UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTA' DI MEDICINA E CHIRURGIA

DIPARTIMENTO DI FISIOPATOLOGIA MEDICO-CHIRURGICA E DEI TRAPIANTI

CORSO DI DOTTORATO IN GASTROENTEROLOGIA
CICLO XXV

INTER-OBSERVER AND INTRA-OBSERVER AGREEMENT IN THE ANALYSIS OF OESOPHAGEAL pH-IMPEDANCE TRACINGS

Settore scientifico-disciplinare: Med/12

Tesi di Dottorato della Dott.ssa IVANA BRAVI
Matricola R08682

TUTOR: Chiar.mo Prof. ROBERTO PENAGINI

COORDINATORE DEL DOTTORATO: Chiar.mo Prof. DARIO CONTE

ANNO ACCADEMICO 2011/2012

INDEX

ABSTRACT

AIM

Information resulting from oesophageal 24h pH-impedance monitoring (pH-MII) may have important consequences for patients' management.

Aims of this study were 1) to evaluate agreement for number of reflux episodes and symptom/reflux association indexes in MII-pH tracings analysis between and within three experienced observers working in different European Centers, 2) to evaluate the predictivity of a shorter manual analysis targeted to the two-minutes window periods before each symptomatic episode with regards to the traditional.

MATERIALS AND METHODS

Forty consecutive pH-MII tracings of patients (23 women, median age 55 years) referred for suspected oesophageal (n=24) or extra-oesophageal (n=16) GORD symptoms in two Northern Italian Centers were included (20 tracings/Center). After automatic analysis (MMS, Enschede, The Nederlands), tracings were duplicated and anonymized. Three expert observers blindly performed the traditional manual analysis on the 40 duplicated tracings, erasing or adding reflux episodes according to their judgment based on the Oporto Criteria. The first manual analysis was performed in a sequential order, the second one in a randomly assigned order. Subsequently a short manual analysis was also blindly performed. Values of both symptom association indexes (Symptom Index=S.I. and Symptom Association Probability=S.A.P.) were transformed into binary response (i.e. positive or negative). Inter- and intra-observer agreement was calculated with Cohen's Coefficient (k) and/or with percentage of agreement.

RESULTS

Inter-observer agreement on acid reflux was excellent between the three observers for both S.I. and S.A.P. (K=1.0; 100%), whereas considering non-acid reflux it was high for S.I. (95%, 92.5% and 95% of concordance for observer 1,2 and 3 respectively) and moderate for S.A.P. (K=0.35; 85%

between observer 1 and 2, K=0.36; 82.5% between observer 1 and 3, K=0.23; 87.5% between observer 2 and 3).

Intra-observer agreement on acid reflux was good to excellent for both S.I. (K=0.77; 95% for observer 1 and 2 and K=1.0; 100% for observer 3), and S.A.P. (K=0.80; 95% , K=0.72; 92.5% and K=1.0; 100% for observer 1, 2 and 3 respectively). Intra-observer agreement on non-acid reflux was high for S.I. (97.5%, 95% and 97.5% of concordance for observer 1, 2 and 3 respectively) and moderate for S.A.P. (K=0.48; 85% and K=0.54; 90% for observer 1 and 3 and good for observer 2, K=0.79; 2.5%).

Considering all the three observers the short analysis was predictive of the traditional one for both S.I., 88% and 98% of the tracings for acid and non-acid reflux, respectively and S.A.P., 99% and 97% of the tracings for acid and non-acid reflux, respectively.

CONCLUSIONS

We found a good inter- and intra-observer agreement for symptom/reflux association indexes, especially when considering acid reflux. Moreover we demonstrated that short analysis was highly predictive of the traditional manual analysis, suggesting that the short one could be used in routine clinical practice with considerable savings.

GASTRO-OESOPHAGEAL REFLUX DISEASE

DEFINITION, EPIDEMIOLOGY AND CLINICAL MANIFESTATIONS

Gastro-oesophageal reflux disease (GORD) is a condition which develops when reflux of gastric contents causes troublesome symptoms and/or complications (1). While this is still predominantly a disease of the Western population (prevalence at 10-15%), changing dietary patterns, the global increase in obesity and other causes have led to a pronounced increase in the prevalence of GORD in Asia. This disease therefore has a global impact on health and impairs health related quality of life of a substantial proportion of the world's population. According to the Montreal definition and classification (1), the clinical manifestations of GORD include:

- typical reflux syndrome (heartburn and/or regurgitation)
- reflux chest pain syndrome
- extra-oesophageal syndrome (chronic cough, chronic laryngitis, asthma, dental erosions)
- syndromes with oesophageal injury: erosive oesophagitis, reflux strictures, Barrett's oesophagus, oesophageal adenocarcinoma.

Symptoms of reflux disease are not completely specific and often co-exist with other gastrointestinal symptoms. Identification of the dominant symptom by either patient or clinician may be difficult. The accuracy of heartburn as a diagnostic symptom for reflux disease is somewhat unclear, however results of empirical PPI therapy in patients with reflux disease and heartburn as their predominant symptom suggest that the sensitivity ranges from 70% to 80% with a specificity of 55-65% (2, 3). The presence of regurgitation appears to strengthen the diagnostic accuracy of heartburn (4). The prevalence of chest pain in patients with reflux disease is unknown, however the presence of non-cardiac chest pain is highly predictive of underlying GORD, with 21% of such patients having oesophagitis on endoscopy, 42% having abnormal acid exposure time, and 39% having a positive association between chest pain and reflux episodes during pH monitoring (5-8). The value of extra-oesophageal symptoms in the diagnosis of reflux disease is questionable. Current

evidence suggests that only in patients with typical reflux symptoms, in addition to extra-oesophageal symptoms, extra-eoesophageal symptoms respond to antireflux therapy (9, 10).

