Radiological evaluation of esophageal function in dysphagia with special emphasis on achalasia

Mats Andersson

Department of Radiology, Institute of Clinical Sciences,

The Sahlgrenska Academy, University of Gothenburg, Sweden

Gothenburg 2008

Correspondence to:

Dr Mats Andersson
Department of Radiology
Sahlgrenska University Hospital
SE-413 45 Göteborg, Sweden
mats.n.andersson@vgregion.se

RADIOLOGICAL EVALUATION OF ESOPHAGEAL FUNCTION IN DYSPHAGIA WITH SPECIAL EMPHASIS ON ACHALASIA

Mats Andersson, MD

Department of Radiology, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, 2008.

Abstract:

In idiopathic achalasia, degeneration of the inhibitory innervation of the esophageal smooth muscle results in absence of primary peristalsis and in incomplete relaxation of the lower esophageal sphincter (LES). All treatments for achalasia aim at reducing the pressure gradient across the LES, thus facilitating esophageal emptying by gravity. Objective evaluation of the response to treatment is important, since persistent poor emptying may lead to progressive deterioration of esophageal function. The timed barium esophagogram (TBE) has been introduced as a standardized technique for evaluating esophageal emptying in patients with achalasia and the aim of the present thesis was to validate this new diagnostic test.

I. In order to investigate the reproducibility and observer variation of TBE, 21 patients with achalasia were examined by repeat TBE median 8 days apart. Radiographs of the esophagus were taken 1, 2 and 5 minutes after patients had ingested 250 ml of barium. The height and width of the barium column and the rate of change over time were recorded. The static parameters were reproducible between studies, but the dynamic data were not (correlation coefficient of only 0.50). There was excellent intra- and interobserver agreement for all measured variables. Control subjects (n= 8) uniformly achieved complete esophageal emptying within 2 minutes.

II. To describe TBE characteristics in patients with newly diagnosed achalasia, and to correlate these to clinical and manometric variables, 46 patients were examined. All patients showed markedly delayed emptying of barium from the esophagus. Emptying, expressed as volume of barium, showed significant inverse correlation with the resting and the maximal relaxing pressure of the LES at manometry (r= -0.34 and r= -0.54, respectively) and with the duration of symptoms (r= -0.36).

III. TBE was prospectively applied in a randomized trial comparing pneumatic dilatation (n= 26) and laparoscopic myotomy (n= 25) in patients with newly diagnosed achalasia. Following therapy, TBE parameters did not differ significantly between treatment groups. Significant correlations were found between the height of the barium column at 1 minute and the symptom scores for "dysphagia for liquids" (r= 0.47), "chest pain" (r= 0.42) and the "Watson dysphagia score" (r= 0.46) at the end of follow-up (median 18 months). Patients with less than 50% improvement in barium column height at 1 minute had a 40% risk of treatment failure during follow-up.

IV. A modified TBE-technique was applied in a case series of 7 patients operated for hypopharyngeal or proximal esophageal cancer with radical resection and reconstruction with a free jejunal transplant. Radiographic signs of disturbed bolus transport through the jejunal transplant were found in all patients, but the patients only reported mild dysphagia symptoms on clinical assessment. One possible explanation for this discrepancy might be diminished visceral sensation in the denervated jejunal transplant.

In conclusion, we found that TBE is an easily performed and reproducible technique for the objective evaluation of esophageal emptying before and after treatment for achalasia. However, the impact of routinely performing TBE on the long-term outcome of achalasia patients needs to be studied in further prospective trials.

Keywords: Achalasia, dysphagia, radiography, barium esophagogram, reproducibility of findings, observer variation, esophageal neoplasms, laparoscopic myotomy, balloon dilatation.

ISBN 978-91-628-7566-4 Gothenburg 2008

List of papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I. Kostic S, Andersson M, Hellström M, Lönroth H, Lundell L.
Timed barium esophagogram in the assessment of patients with achalasia: Reproducibility and observer variation.
Diseases of the Esophagus 2005; 18: 96-103.

II. Andersson M, Kostic S, Ruth M, Lönroth H, Kjellin A, Hellström M, Lundell L.
Characteristics of the timed barium esophagogram in newly diagnosed idiopathic achalasia. Clinical and manometric correlates.
Acta Radiol 2007; 48: 2-9.

