Inter- and intra-rater reliability of the Chicago Classification in pediatric high-resolution esophageal manometry recordings

Maartje MJ Singendonk (BSc)[1ab], Marije J Smits (MSc)[1a], Ilja E Heijting (BSc)[a], Michiel P. van Wijk (MD, PhD)[a], Samuel Nurko (MD)[c], Rachel Rosen (MD)[c], Pim W Wijenborg (MD)[f], Rammy Abu-Assi (MD)[b], Daniel R Hoekman (MD)[a], Sophie Kuizenga-Wessel (MD)[a], Grace Seiboth[b], Marc A. Benninga (MD, PhD)[a], Taher I Omari (PhD)[b,d,e], Stamatiki Kritas[b]

[1]Both authors contributed equally to this paper

Affiliations: [a]Department of Pediatric Gastroenterology and Nutrition, Emma Children's Hospital AMC, Amsterdam, The Netherlands; [b]Gastroenterology Unit, Women's and Children's Health Network, North Adelaide, Australia; [c]Center for Motility and Functional Gastrointestinal Disorders, Division of Gastroenterology, Children's Hospital Boston, Boston, USA; [d]School of Medicine, Flinders University, Bedford Park, South Australia; [e]Translational Research Center for Gastrointestinal Diseases, University of Leuven, Belgium; [f]Department of Gastroenterology and Hepatology AMC, Amsterdam, The Netherlands

Running title: Inter- and intra-rater reliability of pediatric HRM

Address correspondence to:

Maartje Singendonk or Marije Smits, Department of Pediatric Gastroenterology and nutrition, Emma Children's Hospital, AMC

Room C2-312, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

[m.m.j.singendonk@students.uu.nl], [m.j.smits@amc.nl], 0031-20-5665270.

Keywords: Chicago Classification – pediatrics – esophageal motility – high resolution impedance manometry – reliability

Abstract

Background: The Chicago Classification (CC) facilitates interpretation of high-resolution manometry (HRM) recordings. The CC algorithm is based upon adult data and its application to the pediatric population is unknown. We assessed reliability of software-based CC diagnosis in a pediatric cohort.

Methods: Thirty pediatric solid state HRM recordings (13M; mean age 12.1 SD 5.1 years) including 10 liquid swallows per patient were analyzed twice by 11 raters (6 experts, 5 non-experts). Software-placed anatomical landmarks required manual adjustment or removal. Integrated Relaxation Pressure (IRP4s), Distal Contractile Integral (DCI), Contractile Front Velocity (CFV), Distal Latency (DL) and Break size (BS) and a CC diagnosis were software-generated. Additionally, raters provided their subjective CC diagnosis. Agreement was calculated with Cohen's and Fleiss' kappa (κ) and intraclass correlation coefficient (ICC).

Results: Intra- and inter-rater reliability of software generated CC diagnosis was *substantial* (mean κ=0.69 and 0.77 respectively) and *moderate-substantial* for subjective CC diagnosis (mean κ=0.70 and 0.58 respectively). Agreement on software-generated and subjective diagnosis of Normal motility was high (k=0.81 and k=0.79). Intra- and inter-rater agreement was *excellent* for IRP4s, DCI and BS. Experts had higher inter-rater agreement than non-experts for DL (ICC=0.65 vs ICC=0.36 respectively) and the software-generated diagnosis diffuse esophageal spasm (DES, κ=0.64 vs κ=0.30). Amongst experts the agreement for the subjective diagnosis of achalasia and EGJ outflow obstruction was *moderate-substantial* (κ=0.45 - 0.82).

Conclusion: Inter- and intra-rater reliability of software-based CC diagnosis of pediatric HRM recordings was high overall, however experience was a factor influencing the diagnosis of some disorders, particularly DES and achalasia.

List of abbreviations

BS	Break size
CC	Chicago Classification
CDP	Contractile deceleration point
CFV	Contractile front velocity
DCI	Distal contractile integral
DL	Distal latency
EGJ	Esophageal gastric junction
EPT	Esophageal pressure topography
HRM	High resolution manometry
HRIM	High resolution impedance manometry
ICC	Intraclass correlation coefficient
IRP4s	Integrated relaxation pressure
LES	Lower esophageal sphincter
UES	Upper esophageal sphincter

Key Messages

- There was excellent intra- and inter-rater agreement amongst experienced and inexperienced raters for the EPT metrics IRP4, DCI and BS and moderate/substantial for CFV and DL. Fair to substantial agreement was reached for the overall CC based diagnosis (software derived and subjective). For individual diagnoses, high inter-rater agreement was reached on normal motility.

