10, 2005; accepted January 17, 2006.

Address correspondence and reprint requests to F. de Lorijin, Department of Paediatric Gastroentorology and Nutrition, Emma Children's Hospital/Academic Medical Centre, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. (e-mail: f.delorigin@amc.uva.nl).

(AChE) activity and an aganglionosis in case of HD. The gold standard, however, remains a full-thickness biopsy (FTB) of the rectum and provides the most definitive answer. The classical method of rectal biopsy involves an FTB of rectal mucosa and underlying muscle. However, this requires general anaesthesia and suturing of the biopsy site (5). Possible complications of this procedure are perforation, bleeding and infection. Absence of ganglion cells in an FTB confirms the diagnosis of HD.

There has been considerable debate about the most appropriate initial test for diagnosing HD because CE, RSB and ARM can all produce false-negative and false-positive test results (6–8). Each of these tests has both advantages and disadvantages in availability, technical difficulty, radiation exposure and invasiveness. Consequently, the choice of screening test and order of investigations for HD differs among medical centres.

We conducted a systematic review to determine and compare the diagnostic accuracy between CE, ARM and RSB in infants suspected of having HD.

METHODS

Search Strategy

A comprehensive literature search was performed to identify relevant publications examining the diagnostic accuracy of at least one of the following tests: CE, ARM or RSB. The Medline, EMBASE.com and the Cochrane Controlled Trials Register databases were searched using 3 different queries: (a) *Hirschsprung* [mesh or textword] AND (*barium* [mesh or text word] OR *enema* [mesh or text word]); (b) *Hirschsprung* [mesh or text word] AND *manometry* [mesh or text word]; (c) *Hirschsprung* [mesh or text word] AND (*pathology* [subheading] OR *biopsy* [mesh or text word] OR *acetylcholinesterase* [mesh or text word]). Searches were limited to articles published after 1966 in PubMed and after 1980 in EMBASE.com, and to studies involving humans. No language restriction was applied. Searches were performed in May 2003.

Selection of Studies

Eligible studies had to report on the diagnostic accuracy of at least one of the following tests: CE, ARM or RSB in a study population suspected of having HD. Staining for AChE activity or staining with hematoxylin and eosin (H&E) should have been used to evaluate the RSB. We posed no restriction on the design of the studies and included case series, case control, cohort studies and studies in which at least one reference standard was described. For a positive-index test, we defined the reference standard to be an aganglionosis on FTB. For a negative-index test, we defined the reference standard to be the presence of ganglion cells on FTB or on RSB or disappearance of symptoms during follow-up. Studies in which we were unable to reproduce the 2×2 table were excluded. Articles were independently selected and reviewed by 2 authors (F.d.L. and M.B.). Disagreements were resolved by consensus.

Methodological Quality and Data Extraction

Two authors (F.d.L. and M.B.) independently assessed the design of the study, blinding and prevention of verification bias, methods of data collection and patient characteristics (9). The other 2 authors (L.K. and J.R.) assessed those results and disagreements were resolved by consensus. For each article, we recorded the number of true-positive, false-positive, true-negative and false-negative results for each index test that was examined.

Data Synthesis

We used Forrest plots to display the precision by which sensitivity and specificity had been measured in each study, and to illustrate the variation in estimates between studies. The exact binomial method was used to calculate the 95% CIs. We used a bivariate metaregression model to meta-analyse the estimates of sensitivity and specificity (10,11). Rather than using a single outcome measure per study, like the diagnostic odds ratio in the summary receiver operating characteristic approach, the bivariate model preserves the 2-dimensional nature of diagnostic data by directly analysing the logit transformed sensitivity log(sens/[1−sens]) and specificity log(spec/[1−spec]) of each study in a single model. This model estimates and incorporates the correlation that might exist between logit sensitivity and specificity within studies caused by possible differences in threshold between studies. The bivariate model uses a random effects approach for both sensitivity and specificity, allowing for heterogeneity beyond chance caused by clinical or methodological differences between studies. In addition, the model acknowledges the difference in precision by which sensitivity and specificity have been measured in each study. This means that studies with a larger number of patients with the target condition receive more weight in the calculation of the summary estimate of sensitivity; target studies with more patients without the qualifying condition are more influential in the pooling of specificity.

The model requires logit transformation of the sensitivity and specificity. A standard correction of adding 0.5 to all 4 cells of the 2×2 table was applied when either sensitivity or specificity was 100%. The model produces the following results: a random effect estimate of the mean sensitivity and specificity with corresponding 95% CIs, the amount of between-study variation for sensitivity and specificity separately and the strength and shape of the correlation between sensitivity and specificity. Using these results, we calculated a 95% confidence ellipse around the summary estimate of sensitivity and specificity. All the results have been transformed back (antilogit) to the original scale, and plotted in ROC space.

Covariates indicating the type of index test can be added to the model to test explicitly whether either sensitivity or specificity or both are different between index tests.