III. Andersson M, Lundell L, Kostic S, Ruth M, Lönroth H, Kjellin A, Hellström M.
Evaluation of the response to treatment in patients with idiopathic achalasia by the timed barium esophagogram: results from a randomized clinical trial.
Provisionally accepted for publication in Diseases of the Esophagus.

IV. Bergquist H, Andersson M, Ejnell H, Hellström M, Lundell L, Ruth M.
Functional and radiological evaluation of free jejunal transplant reconstructions after radical resection of hypopharyngeal or proximal esophageal cancer.
World Journal of Surgery 2007; 31: 1988-1995.

Contents

Abbreviations

CV	Coefficient of Variation
DES	Diffuse Esophageal Spasm
DMN	Dorsal Motor Nucleus
EBP	Evidence Based Practice
EORTC	European Organisation for Research and Treatment of Cancer
FEES	Fiberoptic Endoscopic Examination of Swallowing
GEJ	Gastro-Esophageal Junction
GERD	Gastro-Esophageal Reflux Disease
GSRS	Gastrointestinal Symptom Rating Scale
HM	Heller Myotomy
HRQL	Health Related Quality of Life
IBS	Irritable Bowel Syndrome
IEM	Ineffective esophageal motility
KPSSI	Karnofsky Performance Status Scale Index
LES	Lower Esophageal Sphincter
LoA	Limits of Agreement
MII	Multichannel Intraluminal Impedance
MRI	Magnetic Resonance Imaging
NE	Nutcracker Esophagus
NO	Nitric Oxide
NSEMD	Nonspecific Esophageal Motility Disorder
PLE	Pharyngo-Laryngo-Esophagectomy
PD	Pneumatic Dilatation
QLQ-C30	Quality of Life Questionnaire-Core 30
QLQ-OES18	Quality of Life Questionnaire-Oesophageal module 18
RCT	Randomized Controlled Trial
SD	Standard Deviation
TBE	Timed Barium Esophagogram
VIP	Vasoactive Intestinal Polypeptide

Introduction

Swallowing, known scientifically as deglutition, is the process by which food is transported from the mouth to the stomach. When functioning properly, we are seldom aware of the act of swallowing. On average, a normal individual swallows roughly between 600 and 1500 times every day. However, when dysfunction of swallowing occurs, this may lead to serious medical problems due to malnutrition, dehydration, aspiration pneumonia, or choking episodes. Swallowing impairment not only affects the individual's ability to enjoy food, but is also associated with psychosocial consequences. The act of dining is a social process often shared with others and problems with chewing and swallowing may affect the interaction with friends and family, productivity at work and the overall quality of life.

Dysphagia

The word dysphagia is derived from the Greek roots dys (with difficulty) and phagia (to eat) and is defined as the subjective sensation of swallowing difficulty during passage of a solid or liquid bolus from the mouth to the stomach. Complaints of dysphagia are common, especially in elderly people (1). Approximately 15% of adults older than 65 years have dysphagia and up to 30% of hospitalized patients experience swallowing problems (2, 3). Dysphagia can be classified into two major types: oropharyngeal dysphagia and esophageal dysphagia (4). A carefully conducted patient history will enable the physician to narrow the differential diagnoses to an anatomic or pathophysiologic-related diagnosis in the majority of patients. Patients with *oropharyngeal dysphagia* often present with difficulty in initiating swallowing, nasal regurgitation, choking, and associated coughing. Although the cause in some instances may be a local structural lesion such as a diverticulum, web or a tumor, oropharyngeal dysphagia is most often associated with diseases in the central nervous system, such as stroke and Parkinson's disease, or other chronic neuromuscular disorders. Patients with *esophageal dysphagia* present with a sensation of blockage and a feeling that food stops or "sticks" after swallowing. However, the patient's subjective assessment of the site of dysphagia does not always correlate with the site of the actual pathology. Not infrequently, abnormalities of the mid or distal esophagus may cause referred dysphagia to the pharynx or upper thorax (5). Mechanical obstruction of the esophagus is typically associated with dysphagia of solid food but not liquids. Peptic stricture, carcinoma and a lower esophageal ring (Schatzki's ring) are the most common obstructive lesions. A long history of intermittent, solid-bolus dysphagia is highly suggestive of an esophageal mucosal ring. However, in a young male patient this presentation is commonly attributable to a multiringed esophagus associated with eosinophilic esophagitis (6). Malignant dysphagia usually presents with a short history of progressive dysphagia that is frequently associated with weight loss. Patients with motility disorders most often experience gradually progressive difficulty in swallowing both solid food and liquids (4).

