

Interobserver variability in antroduodenal manometry

F. L. CONNOR,* P. E. HYMAN,† C. FAURE,‡ T. TOMOMASA,§ N. PEHLIVANOV,† J. JANOSKY,¶ C. RUDOLPH,** O. LIEM†† & C. DI LORENZO††

*Department of Gastroenterology, Hepatology and Nutrition, Royal Children's Hospital, Brisbane, Australia

†Department of Pediatric Gastroenterology, Kansas University Children's Center, Kansas City, KS, USA

‡Pediatric Gastroenterology Unit, Sainte Justine Hospital, Montreal, QC, Canada

§Department of Pediatrics at Gunma University Faculty of Medicine, Maebashi, Gunma, Japan

¶Division of Gastroenterology, Children's Hospital of Pittsburgh, Pittsburgh, PA, USA

**Section of Pediatric Gastroenterology, Medical College and Children's Hospital of Wisconsin, Milwaukee, WI, USA

††Division of Pediatric Gastroenterology, Nationwide Children's Hospital, Columbus, OH, USA

Abstract Interobserver variability affects investigations involving assessment of complex visual data, such as histopathology, radiology and motility. This study assessed interobserver variation for interpretation of antroduodenal manometry (ADM), as this has not been previously investigated. Thirty-five ADM recordings from children aged 0.3–18 years were independently evaluated by five experienced paediatric gastroenterologists who were blinded to cases' clinical histories. Intra-class correlation (ICC) was analysed for detection and measurement of phase three of the migrating motor complex (MMC) and Cohen's kappa statistic was calculated between observer pairs for detection of specific motility features and final diagnosis. Observers were unanimous on the differentiation of normal and abnormal motility in 63% of cases. There was excellent interobserver agreement for the number of phase three of the MMC in fasting (ICC = 0.82, $P < 0.0001$) and for measurements of phase three of the MMC (ICC = 0.9999, $P < 0.0001$). Detection of other normal and abnormal motility patterns varied more. Objective findings such as the presence of phase three of the MMC correlated more closely than findings that involved the integration of several variables, such as final diagnosis. However, these data overall indicate that agreement between expert observers for the distinction of normal

and abnormal antroduodenal motility compares favourably with other standard medical assessments.

Keywords antroduodenal manometry, children, functional gastrointestinal disorders, interobserver variation, intra-class correlation, kappa values.

Abbreviations: ADM, antroduodenal manometry; IBD, inflammatory bowel disease.

INTRODUCTION

Diagnostic criteria for paediatric antroduodenal manometry (ADM) studies have been published,^{1–3} but the impact of these criteria on interobserver variability in interpretation of these studies has not been assessed. In other disciplines involving the interpretation of complex visual data, such as radiology and histopathology, identification of areas of maximal interobserver variation has led to modification of diagnostic criteria and increased reliability.^{4–9} The objectives of this study were to assess the degree of inter-observer variation in the interpretation of paediatric ADM studies and to identify sources of inter-observer variation to provide recommendations for future guidelines on interpretation. We hypothesized that significant interobserver variation occurs in the interpretation of such studies and that the variability will be less for objective criteria, such as the presence of phase III of the migrating motor complex (MMC) than for subjective assessments involving the integration of objective information such as final diagnosis.

MATERIALS AND METHODS

This study was approved by the human rights committee of all participating hospitals. Water perfused ADM

Address for correspondence

Frances Connor, MBBS, FRACP, Department of Gastroenterology, Hepatology and Nutrition Royal Children's Hospital, Herston Rd, Herston, Qld 4029, Australia.

Tel: +61 7 3852-1332; fax: +61 7 3666 0250;

e-mail: flconnor@tpgi.com.au

Received: 27 April 2008

Accepted for publication: 13 May 2008

recordings from 35 children (19 female, age range 4 months to 18 years, median 6 years,) previously studied for clinical indications were selected by research assistants in participating hospitals to represent a variety of motility abnormalities encountered in clinical practice. The de-identified recordings were independently evaluated by five paediatric gastroenterologists, experienced in the interpretation of paediatric ADM, who were blinded to cases' clinical histories. Original clinicians' motility diagnoses for the patients, incorporating clinical history, physical examination, results of ADM and other investigations, were: normal ($n = 8$), intestinal myopathy ($n = 5$), intestinal neuropathy

($n = 10$), rumination ($n = 3$), postprandial hypomotility ($n = 6$), low amplitudes due to duodenal dilatation ($n = 2$) and non-specific abnormalities ($n = 6$). Some cases had more than one diagnosis. All recordings conformed to published minimum standards for paediatric manometry.¹⁰ Specifically, all included at least one recording channel in the gastric antrum and three in the small intestine at all times. After the test meal, retrograde catheter migration occurred in 16 cases, placing between one and three additional channels in the antrum or gastric body. Recordings had five ($n = 1$), six ($n = 5$), seven ($n = 3$) or eight ($n = 26$) channels. Each study included at least 3 h of fasting and one postprandial hour. Median total duration was 319 min (range 250–500 min). Drugs were administered in 20 cases; erythromycin was used in all 20 cases, octreotide in two cases and cisapride in one case.