We performed a subgroup analysis comparing studies with a substandard verification protocol with studies with an adequate protocol. An adequate verification protocol was defined as a study in which a predefined reference standard was performed for both a positive-index test and a negative-index test. Case series in which the reference standard was performed for a positive-index test were also defined as an

J Pediatr Gastroenterol Nutr, Vol. 42, No. 5, May 2006

TABLE 1. *Evidence table showing clinical and design characteristics of individual studies*

First author, year	Design	Verification		Blind	Data collection	Population		TP	FP	TN	FN	Inconclusive
		Index +	Index −			Age	% Boys					
Davis (12), 1972	Case series	FTB	NA	NR	Retro	NR	79	22	NA	NA	6	0
Mahboubi (13), 1978	Case series	FTB	NA	NR	Retro	25/27 ≤2 yrs	67	22	NA	NA	3	0
Kekomaki (14), 1979	Cohort	FTB in 10 infants	Unclear	NR	Retro	Range, 2 wks–11 yrs	Unclear	10	5	45	0	0
De Campo (15), 1984	Case series	FTB	NA	NR	Retro	Unclear	54	2	NA	NA	11	0
Lanfranchi (16), 1984	Cohort	FTB	Follow-up of at least 6 mos	NR	Retro	Mean, 4.8 yrs; range, 17 d–16 yrs	62	13	0	15	6	0
Rosenfield (17), 1984	Cohort	FTB	FTB	NR	Retro	Unclear	Unclear	29	4	16	13	0
Loening-Baucke (18), 1985	Cohort	Aganglionosis on RSB	Ganglioncells on RSB and follow-up	NR	Unclear	Mean, 14 d; range, 2–28 d	52	3	7	14	1	0
Taxman (19), 1986	Cohort	FTB	Follow-up or RSB with ganglioncells	NR	Retro	Range, 1 d–1 yr	NR	12	9	29	3	0
Smith (20), 1991	Case series	FTB	NA	NR	Retro	Median, 2–5 d	NR	13	NA	NA	4	0
O'Donovan (21), 1996	Cohort	FTB	Combination of ARM + CE + follow-up	NR	Retro	Range, 2 d–9 mos	61	13	NR	NR	5	0
Osatakul (7), 1999	Cohort	FTB	Follow-up	NR	Unclear	Range, 0.2–120 mos	64	21	0	5	10	0
Reid (22), 2000	Cohort	FTB	ARM and/or ganglioncells on RSB	NR	Retro	Mean, 3 yrs; range, 4 wks–15 yrs	56	5	1	47	1	0

Index test is CE. NR = not reported; FTB = full thickness biopsy; TP = true-positive test results; FP = false-positive test results; FN = false-negative test results; TN = true-negative test results; NA = not applicable.

adequate verification protocol. Unclear descriptions of reference tests were classified as a substandard verification.

The PROC MIXED statistical procedure in SAS version 8.2 for Windows (SAS Institute Inc, Cary, NC) was used to fit the various bivariate models.

RESULTS

Literature Search

We identified 587 studies from the electronic search; after initial evaluation, 54 of them were judged potentially relevant. Twenty-eight studies were excluded because no reference standard was specified and 2 studies were excluded because no AChE or H&E staining was used; the remaining 24 studies met our inclusion criteria. Twelve studies evaluated the diagnostic value of CE, 9 studies evaluated ARM and 14 studies evaluated RSB. The median sample size of the 24 studies was 45 patients (range, 9–293 patients). Two of these 24 articles were excluded from the statistical analysis because data were presented unclear and we were unable to produce a 2 × 2 table.

Quality Assessment and Data Extraction

Tables 1–3 list the 24 included studies and their clinical and design characteristics. The design of studies varied: 16 studies were cohort studies, 4 studies were case series, and 1 study used a case-control design. The design was unclear in 3 studies. Fourteen of the 24 studies used retrospective data collection, whereas the method of data collection was unclear in the other 10 studies. None of the studies reported whether the readers of the index tests were blinded to the results of the other index tests. Eleven of the 12 studies evaluating CE used FTB as the reference standard to verify positive CE results. However, only 7 studies used one of our predefined reference standards to verify negative CE results. An FTB was used as reference standard for positive ARM results in 8 of the 9 studies evaluating ARM, but verification of negative ARM results was suboptimal in 3 studies. Thirteen of the 14 RSB studies used FTB as the reference standard in case of a positive RSB result. In 8 RSB studies, the verification of negative test results was inadequate. Ten of the 14 RSB studies were stained with AChE, whereas 4 studies were stained with H&E. Because of the small number of studies using H&E staining, the results were not compared with the other index tests. In 2 studies using the AChE staining, it was not possible to reproduce the 2 × 2 table; these studies were therefore not included the analysis.