Further testing is often indicated to confirm the diagnosis and the choice of diagnostic technique depends upon the presenting clinical features. A barium esophagogram identifies structural obstructive lesions and has the advantage of assessing motility better than endoscopy. On the other hand, the best assessment of the esophageal mucosa is provided by endoscopy (7). Any mass or other lesion identified at a barium study should initiate endoscopy with biopsy and cytology (4). Manometry assesses motor function of the esophagus and is indicated if no abnormality is identified by barium study or endoscopy (8). Radionuclide studies may also be used to evaluate transit function through the esophagus. In recent years, magnetic resonance imaging (MRI) has been used to assess esophageal function, and further advances can be anticipated with this investigation modality (9). Functional aspects of swallowing in patients with oropharyngeal dysphagia can be assessed by a specialized videoendoscopic technique known as "fiberoptic endoscopic examination of swallowing" (FEES) (10). Patients at risk for silent aspiration may also benefit from videoradiographic studies that are performed by a team composed of a radiologist, an otolaryngologist and a speech pathologist with expertise in swallowing disorders. This examination helps to objectively identify the nature of the swallowing problems and is useful to assess treatment options, such as postural techniques and swallowing maneuvers (11).

The esophagus

The adult human esophagus is an approximately 25-cm long muscular tube that has cervical, thoracic, and abdominal parts. The esophagus wall is composed of striated muscle in the upper part, smooth muscle in the lower part, and a mixture of the two in the middle. The muscular coat (muscularis propria) consists of an internal layer of circular fibers and an external layer of longitudinal fibers. There is also a less prominent layer of muscle oriented longitudinally and found between the mucosa and the muscularis propria called the muscularis mucosa. The lower esophageal sphincter (LES) is a high-pressure zone located where the esophagus merges with the stomach. The LES is a functional sphincter composed of an intrinsic and an extrinsic component. The extrinsic component consists of the diaphragm muscle, which functions as an adjunctive external sphincter (12).

The motor innervation of the esophagus is predominantly via the vagus nerve. The smooth muscle of the distal esophagus and the LES is innervated by preganglionic, cholinergic fibers that originate in the dorsal motor nucleus (DMN) in the brainstem and terminate in the myenteric (Auerbach's) plexus. The ganglia of the myenteric plexus lie between the longitudinal and the circular muscle layers and postganglionic neurons subsequently innervate the esophageal wall and LES. Postganglionic excitatory neurons release acetylcholine while postganglionic inhibitory neurons release nitric oxide (NO) and vasoactive intestinal polypeptide (VIP) (13). Under resting conditions (i.e. in between swallows) the LES is in a tonic contractile state. The act of swallowing is associated with the activation of the involuntary swallowing reflex. Once activated by this reflex, the swallowing center neurons send patterned discharges of inhibition and excitation to motor nuclei of the cranial nerves. The inhibitory pathway neurons are activated first, which result in inhibition of all ongoing activity in the esophagus and relaxation of the LES. Peristalsis is the result of the coordinated relaxation and contraction mediated by the inhibitory and excitatory myenteric plexus neurons along the length of the esophagus (14).

Motility disorders of the esophagus

Abnormal esophageal motility can cause esophageal dysphagia, although significant primary motility disorders are a far less common cause of dysphagia than mechanical obstruction (15). However, the relation between symptoms and manometric patterns in motility disorders is poorly defined and the response to therapy unpredictable (16). In patients with diffuse esophageal spasm (DES), intermittent chest pain and dysphagia are the most common presenting symptoms. The diagnosis is established by manometry and the results of radiologic studies are variable. Intermittently weakened or absent primary esophageal peristalsis and multiple simultaneous, nonperistaltic contractions of varying severity may suggest the diagnosis (17). Patients with "nutcracker esophagus" (NE) usually present with chest pain. Manometry typically shows peristaltic waves with significantly elevated amplitude (> 180 mm Hg). As these patients have normal peristalsis, barium studies are usually normal. Treatment is similar to that of DES and is primarily medical (16). Nonspecific esophageal motility disorder (NSEMD) is diagnosed in a substantial number of patients, having motor abnormalities not sufficiently characteristic to be classified as any of the specific motor disorders previously described. These motor abnormalities include nonpropulsive, tertiary waves, or interrupted/ retrograde primary wave, or mild delay in transit (18). It has been shown that the ineffective esophageal motility in most of these patients is the result of esophageal hypocontraction (contraction amplitudes of < 30 mm Hg) and it has therefore been recommended to replace the term NSEMD by the more accurate term "ineffective esophageal motility" (IEM) (19).