DIAGNOSIS

Upper GI endoscopy is the gold standard to identify oesophageal erosions, the hall-mark feature of erosive reflux disease. Recent studies however document that up to 70% of patients with oesophageal symptoms have normal endoscopic findings in the oesophagus (11). In addition, empiric proton pump inhibitor (PPI) therapy trials are used in the primary care setting as the first step to evaluate patients with symptoms suggestive of GORD (12). Symptom response to PPI therapy establishes the diagnosis of GORD with a sensitivity of 75-92% and specificity of 55-90% (7, 13, 14). Patients with persistent oesophageal symptoms on acid suppressive therapy are considered patients with refractory GORD (15) and fall into one of the 3 categories: 1) patients with inadequate acid suppression on standard dose PPI in whom symptoms are related to acid reflux, 2) patients with adequate acid suppression on standard dose PPI but with an abnormal number of reflux episodes in whom symptoms are related to non-acid reflux and 3) patients with oesophageal symptoms independent of gastro-oesophageal reflux. These patients deserve laboratory based investigations, which are performed in specialized GI function laboratories and aim at identifying abnormal amounts of gastro-oesophageal reflux in the oesophagus and/or clarify the relationship between oesophageal symptoms and gastro-oesophageal reflux episodes. Most commonly used oesophageal reflux monitoring investigations include oesophageal pH monitoring (catheter-based and wireless) and the more recently developed impedance-pH monitoring (the latter one will be described in the next chapter).

Oesophageal pH monitoring

Despite methodological limitations, conventional pH monitoring is still very accurate in quantifying distal oesophageal acid exposure. In patients off acid suppressive therapy this is still considered the

gold standard to diagnose or exclude GORD. Another important indication is the oesophageal pH testing in the pre-operative setting, i.e. before surgical or endoscopic therapy for reflux.

Prior to reflux monitoring, calibration of pH sensors using reference buffer solutions needs to be performed. Two calibration substances are usually used – one in the acidic range (pH 1-4) and one in the neutral range (pH 6-7).

By convention, catheter-based pH electrodes are positioned five cm above the proximal border of the lower oesophageal sphincter (LOS) (16, 17). After the correct placement of pH electrodes, patients are asked to undergo the same daily routines as usual and to reproduce situations that cause the symptoms for which they have undergone reflux monitoring, in order to maximise the ability of pH monitoring to evaluate the relationship between symptoms and reflux.

The capsule pH system is a catheter-free oesophageal pH monitoring system, which consists of an antimony pH electrode, a radio transmitter and a battery contained in a capsule. The data from the capsule are transmitted by telemetry to a portable digital recorder clipped onto the patient's belt. Capsule placement is accomplished with the aid of a delivery device, usually during an upper endoscopy and conventionally placed six cm above the squamocolumnar junction. The two main advantages of this device are 1) no catheters in the patients throat or nose, thus reducing discomfort and improving adherence to daily activities; 2) longer monitoring (usually 48 hours and up to 96 hours) (18).

Intraoesophageal pH monitoring, regardless whether catheter-based or wireless, identifies gastro-oesophageal reflux episodes as abrupt drops in pH from above to below 4.0. The parameters more often analysed during oesophageal pH monitoring, in order to determine if a patient has abnormal oesophageal reflux, are:

- total oesophageal acid exposure time (percentage of time at pH < 4.0)
- upright oesophageal acid exposure time (percentage of time at pH < 4.0)
- supine oesophageal acid exposure time (percentage of time at pH < 4.0)

Moreover, symptom indices are important components of assessing the relationship between reflux and patients symptoms. The two main symptom indices include the Symptom Index (S.I.) i.e. the percentage of symptom episodes that are related to reflux, which is considered positive if > 50% (19), and the Symptom Association Probability (S.A.P.), which is a more complex index indicating that the association between reflux and symptoms does not occur by chance and is considered positive if > 95 (20).

TREATMENT

PPIs represent the main treatment in GORD and cause symptom relief in 70-80% of patients. Because most reflux episodes, both acidic and non-acid, result from transient LOS relaxation (TLOSR), i.e. complete and long lasting relaxations of the LOS that are not preceded by swallows, controlling occurrence of TLOSRs appears to be a relevant therapeutic target in GORD patients refractory to PPIs. The most popular available anti-reflux therapy which is able to reduce TLOSRs occurrence is surgery (i.e. fundoplication). Regarding pharmacological compounds, GABAb agonists and the mGluR5 antagonists were the most promising, however they have not been introduced in clinical practice because of both the unsatisfactory performance in controlling GORD symptoms and their important side effects.

COMBINED 24-h MULTICHANNEL INTRALUMINAL IMPEDANCE-pH MONITORING

In recent years the increasing number of patients with persistent symptoms on acid suppressive therapy revealed the limitations of single pH monitoring in order to detect gastro-oesophageal reflux. A further development has been multichannel intraluminal impedance (MII), first described by Silny et al in 1991 (21). MII monitoring detects changes in intraluminal electrical conductivity of the oesophageal content providing information on presence of liquids and gas in the oesophagus. The basic component of MII technology is an impedance circuit. Alternating electrical current is applied between two metal rings mounted on a catheter which acts as an electrical isolator. In an empty oesophagus the system will measure a baseline oesophageal impedance of 1500-2000 Ohm. The appearance of liquid bolus in the impedance measuring segment leads to a decline in impedance from baseline, decline that persists as long as the bolus is present between the pair of metal rings. Once the liquid has been cleared from this segment, impedance returns back to its baseline value. Incorporating multiple impedance measuring segments allows detecting the direction of bolus movement: a rapid decline in intra-oesophageal impedance starting in the proximal channels and advancing over time to the distal channel indicates bolus movement in anterograd (aboral) movement. Vice versa a decline starting in the distal oesophagus and moving towards the proximal oesophagus indicates bolus transit in retrograde (oral) direction. The differentiation in direction of bolus movement allows identification of swallows (anterograde) and reflux events (retrograde). Combining multichannel intraluminal impedance with pH monitoring (i.e. impedance-pH) becomes an important tool to identify and validate gastro-oesophageal reflux episodes of all types, independently of their acidity (22). In particular, gastro-oesophageal reflux episodes detected by MII may be differentiated in acidic (if pH drops below 4.0), weakly acidic (if pH is between 4.0 and 7.0) and weakly alkaline (if intra-oesophageal pH during an MII-detected reflux episode is above 7.0) (23). Further studies are required to clarify the clinical relevance of separating non-acid reflux into weakly acidic and weakly alkaline.