Diagnostic Accuracy

Figures 1 and 2 show the sensitivity and the specificity, with their 95% confidence interval (CI),

TABLE 2. *Evidence table showing clinical and design characteristics of individual studies*

First author, year	Design	Verification		Blind	Data collection	Population		TP	FP	TN	FN	Inconclusive
		Index +	Index −			Age	% Boys					
Frenckner (23), 1978	Cohort	FTB	Follow-up of at least 4 mos	NR	Unclear	Median, 5 mos; range, 1–12mos	NR	3	0	5	0	1
Mahboubi (13), 1978	Case series	FTB	FTB	NR	Retro	25/27 ≤2 yrs	67	22	NA	NA	3	0
Iwai (24), 1979	Case control	FTB	Follow-up, negative CE and 30% FTB	NR	Unclear	Range, 4 mos–19 yrs	NR	9	0	10	0	0
Kekomaki (14), 1979	Cohort	FTB in 10 infants	NR	NR	Retro	Range, 2 wks–11 yrs	Unclear	8	1	47	1	3
Lanfranchi (16), 1984	Cohort	FTB	Follow-up of at least 6 mos	NR	Retro	Mean, 4.8 yrs; range, 17 d–16 yrs)	62	17	0	15	2	0
Loening-Baucke (18), 1985	Cohort	Aganglionosis on RSB	Ganglioncells on RSB and follow-up	NR	Unclear	Mean, 14 d; range, 2-28 d	52	3	1	20	1	0
Yaxoing (25), 1986	Unclear	FTB	NR	NR	Unclear	Range, 30 d–12 yrs	59	63	5	58	0	0
Emir (6), 1997	Cohort	FTB	38% FTB or followu-up at least 4 mos	NR	Unclear	Mean 30,3 d; range, 2 d-3mos	63	22	1	32	1	0
Osatakul (7), 1999	Cohort	FTB	Follow-up	NR	Unclear	Range, 0,2–120 mos	64	31	2	12	0	0

Index test is ARM. NR = not reported; FTB = full thickness biopsy; TP = true-positive test results; FP = false-positive test results; FN = false-negative test results; TN = true-negative test results; NA = not applicable.

TABLE 3. *Evidence table showing clinical and design characteristics of individual studies*

First author, year	Design	Verification		Blind	Data collection	Population		TP	FP	TN	FN	Inconclusive
		Index +	Index −			Age	% Boys					
Campbell (26), 1969	Unclear	FTB	NR	NR	Unclear	Unclear	NR	NR	NR	NR	NR	NR
Kekomaki (14), 1979	Cohort	FTB in 10 infants	NR	NR	Retro	Range, 2 wks–11 yrs	Unclear	10	2	45	0	3
Hamoudi (27), 1982	Unclear	FTB	NR	NR	Unclear	NR	NR	NR	NR	NR	NR	NR
Huntley (23), 1982	Cohort	FTB	Ganglioncells on RSB	NR	Retro	Range, 4 d–12 yrs	NR	22	1	32	0	3
Barr (29), 1985	Cohort	FTB	Follow-up of 18 mos	NR	Retro	Nm	NR	19	0	79	3	0
Loening-Baucke (18), 1985	Cohort	Aganglionosis on RSB	Ganglioncells on RSB and follow-up of at least 6 mos	NR	Unclear	Mean, 14 d; range, 2–28 d	52	4	0	21	0	0
Kurer (5), 1986	Cohort	FTB	NR	NR	Unclear	Range, few wks-17 yrs	NR	17	0	54	1	0
Polley (30), 1986	Cohort	FTB	RSB with ganglioncells	NR	Retro	Mean, 14.4 mos	NR	41	0	251	1	0
Taxman (19), 1986	Cohort	FTB	Follow-up or RSB with ganglioncells	NR	Retro	Range, 1d–1yr	NR	11	0	40	0	7
Yaxoing (25), 1986	Unclear	FTB	NR	NR	Unclear	Range, 30 d-12 yrs	59	56	0	11	2	0
Bonham (31), 1987	Cohort	FTB	NR	NR	Retro	Range, 2 d–14 yrs	NR	45	0	164	4	0
Chen (32), 1987	Cohort	FTB	FTB or follow-up of at least 1 mo.	NR	Unclear	Range, 3 d–10 yrs	NR	17	1	65	0	0
Smith (20), 1991	Case series	FTB	NR	NR	Retro	Median, 2–5 d	NR	21	NA	NA	0	3
Park (33), 1992	Cohort	FTB in 13 infants	NR	NR	Unclear	Range, 3 d–17 yrs	Unclear	12	0	24	1	0

Index test is RSB. NR = not reported; FTB = full thickness biopsy; TP = true-positive test results; FP = false-positive test results; FN = false-negative test results; TN = true-negative test results; NA = not applicable.