Clinicians were provided with patient age, study duration, relevant events including fasting, feeding, medication and symptoms during the study and catheter details. Composition of test meals was provided. In 27 cases, the test meal provided at least 10 kcal kg⁻¹ and at least 30% of calories as fat calories. In the remainder, caloric intake was limited by the patients' clinical condition. Diagnostic criteria for normal and abnormal motility, including specific abnormal patterns were not provided.

Recordings were coded for 20 manometric features (Table 1). Categories were based upon published criteria^{1–3} and included final diagnosis (Table 2), and presence and characteristics of the phase III of the MMC. For each item, the option of 'judgment not possible from this recording' was available.

Statistics

To determine sample size, the level of agreement between two comparable raters was estimated to be

Table 1 Manometric features coded in report form

Phase III of the migrating motor complex	
Presence in fasting trace and number found	
Presence in postprandial trace and number found	
Presence after drug provocation and number found	
Entirely normal or present with abnormalities	
Amplitude	
Migration	
Interval	
Presence of abnormal patterns	
Sustained tonic-phasic contractions	
Postprandial non-migrating clusters	
Other abnormalities found	
Postprandial response	
Presence of a change to postprandial motility pattern (interruption of MMC by meal)	
Postprandial motility, gastric antrum (normal, reduced, increased)	
Postprandial motility, duodenum (normal, reduced, increased)	
Overall Assessment	
Motility diagnosis possible on recording provided	
Motility normal or not	
Specific motility diagnosis	
Severity of motility abnormality	
Comments	

Table 2 Manometric features of motility disorders

Motility disorders	Main characteristics
Intestinal myopathy	Persistently low-amplitude (<20 mmHg), coordinated contractions in the absence of dilated bowel
Intestinal neuropathy	Normal amplitude but abnormal patterns of contractions
Rumination	Pattern of brief, simultaneous pressure increase in all recording sites associated with regurgitation
Postprandial hypomotility	Motility index <600 mmHg/30 min after ingestion of meal
Failure to induce fed pattern	Occurrence of phase III of the MMC within 2 h after ingestion of meal
Mechanical intestinal obstruction	Prolonged simultaneous non-propagating contractions Postprandial clustered contractions lasting >30 min Neonates: High amplitude retrograde propagating contractions

0.80. To detect this level of agreement with an error bound of 5% and a power of 0.80, 35 cases must be assessed by each of the five paediatric gastroenterologists with a comparable level of experience. These values provided for adequate precision and power for detecting agreement between pairs of raters and an overall generalizability estimate of agreement. The level of precision and power was lower for different types of cases due to sub-grouping.

For the assessment of agreement for continuous variables, intra-class correlation (ICC) coefficients were calculated. For the assessment of pairwise rater agreement for categorical variables (between clinicians A and B, B and C, etc), Cohen's kappa values were calculated. Kappa values above 0.4 indicate good agreement; above 0.75 denotes excellent agreement.¹¹ Other categorical data were compared with chi-squared tests. For non-parametric continuous numeric variables, Mann-Whitney *U*-test or Wilcoxon signed rank sum tests were used. For the assessment of relationships for continuous variables, Spearman's correlation coefficient was calculated. For the overall assessment of agreement (i.e. all five raters concomitantly), a generalizability coefficient was calculated. A *P*-value of <0.05 was taken as significant.

RESULTS

Agreement on major categories

Kappa tests for all categories are shown in Table 3. Detection of phase III ranked best, followed by differentiation of normal and abnormal motility, followed by

Table 3 Kappa values: pairwise interobserver agreement

Value	Median	IQR	Range
Normal vs abnormal motility	0.5685	0.134	0.69
Detection of phase III	0.7065	0.1527	0.44
Detection of change to postprandial motility pattern	0.478	0.099	0.217
Abnormal patterns	0.2715	0.1877	0.71
Diagnosis	0.44	0.3197	0.9
Normal motility	0.5685	0.134	0.691
Myopathy	0.4515	0.28525	0.541
Neuropathy	0.267	0.163	0.356
Phase III normal	0.5255	0.2875	0.752
Phase III amplitude	0.6015	0.28575	0.816
Phase III interval	0.6	0.284	1.154
Phase III migration	0.3185	0.1475	0.611
Postprandial antral hypomotility	0.2975	0.5115	0.701
Postprandial duodenal hypomotility	0.4245	0.38025	0.609

diagnosis, meal response and detection of abnormal patterns.