Oesophageal impedance-pH monitoring systems use a thin flexible catheter similarly to conventional pH monitoring. The procedure of transnasal insertion of the catheter with intubation of oesophagus and stomach, duration of measurement and patient activities does not differ from standard 24-h pH monitoring. Besides a pH sensor, multiple rings are mounted on the catheter and electrical impedance data are measured between two adjacent rings. The most common design used in clinical practice incorporates impedance rings placed at 2, 4, 6, 8, 10, 14, 16 and 18 cm and a pH sensor at 5 cm from the tip of the catheter. This catheter therefore collects impedance data at 3, 5, 7, 9, 15 and 17 cm above the LOS in addition to pH data at the usual position 5 cm above the LOS (figure 1). During monitoring, the MII-pH catheter is connected to an external data logger and data are sampled at 50 Hz. Data are stored on a flash memory card and transferred to a computer work station with dedicated interpretation software after the monitoring period.

Combined MII-pH monitoring provides information on: 1) the number of gastro-oesophageal reflux episodes (both acid and non-acid), 2) the physical condition of reflux episodes (i.e. liquid, gas or mixed liquid-gas), 3) the height of the reflux column inside the oesophagus and 4) the association between symptoms and reflux episodes (i.e. using S.I. or S.A.P.).

The importance of gastric acid in the development of oesophageal mucosal erosions is well established. Reducing the acid content as done by PPI therapy has shown healing rates close to 90% in GORD (24). On the other hand, non-acid reflux is currently considered to be a relevant cause of symptoms in patients refractory to PPIs rather than having a major role in causing oesophagitis. This is important as one third of patients diagnosed with erosive GORD (24) and up to 65% of patients with non-erosive GORD have persistent symptoms on daily dose of PPI independent of endoscopic integrity (25).

Maine et al (26) and Zerbib et al (27) documented that oesophageal symptoms during 24h combined MII-pH monitoring were associated with ongoing gastro-oesophageal reflux in up to 50% of patients on therapy with PPI twice daily. Recently Backer et al (28) have evaluated patients with ongoing symptoms under acid suppressive therapy and found abnormal MII-pH findings (abnormal

distal oesophageal acid exposure or abnormal number of MII-detected reflux episodes) in 39% of patients taking standard dose PPI once daily. Patients were followed up for at least 3 months after the dose of PPI was increased. The authors were able to document a significantly better symptomatic relieve in patients with abnormal MII-pH results (90%) compared to patients with normal MII-pH findings (43%), concluding that combined MII-pH monitoring facilitates a more focused therapeutic approach to patients with PPI-resistant GORD possibly avoiding PPI overuse.

Whereas pH-monitoring analysis is automatic and very quick, analysis of MII-pH tracings is much more time consuming because it needs manual revision of tracings after the automatic analysis as events other than reflux are included among reflux episodes by the software. In particular automatic analysis overestimates the number of non-acid reflux events resulting in a lower sensitivity and specificity of a positive symptom/reflux association index (S.I. and/or S.A.P.) compared to visual analysis (29). Moreover, the presence of low baseline impedance, as observed in presence of erosive oesophagitis or Barrett's oesophagus (30, 31), makes the analysis more difficult and mistakes easier to occur. These considerations are of relevance because information resulting from MII-pH is important, guiding medical and eventually surgical treatment. However, studies on inter- and intra-observer agreement of manual analysis are scanty in the adult population (32), whereas a few data are available in children (33, 34) and they have not focused on symptom/reflux association analysis. Finally the introduction of a shorter manual analysis would be a valid attempt to save physician's time and Health Care System money. One way to do it would be to concentrate own attention on the short window periods before each symptomatic episode, however so far there are no data.

AIM

Therefore, aims of this study were: 1) to evaluate agreement for number of reflux episodes and symptom/reflux association indexes in MII-pH tracings analysis between and within three experienced observers working in different European Centers, 2) to evaluate the predictivity of a shorter manual analysis targeted to two-minutes window periods before each symptomatic episodes with regards to the traditional analysis.

MATERIALS AND METHODS

PATIENTS POPULATION

Between September 2011 and January 2012 forty consecutive patients off PPI therapy with typical (i.e. heartburn and regurgitation) and/or atypical (i.e. chest pain) oesophageal or extra-oesophageal (i.e. cough and hoarseness) symptoms possibly related to GORD, who have undergone 24h MII-pH in two Centers in Northern Italy and have reported symptoms during the test, were prospectively enrolled. Each Center has provided 20 MII-pH tracings. The study protocol has been approved by the Ethics Committees of both hospitals.

IMPEDANCE-pH EQUIPMENT

Oesophageal impedance-pH monitoring was performed using a multi-channel intraluminal impedance-pH ambulatory system, including an MII-pH catheter (pHersaflex®**,** Sierra Scientific Instruments, CA, USA) containing one distal antimony pH electrode and eight impedance electrodes rings at 2, 4, 6, 8, 10, 14, 16 and 18 cm from the tip of the catheter, and a portable data-logger with impedance amplifier (Medical Measurement Systems, Enschede, The Netherlands).

STUDY PROTOCOL

After an overnight fast, patients attended the Upper GI Physiology Unit of both Centers. Patient's medical history was collected and informed consent was signed. The LOS was located by oesophageal manometry and the impedance-pH catheter was passed transnasally under topical anaesthesia and positioned with the pH electrode 5 cm above the upper border of the LOS. During the MII-pH monitoring, patients were asked to report timing of meals and periods spent in recumbent position on a daily diary card; when a symptom occurred patients were asked to push a button on the portable receiver and to report the exact time on the diary card. When many symptoms were reported, only the principal symptom was taken into account. During the recording period patients were allowed to have a free diet, except for known acidic food and beverages, and to

continue their usual daily activities. Patients returned to the Upper GI Physiology Unit on the following morning for catheter removal.