In patients with secondary motility disorders, the esophageal motor disturbance is a manifestation of a systemic disease or the result of medication. Classic examples of systemic conditions that can give rise to esophageal motility disorders are scleroderma, diabetes mellitus and alcoholism. There is some evidence that very old patients in their 80s and 90s can develop a condition called "presbyesophagus" that is characterized by a decrease in peristaltic amplitude and an increased frequency of nonpropulsive contractions (20).

Achalasia

Achalasia is the only primary motility disorder of the esophagus with an established pathology. Achalasia is uncommon, with an incidence in the western hemisphere in the range of 0.4-1.1/ 100.000 (21). Historically, the oldest available description of achalasia is from 1672 by Sir Thomas Willis (22). He described a 38-year old man from Oxford with severe swallowing difficulties. Sir Willis came to the conclusion that the patient's problem was caused by lower esophageal narrowing leading to a massive dilatation of the esophagus. He successfully treated the patient with a dilatator made of whalebone, with a sponge at the distal end, with which the patient forced food into the stomach after each meal. The disease was first termed achalasia (Greek for "failure to relax") by Arthur Hurst in 1927 (23). Achalasia affects men and women equally and may occur at any age. However there seems to be two incidence peaks, one minor peak is seen in the 20-40 age range and the predominant one in the seventh decade (24). Less than 5% of cases occur in children <15 years of age (25). The most common symptoms include dysphagia to both solids and liquids (82-100%), regurgitation (56-97%), weight loss (30-91%), chest pain (17-95%), and heartburn (27-42%) (26).

The physiologic alterations in achalasia result from damaged innervation and neuroanatomic data suggest the esophageal myenteric plexus as the primary neurologic target. However, the superficial nature of esophageal biopsies is not suitable for the evaluation of the myenteric plexus, which is situated deep in the muscularis propria (26). Pathological changes identified at necropsy or from myotomy include a patchy inflammatory response consisting of T-lymphocytes, eosinophils and mast cells, loss of ganglion cells, and some degree of myenteric neural fibrosis (27). The end result of these inflammatory changes is a selective loss of post-ganglionic inhibitory neurons containing NO and VIP. Since post-ganglionic cholinergic neurons are spared, cholinergic stimulation continues unopposed, resulting in insufficient LES relaxation (28). Aperistalsis is caused by loss of the latency gradient that permits sequential contractions along the esophageal body, a process mediated by NO (16). Although not proven, current evidence suggests that some initial insult to the esophagus, perhaps a viral infection, results in myenteric plexus inflammation. This inflammation then leads to an autoimmune response in susceptible individuals who may be genetically predisposed. Subsequently, chronic inflammation leads to destruction of the inhibitory myenteric ganglion cells resulting in the clinical syndrome of achalasia (29).

Although little is known concerning the natural history of the disease, it is assumed that long-term poor esophageal emptying and food retention leads to progressive esophageal dilatation. In a retrospective study, Csendes et al studied 14 patients who refused treatment at the time of diagnosis (30). The patients came for clinical consultations mean 5 years after the initial diagnosis and at that time, all had severe dysphagia and poor quality of life, with a mean loss of weight of 12 kg. By radiography, the maximal diameter of the esophageal body had increased from mean 29.2 mm at presentation to mean 59.5 mm, corresponding to a rate of "dilatation" of 6.1 mm/year over a period of 5 years. With increasing degree of esophageal dilatation, regurgitation, especially nocturnal regurgitation, becomes a more prominent symptom, with the potential of pulmonary complications due to chronic aspiration (31, 32). Although there are no data on the prevention of complications, theoretically, early and adequate therapy should minimize these problems by its demonstrated ability to halt and even reverse the progression of esophageal dilatation (33, 34). On the other hand, failure to improve esophageal emptying by relieving the obstruction at the LES may lead to further detoriation and the development of end-stage "mega-esophagus". In this situation, esophagectomy may become the only treatment option (35).