for each study. CIs are wide, especially for estimates of sensitivity, because the number of patients with HD is low in most studies. The estimates of mean sensitivity and specificity of the 3 index tests are given in Table 4. The joint CI around the mean sensitivity and specificity is depicted in Figure 3. This figure clearly shows that the accuracy of CE is substantially poorer compared with the other tests. Formal statistical testing for differences between sensitivity and specificity reveal that RSB stained for AChE activity was the most accurate test, having both the highest mean sensitivity (93%; 95% CI, 88%–95%) and mean specificity (98%; 95% CI, 95%–99%). The RSB studies stained for H&E showed a comparable sensitivity and specificity (96% and 98%, respectively). Sensitivity and specificity of ARM (9 studies for a total of 400 patients) were similar to those of RSB stained for AChE activity (91% vs 93%, $P = 0.73$ and 94% vs 98%, $P = 0.08$, respectively). Sensitivity and specificity of CE were significantly lower than those of RSB stained for AChE activity and ARM, with sensitivity and specificity of 70% and 83%, respectively.

Subgroup analyses showed that there was no significant difference in either sensitivity or specificity between studies with a low-quality and high-quality verification protocol. However, a trend toward an exaggerated estimate of sensitivity was found in low-quality (n = 2) versus high-quality (n = 10) studies

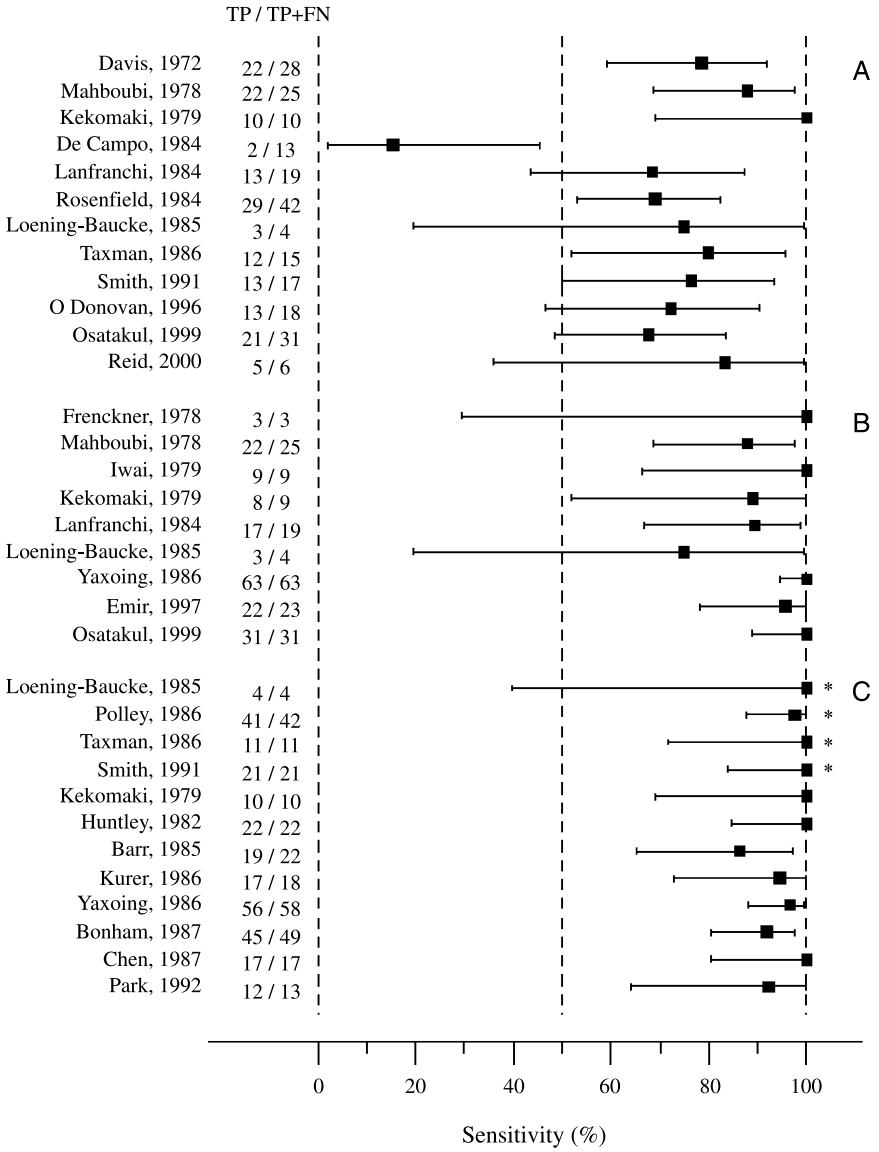

FIG. 1. Forrest plot of sensitivity in studies examining the accuracy of CE (A), ARM (B), and RSB (C). TP = true-positive test result; FN = false-negative test result; asterisk (*) = RSB studies stained with H&E.