- Automated software-based Chicago Classification (CC) diagnoses pediatric esophageal motility disorders with high inter- and intra-rater reliability. However, the clinical diagnosis of disorders such as achalasia and distal esophageal spasm is less reliable. This study highlights the challenges of CC based diagnosis of pediatric esophageal motility disorders when software is used.

Introduction

The recent introduction of high-resolution manometry (HRM) with esophageal pressure topography (EPT) into clinical practice has allowed for better characterization of esophageal motor function and uniform consensus on diagnosis of esophageal motility disorders. (1,2,3) The 2012 Chicago Classification (CC) algorithm for esophageal motility, with application through interactive analysis software, facilitates this diagnostic interpretation of pressure recordings. The CC uses five EPT metrics based on ten liquid swallows and characterizes motor dysfunction into four main categories in order of severity, i.e. achalasia (Category 1), EGJ outflow obstruction (Category 2), disorders never observed in healthy individuals (Category 3; absent peristalsis, diffuse esophageal spasm or hypercontractile esophagus and motor patterns outside the normal range (Category 4; weak peristalsis, frequent failed peristalsis, hypertensive peristalsis or rapid contraction).(4,5)

In the pediatric population the spectrum of esophageal motility disorders resembles that seen in adults. (6,7,8) However, implementation of the CC for use in the pediatric population is not without its challenges. There are no established pediatric normative ranges for EPT metrics and some metrics, such as IRP4 and DL, have been shown to be significantly influenced by patient age- and size (9-12). Hence, if adult diagnostic CC criteria are not adjusted to account for these effects it is likely that esophageal motility disorders, particularly EGJ outflow obstruction and DES, will be over diagnosed (9). Pediatric HRM studies are also more challenging to perform with a greater incidence of incomplete studies with fewer than the requisite 10 liquid swallows. Finally, manometric recordings from children may be harder to interpret due to multiple swallowing and artifact due to body movement and crying (6,13,15).

Derivation of EPT metrics and a CC diagnosis is reproducible and reliable when applied to the adult population (12,16,17). However, there are no equivalent data available based on the analysis of more challenging pediatric studies. Therefore, the aim of this study was to assess inter- and intra-rater reliability of interactive CC analysis software for the diagnosis of esophageal motility disorders in a pediatric cohort.

Methods

Study Database

Combined high-resolution impedance and manometry measurements (HRIM) pediatric patients were extracted from an extensive database of studies conducted at the Gastroenterology units of the Women's and Children's Hospital Adelaide, Boston Children's Hospital and the Amsterdam Medical Center AMC between December 2008 and September 2013. The typical manometric protocol used a 3.2mm diameter solid state HRIM catheter incorporating 25 or 36 1cm-spaced pressure sensors and 12 adjoining impedance segments, each of 2 cm (Unisensor USA Inc, Portsmouth, NH). If the pressure-impedance sensor array was not large enough to accommodate the entire region from UES to EGJ, the catheter was positioned with sensors straddling the distal esophagus from transition zone to stomach. Patients were studied sitting in the supine or semi-supine posture with a standard protocol including 10 x 3, 5 or 10ml swallows administered via syringe at ≥ 30s intervals.

Patient measurements were only considered for inclusion if they met the following criteria: (i) 10 liquid swallows performed, (ii) adequate catheter position to resolve EGJ pressures and, (iii) no technical errors, e.g. pressure or impedance channel failure.

From these potential studies, a database of 30 de-identified studies was created to assess intra/inter-rater reliability. Using the original diagnostic findings as a guide, all four main Categories of Chicago Classification disorders were represented in the database. Furthermore, the distribution of disorders in the study database was designed to be consistent with the overall distribution within the more extensive database. Based on the original expert defined diagnostic analysis, the study database was constructed as follows: 10 patients (33%) with normal peristalsis, seven (23%) weak peristalsis (five large breaks), six (20%) with EGJ outflow obstruction, two (7%) frequent failed peristalsis, two (7%) distal esophageal spasm and two (7%) achalasia (type I and type II) and one (3%) absent peristalsis.