DATA ANALYSIS

Data stored on the Compact Flash Card were downloaded into a personal computer. Markers of meal periods and of timing in recumbent position were manually inserted. Data were analysed by using an automated reflux detection algorithm (Medical Measurement Systems, Enschede, The Netherlands) and meal periods were excluded from the analysis. Original tracings were anonimyzed and numbered from 01 to 20 (provided by Milano) and from 41 to 60 (provided by Verona) for inter-observer agreement analysis (first analysis). These tracings were subsequently duplicated and numbered in a randomised order from 21 to 40 and from 61 to 80 for intra-observer agreement analysis (second analysis). Each tracing was named adding a code identifying the Center in order to distinguish those reviewed by each observer (observer 1 from Milano, observer 2 from Verona and observer 3 from London). In order to identify tracings difficult to analyse, before performing manual analysis, baseline impedance of each tracing was measured. Oesophageal baseline impedance was assessed as a mean baseline at the two most distal impedance channels (situated at 3 and 5 cm above the LOS), considering a five-minutes window period during the night. Manual analysis (traditional analysis) was performed as follows. At first each observer went through every reflux episode; when the observer did not agree with the automatic analysis, the reflux episode was erased. Afterwards the observer went through the two-minutes window period preceding each symptom marked by the patient in order to identify reflux episodes possibly missed by the automatic analysis. A short manual analysis was also performed as follows: for every tracing, using a separate automated analysis file, each observer erased or added only the reflux episodes in the two-minutes window period preceding each symptom, on the basis of the traditional analysis he/she had already done.

The three observers were all experienced in analyzing MII-pH tracings.

DEFINITIONS OF REFLUX EPISODES

Liquid reflux was defined as a retrograde 50% drop in impedance, starting at the level of the LOS and propagating to at least the next two more proximal impedance-measuring segments. Only liquid reflux lasting at least 3 seconds were taken into account. Gas reflux was defined as a rapid (3 kΩ/second) increase in impedance > 5000 Ω, occurring simultaneously at least in two oesophageal measuring segments, in the absence of swallowing. Mixed liquid-gas reflux was defined as a gas reflux occurring immediately before or during a liquid reflux.

Reflux episodes were characterized by pH drop nadir in: (1) Acid reflux: impedance-detected reflux event with a nadir pH less than 4 (2) Weakly acid reflux: impedance-detected reflux event with a nadir pH between 4 and 7 (3) Weakly alkaline reflux: impedance-detected reflux event with a nadir pH above 7. This classification is according to a recently published consensus report (22). As weakly alkaline refluxes are very unfrequent, in the analysis they were merged with weakly acidic refluxes and considered as non-acid reflux.

GASTROESOPHAGEAL REFLUX PARAMETERS

Impedance and pH data were analysed during both upright and supine positions; meals were excluded from the analysis. Data were used to calculate the number and type of reflux events.

Total number of reflux episodes was considered pathological when $\geq$ 75/24 hours, whereas was considered normal when < 75 (35).

SYMPTOM/REFLUX ASSOCIATION ANALYSIS

S.I. and S.A.P. have been automatically calculated by the software in each patient. Only the association between the principal symptom reported by the patient and acid and non-acid reflux were reported. S.I. and S.A.P. were defined according to Wiener et al. and Weusten et al, respectively (19, 20). S.I. was considered positive when > 50% and S.A.P. when > 95.

STATISTICAL ANALYSIS

All data processing and analyses were carried out with SAS statistical software (version 9.2; SAS Institute, Cary, NC, USA).

Data are presented as median (interquartile range). Inter-observer and intra-observer agreement were assessed by using Cohen's kappa (k) with standard error (SE) and/or expressed as percentages of agreement when the table was not squared . By convention, a kappa between 0.81 and 1.00 is interpreted as indicating excellent agreement. Values of <0.20, 0.21-0.40, 0.41-0.60 and 0.61-0.80 are interpreted as showing poor, fair, moderate and good agreement, respectively.

RESULTS

PATIENTS CHARACTERISTICS

Twenty three of the forty enrolled patients were women and the median age was 55 years (range: 19-71). All the patients completed the study and the recording period was more than 23 hours in all of them. Seventeen patients (42.5%), 7 patients (17.5%) and 16 patients (40%) experienced typical oesophageal, atypical oesophageal and extra-oesophageal symptoms, respectively. Concerning symptoms frequency, half of the patients (20/40; 50%) reported less than 4 symptoms and half of them more or equal than 4 symptoms during the recording period. No patients had a low oesophageal impedance baseline, the median value being 2487 Ω (range: 662-5548 Ω). Table 1 shows variables of the 40 MII-pH tracings as assessed by the 3 observers. The total number of reflux episodes was more than 75 in 5/40 (12.5%), 2/40 (5%) and 5/40 (12.5%) tracings for observer 1, 2, 3, respectively.

AGREEMENT FOR DETECTION OF REFLUX EPISODES

Inter-observer (table 2)

When considering all reflux episodes together agreement between the observers was high for acid refluxes and lower for non-acid ones. Table 2 shows the inter-observer agreement between the observers with regards to the number of reflux episodes divided into acid/non acid and liquid/mixed refluxes. When considering acid reflux, inter-observer agreement was good to excellent, K values being 0.69 between observer 1 and 2, 0.72 between observer 2 and 3 and 0.89 between observer 1 and 3, respectively; when considering non-acid reflux, inter-observer agreement was moderate, K=0.32 between observer 1 and 2, K= 0.36 between observer 2 and 3 and K=0.40 between observer 1 and 3, respectively. Inter-observer agreement for number of liquid and mixed reflux episodes was similar.

Intra-observer (table 3)

When considering all reflux episodes together intra-observer agreement was moderate for observer 1, K=0.40, whereas it was good for observer 2 and 3, K=0.84 and K=0.77, respectively. Table 3 shows the intra-observer agreement between the first and the second analysis with regards to the number of reflux episodes divided into acid/non acid and liquid/mixed refluxes. When considering acid reflux, intra-observer agreement was good to excellent, K values being 0.64, 0.68 and 0.85 for observer 1, 2 and 3, respectively; when considering non-acid reflux, intra-observer agreement was excellent, K=0.84 for observer 2 whereas was moderate for observer 1 and 3, K=0.32 and K=0.40, respectively. Intra-observer agreement for number of liquid and mixed reflux episodes was similarly moderate for observer 1 and 3 and good for the observer 2.