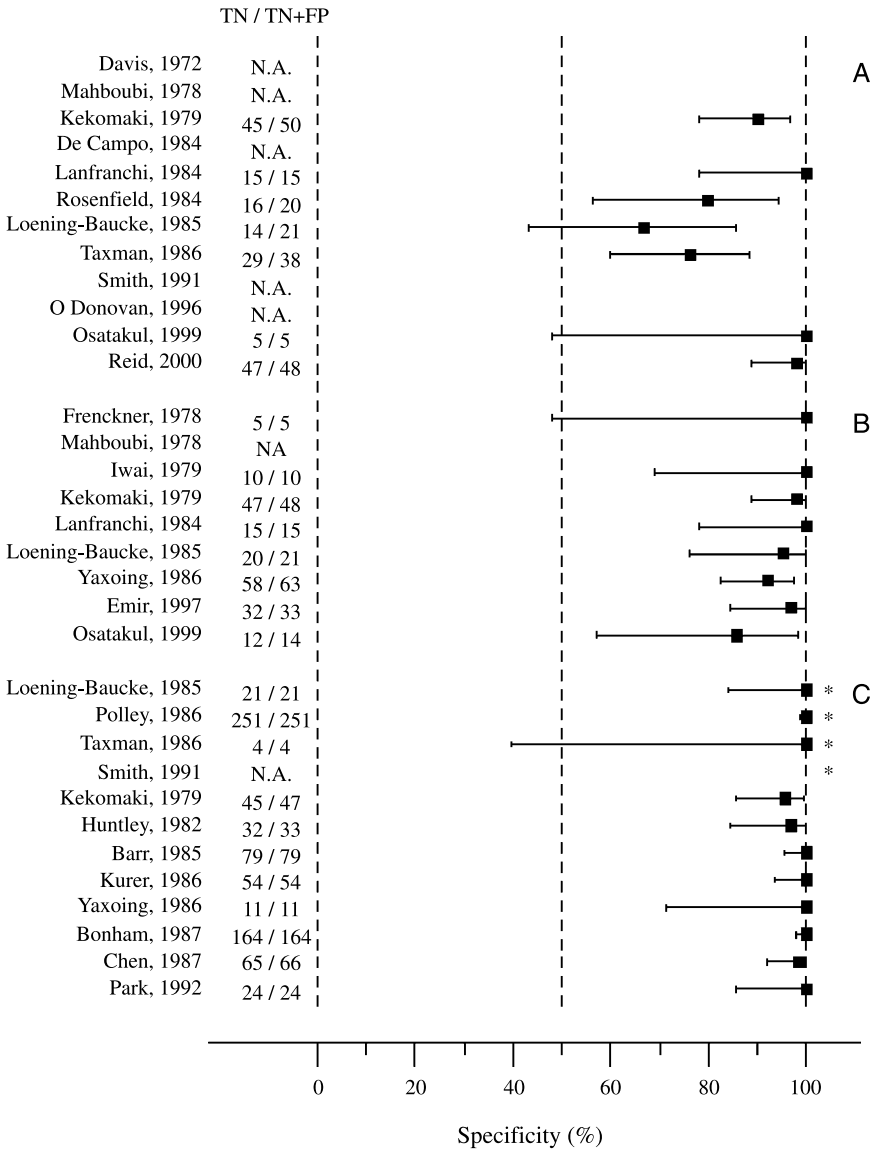

FIG. 2. Forrest plot of specificity in studies examining the accuracy of CE (A), ARM (B), and RSB (C). TN = true-negative test result; FP = false-positive test result; NA = not applicable, case series that only included patients with Hirschsprung; asterisk (*) = RSB studies stained with H&E.

evaluating CE (mean sensitivity: low-quality studies, 86%; high-quality studies, 70%), but this difference was not statistically significant.

A subgroup analysis was performed to evaluate the value of ARM in infants younger than 6 months. Only 4 studies in which the age of the infants studied was specified were available. In these studies, the mean sensitivity of ARM was 88% (95% CI, 66%–96%), whereas the mean specificity was 89% (95% CI, 57%–98%) in infants younger than 6 months. This result was not significantly different compared with the total group of children who underwent an ARM (mean sensitivity, 91%; mean specificity, 94%).

Inconclusive tests were most frequently seen in studies evaluating RSB; 4 studies showed 16 inconclusive tests, and the number of inconclusive tests was unclear in 2 studies. ARM showed 4 inconclusive tests in 2 studies and CE did not show inconclusive tests.

DISCUSSION

Our results show that RSB stained for AChE activity is the most accurate test in the diagnostic workup of HD. Mean sensitivity of ARM was similar to that of RSB, but its specificity was significantly lower. Mean

TABLE 4. *Summary estimates for sensitivity and specificity from the bivariate model*

Index test	Parameter	
	Mean sensitivity, % (95% CI)	Mean specificity, % (95% CI)
CE	70 (64–76)	83 (74–90)
ARM	91 (85–95)	94 (89–97)
RSB	93 (88–95)	98 (95–99)
P-value CE vs ARM	<0.0001	0.014
P-value CE vs RSB	<0.0001	<0.0001
P-value ARM vs RSB	0.730	0.085

Comparison between CE, ARM and RSB stained for AChE activity.

sensitivity and specificity of CE were significantly lower than those of RSB and ARM.