Data analysis

Each rater was provided with reference literature regarding the assessment of esophageal motility based on EPT metrics and the Chicago Classification. (3,4,9) All raters also viewed an introductory Powerpoint tutorial explaining the correct use of the MMS automated analysis software and completed a practice run of a patient study to confirm they were proficient. Factsheets detailing the principle steps of software analysis and the CC algorithm could be referenced at any stage of analysis. Raters with varying levels of experience with esophageal manometry were invited to participate. Raters with experience from ≥200 HRM analyses were considered 'experts'. To assess intra-rater reliability, each rater analyzed the data set twice, with at least seven days between repeat analyses. To avoid the potential for sequence bias, the order of studies was randomized between raters and between repeat analyses.

Patient studies were analyzed using the MMS analysis software, version 8.23 (MMS, Enschede, The Netherlands). Raters were instructed to manually place or adjust the automatically populated landmarks. These included gastric position, EGJ proximal and distal margin, UES margins, transition zone, swallow onset, distal contractile integral (DCI) box and contractile deceleration point (CDP). Swallow onset was defined by the relaxation of the UES. If UES pressures were not visible, the onset of impedance drop in the most proximal impedance segment was used. Raters were instructed to delete analysis landmarks if they considered them to be not applicable to the swallow (e.g. CDP and DCI box in circumstances of failed peristalsis). Following completion of analysis, the standard EPT metrics (per swallow and mean of 10 swallows) were derived by the software. These were, (i) Integrated Relaxation Pressure (IRP4), (ii) Contractile Front Velocity (CFV, cm/sec), (iii) Distal Contractile Integral (DCI, mmHg/cm/sec), (iv) Distal Latency (DL, sec) and (v) peristaltic 20mmHg isocontour break size (BS, cm). (4) An overall CC diagnosis per study was automatically software-generated based on these metrics. In addition to the software-based CC diagnosis, raters were asked to provide their own subjective opinion on the CC diagnosis for each patient.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics 20. For categorical data, inter- and intra-rater agreement was calculated using Cohen's kappa (2 raters, kappa further annotated as κ) and Fleiss' κ (>2 raters). For ordinal data, the intraclass correlation coefficient (ICC) was used. The first session of analysis was used to determine inter-rater reliability. We additionally compared inter-rater reliability between the two sessions. Fleiss' κ was calculated by using a pre-made syntax for SPSS (available from corresponding author). Statistical analysis on EPT metrics was performed based on mean values. In circumstances where landmarks were removed preventing an EPT metric average being based on all 10 swallows then these data were excluded from analysis of reliability. Mean values for κ and ICC were calculated by using the Fisher's *Z*-transformation ($Z = \text{arctanh}(\kappa)$). We applied the common scale for κ and ICC values: 0.00 = no agreement, 0.01 to 0.20 = slight agreement, 0.21 to 0.40 = fair agreement, 0.41 to 0.60 = moderate agreement, 0.61 to 0.80 = substantial agreement, 0.81 to 0.99 = excellent agreement and 1.00 = perfect agreement.

Results

Analysis of all 30 manometric measurements was completed twice by 11 raters. Six raters were considered experts (Group 1; three gastroenterologists (two pediatric) and three medical scientists) and five raters were naïve to HRM analyses (Group 2; one pediatric gastroenterologist, two PhD candidates undertaking pediatric research, one medical student and one technician performing diagnostic reflux monitoring).

Intra-rater reliability of software derived EPT metrics

The mean Cohen's κ statistics for intra-rater agreement between the two sessions are shown in table 1. Overall, *excellent* intra-rater agreement was noted for the metrics IRP4, DCI, DL and BS and *substantial for CFV*. The inta-rater reliability was generally similar for experienced and in-experienced raters, however DL was, paradoxically, less reliably scored by experts.

Inter-rater reliability of EPT metrics

Amongst all raters, *excellent* agreement was reached for the metrics IRP4s, DCI and BS and level of agreement appeared to be independent of level of experience (table 1). Reliability for CFV and DL was *moderate* and depended strongly on level of experience, with higher levels of agreement amongst the group of experienced raters when compared to the inexperienced raters (ICC = 0.80 vs ICC = 0.39 respectively for CFV and ICC = 0.65 vs ICC = 0.36 respectively for DL). The agreement amongst inexperienced raters on DL was higher for the second session of analysis (κ = 0.56), suggesting a training effect. There were no marked differences in agreement between sessions in relation to the other metrics.