Motility diagnosis

When cases of 'judgment not possible' were excluded, observers were unanimous on the differentiation of normal and abnormal motility in 63% of cases. When 'judgment not possible' was included as a diagnostic category, this agreement was 46%. The 'judgment not possible' option was used for motility diagnosis up to nine times (mean 3.6) by the observers. Observers were unwilling or reluctant to assign diagnoses in many cases due to lack of clinical information or concerns about technical aspects of the studies. Vascular compression of the manometry catheter was suspected in three (9%) cases. Problems with catheter position or connections were queried in six (17%) cases. The median agreement for the distinction between normal and abnormal was $K = 0.57$. Levels of agreement varied between observer pairs, ranging from $K = 0.92$ ($P < 0.001$) for observers A and B to $K = 0.2$ ($P = \text{ns}$) for observers C and D. Levels of agreement for specific diagnoses are indicated in Table 3. After normal motility, the highest agreement was for the diagnosis of intestinal myopathy, median $K = 0.45$ (range: $K = 0.21$, $P = 0.055$ to $K = 0.75$, $P < 0.001$). In the case where well-organized, low-amplitude phase III of the MMC were clearly visible, agreement on the diagnosis of myopathy was universal. In other cases, observers noted that either myopathy or severe neuropathy could produce similar manometric appearances. In the case of combined myopathy and neuropathy, agreement was less (17% of observer pairs) than in the four cases of pure myopathy (58% observer pairs). As in many other instances, observers were reluctant to make a diagnosis without clinical information. Only three of the five observers diagnosed rumination, highlighting the need for patient's observation during manometry to diagnose this condition. For the observers who made this diagnosis, agreements ranged from $K = 0.356$ ($P = 0.03$) to $K = 0.839$ ($P < 0.001$), median 0.37. There were insufficient cases assigned to other diagnostic categories for kappa analysis. Intraobserver agreement was assessed by comparing the original clinician's diagnosis (incorporating clinical history, examination and other investigations) with the same clinician's findings when examining the de-identified files used in the study. Original clinician's differentiation of normal vs abnormal was reproduced in 62.5–100% of cases (kappa=ns due to insufficient cases). Using logistic

regression analysis, the only factor independently related to consensus for the diagnosis of normal vs abnormal motility was agreement on the presence of at least one phase III of the MMC. Therefore, factors affecting detection of phase III were examined in detail.

Fasting motility

Agreements for characteristics of the MMC are shown in Table 3. Because only 13 recordings had two or more phase III of the MMC, case numbers were insufficient to determine interobserver agreement for characterization of the interval between phase III in some observer pairs. Identification of reduced contraction amplitude agreed well, with $K = 0.41$ ($P = 0.04$) to $K = 1.0$ ($P < 0.001$), median $K = 0.65$. Generalizability coefficient summarizing kappa values for determination of amplitude (normal or reduced) was 0.60, $P = 0.0003$. In contrast, there was marked variation in classification of migration of phase III, with $K = 0.04$ ($P = 0.86$) to 0.65 ($P = 0.001$), median $K = 0.32$. The generalizability coefficient summarizing kappa values for determination of amplitude, migration and interval for phase III of the MMC was 0.53, $P = 0.0019$.

Because the majority of recordings demonstrated abnormal motility, characterization of phase III of the MMC was sometimes difficult. Agreement was influenced by whether or not observers classified groups of contractions as phase III or merely as clusters. Examining only the 44 (57%) phase IIIs that all observers agreed were present, agreement for all characteristics was close to perfect, although subgroups were often too small to perform kappa analysis.

Motility response to meal

Agreements for assessment of motility response to feeding are shown in Table 3. All agreed on change to postprandial motility pattern in 63% of relevant cases ($n = 27$ with meal stimulus adequate to induce fed pattern), excluding 'judgment not possible' or 41% of cases if this was included as a separate category. However, agreement for assessment of postprandial hypomotility was highly variable. For antral postprandial hypomotility, values ranged from $K = 0.02$ ($P = \text{ns}$) to $K = 0.72$ ($P < 0.001$), median $K = 0.30$. Where two antral sensors were present ($n = 5$), 86% of observer pairs agreed vs 67% for cases with one antral sensor ($P = \text{ns}$). For duodenal postprandial hypomotility, agreement ranged from $K = 0.02$ ($P = \text{ns}$) to $K = 0.63$ ($P = 0.001$), median $K = 0.42$. The generalizability coefficient

summarizing kappa values for postprandial hypomotility was 0.327, $P = 0.068$.