To our knowledge, this is the first systematic review evaluating the accuracy of tests for the diagnosis of HD. Because it is not clear what test is the most accurate in the diagnostic workup of HD, different approaches are used in different hospitals. Only 1 report in this systematic review compared all 3 possible tests in a single study (18). Therefore, we used indirect comparisons to examine the differences in sensitivities and specificities between the 3 tests.

All studies included in this systematic review had their limitations. The way patients were selected was poorly described in all studies. Therefore, it was unclear whether all consecutive series of patients suspected of having HD were included or not. Furthermore, a description of the clinical symptoms of patients suspected of having HD was lacking. For most studies, it was unclear whether the readers of the index tests were blinded to the results of the other index tests or reference standard. Another drawback is that all studies provided sparse details about the thoroughness of their clinical follow-up. Essential information, like the duration of follow-up, the number of patients lost to follow-up and the way in which patients were contacted, were missing. All studies retrospectively collected their data. The disadvantages of collecting data retrospectively are the uncertainty about the inclusion criteria, the nonuniformity of data collection, the evaluation of known test results and the quality of follow-up being limited. In almost all studies (23 studies) included in this systematic review, an FTB was performed in those patients with a positive-index test. However, several studies did not verify the patients with negative test results and classified these patients as true negatives. Patients with a false-negative test result may stay undetected if follow-up is not adequate or if patients with persistent symptoms return to a different hospital. In that case, both sensitivity and specificity will be overestimated, although the relative effect will be greater on sensitivity because of the lower number of patients with HD.

In this systematic review, CE had the lowest sensitivity compared with RSB and ARM (70% vs 93% and 91%, respectively), indicating that false-negative test results were more common in CE. Several explanations have been described for false-negative results of CE. Especially in young infants suspected of having HD, a transition zone is difficult to demonstrate (19). Furthermore, De Campo et al. (15) demonstrated a normal-caliber colon in 75% of children with total aganglionosis. Rectal washouts and even digital rectal examinations may also lead to false-negative test results (34). RSB and ARM showed false-negative test results in rare cases. It has been suggested that false-negative test results of ARM occur because of the displacement of the transducer probe or as a consequence of relaxation of the external anal sphincter rather than the internal anal sphincter (35). The age at which ARM was performed played no significant role with respect to sensitivity and specificity. However, the number of studies included (n = 4) and the number of patients included in the studies (n = 104) were low. In addition, both sensitivity and specificity showed a wide CI. The possible causes for false-negative test results in RSB are the variability in the biopsy site (correct site is 2–3 cm above the mucocutaneous junction), biopsy material that is too superficial and does not contain muscularis mucosa, biopsy material with bleeding artefacts, the technical variations in performance of the stain and the experience of individual pathologists. Furthermore, a

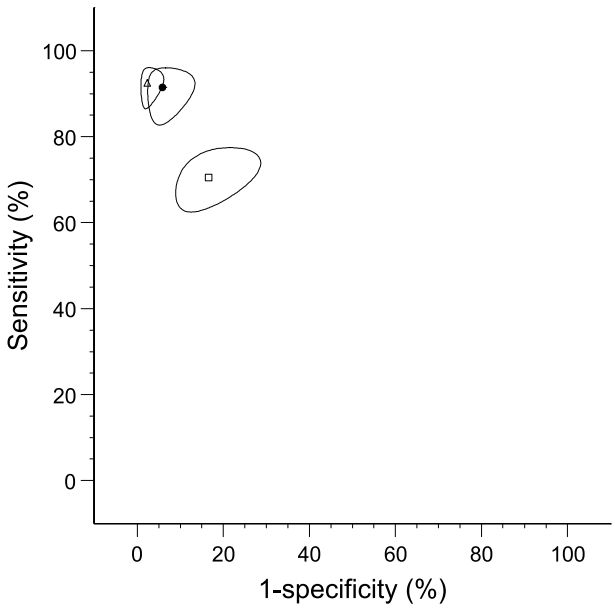

FIG. 3. Bivariate random summary estimates of sensitivity and specificity for each of the 3 index tests (square = CE; dot = ARM; triangle = RSB stained for AChE activity) and the corresponding 95% confidence ellipse around these means.

false-negative reaction may occur in neonates, possibly because of the immaturity of the enzyme system (34).