Intra- and inter-rater reliability of software generated and subjective Chicago Classification diagnosis

Both intra- and inter-rater reliability of the software generated CC diagnosis were *substantial* amongst all raters and did not depend on level of experience (table 2). The initial software generated diagnosis was changed according to the raters' personal opinion in 32.1% of all studies. Experienced raters were more likely to change the software generated diagnosis in comparison to the inexperienced raters (34.4% vs 29.3%). Both intra- and inter-rater reliability were lower for the subjective CC diagnosis, but remained *fair* to *substantial* (table 2). Amongst experienced raters, a higher level of inter-rater agreement on the subjective CC diagnosis was reached when compared to the inexperienced raters ($\kappa = 0.56$ vs $\kappa = 0.48$). We observed an increase in inter-rater agreement of subjective CC diagnosis in the second session of analysis amongst inexperienced raters only ($\kappa = 0.62$). Agreement on software generated diagnosis did not differ between sessions.

Inter-rater reliability of the individual software generated and subjective Chicago Classification diagnoses

Results on inter-rater agreement for all individual software generated CC diagnoses are displayed in table 3a. Highest agreement amongst all raters was reached for the diagnoses of normal motility, EGJ outflow obstruction and absent peristalsis ($\kappa = 0.81$, $\kappa = 0.79$ and $\kappa = 0.84$ respectively), and level of agreement appeared to be independent of experience. Level of agreement for the software-generated diagnosis DES was higher in the group of experienced raters ($\kappa = 0.64$ vs $\kappa = 0.30$ for non-experts).

Table 3b displays the levels of inter-rater agreement for the subjective Chicago Classification diagnoses based on raters' personal opinions. Although overall agreement on the CC diagnosis of normal motility was *substantial,* the level of agreement amongst experienced raters was substantially higher when compared to the group of inexperienced raters ($\kappa = 0.79$ vs $\kappa = 0.49$ respectively). Agreement based on subjective diagnosis of the two most severe CC diagnoses, achalasia and EGJ outflow obstruction, ranged from *moderate* to *substantial* ($\kappa = 0.45$ to $\kappa = 0.82$) amongst expert raters.

Influence of visibility of the UES on reliability

We performed sub-analysis to determine the possible influence of the lack of visibility of the UES during analysis of patient studies (n=12). Visibility of the UES did not influence the level of inter-rater reliability for all metrics, apart from DL. For DL, inter-rater reliability appeared to be higher in studies in which the UES was completely in view amongst inexperienced raters albeit on a background of relatively poor reliability overall (κ = 0.19 vs κ = 0.44). However when the level of experience was taken into account the overall level of reliability for determining DL was far better in experienced raters and visibility of the UES had no effect (amongst experts κ = 0.69 vs κ = 0.64 when UES was absent or present respectively). For the overall CC diagnosis, there were no differences in reliability between studies with and without UES high pressure zone in view suggesting that variability in placement of the CDP, rather than swallow onset, was the main driver of variability in the DL measurement (Figure 1).

Discussion

This study is the first to report data on inter- and intra-rater reliability for both software and personal opinion based CC diagnosis of esophageal motility disorders by HRM in a pediatric cohort. Using semi-automated software-assisted analysis we demonstrated high reliability and reproducibility of CC-based diagnoses of esophageal motility disorders amongst both experienced and inexperienced raters. Our findings are in line with earlier studies in adult cohorts, indicating that the analysis software is easy to learn and can be easily implemented (17,18). As would be predicted, agreement was generally lower amongst inexperienced raters showing that training courses are needed to ensure reliability of diagnosis even though the software now available is simple to use and easy to apply. The findings of our study support the clinical utility of HRM in the objective CC-based diagnosis of esophageal motor disorders in pediatric patients, but furthermore also highlight diagnostic challenges specific to the pediatric population.