Diagnosis of achalasia

Most patients with idiopathic achalasia are symptomatic for years before seeking medical attention, although some patients present early with severe symptoms. In the early stages of the disease, symptoms may be subtle and atypical (36). Although dysphagia is the dominant symptom at presentation, associated symptoms such regurgitation, chest pain and heartburn may lead the clinician to suspect more common entities such as gastro-esophageal reflux disease (GERD), dyspepsia or stress-related symptoms (31, 37). Achalasia has also been confused with eating disorders, such as anorexia nervosa (38). As a consequence, the diagnosis of achalasia is often delayed, with the mean duration of symptoms before treatment reported to be from 4.5 to 7.6 years (39). As a diagnosis based on clinical symptomatology alone is difficult, functional studies such as a barium esophagogram and/or esophageal manometry are very useful.

Barium investigation in patients with suspicion of achalasia is attractive due to its availability in most medical institutions where these patients first seek medical attention and should be done immediately when achalasia is suspected (16). Early in the disease, the esophagus is normal in diameter but will show loss of primary peristalsis in its distal two-thirds, when examined in the recumbent position. There is a typical, smooth tapering of the lower esophagus down to the closed LES, resembling a "bird's beak" (32). As the disease progresses, the esophagus becomes more dilated and tortuous and does not empty. With the patient upright, barium builds up to a point where the hydrostatic pressure of the barium overcomes the LES pressure (Hurst phenomenon). The presence of an epiphrenic diverticulum suggests the diagnosis of achalasia (40). In end-stages, the esophagus assumes a sigmoid shape, and becomes noncompliant and non-functional.

Manometry is considered to be the gold standard for the diagnosis of achalasia (39, 41). Aperistalsis is always present, meaning that all wet swallows are followed by simultaneous contractions over the length of the esophageal body. The contractions are classically identical to each other due to a common cavity or closed chamber phenomenon. The contractile amplitudes are typically low (10-40 mm Hg) and repetitive. Some manometric abnormality of the LES is always present in patients with achalasia. The LES pressure is usually elevated but may be normal (10-45 mm Hg) in up to 45% of patients. About 70-80% of patients with achalasia have absent or incomplete LES relaxation with wet swallows. In the remaining 20-30%, the relaxations are complete to the gastric baseline but are of short duration (usually <6s) and functionally inadequate as assessed by barium and radionuclide emptying studies (42, 43).

Endoscopy is normal in many cases of achalasia, but is always indicated to exclude a tumor of the gastro-esophageal junction (26). Such a tumor can produce a clinical syndrome similar to achalasia called pseudoachalasia. Approximately 2-4% of patients suspected of achalasia suffer from pseudoachalasia (35). Nearly 75% of patients with this condition are found to have underlying carcinoma of the cardia (44), but other responsible malignancies such as bronchogenic carcinoma, lymphoma and pancreatic carcinoma have been described.

Therapy for achalasia

Ideally, successful treatment of achalasia should reverse the aperistalsis and restore LES function; however, these objectives are almost never achieved. All currently available treatments for achalasia are aimed at improving symptoms by reducing the functional barrier at the LES, thus facilitating esophageal emptying by gravity. Different treatment modalities can be evaluated and compared by their efficacy, incidence of side effects, morbidity and mortality. The two most effective treatment options are graded pneumatic dilatation and surgical division of the smooth muscle of the LES (Heller myotomy). The short term success rates of these therapies are 70-90% (45). Most reports of the outcomes of achalasia treatment are produced by single centers and describe the effectiveness of only one treatment method. In the only randomized trial comparing the outcomes of pneumatic dilatation and surgical myotomy, 95% of patients treated with surgical myotomy had a good long-term result compared with 65% of patients treated with pneumatic dilatation (34). However, this study is frequently criticized for the technique of dilatation and patient selection. Recent widespread use of laparoscopic surgery has resulted in a greater role for surgical myotomy in the initial management of achalasia. Compared to the open approach, laparoscopic surgery is less invasive, reduces the pain and postoperative disability, and allows a shorter hospital stay. Short-term follow-up studies have reported results comparable to conventional Heller myotomy with success rates of 90-94%, but the long-term outcome of patients undergoing laparoscopic myotomy is unknown (45). Our group from the Sahlgrenska and Karolinska University Hospitals recently reported the short term results of the first randomized prospective study comparing pneumatic dilatation with laparoscopic myotomy, suggesting a superiority of the latter therapeutic approach (46).