A high specificity is also desirable for tests in the workup of patients suspected of having HD; otherwise, too many patients unnecessarily have to undergo the invasive definitive test because of false-positive test results. False-positive test results were also most frequent in CE. Rosenfield et al. (17) described 4 newborn infants with a meconium plug syndrome leading to a transition zone at the splenic flexure, mimicking the finding in HD. Other CE studies did not provide a good explanation for the occurrence of false-positive test results. The literature contains conflicting views whether the accuracy of ARM is lower in neonates. Some studies suggest that failure of the reflex might be caused by the physiological immaturity of anorectal function (36,37). Nevertheless, we demonstrated that term and premature infants older than 26 weeks postmenstrual age have a well-developed RAIR upon rectal distension (38,39). The false-positive test results in ARM might be caused by technical problems, such as an air leak in the circuit or insufficient inflation of the balloon (24). False-positive results in RSB are rare. Only 3 studies showed patients with false-positive test results (28,14,32). The reason for this is unclear.

The highest frequency of inconclusive tests results were seen in studies evaluating RSB; 4 of the 14 studies showed 16 inconclusive tests. The explanation is that biopsies are taken too superficially (containing insufficient muscularis mucosa) for proper evaluation. In contrast, CE did not show inconclusive tests.

In this systematic review, CE showed a wide range in sensitivity and specificity. The wide range in sensitivity and specificity values might be due to difference in quality of studies and/or variation due to chance because of the low number of included patients. CE showed 1 outlier with a sensitivity of only 15% (15). The reason of this low sensitivity can be the inclusion of only 13 patients, all with total aganglionosis, which often leads to false-negative test results on CE. Specificity of CE studies showed also a wide range and could be calculated in only 7 studies. The range of sensitivity and specificity in RSB studies is small. This is also true for the sensitivity and specificity in ARM studies.

In conclusion, this systematic review shows that RSB stained for AChE activity and ARM are the most accurate tests in the diagnostic workup of patients suspected of having HD. ARM is a noninvasive diagnostic test and is easy to perform in children older than 1 year. It therefore has often been suggested as an ideal screening tool. However, the equipment is expensive, and extensive experience is necessary to perform this procedure to evaluate the test results in infants younger than 1 year. In contrast, an RSB is a relatively simple, efficient and, mostly, an incident-free procedure, although rectal bleeding, perforation or sepsis caused by RSB has been described (19,40). Because verification of index tests was not optimal in all studies included in this systematic review and because all studies showed methodological limitations, the value of the accuracy of these tests need to be confirmed. Large blinded prospective studies with well-defined inclusion criteria and long-term clinical follow-up should be performed to evaluate the accuracy of the different diagnostic tests simultaneously in infants suspected of having HD.

REFERENCES

1. Loening-Baucke V. Chronic constipation in children. *Gastroenterology* 1993;105(5):1557–64. [Review, 61 refs].
2. Puri P. Hirschsprung's disease: clinical generalities. In: Holschneider AM, Puri P, eds. *Hirschsprung's Disease and Allied Disorders*, 2nd ed, 2000:129–35.
3. Ikeda K, Goto S. Diagnosis and treatment of Hirschsprung's disease in Japan. An analysis of 1628 patients. *Ann Surg* Apr 1984;199(4):400–5.
4. Kleinhaus S, Boley SJ, Sheran M, Sieber WK. Hirschsprung's disease—a survey of the members of the Surgical Section of the American Academy of Pediatrics. *J Pediatr Surg* Oct 1979;14(5): 588–97.
5. Kurer MH, Lawson JO, Pambakian H. Suction biopsy in Hirschsprung's disease. *Arch Dis Child* Jan 1986;61(1):83–4.
6. Emir H, Akman M, Sarimurat N, et al. Anorectal manometry during the neonatal period: its specificity in the diagnosis of Hirschsprung's disease. *Eur J Pediatr Surg* Apr 1999;9(2): 101–3.
7. Osatakul S, Patrapinyokul S, Osatakul N. The diagnostic value of anorectal manometry as a screening test for Hirschsprung's disease. *J Med Assoc Thai* Nov 1999;82(11):1100–5.
8. De Campo JF, Mayne V, Boldt DW, et al. Radiological findings in total aganglionosis coli. *Pediatr Radiol* 1984;14(4):205–9.
9. Bossuyt P, Reitsma J, Bruns D, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* Jan 2003;7(138):w1–w12.
10. van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Stat Med* 1993;12:2273–84.
11. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21:589–624.
12. Davis PW, Foster DB. Hirschsprung's disease. A clinical review. *Br J Surg* Jan 1972;59(1):19–26.
13. Mahboubi S, Schnaufer L. The barium-enema examination and rectal manometry in Hirschsprung disease. *Radiology* Mar 1978;130(3):643–7.
14. Kekomaki M, Rapola J, Louhimo I. Diagnosis of Hirschsprung's disease. *Acta Paediatr Scand* Nov 1979;68(6):893–7.
15. De Campo JF, Mayne V, Boldt DW, et al. Radiological findings in total aganglionosis coli. *Pediatr Radiol* 1984;14(4):205–9.
16. Lanfranchi GA, Bazzocchi G, Federici S, et al. Anorectal manometry in the diagnosis of Hirschsprung's disease—comparison with clinical and radiological criteria. *Am J Gastroenterol* Apr 1984;79(4):270–5.
17. Rosenfield NS, Ablow RC, Markowitz RI, et al. Hirschsprung disease: accuracy of the barium enema examination. *Radiology* Feb 1984;150(2):393–400.
18. Loening-Baucke V, Pringle KC, Ekwo EE. Anorectal manometry for the exclusion of Hirschsprung's disease in neonates. *J Pediatr Gastroenterol Nutr* Aug 1985;4(4):596–603.
19. Taxman TL, Yulish BS, Rothstein FC. How useful is the barium enema in the diagnosis of infantile Hirschsprung's disease? *Am J Dis Child* Sep 1986;140(9):881–4.
20. Smith GHH, Cass D. Infantile Hirschsprung's disease—is a barium enema useful? *Pediatr Surg Int* 1991;6(4–5):318–21.