Abnormal patterns

The presence or absence of sustained tonic-phasic contractions was agreed in 60% and postprandial non-migrating clusters in 57% of cases. Authors most commonly agreed when abnormal patterns were absent. Agreements ranged from $K = 0.06$ ($P = \text{ns}$) to $K = 0.74$ ($P < 0.001$), median 0.25 for sustained tonic-phasic contractions to $K = 0.09$ ($P = \text{NS}$) to 0.59 ($P = 0.002$), median 0.29 for postprandial non-migrating clusters.

Kappa testing between observer pairs

Kappa values for all variables assessed were combined into an overall mean kappa value for each observer pair. Mean overall kappa values were 0.47, 0.48, 0.47, 0.38 and 0.47 for observer pairs involving observers A, B, C, D and E, respectively. Mean overall kappa values for observer pairs were not significantly different, except for those pairs including observer D. Mean kappa values for all variables for pairs including observer D were significantly lower than those for pairs not involving this observer ($P \leq 0.003$ for all pairs).

Detection of phase III of the MMC

Interobserver agreement for the number of phase III of the MMC during fasting was excellent (ICC = 0.82, $P < 0.0001$). Intra-class correlations for number of phase III of the MMC after drug provocation and in the postprandial period were 0.51 ($P < 0.0001$) and 0.67 ($P < 0.0001$), respectively. For the overall detection of phase III, ICC was 0.75 ($P < 0.0001$). Kappa values for presence of phase III were also determined, to compare agreement for phase III detection with agreement in other categories subsequently. Kappa values for presence of phase III in fasting, postprandial and post-drug periods were 0.69 (median, range: 0.57–0.78), 0.74 (0.67–0.80) and 0.64 (0.41–0.85), respectively. The generalizability coefficient summarizing detection of phase III in these periods among all observers was 0.69, $P = 0.0001$.

There were 77 phase III of the MMCs measured by at least one observer. Fifty-seven per cent of the phase IIIs were marked by all five observers, 13% by four, 12% by three, 1% by two and 17% by one observer.

Because of the importance of agreement on phase III of the MMC as a predictor of consensus for overall

Table 4 Factors independently associated with agreement for detection of phase III of the migrating motor complex (MMC)

Feature	Agreed	Disagreed	P*
Duration (mean in min)	5.4	3.9	<0.001
Number of channels	5.7	4.5	<0.001
Absence of phase II of the MMC	35/220 (16%)	5/82 (6%)	<0.001
Low amplitude	12/44 (27%)	17/33 (52%)	<0.001
Continuous adjacent activity	10/44 (23%)	17/33 (52%)	<0.001
Emesis on same page	20/220 (9%)	19/83 (23%)	<0.001

*Logistic regression analysis.

diagnosis, factors influencing phase III detection were assessed. Examining phase IIIs marked on tracings indicated that some clusters of contractions marked as phase III by some observers, but were not classified as phase III by other observers. For instance, some observers classified an 'interrupted' MMC (described in some cases of pseudo-obstruction and in myotonic dystrophy) as a single phase III of the MMC, whereas others counted two phase III of the MMCs.

When traces were analysed for possible factors contributing to the disparity in numbers of phase III of the MMC detected, several factors emerged as significant ($P < 0.05$) in logistic regression analyses (Table 4). Factors associated with increased agreement on phase III were longer duration of phase III, presence of phase III in more channels, and the absence of phase II, that is, a quiescent tracing prior to phase III. Factors associated with greater disagreement on the presence of phase IIIs included low-amplitude contractions, continuous adjacent manometric activity in one or more channels and emesis on the same page as phase III.

Measurements of phase III of the MMC

Measurements of the 44 phase IIIs marked by all observers were used to calculate ICC for both start and finish times. This yielded 440 start and finish times in up to eight channels or 1259 measurements for both start and finish. Agreement was extremely strong, with ICC = 0.9999, $P < 0.0001$ for both start and finish.

DISCUSSION

The Research Agenda for Pediatric Gastroenterology, Hepatology and Nutrition notes that diagnostic tests and techniques used to evaluate motility disorders and functional gastrointestinal (GI) disorders would be more widely accepted if methods of interpretation were validated.¹² The aim of this study was to assess paediatric ADM by gauging existing levels of observer

agreement. By pinpointing factors associated with interobserver variation, this study identifies areas where variation may be possibly reduced.