Intra-observer agreement for judging a study normal or pathological on the basis of the number of reflux episodes was almost perfect for all the observers, as the number of studies with a pathological number of reflux episodes remained the same for observer 2 and increased from 5 to 6/40 for observer 1 and 3 in the second analysis.

AGREEMENT FOR SYMPTOM/REFLUX ASSOCIATION ANALYSIS

Inter-observer (table 4)

Agreement between the three observers was high for acid refluxes and lower for non acid ones (table 4). In particular when considering acid reflux, inter-observer agreement between the observers was excellent for both S.I. and S.A.P., being K=1.0. When considering non-acid reflux, inter-observer agreement between the observers was moderate for S.A.P., being K=0.35 between observer 1 and 2, K=0.36 between observer 1 and 3, K=0.23 between observer 2 and 3. K value for inter-observer agreement between the observers for S.I. was not measurable because the table was not squared; however the percentage of agreement between the observers was $\geq$ 92.5%.

Intra-observer (table 5)

Agreement between the first and the second analysis was generally high for all three observers (table 5). When considering acid reflux, intra-observer agreement was good to excellent for both S.I., K=0.77 for observer 1 and 2 and K=1.0 for observer 3, and S.A.P., K=0.80, K=0.72 and K=1.0 for observer 1, 2 and 3, respectively. When considering non acid reflux intra-observer agreement for S.I. was not measurable because the table was not squared; however the percentage of agreement was $\geq$ 95% for each observer. Intra-observer agreement for S.A.P. was moderate for observer 1 and 3, K=0.48 and K=0.54 and remained good for observer 2, K=0.79.

PREDICTIVITY OF A SHORT ANALYSIS

Short analysis showed a high predictivity of the traditional one although a bit lower for S.I. for acid reflux (table 6). Regarding S.I. the percentage of agreement for acid reflux was 87.5%, 90% and 87.5% for observer 1, 2 and 3, respectively, whereas it was higher for non acid reflux, being 100%, 95%, and 100% for observer 1, 2 and 3 respectively. Regarding S.A.P. the percentage of agreement for acid reflux were 100%, 97.5% and 100% for observer 1, 2 and 3 respectively, and for non acid reflux were 97.5 %, 92.5 % and 100 % for observer 1, 2 and, 3 respectively.

DISCUSSION

The major finding of this study was that, in MII-pH tracings analysis of patients with GORD, inter- and intra-observer agreement for number of acid reflux episodes and for symptom/reflux association indexes (i.e. S.I. and S.A.P.) for acid reflux was high, whereas was lower for non acid reflux. Another important finding of this study was that a short manual analysis was highly predictive of the traditional one.

So far only few studies are available on inter- and intra- observer agreement analysis and all of them focused on agreement for number of reflux episodes; no studies on agreement for symptom/reflux association indexes and on predictivity of a short versus the traditional analysis are reported.

Our study demonstrated that inter-observer agreement for number of reflux episodes was good. When considering separately acid and non-acid reflux, we observed that agreement remained good for acid reflux but it became poorer for non-acid reflux. A possible explanation of this finding could be that non-acid refluxes are more challenging to be evaluated, therefore the observer is tempted to erase them more frequently than acid refluxes which are easier to be interpreted for the presence of pH drop below 4. In particular the observer of the Verona Center was more restrictive in the acceptance of an event as non acid reflux. In the multi-center pediatric study conducted by Loots et al (34), inter-observer agreement for number of reflux episodes was lower (K=0.46) than ours as only 42% of the reflux episodes were confirmed by the majority of the ten observers; moreover the authors did not evaluate whether inter-observer agreement differed between acid and non-acid reflux. In another multi-center study conducted in the pediatric population four observers reviewed 24 tracings and the authors found an excellent inter-observer agreement (k=0.84), however each tracing was analyzed only by two reviewers and again differences between acid and non acid reflux were not evaluated (33). The only study performed in the adult population showed a good mean inter-observer agreement (K=0.72) for number of reflux episodes (32); an important limit of this

study was that each tracing was assessed by four reviewers from the same institution and probably trained by the same person, thus introducing a potential bias in agreement results.

Concerning intra-observer agreement for number of reflux episodes, we found a moderate to excellent agreement depending on observers. Intra-observer agreement was higher for acid refluxes than for non acid refluxes for observer 1 and 3, whereas it was the contrary for observer 2. Loots et al. (34) evaluated the intra-observer agreement for number of reflux episodes only in three of the ten observers and they showed agreements' values similar to ours but, as previously reported, a subanalysis on acid and non acid refluxes was not performed. Pilic et al. (33) found a good median intra-observer agreement (k=0.88), although it was calculated in only 6 of the 24 tracings performed, whereas the only other study conducted in the adult population did not evaluate intra-observer agreement (32).

Agreement for judging a study normal or pathological on the basis of number of reflux episodes was high in our study. However this is only a descriptive report that should be cautiously interpreted because in the majority of the patients the number of reflux episodes detected was lower than 75. In the series reported by Loots et al. (34) half of the ten tracings analyzed by their observers had the number of reflux episodes lower than 75, although a substantial agreement (k= 0.70) in defining a study normal or pathological was found. However the authors pointed out that this result can be considered poor when being used to guide clinical decision making.

So far, our is the first multi-center study reporting the agreement on symptom/reflux association indexes for both acid and non acid reflux. As it could be expected inter-observer agreement between the observers for S.I. and S.A.P. for acid reflux was excellent. Conversely, inter-observer agreement for non acid reflux was fair, reflecting the lower agreement in evaluation of number of non acid refluxes than acid ones. Moreover, intra-observer agreement for S.I. and S.A.P. for acid reflux was good to excellent, remaining good for S.A.P. for observer 2 but becoming moderate for observer 1 and 3 for non-acid reflux. This trend reflects the intra-observer agreement for number of reflux episodes previously described and this finding support the assumption that non acid refluxes are

more challenging to be evaluated even by experienced observers. This evidence should move the factories to improve the software used for automatic analysis on one hand and to improve the training of the new reviewers who wants to approach this technique on the other. Concerning the last point, consensus among experts with the purpose to better define MII-pH criteria in order to evaluate reflux episodes should be planned (34). A possible limitation of analysis for agreement on symptom/reflux association indexes could be the low number of patients with positive S.I. or S.A.P. As a matter of fact S.I. was positive in 17.5%, 12.5% and 15% of the tracings for observer 1, 2 and 3, respectively and S.A.P. in 35%, 17.5% and 27.5% of the tracings for observer 1, 2 and 3, respectively. However it needs to be pointed out that our patients are representative of those referred to a Tertiary Center and they were similar between the two Centers regarding S.I. and S.A.P. variables .