21. O'Donovan AN, Habra G, Somers S, et al. Diagnosis of Hirschsprung's disease. *AJR Am J Roentgenol* Aug 1996;167(2): 517–20.
22. Reid JR, Buonomo C, Moreira C, et al. The barium enema in constipation: comparison with rectal manometry and biopsy to exclude Hirschsprung's disease after the neonatal period. *Pediatr Radiol* 2000;30(10):681–4.
23. Frenckner B. Ano-rectal manometry in the diagnosis of Hirschsprung's disease in infants. *Acta Paediatr Scand* 1978;67(2):187–92.
24. Iwai N, Ogita S, Kida M, et al. A manometric assessment of anorectal pressures and its significance in the diagnosis of Hirschsprung's disease and idiopathic megacolon. *Jpn J Surg* Sep 1979;9(3):234–40.
25. Yaxiong S, Chengren S. Clinical evaluation of diagnostic methods for Hirschsprung's disease. *Pediatr Surg Int* 1986;1(4):218–22.
26. Campbell PE, Noblett HR. Experience with rectal suction biopsy in the diagnosis of Hirschsprung's disease. *J Pediatr Surg* Aug 1969;4(4):410–5.
27. Hamoudi AB, Reiner CB, Boles ET Jr, et al. Acetylthiocholinesterase staining activity of rectal mucosa. Its use in the diagnosis of Hirschsprung's disease. *Arch Pathol Lab Med* Dec 1982;106(3): 670–2.
28. Huntley CC, Shaffner LD, Challa VR, et al. Histochemical diagnosis of Hirschsprung disease. *Pediatrics* Jun 1982;69(6):755–61.
29. Barr LC, Booth J, Filipe MI, et al. Clinical evaluation of the histochemical diagnosis of Hirschsprung's disease. *Gut* Apr 1985;26(4):393–9.
30. Polley TZ Jr, Coran AG, Heidelberger KP, et al. Suction rectal biopsy in the diagnosis of Hirschsprung's disease and chronic constipation. *Pediatr Surg Int* 1986;1(2):84–9.
31. Bonham JR, Dale G, Scott DJ, et al. A 7-year study of the diagnostic value of rectal mucosal acetylcholinesterase measurement in Hirschsprung's disease. *J Pediatr Surg* Feb 1987;22(2): 150–2.
32. Chen CL, Hsu HC, Chen CC, et al. Acetylcholinesterase activity in rectal suction biopsy for the diagnosis of Hirschsprung's disease. *Taiwan yi zhi* Jul 1987;86(7):723–7.
33. Park WH, Choi SO, Kwon KY, et al. Acetylcholinesterase histochemistry of rectal suction biopsies in the diagnosis of Hirschsprung's disease. *J Korean Med Sci* Dec 1992;7(4):353–9.
34. Blake NS. *Diagnosis of Hirschsprung's Disease and Allied Disorders. Hirschsprung's Disease and Allied Disorders*, 2 ed, 2000:223–90.
35. Aaronson I, Nixon HH. A clinical evaluation of anorectal pressure studies in the diagnosis of Hirschsprung's disease. *Gut* 1972;13:138–46.
36. Holschneider AM, Kellner E, Streibl P, et al. The development of anorectal continence and its significance in the diagnosis of Hirschsprung's disease. *J Pediatr Surg* Apr 1976;11(2):151–6.
37. Ito Y, Donahoe PK, Hendren WH. Maturation of the rectoanal response in premature and perinatal infants. *J Pediatr Surg* Jun 1977;12(3):477–82.
38. Benninga MA, Omari TI, Haslam RR, et al. Characterization of anorectal pressure and the anorectal inhibitory reflex in healthy preterm and term infants. *J Pediatr* Aug 2001;139(2):233–7.
39. de Lorijn F, Omari TI, Kok JH, et al. Maturation of the rectoanal inhibitory reflex in very premature infants. *J Pediatr* Nov 2003;143(5):630–3.
40. Rees BI, Azmy A, Nigam M, et al. Complications of rectal suction biopsy. *J Pediatr Surg* 1983;18(3):273–5.