Evaluating response to treatment

As no therapy for achalasia reverses the underlying esophageal pathology, therapeutic efficacy can only be assessed by symptomatic and/or functional improvement. However, several researchers have noted that evaluating response to treatment by symptom relief alone may be inacurrate (47-49). Some patients report good or excellent relief of symptoms despite very abnormal esophageal studies. Compared to the pre-therapy situation, these patients often experience a significant improvement, but a detailed history may disclose that the patients indeed have some residual symptoms (26). An explanation for this phenomenon may be that patients with chronic symptoms subjectively interpret even minimal improvement in esophageal emptying as dramatic (50). Other patients may have unrealistic expectations of therapy and report little symptomatic improvement despite normal emptying by barium esophagogram (51). Vaezi et al reported that a subset of patients (8 of 26) reported near complete symptom resolution after pneumatic dilatation for achalasia despite having poor emptying on barium examination (timed barium esophagogram) (49). Compared with the 16 patients with near complete improvement in both symptoms and barium emptying, these patients were significantly older. There is some evidence of diminished visceral sensitivity with aging (52) and alteration in esophageal sensation may well be an important factor for the subjective perception of symptoms in achalasia. Diminished visceral sensation in achalasia patients compared with controls has also been demonstrated and may explain poor perception of esophageal distension and retention in achalasia (53). It has also been suggested that symptoms in achalasia patients are related to the patients' tolerance and dietary adaptations and the degree of LES relaxation (54). Patients with longer duration of symptoms may have adapted to their dysphagia better than those of shorter duration, thus accounting for less symptomatic relief after treatment in patients with longer duration of achalasia (55).

Despite these limitations, assessment of treatment success in achalasia has usually been based on symptom improvement (56). Many series use "excellent" or "good" relief of dysphagia as the endpoint, but the meaning of such descriptions obviously varies with the initial symptom severity (31). Comparisons across studies are also difficult because definitions of success may vary from strict criteria, like symptoms once a week or less (57, 58) to more liberal endpoints such as the lack of need for repeat treatment (59). Several disease-specific scoring systems in achalasia exist (60) but the measurement properties of these are often poorly defined. A recently developed measure of achalasia-specific health-related quality of life (HRQL) may cover the multidimensional aspects of achalasia more completely, meeting the criteria of a valid score system (61).

These difficulties in symptomatic evaluation underscore the need for a simple test to objectively assess the response to treatment in achalasia. The principal aim of the present dissertation was to investigate the reliability, validity and clinical value of such a test, the timed barium esophagogram.

Timed barium esophagogram (TBE)

A barium esophagogram has the potential to quantify the functional disturbance in achalasia, namely impaired esophageal emptying. It is also a simple and widely available technique and has accordingly been used in the objective evaluation of patients pre- and post-therapy since the late 1960s. Although some authors have found a barium esophagogram useful in the evaluation of patients post-dilatation (50, 62), other studies have demonstrated poor correlation between symptom improvement and radiographic findings (55, 63, 64). However, these studies have relied on the height or the width of the barium column as the only measure of esophageal emptying. This may not be appropriate, as with progression of the disease, the esophagus often widens, leading to a decrease in the height of the barium column instead of the expected increase. The timed barium esophagogram (TBE) was introduced in 1997 (47) in an attempt to clarify the role of a standardized radiographic technique, using both the height and the width in the calculation of esophageal emptying. The technique is as follows: while standing, the patient ingests a low-density barium sulphate suspension (45% weight in volume). Patients are instructed to drink up to 250 ml within one minute, but the volume ingested is based on patient tolerance and the ingested volume is recorded. Three antero-posterior radiographs are obtained 1, 2, and 5 minutes after the start of ingestion, with the patient standing in a left posterior oblique position to avoid over-projection of the esophagus over the spine. The distance of the fluoroscope carriage from the patient is kept constant during the examination. The 2-minute film is optional, but fluoroscopy is performed at 2 minutes to determine the state of emptying. In the original description of the technique, it was suggested to estimate the degree of emptying quantitatively by calculating the rough area of the barium column on the 1- and 5-minute films by using the product height times width, and then determining the percentage change in this area. However, estimating the degree of emptying qualitatively by comparing the 1- and 5 minute films subjectively resulted in good agreement with the quantitative method. As this subjective method was easier and quicker to perform, it was recommended for routine use in TBE. Although obviously an improvement compared to previous attempts to evaluate esophageal emptying by barium studies, the described TBE concept has up till now not been validated.

Aims of the study

The general aim of this thesis was to explore the reliability, validity, and the clinical usefulness of timed barium esophagogram (TBE) in the pre- and post-treatment evaluation of patients with achalasia and after surgical treatment of hypopharyngeal and proximal esophageal cancer.