In our study no tracings difficult to analyse were found based solely on impedance baseline, as median impedance baseline was 2487 Ω (range: 662-5548 Ω) because none of the patients had severe erosive oesophagitis or Barrett's oesophagus. Impedance baseline was measured considering a five-minutes window period during night-time, as Kessing et al. showed that impedance baseline did not differ between daily and night period (31). Loots et al reported a better agreement for number of reflux episodes when easy tracings were compared to challenging ones (i.e. the ones with low baseline impedance). Probably other impedance variables can be assessed in order to consider a tracing difficult to be analyzed (i.e. unstable baseline) as recently proposed by Loots et al. (34) but they were not be considered in our study.

Finally, to our knowledge this is the first study evaluating the predictivity of a short analysis compared to the traditional one. Short analysis showed high concordance with the traditional one for both S.I. and S.A.P. for acid and non acid reflux. This is a very important finding because manual MII-pH analysis is a time consuming procedure.

In conclusion our study has demonstrated a good inter- and intra-observer agreement for number of reflux episodes detection and for symptom/reflux association indexes, especially when considering

acid reflux. Moreover, high concordance was found for symptom/reflux association indexes between a short manual analysis, less time-consuming, and the longer time-consuming traditional analysis, suggesting that the short one could be used in routine clinical practice to save physician's time and Health Care System money.

TABLES

Table 1. Variables of the 40 MII-pH tracings as assessed by the three observers.

	Observer 1 (Milano)	Observer 2 (Verona)	Observer 3 (London)
AC reflux episodes (n)	25 (1-90)	22 (0-83)	25 (1-91)
NA reflux episodes (n)	19 (2-89)	8 (1-76)	21 (2-99)
Total reflux episodes (n)	44 (1-90)	30 (0-83)	46 (1-99)
Positive SI for AC, (%)	5 (12.5)	5 (12.5)	5 (12.5)
Positive SI for NA, (%)	2 (5)	0 (0)	1 (2.5)
Positive SAP for AC, (%)	6 (15)	5 (12.5)	6 (15)
Positive SAP for NA, (%)	8 (20)	2 (5)	5 (12.5)

AC: acid; NA: non acid; n: expressed as median and range

Table 2. K values with standard error (SE) and percentage of discordance (in brackets) between observers with regards to the number of reflux episodes divided into acid/non acid and mixed (liquid + gas).

	Number of reflux episodes			
	Acid	**Non Acid**	**Liquid**	**Mixed**
Observer 1 Vs Observer 2	0.69; SE 0.02 (11%)	0.32; SE 0.19 (14%)	0.74; SE 0.12 (9%)	0.40; SE 0.19 (16%)
Observer 1 Vs Observer 3	0.89; SE 0.18 (4%)	0.40; SE 0.23 (16%)	0.82; SE 0.18 (6%)	0.36; SE 0.19 (17.5%)
Observer 2 Vs Observer 3	0.72; SE 0.18 (9.5%)	0.36; SE 0.12 (17.5%)	0.70; SE 0.20 (9.5%)	0.30; SE 0.23 (14.5%)

Table 3. K values with standard error (SE) and percentage of discordance (in brackets) between the first and the second analysis with regards to the number of reflux episodes divided into acid/non acid and mixed (liquid + gas).

	Number of reflux episodes			
	Acid	**Non Acid**	**Liquid**	**Mixed**
Observer 1				
	0.64; SE 0.08 (1.8%)	0.32; SE 0.04 (17.9%)	0.36; SE 0.05 (13.7%)	0.41; SE 0.04 (9.2%)
Observer 2				
	0.68; SE 0.03 (7.6%)	0.84; SE 0.01 (7.6%)	0.84; SE 0.02 (8.1%)	0.83; SE 0.01 (7.4%)
Observer 3				
	0.85; SE 0.07 (0.4%)	0.4; SE 0.05 (7.7%)	0.47; SE 0.09 (4.4%)	0.49; SE 0.05 (4.3%)

Table 4. K values with standard error (SE) and percentage of discordance (in brackets) between observers, with regards to S.I. and S.A.P. divided into acid/non acid

	S.I.		**S.A.P.**	
	Acid	**Non Acid**	**Acid**	**Non Acid**
Observer 1 Vs Observer 2	1.0; SE 0 (0%)	- (5%)	1.0; SE 0 (0%)	0.35; SE 0.19 (15%)
Observer 1 Vs Observer 3	1.0; SE 0 (0%)	- (7.5%)	1.0; SE 0 (0%)	0.36; SE 0.19 (17.5%)
Observer 2 Vs Observer 3	1.0; SE 0 (0%)	- (5%)	1.0; SE 0 (0%)	0.23; SE 0.23 (12.5%)

Table 5. K values with standard error (SE) and percentage of discordance (in brackets) between the first and the second analysis, with regards to S.I. and S.A.P. divided into acid/non acid