Previously, interobserver variation in GI manometry has been studied incidentally during attempts to develop computer programs for the analysis of motility data. In this context, computer programs have been evaluated against the combined responses of teams of trained observers.^{13–17} Alternatively, assessment of observer characterization of motility traces has been used in attempts to generate 'learning sets' from which computer pattern recognition software can be devised.¹⁸ Interobserver variation has also been assessed in the evaluation of pull through techniques for the measurement of lower oesophageal sphincter pressure.^{19,20} Interlaboratory differences have been noted in the analysis of ADM with respect to the identification of 'late phase III' of the MMC.²¹ However, few studies have set out with the primary aim of assessing interobserver variability in the assessment of manometric data.^{22–25}

Previous studies of interobserver variability in GI manometry assessed *detection* of contractile events, such as antroduodenal contractions or transient lower oesophageal sphincter relaxations. No previous study has addressed *pattern recognition, interpretation of findings* or *final diagnosis*.

In regard to *detection* of contractile events, Anderson et al¹⁸ analysed detection of contractions by five observers in ADM and found 60% agreement overall (range 72–97% between observer pairs). They did not assess observers' measurement of contractions, pattern recognition or final diagnosis. The agreement for detection of phase III of the MMC found in this study among five observers is consistent with the above.

In our study, factors which influenced the detection of phase III of the MMC included low amplitude, brief duration, limited propagation, and artefacts such as continuous manometric activity. For many of these factors, increased awareness is likely to reduce variation. For others, such as duration of phase III, extent of propagation and approach to interrupted phase IIIs,

consensus is required on a uniform approach. Variability in the detection of low-amplitude phase IIIs was exaggerated by the study design as only paper tracings were provided, preventing observers from expanding the amplitude (y) axis to enhance visibility.

In general, interobserver variation is greater for *detection* of events than for *measurement* of defined contractile events as shown in a study of interobserver variability in measurement of mean lower oesophageal sphincter pressure by Van Herwaarden *et al.*¹⁶ Similarly, in this study correlation was even better for measurement of phase III of the MMC than for its detection.

In regard to postprandial hypomotility, there was a trend to more disagreement in cases where meal stimulus was regarded as inadequate. In normal children, duodenal motility index doubles for 40 min after a meal.³ However, motility index is generally regarded

as a research tool and is not usually calculated in clinical studies. In clinical paediatric practice, frequent artefacts from movement and straining often prevent reliable computerized calculation of motility index. For these reasons, postprandial motility was assessed visually in the usual manner for clinical studies. The variability encountered in the assessment of postprandial response suggests that when possible motility index should be calculated.

The interobserver variability for detection of the abnormal patterns, sustained tonic-phasic contractions and postprandial non-migrating clusters was greater than expected, given that both patterns have been associated with abnormal motility in children.^{1,2} The sustained tonic-phasic contraction was defined as a cluster of contractions which had tonic components and lasted more than 10 min in only one recording site, with normal motility patterns at the other sites.²

Table 5 Comparison of interobserver variability in clinical assessments and investigations in gastroenterology

Clinical assessment or condition	Aspect analysed for interobserver agreement	Kappa value	
Abdominal X-rays in emergency department ³²	Normal vs abnormal	K = 0.5–0.78	
Antroduodenal motility (current study)	Antroduodenal manometry: normal vs abnormal	K = 0.2–0.92 (median 0.57)	
Reflux esophagitis ³³	Endoscopic grading: severity of oesophagitis using Los Angeles scale	K = 0.56	
Functional gastrointestinal disorders ³⁴	Reliability of the paediatric Rome II criteria	K = 0.37	Specialist
		K = 0.41	Trainees
		K = 0.37	Physicians with special interest
		K = 0.38	Other specialists
		K = 0.2	constipation
		K = 0.3	pain
Inflammatory bowel disease ³⁵	Colonoscopic diagnosis of inflammatory bowel disease (video recorded studies)	K = 0.375–1	
Inflammatory bowel disease ³⁶	Histological diagnosis of inflammatory bowel disease	K = 0.47	Normal
		K = 0.43	Inflammatory bowel disease
		K = 0.2	Crohn's disease
		K = 0.19	Non-specific inflammation
Inflammatory bowel disease ³⁷	Histological diagnosis of inflammatory bowel disease	K = 0.43	Normal
		K = 0.37	Ulcerative colitis
		K = 0.2	Crohn's disease
		K = 0.19	Non-specific
Jaundiced patients ³⁸	Physical examination of jaundiced patients	K = 0.32	Spleen palpable
		K = 0.26	Character of liver edge
Necrotizing enterocolitis ³⁹	Presence of necrotizing enterocolitis on abdominal radiographs	K = 0.31	
Liver histology ⁴⁰	Liver biopsies: final histological diagnosis	K = 0.59	Cirrhosis
		K = 0.14	Acute hepatitis
		K = 0.09	Chronic hepatitis
		K = 0.05	Necrosis
Neuronal intestinal dysplasia ²⁸	Histological diagnosis of neuronal intestinal dysplasia	K = 0.11–0.26	

Kappa values measure interobserver agreement, corrected for chance; kappa=1 indicates 100% agreement.