Our study identified areas of diminished reliability of the analysis. This was notable in regard to the reduced reliability and reproducibility of derivation of the DL metric which, in turn, contributes to a poorer agreement for diagnosis of DES, particularly amongst inexperience raters. We have recently shown that in pediatric patients DL varies in an age- and esophageal length dependent manner (9). Deriving the DL might therefore be particularly challenging in pediatric patients and may explain the low levels of agreement on DL and DL-driven diagnosis DES amongst raters naïve to HRM analysis. The DL is calculated from the swallow onset to the CDP. Manually reviewing all studies, we observed consistent placement of the swallow onset amongst and between raters. In challenging studies, large variations in the placement of the CDP were observed (Figure 1). This illustrates that consistent placement of the CDP is particularly challenging, even amongst raters experienced to HRM.

We additionally found low levels of agreement for software-derived diagnosis of achalasia subtypes, whereas subjective diagnosis was *moderate* to *substantial* amongst expert raters. To further explore this discrepancy, we manually reviewed the study database. Two patients in this cohort were given an initial diagnosis of achalasia based on expert consensus, prior to the study. Marked differences were noted in the diagnosis of achalasia, when comparing the software-based and subjective diagnoses for individual raters. Patient studies that were classified with an achalasia disorder based on subjective diagnosis, were either allocated the diagnosis EGJ outflow obstruction or absent peristalsis by software-driven analysis. Both achalasia and EGJ outflow obstruction require an IRP4 ≥ 15 mmHg, but differ in that, for EGJ outflow obstruction, there is some evidence of an esophageal peristaltic wave which demonstrates normal latency. Retrospective analysis of studies that were allocated a software-driven diagnosis of EGJ outflow obstruction showed that studies subjectively classified as having pan-esophageal pressurization patterns, were allocated a 'normal peristaltic pattern' by the software. The diagnosis of absent peristalsis requires an IRP4s < 15mmHg in combination with 100% failed swallows. (4) Retrospective analysis of EPT metrics revealed that the criterion of an abnormal IRP4s was ignored in the subjective diagnosis of achalasia by three of the expert raters. This finding might indicate that the decision to change the software-derived diagnosis of absent peristalsis towards an achalasia diagnosis was rather based upon pattern recognition and clinical expertise, than on EPT metrics and stresses the importance of careful review of the motility studies before a diagnosis is made. The diagnosis of achalasia is changing and it has been suggested to reduce the IRP4s criteria for Type 1 achalasia to 10mmHg (19). Applying this criterion might potentially enhance diagnostic accuracy, leading to a more uniform diagnosis of achalasia and achalasia subtypes.

Another diagnostic category with low reliability is the small or large peristaltic breaks. There are several possibilities for this finding. First, these diagnoses did not exist with standard manometry, so incorporation of these findings into routine manometry readings lag behind. A second, more likely explanation is that clinical significance of these diagnoses remains uncertain. While adult studies have shown that the length of esophageal breaks correlates with incomplete bolus transit, no comparable data exists in children (20, 21). Furthermore, early studies with esophageal function testing suggest that there is an imperfect correlation between bolus transit and some peristaltic defects, particularly non-specific motor disorders (22, 23). Therefore, while the manometric findings may show breaks in peristalsis, a normal bolus clearance pattern combined with the lack of data to show clinical significance of breaks in children may result in underreporting of a diagnosis with unknown significance.

Our study has several limitations. Firstly, part of the included HRM studies did not incorporate the UES due to limited pressure sensor array length (n=12, 40%). In the absence of visual confirmation, swallow onset is more difficult to determine. Direct visualization of the UES relaxation onset did not appear to influence reliability of overall CC diagnosis. However, it could have influenced reliability and reproducibility of IRP4s and DL, since the calculation of these metrics relates on the position of the swallow onset marker. In our experience IRP4s calculation is very resilient to swallow onset difference of up to ±1sec and reliability of IRP4s appeared to be high. Therefore localizing of swallow onset was more a potential factor for the calculation of distal latency. However, sub-analysis that compared studies with the UES high pressure zone present or absent suggested that the main driver of variability in the DL was location of the CDP, rather than the timing of swallow onset (as illustrated in Figure 1). Failure of capture of the UES can occur, importantly or data show that studies can still be analyzed accurately if this happens. A second limitation was that raters were instructed to delete metrics from analysis if considered inapplicable to a swallow (i.e. in the absence of a peristaltic contraction pattern). This approach influenced statistical analysis of EPT metrics CFV, DL and DCI, as patient studies were pairwise excluded from analysis when metrics were not uniformly obtained. We therefore also assessed whether raters would consistently apply these landmarks to a swallow. The combined finding of low reliability on both the applicability and the values of the CFV and DL, points out that these metrics might be particularly challenging. Lastly, reliability between first and second analysis was tested after a minimum of seven days. Although raters were blinded from patient characteristics and studies were presented in a randomized order for both sessions, this short period could have resulted in raters recognizing some of the patient studies from their initial session. Another limitation is that the limited number of achalasia patients included might have resulted in the lower reliability for its diagnosis, so future studies that focus on this entity will be needed to establish if the CC criteria for the diagnosis of achalasia in children need to be modified.