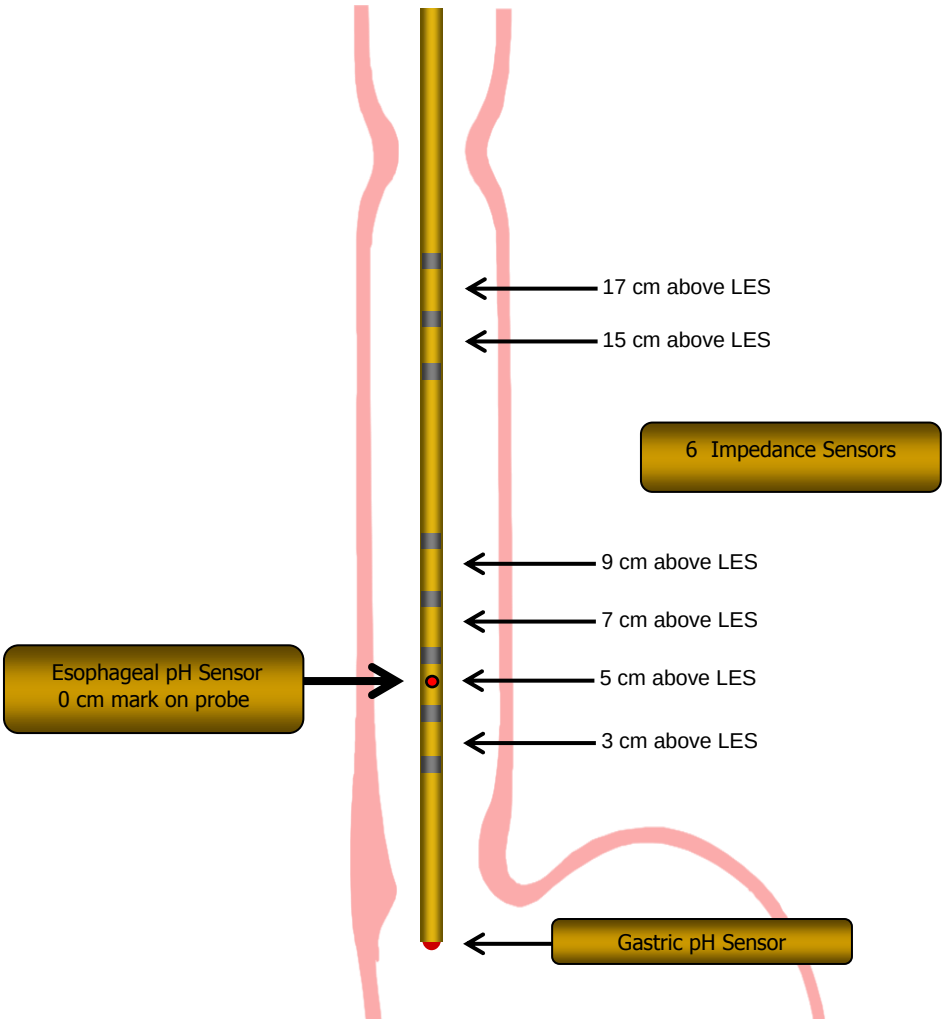
	S.I.		**S.A.P.**	
	Acid	**Non Acid**	**Acid**	**Non Acid**
Observer 1	0.77; SE 0.15 (5%)	0.65; SE 0.31 (2.5%)	0.80; SE 0.13 (5%)	0.48; SE 0.18 (15%)
Observer 2	0.77; SE 0.15 (5%)	- (5%)	0.72; SE 0.15 (7.5%)	0.79; SE 0.20 (2.5%)
Observer 3	1.0; SE 0 (0%)	- (2.5%)	1.0; SE 0 (0%)	0.54; SE 0.20 (10%)

Table 6. Percentage of agreement with confidence interval (in brackets) between short and traditional analysis.

	S.I.		**S.A.P.**	
	Acid	**Non Acid**	**Acid**	**Non Acid**
Observer 1	87.5 (72-95)	100 (9-100)	100 (91-100)	97.5 (85-100)
Observer 2	90.0 (75-97)	95 (82-99)	97.5 (85-100)	92.5 (78-98)
Observer 3	87.5 (72-95)	100 (91-100)	100 (91-100)	100 (91-100.0)

FIGURES

Figure. 1. Oesophageal impedance-pH catheter.



REFERENCES

1. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. Am J Gastroenterol 2006;101(8):1900-20; quiz 1943.
2. Johnsson F, Weywadt L, Solhaug JH, Hernqvist H, Bengtsson L. One-week omeprazole treatment in the diagnosis of gastro-oesophageal reflux disease. Scand J Gastroenterol 1998;33(1):15-20.
3. Fass R, Ofman JJ, Gralnek IM, Johnson C, Camargo E, Sampliner RE, et al. Clinical and economic assessment of the omeprazole test in patients with symptoms suggestive of gastroesophageal reflux disease. Arch Intern Med 1999;159(18):2161-8.
4. Johnsson F, Joelsson B, Gudmundsson K, Greiff L. Symptoms and endoscopic findings in the diagnosis of gastroesophageal reflux disease. Scand J Gastroenterol 1987;22(6):714-8.
5. Hewson EG, Sinclair JW, Dalton CB, Richter JE. Twenty-four-hour esophageal pH monitoring: the most useful test for evaluating noncardiac chest pain. Am J Med 1991;90(5):576-83.
6. Kahrilas PJ, Quigley EM. Clinical esophageal pH recording: a technical review for practice guideline development. Gastroenterology 1996;110(6):1982-96.
7. Fass R, Fennerty MB, Ofman JJ, Gralnek IM, Johnson C, Camargo E, et al. The clinical and economic value of a short course of omeprazole in patients with noncardiac chest pain. Gastroenterology 1998;115(1):42-9.
8. Ho KY, Ng WL, Kang JY, Yeoh KG. Gastroesophageal reflux disease is a common cause of noncardiac chest pain in a country with a low prevalence of reflux esophagitis. Dig Dis Sci 1998;43(9):1991-7.
9. El-Serag HB, Lee P, Buchner A, Inadomi JM, Gavin M, McCarthy DM. Lansoprazole treatment of patients with chronic idiopathic laryngitis: a placebo-controlled trial. Am J Gastroenterol 2001;96(4):979-83.

10. Qadeer MA, Phillips CO, Lopez AR, Steward DL, Noordzij JP, Wo JM, et al. Proton pump inhibitor therapy for suspected GERD-related chronic laryngitis: a meta-analysis of randomized controlled trials. Am J Gastroenterol 2006;101(11):2646-54.