To achieve this, the following specific aims were defined:

- To assess the reliability of TBE by determining the day-to-day variability of the TBE variables in patients with newly diagnosed, and previously treated achalasia.

- To further assess the reliability of TBE by determining the intra- and inter-observer agreement.

- To describe TBE characteristics in patients with newly diagnosed idiopathic achalasia.

- To estimate the validity of TBE by correlating these characteristics to manometric and clinical variables in patients with newly diagnosed achalasia.

- To assess which of the TBE variables that correlates closest with the degree of functional impairment, using the manometric resting and relaxing pressure of the LES as the reference.

- To prospectively apply TBE in a randomized trial comparing pneumatic dilatation and laparoscopic myotomy and to determine whether TBE reveals any difference in esophageal emptying after respective treatment.

- To further validate TBE by assessing the ability of TBE to predict symptoms and the frequency of treatment failures after treatment for achalasia.

- To apply a modified TBE technique in the radiological evaluation of patients who have undergone radical resection of hypopharyngeal or proximal esophageal cancer and reconstruction with a free jejunal transplant, and to correlate results of the radiological evaluation with the functional outcome.

Patients and methods

Patients

Study I

To determine the day-to-day variability of TBE, 21 patients managed for achalasia at the Department of Surgery, Sahlgrenska University Hospital, from March 2001 to May 2003 were examined with repeated TBE. The TBE examinations in each subject were performed approximately one week apart (median 8 days, range 3-100 days). There were 12 men and 9 women with a mean age of 44 years (range, 20-77 years). Five of the patients had newly diagnosed achalasia. Nine patients had previously been treated with one or two balloon dilatations and 7 patients with a Heller myotomy. The diagnosis of achalasia was based on classical clinical, manometric and radiographic criteria.

In order to test the interobserver agreement, 30 TBE examinations performed between March 2001 and May 2003 were randomly selected. Two observers independently measured all TBE parameters in these examinations and the observers were blinded to all patient data. One of the observers also re-assessed 21 of these TBE examinations after a time interval of more than 3 months, to provide evaluation of intra-observer agreement.

To test if TBE results in patients with achalasia differ from those in normal persons, eight healthy volunteers were recruited. They reported no current or previous esophageal or gastrointestinal symptoms. None of them took any medication known to have effects on esophageal or gastrointestinal function. In 6 subjects the examination was repeated after about one week. Thus, there were a total of 14 examinations in healthy subjects available for assessment.

Study II

The aim of the study was to describe TBE characteristics in patients with newly diagnosed idiopathic achalasia and to correlate these to clinical and manometric variables. From January 2001 until January 2005 consecutive patients who received the index diagnosis of idiopathic achalasia in the region of Västra Götaland (population 1,600,000) and Jönköping County (population 300,000), Sweden, were prospectively recruited into the study. During the last two years of the enrolment period, similar patients investigated at the GI motility laboratory at Karolinska University Hospital, Huddinge, Sweden, were also included. During the defined time period, 56 patients were found to fulfil the inclusion criteria. Forty-six of those underwent a TBE examination before therapy and they comprised our study group. There were 22 males and 24 females whose median age was 41 years (range, 16-78 years).

The diagnosis of achalasia was based on a classical history without endoscopic evidence of other specific causes, combined with incomplete swallow-induced relaxation of the LES at manometry. None of the patients had previously been treated with a pneumatic dilatation or other specific therapeutic interventions.

Study III

In the period from January 2001 to March 2005, 51 patients with newly diagnosed achalasia were included in a prospective, comparative study (46) and randomized to pneumatic dilatation (n= 26) or laparoscopic myotomy (n= 25). Evaluation with TBE was performed before (n= 46) and after treatment (n= 43). The primary objective was to elucidate whether TBE parameters predict the subsequent symptomatic response to respective therapy, including therapeutic failure. We also wanted to determine whether esophageal emptying differs after respective treatment. Thirty-five of the patients were examined at the Department of Radiology, Sahlgrenska University Hospital, Göteborg, and eleven were examined at the Department of Radiology, Karolinska University Hospital, Huddinge, Sweden.