One of the strengths of our study is that we tested reproducibility of CC-based diagnosis of pediatric HRM recordings in a large cohort of patients and selected a variety of observers with different background and varying experience. Patient studies were selected in such way that the distribution of the study database matched the distribution of the four broad CC categories earlier reported in a large cohort of pediatric patients referred for manometry. (9) Finally, we incorporated software derived diagnosis as well as the personal opinion of raters (subjective diagnosis). Differences between software and subjective diagnosis might be more substantial in clinical practice due to awareness of patients' symptoms, clinical history.

In conclusion, automated software based CC diagnosis of pediatric esophageal motility disorders shows to be of high inter- and intra-rater reliability amongst experts and non-experts. However, the application of CC-based semi-automated software is least reliable and most influenced by rater's expertise for the diagnosis of disorders such as DES and achalasia.

Contributors' statement

All authors accept responsibility for the reported content and declare they all have participated in the concept, design, draft and revision of this manuscript. All authors have approved this version of the manuscript.

Maartje MJ Singendonk and Marije J. Smits: conceptualized and designed the study, build the data base, initialized the study, extracted and analyzed the data, drafted the initial manuscript, and approved the final manuscript as submitted.

Ilja E Heijting extracted and analyzed the data, drafted parts of the initial manuscript, reviewed and revised the manuscript, and approved the final manuscript as submitted.

Michiel P. van Wijk, Marc A. Benninga, Taher I. Omari, Rachel Rosen and Samuel Nurko reviewed and revised the manuscript, and approved the final manuscript as submitted.

Maartje MJ Singendonk, Ilja E Heijting, Samuel Nurko, Rachel Rosen, Pim W Wijenborg, Rammy Abu-Assi , Daniel R Hoekman, Sophie Kuizenga-Wessel, Grace Seiboth, Taher I Omari, Stamatiki Kritas analyzed the selected patients studies

Stamatiki Kritas, as senior author, reviewed and revised the manuscript, and approved the final manuscript as submitted.

Funding Source: No funding was secured for this study.