11. Tack J, Fass R. Review article: approaches to endoscopic-negative reflux disease: part of the GERD spectrum or a unique acid-related disorder? Aliment Pharmacol Ther 2004;19 Suppl 1:28-34.

12. Hirano I, Richter JE. ACG practice guidelines: esophageal reflux testing. Am J Gastroenterol 2007;102(3):668-85.

13. Bautista J, Fullerton H, Briseno M, Cui H, Fass R. The effect of an empirical trial of high-dose lansoprazole on symptom response of patients with non-cardiac chest pain--a randomized, double-blind, placebo-controlled, crossover trial. Aliment Pharmacol Ther 2004;19(10):1123-30.

14. Dickman R, Emmons S, Cui H, Sewell J, Hernandez D, Esquivel RF, et al. The effect of a therapeutic trial of high-dose rabeprazole on symptom response of patients with non-cardiac chest pain: a randomized, double-blind, placebo-controlled, crossover trial. Aliment Pharmacol Ther 2005;22(6):547-55.

15. Bredenoord AJ, Dent J. Proton pump inhibitor-therapy refractory gastro-oesophageal reflux disease patients, who are they? Gut 2007;56(4):593-4.

16. Kahrilas PJ, Quigley EMM. American Gastroenterological Association medical position statement: guidelines on the use of esophageal pH recording. Gastroenterology 1996;110:1981-1996.

17. DeVault KR, Castell DO. Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. The Practice Parameters Committee of the American College of Gastroenterology. Am J Gastroenterol 1999;94(6):1434-42.

18. Pandolfino JE, Richter JE, Ours T, Guardino JM, Chapman J, Kahrilas PJ. Ambulatory esophageal pH monitoring using a wireless system. Am J Gastroenterol 2003;98(4):740-9.

19. Wiener GJ, Richter JE, Copper JB, Wu WC, Castell DO. The symptom index: a clinically important parameter of ambulatory 24-hour esophageal pH monitoring. Am J Gastroenterol 1988;83(4):358-61.

20. Weusten BL, Roelofs JM, Akkermans LM, Van Berge-Henegouwen GP, Smout AJ. The symptom-association probability: an improved method for symptom analysis of 24-hour esophageal pH data. Gastroenterology 1994;107(6):1741-5.

21. Silny J. Intraluminal multiple electric impedance procedure for measurement of gastrointestinal motility. J Gastrointest Motil 1991;3:151-62.

22. Sifrim D, Castell D, Dent J, Kahrilas PJ. Gastro-oesophageal reflux monitoring: review and consensus report on detection and definitions of acid, non-acid, and gas reflux. Gut 2004;53(7):1024-31.

23. Tack J, Koek G, Demedts I, Sifrim D, Janssens J. Gastroesophageal reflux disease poorly responsive to single-dose proton pump inhibitors in patients without Barrett's esophagus: acid reflux, bile reflux, or both? Am J Gastroenterol 2004;99(6):981-8.

24. Castell DO, Kahrilas PJ, Richter JE, Vakil NB, Johnson DA, Zuckerman S, et al. Esomeprazole (40 mg) compared with lansoprazole (30 mg) in the treatment of erosive esophagitis. Am J Gastroenterol 2002;97(3):575-83.

25. Dean BB, Gano AD, Jr., Knight K, Ofman JJ, Fass R. Effectiveness of proton pump inhibitors in nonerosive reflux disease. Clin Gastroenterol Hepatol 2004;2(8):656-64.

26. Mainie I, Tutuian R, Shay S, Vela M, Zhang X, Sifrim D, et al. Acid and non-acid reflux in patients with persistent symptoms despite acid suppressive therapy: a multicentre study using combined ambulatory impedance-pH monitoring. Gut 2006;55(10):1398-402.

27. Zerbib F, Roman S, Ropert A, des Varannes SB, Pouderoux P, Chaput U, et al. Esophageal pH-impedance monitoring and symptom analysis in GERD: a study in patients off and on therapy. Am J Gastroenterol 2006;101(9):1956-63.

28. Becker V, Bajbouj M, Waller K, Schmid RM, Meining A. Clinical trial: persistent gastro-oesophageal reflux symptoms despite standard therapy with proton pump inhibitors - a follow-up study of intraluminal-impedance guided therapy. Aliment Pharmacol Ther 2007;26(10):1355-60.

29. Roman S, Bruley des Varannes S, Pouderoux P, Chaput U, Mion F, Galmiche JP, et al. Ambulatory 24-h oesophageal impedance-pH recordings: reliability of automatic analysis for gastro-oesophageal reflux assessment. Neurogastroenterol Motil 2006;18(11):978-86.

30. Farre R, Blondeau K, Clement D, Vicario M, Cardozo L, Vieth M, et al. Evaluation of oesophageal mucosa integrity by the intraluminal impedance technique. Gut 2011;60(7):885-92.

31. Kessing BF, Bredenoord AJ, Weijenborg PW, Hemmink GJ, Loots CM, Smout AJ. Esophageal acid exposure decreases intraluminal baseline impedance levels. Am J Gastroenterol 2011;106(12):2093-7.

32. Ravi K, DeVault KR, Murray JA, Bouras EP, Francis DL. Inter-observer agreement for multichannel intraluminal impedance-pH testing. Dis Esophagus 2010;23(7):540-4.

33. Pilic D, Hofs C, Weitmann S, Noh F, Frohlich T, Skopnik H, et al. Inter- and intraobserver agreement in 24-hour combined multiple intraluminal impedance and pH measurement in children. J Pediatr Gastroenterol Nutr 2011;53(3):255-9.

34. Loots CM, van Wijk MP, Blondeau K, Dalby K, Peeters L, Rosen R, et al. Interobserver and intraobserver variability in pH-impedance analysis between 10 experts and automated analysis. J Pediatr 2012;160(3):441-446 e1.

35. Zerbib F, des Varannes SB, Roman S, Pouderoux P, Artigue F, Chaput U, et al. Normal values and day-to-day variability of 24-h ambulatory oesophageal impedance-pH monitoring in a Belgian-French cohort of healthy subjects. Aliment Pharmacol Ther 2005;22(10):1011-21.