Similar features have been attributed to artefact due to compression of the manometry catheter in the duodenum by the superior mesenteric artery.²⁶ This study is the first to examine interobserver variability for *diagnosis* in GI manometry. For other diagnostic tests in gastroenterology, agreement is highly variable. In a study of endoscopic ultrasound, agreement for the diagnosis of chronic pancreatitis was moderate ($K = 0.45$).²⁷ In contrast, a landmark study of rectal biopsy for diagnosis of intestinal neuronal dysplasia demonstrated levels of agreement among three experienced histopathologists which were close to zero.²⁸ The median agreement detected in this study for the distinction of normal and abnormal was $K = 0.57$. For the other major diagnostic categories, it was lower (median kappa values for myopathy: $K = 0.45$, neuropathy: $K = 0.25$).

In this study, expert observers were deliberately chosen from various cultural and geographic backgrounds. There was a possible 'center effect', in that agreement between observer pairs was greatest for those practicing in North America. Differences in clinical case load between centers may also have contributed to the observed variation.

The design of this study exaggerated disagreement for diagnosis, as observers were given no information on history, physical examination or results of previous investigations to avoid biasing the observers' assessments of the tracings. As with most medical investigations, these factors are crucial to interpreting test results. On their own, many manometric findings are non-specific.²⁹ In clinical practice, it is highly unlikely that any specific diagnosis is ever made based solely on ADM. In particular, correlation with results of radiology and validated gastric emptying tests is essential for accurate interpretation. Therefore, the resulting variability detected for diagnosis probably overestimates the variability in clinical practice. Importantly, the overall assessment of normal vs abnormal showed good agreement (median kappa 0.57). This compares favourably with other standard medical investigations. For example, a study found that interobserver agreement for electrocardiograms in the diagnosis of acute myocardial infarction was only fair, with $kappa=0.42$.³⁰ In another study, 16 doctors disagreed on the electrocardiographic diagnosis of infarction in 70% of cases.³¹ A summary of studies of the reliability of various tests used in gastroenterology is presented in Table 5.

In conclusion, the overall agreement for distinction of normal and abnormal motility compares well with other standard medical tests. However, interobserver variation remains as a significant issue in the interpretation of ADM. ADM should be interpreted in the

Table 6 Manometric features with a clear association with gastrointestinal motility disorders in children are:

Absence of phase III of the MMC after 4 h fasting recording (95% of normal children have phase III within 4 h fasting study) ^{*2}
Abnormal migration of phase III (absence of aboral migration) ²
Short intervals between phase III* (less than 30 min) ²
Persistent low amplitude contractions (peak amplitude of >90% contractions <20 mmHg) ^{*2}
Sustained tonic-phasic contractions ²
Postprandial hypomotility: motility index in antrum or duodenum <600 mmHg/30 mins in the first 30 min after meal ingestion ⁴¹
High amplitude retrograde propagating contractions ⁴²

*These features have been observed in control children occasionally.

setting of the clinical picture and the findings on ancillary and complementary tests, especially radiology and validated gastric emptying tests. Care should be taken to avoid over-interpretation of manometric findings. Objective findings such as the presence of phase III of the MMC correlate more closely than findings that involve the integration of several variables, such as final diagnosis. Some manometric patterns are non-specific and the poor agreement for final diagnosis likely reflects the absence of clinical data. Manometric findings with a clear association with paediatric GI motility disorders are listed in Table 6. Until comprehensive criteria are generated, validated and universally accepted, clinicians and investigators should document their own criteria for phase III of the MMC and include this information in all publications and motility reports.

More importantly, the true value of a diagnostic test depends upon the ability to discriminate between those patients with disease and those that are disease free, and to predict outcomes. This study does not address these aspects of paediatric ADM. Further validation of the test will require long-term follow-up of patients to determine if outcomes can be determined from ADM. In addition, further efforts to characterize the variations in antroduodenal motility among normal infants and children will be necessary prior to being able to truly determine the predictive value of this diagnostic tool.

ACKNOWLEDGMENTS

Dr Connor's Research fellowship was supported by the University of Pittsburgh and by a Grant-in-Aid for Overseas study from the Gastroenterology Society of

Australia and AstraZeneca, which covers three related projects, including the current study.