Financial Disclosure: The authors have no financial relationships relevant to this article to disclose. None of the authors received an honorarium, grant, or other form of payment to produce the manuscript.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Srinivas M, Balakumaran TA, Palaniappan S, Srinivasan V, Batcha M, Venkataraman J. High resolution esophageal manometry-The switch from "intuitive" visual interpretation to Chicago classification. Indian J Gastroenterol. 2013 Oct 9;
2. Kessing BF, Smout AJPM, Bredenoord AJ. Clinical applications of esophageal impedance monitoring and high-resolution manometry. Curr Gastroenterol Rep. 2012 Jun;14(3):197–205.
3. Pandolfino JE, Fox MR, Bredenoord AJ, Kahrilas PJ. High-resolution manometry in clinical practice: utilizing pressure topography to classify oesophageal motility abnormalities. Neurogastroenterol Motil. 2009 Aug;21(8):796–806.
4. Bredenoord AJ, Fox M, Kahrilas PJ, Pandolfino JE, Schwizer W, Smout AJPM. Chicago classification criteria of esophageal motility disorders defined in high resolution esophageal pressure topography. Neurogastroenterol Motil. 2012 Mar;24 Suppl 1:57–65.
5. Kahrilas PJ. Esophageal motor disorders in terms of high-resolution esophageal pressure topography: what has changed? Am J Gastroenterol. Nature Publishing Group; 2010 May;105(5):981–7.
6. Chumpitazi B, Nurko S. Pediatric gastrointestinal motility disorders: challenges and a clinical update. Gastroenterol Hepatol (N Y). 2008;4(2).
7. Rommel N, Gerarduzzi T, Biasotto E, Faleschini E, Martelossi S. The complexity of feeding problems in 700 infants and young children presenting to a tertiary care institution. J Pediatr Gastroenterol Nutr. 2004 Mar;38(3):360–1.
8. Richter JE. Oesophageal motility disorders. Lancet. 2001 Sep 8;358(9284):823–8.
9. Singendonk MMJ, Kritas S, Cock S, Ferris L, McCall L, Rommel N, van Wijk MP, et al. Applying the Chicago Classification criteria of esophageal motility to a pediatric cohort: effects of patient age and size. Neurogastroenterol Motil 2014;
10. Goldani HA, Staiano A, Borelli O, Thapar N, Lindley KJ. Pediatric esophageal high-resolution manometry: utility of a standardized protocol and size-adjusted pressure topography parameters. Am J Gastroenterol. 2010 Feb;105(2):460-467.
11. Weijenborg PW, Kessing BF, Smout AJPM, Bredenoord AJ. Normal values for solid-state esophageal high-resolution manometry in a European population; an overview of all current metrics. Neurogastroenterol Motil. 2014 Feb 7;
12. Bogte a, Bredenoord a J, Oors J, Siersema PD, Smout a JPM. Reproducibility of esophageal high-resolution manometry. Neurogastroenterol Motil. 2011 Jul;23(7):e271–6.
13. Roman S, Damon H, Pellissier PE, Mion F. Does body position modify the results of oesophageal high resolution manometry? Neurogastroenterol Motil. 2010 Mar;22(3):271–5.
14. Vandenplas Y, Rudolph CD, Di Lorenzo C, Hassall E, Liptak G, Mazur L, et al. Pediatric gastroesophageal reflux clinical practice guidelines: joint recommendations of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition (NASPGHAN) and the European Society for Pediatric Gastroenterology, Hepatology, a. J Pediatr Gastroenterol Nutr. 2009 Oct;49(4):498–547.
15. Xiang X, Tu L, Zhang X, Xie X, Hou X. Influence of the catheter diameter on the investigation of the esophageal motility through solid-state high-resolution manometry. Dis Esophagus. 2013;26(7):661–7.
16. Hernandez JC, Ratuapli SK, Burdick GE, Dibaise JK, Crowell MD. Inter-rater and intra-rater agreement of the chicago classification of achalasia subtypes using high-resolution esophageal manometry. Am J Gastroenterol. 2012 Feb;107(2):207–14.

17. Singh ER, Rife C, Clayton S, Naas P, Nietert P, Castell DO. Interobserver variability in esophageal body measurements with high-resolution manometry among new physician users. J Clin Gastroenterol. 2013 Feb;47(2):e12–6.
18. Grübel C, Hiscock R, Hebbard G. Value of spatiotemporal representation of manometric data. Clin Gastroenterol Hepatol. 2008 May;6(5):525–30.
19. Lin Z, Kahrlias PJ, Roman S, Boris L, Carlson D, Pandolfino JE. Refining the criterion for an abnormal integrated relaxation pressure in esophageal pressure topography based on the pattern of esophageal contractility using a classification and regressioin tree model. Neurogastroenterol Motil 2012 Aug;24(8):356-363.
20. Bulseiwicz WJ, Kahrilas PJ, Kwiatek MA, Ghosh SK, Meek A, Pandolfino JE. Esophageal pressure topography criteria indicative of incomplete bolus clearance: a study using high-resolution impedance manometry. *Am J Gastroenterol*. 2009 Nov;104(11):2721-2728.
21. Ghosh SK, Pandolfino JE, Kwiatek MA, Kahrilas PJ. *Neurogastroenterol Motil*. 2008 Dec; 12:1283-1290.
22. Tutuian R, Castell DO. Combined multichannel intraluminal impedance and manometry clarifies esophageal function abnormalities: study in 350 patients. *Am J Gastroenterol*. 2004 Jun: 99(6):1011-1019.
23. Tutuian R, Castell DO. Clarification of the esophageal function defect in patients with manometric ineffective esophageal motility: studies using combined impedance-manometry. *Clin Gastroenterol Hepatol*. 2004 Mar;2(3):230-236.