REFERENCES

- Tomomasa T. Antroduodenal manometry. In: Hyman P, ed. *Pediatric Gastrointestinal Motility Disorders*. New York: Academy Professional International Services, Inc, 1994: 195–214.
- Tomomasa T, DiLorenzo C, Morikawa A, Uc A, Hyman PE. Analysis of fasting antroduodenal manometry in children. *Digestive Diseases and Sciences* 1996; **41**: 2195–203.
- Uc A, Hoon A, Di Lorenzo C, Hyman PE. Antroduodenal manometry in children with no upper gastrointestinal symptoms. *Scandinavian Journal of Gastroenterology* 1997; **32**: 681–5.
- Sabba C, Merkel C, Zoli M *et al.* Interobserver and inter-equipment variability of echo-Doppler examination of the portal vein: effect of a cooperative training program. *Hepatology* 1995; **21**: 428–33.
- Zoli M, Merkel C, Sabba C *et al.* Interobserver and inter-equipment variability of echo-Doppler sonographic evaluation of the superior mesenteric artery. *Journal of Ultrasound in Medicine* 1996; **15**: 99–106.
- Filippi M, Gawne-Cain ML, Gasperini C *et al.* Effect of training and different measurement strategies on the reproducibility of brain MRI lesion load measurements in multiple sclerosis. *Neurology* 1998; **50**: 238–44.
- Meade MO, Cook RJ, Guyatt GH *et al.* Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine* 2000; **161**: 85–90.
- de Vet HC, Koudstaal J, Kwee WS, Willebrand D, Arends JW. Efforts to improve interobserver agreement in histopathological grading. *Journal of Clinical Epidemiology* 1995; **48**: 869–73.
- Rugge M, Correa P, Dixon MF *et al.* Gastric mucosal atrophy: interobserver consistency using new criteria for classification and grading. *Alimentary Pharmacology and Therapeutics* 2002; **16**: 1249–59.
- Rosner B. The kappa statistic. In: Rosner B, ed. *Fundamentals of Biostatistics*, 5th edn. Pacific Grove: Brooks/Cole, 2000: 406–11.
- Di Lorenzo C, Hillemeier C, Hyman P *et al.* Manometry studies in children: minimum standards for procedures. *Neurogastroenterology and Motility* 2002; **14**: 411–20.
- Li BUK, Altschuler SM, Berseth CL, Di Lorenzo C, Rudolph CD, Brent SR. Research agenda for pediatric gastroenterology, hepatology and nutrition: motility disorders and functional gastrointestinal disorders: report of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition for the Children's Digestive Health and Nutrition Foundation. *Journal of Pediatric Gastroenterology and Nutrition* 2002; **35**: S263–7.
- Castell JA, Castell DO. Computer analysis of human esophageal peristalsis and lower esophageal sphincter pressure. II. An interactive system for on-line data collection and analysis. *Digestive Diseases and Sciences* 1986; **31**: 1211–6.
- Parker R, Whitehead WE, Scuster MM. Pattern-recognition program for analysis of colon myoelectric and pressure data. *Digestive Diseases and Sciences* 1987; **32**: 953–61.
- Bremner RM, Costantini M, Hoeft SF *et al.* Manual verification of computer analysis of 24-hour esophageal motility. *Biomedical Instrumentation and Technology* 1993; **27**: 49–55.
- Van Herwaarden M. Computer analysis of prolonged lower oesophageal sphincter pressure recordings. *Neurogastroenterology and Motility* 2001; **13**: 37–44.
- De Schryver AM, Samsom M, Smout AJ. In search of objective manometric criteria for colonic high-amplitude propagated pressure waves. *Neurogastroenterology and Motility* 2002; **14**: 375–81.
- Andersen MB, Stodkilde-Jorgensen H, Kraglund K, Djurhuus JC, Rosenfalck A. Consistency in manual scoring analysis of gastrointestinal pressure recordings. *Scandinavian Journal of Gastroenterology* 1989; **24**: 321–8.
- Dodds WJ, Hogan WJ, Stef JJ, Miller WN, Lydon SB, Arndorfer RC. A rapid pull-through technique for measuring lower esophageal sphincter pressure. *Gastroenterology* 1975; **68**: 437–43.
- Campos GM, Oberg S, Gastal O *et al.* Manometry of the lower esophageal sphincter: inter- and intraindividual variability of slow motorized pull-through versus station pull-through manometry. *Digestive Diseases and Sciences* 2003; **48**: 1057–61.
- Bjornsson E, Castedal M, Abrahamsson H. Interlaboratory differences in method for analysis of manometry explain discrepancies in studies of duodenal phase III peristalsis. *Neurogastroenterology and Motility* 2003; **15**: 331–3.
- Fox JE, Vidins EI, Beck IT. Observer variation in esophageal pressure assessment. *Gastroenterology* 1973; **65**: 884–8.
- Eckardt VF, Elmer T. Reliability of anal pressure measurements. *Diseases of the Colon and Rectum* 1991; **34**: 72–7.
- Freys SM, Fuchs KH, Fein M, Heimbucher J, Sailer M, Thiede A. Inter- and intraindividual reproducibility of anorectal manometry. *Langenbeck's Archives of Surgery* 1998; **383**: 325–9.
- Rogers J, Laurberg S, Misiewicz JJ, Henry MM, Swash M. Anorectal physiology validated: a repeatability study of the motor and sensory tests of anorectal function. *British Journal of Surgery* 1989; **76**: 607–9.
- Schwartz MP, Samsom M, Smout AJ. Manometric artefacts suggesting compression of the duodenum by the superior mesenteric artery in healthy humans. *Neurogastroenterology and Motility* 2001; **13**: 143–9.
- Wallace M, Hawes R, Durkalski V *et al.* The reliability of EUS for the diagnosis of chronic pancreatitis: interobserver agreement among experienced endosonographers. *Gastrointestinal Endoscopy* 2001; **53**: 294–9.
- Koletzko S, Jesch I, Faus-Keb etaler T *et al.* Rectal biopsy for diagnosis of intestinal neuronal dysplasia in children: a prospective multicentre study on interobserver variation and clinical outcome. *Gut* 1999; **44**: 853–61.
- Quigley EM, Donovan JP, Lane MJ, Gallagher TF. Antroduodenal manometry. Usefulness and limitations as an outpatient study. *Digestive Diseases and Sciences* 1992; **37**: 20–8.
- Holmvang L, Hasbak P, Clemmensen P, Wagner G, Grande P. Differences between local investigator and core laboratory interpretation of the admission electrocardiogram in patients with unstable angina pectoris or non-Q-wave myocardial infarction (a Thrombin Inhibition in Myocardial Ischemia [TRIM] substudy). *American Journal of Cardiology* 1998; **82**: 54–60.

- 31 Gjorup T, Kelbaek H, Nielsen D, Kreiner S, Godtfredsen J. Interpretation of the electrocardiogram in suspected myocardial infarction: a randomized controlled study of the effect of a training programme to reduce interobserver variation. *Journal of Internal Medicine* 1992; **231**: 407–12.
- 32 Robinson P, Wilson D, Coral A, Murphy A, Verow P. Variation between experienced observers in the interpretation of accident and emergency radiographs. *British Journal of Radiology* 1999; **72**: 323–30.
- 33 Pandolfino JE, Vakil NB, Kahrilas PJ. Comparison of inter- and intraobserver consistency for grading of esophagitis by expert and trainee endoscopists [see comment]. *Gastrointestinal Endoscopy* 2002; **56**: 639–43.
- 34 Saps M, Di Lorenzo C. Interobserver and intraobserver reliability of the Rome II criteria in children. *American Journal of Gastroenterology* 2005; **100**: 2079–82.
- 35 Paspatis GA, Millar A, Chlouverakis GJ, Jamieson CP, Swain CP. Interobserver variation in describing endoscopic features in inflammatory bowel disease using videotapes of colonoscopy. *American Journal of Gastroenterology* 2000; **95**: 1376–7.
- 36 Bentley E, Jenkins D, Campbell F, Warren B. How could pathologists improve the initial diagnosis of colitis? Evidence from an international workshop. *Journal of Clinical Pathology* 2002; **55**: 955–60.
- 37 Theodossi A, Spiegelhalter DJ, Jass J *et al.* Observer variation and discriminatory value of biopsy features in inflammatory bowel disease. *Gut* 1994; **35**: 961–8.
- 38 Theodossi A, Knill-Jones R, Skene A *et al.* Inter-observer variation of symptoms and signs in jaundice. *Liver* 1981; **1**: 21–32.
- 39 Di Napoli A, Di Lallo D, Perucci C *et al.* Inter-observer reliability of radiological signs of necrotising enterocolitis in a population of high-risk newborns. *Paediatric and Perinatal Epidemiology* 2004; **18**: 80–7.
- 40 Theodossi A, Skene A, Portmann B *et al.* Observer variation in assessment of liver biopsies including analysis by kappa statistics. *Gastroenterology* 1980; **79**: 232–41.
- 41 Sigurdsson L, Flores A, Putnam P *et al.* Postviral gastro-paresis: presentation, treatment and outcome. *Journal of Pediatrics* 1997; **131**: 751–4.
- 42 Jadcherla SR, Sty JR, Rudolph CD. Mechanical small bowel obstruction in premature infants diagnosed by intestinal manometry. *Journal of Pediatric Gastroenterology and Nutrition* 2005; **41**: 247–50.