Study IV

The study aimed at evaluating long-term radiological and functional outcomes in patients who had undergone circumferential pharyngo-laryngectomy and esophagectomy with reconstruction with a free vascularized jejunal transplant due to hypopharyngeal or proximal esophageal cancer. Between June 1995 and October 2005, 13 consecutive patients with proximal esophageal cancer and three patients with hypopharyngeal cancer (15 males and 1 female) underwent circumferential pharyngo-laryngectomy with esophageal resection to which was added a free vascularized jejunal transplant (n=14), colonic transposition (n=1) or a gastric tube (n=1) at the Sahlgrenska University Hospital. At the time of the follow-up (mean 54 months, range 6 –130), 10 of the 16 patients were still alive. The mean age of the survivors was 59 years (range 34-75) and the male:female ratio was 9:1. In two of the survivors, local tumor recurrence was recently diagnosed. Those individuals were not evaluated with radiological examination. One patient participated in the clinical assessment, but did not wish to undergo further radiological investigations. All seven patients that completed the radiographic evaluation had a jejunal transplant. Swallowing function was carefully assessed by a radiological technique utilizing the TBE concept combined with clinical assessment including HRQL questionnaires.

Timed barium esophagogram, study I, II and III

All subjects were examined after having fasted for at least 4 hours. They were tested by full size radiography in an upright, slightly left posterior oblique position. The patients were instructed to drink 250 ml of low-density barium sulphate suspension (45% weight in volume), or if unable, as much as they could tolerate without regurgitation or aspiration. Thereafter, three radiographs of the esophagus were exposed 1, 2 and 5 minutes, after the start of barium ingestion (47). These images were exposed on one X-ray film or on an image plate (35 x 35 cm) to simplify the comparison (Figure 1). The distance of the fluoroscope carriage from the patient was kept constant during the examination. The patients were told not to drink any remaining barium after the exposure of the 1 min film. If barium was completely cleared from the esophagus at the 2 min exposure, the 5 min film was not taken. The ingested volume of barium was recorded.

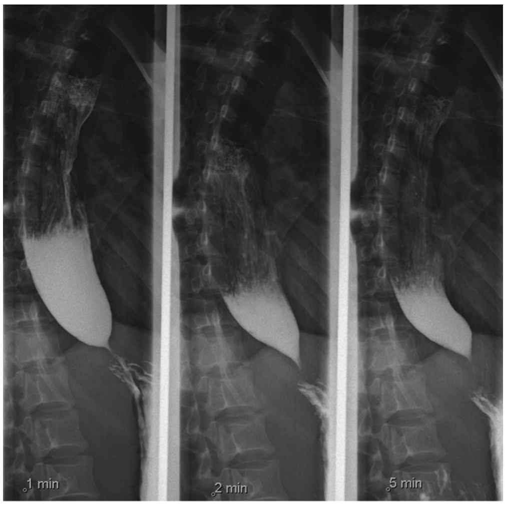

Figure 1. Figure shows technique of timed barium esophagogram (TBE) exposing three radiographs on one spot film at 1 min, 2 min, and 5 min after patient has ingested 250 ml of low-density barium.

Radiographic evaluation, study IV

The barium examinations were carried out in patients fasted for at least 6 hours. The study included both dynamic examination of motility with videofluoroscopy and a series of spot films, to evaluate morphology and emptying of the jejunal graft up to five minutes after swallowing. We used a remote-control fluoroscopy unit (Siemens Polystar, Siemens, Erlangen, Germany) equipped with a video recorder (Sony S-VHS SVO-9500MDP, Sony, Tokyo, Japan). The video recording was obtained with the patient standing in the right lateral position, using a full-field view of the oral cavity and pharynx to show both the oral and pharyngeal phases of swallowing. A coin with a known diameter was attached to the chin to allow correction for the magnification. The patients were asked to take 5 ml of barium ("High-Density", Astratech, Sweden) from a cup and then hold it in the mouth to test for adequacy of containment. They were then asked to swallow on command. Additional swallows of 15 ml "High-Density" contrast and of 5 ml of barium paste were recorded. At least one swallow in frontal projection was also imaged including both the oropharynx and the jejunal graft. Spot films of the jejunal interponate and the native esophagus were exposed, so that the localization of the anastomoses as well as any morphological abnormalities could be determined as accurately as possible. The transit of a bolus of 20 ml of barium through the jejunal segment and the remaining native esophagus was videotaped. As this transit was slow in some patients, we used intermittent fluoroscopy every 15 second for two minutes and then every 30 second up to a total of 5 minutes if appropriate. Using the "Last Image Hold" function and saving the images, transit was documented at the same time as radiation was kept to a minimum.

Image analysis

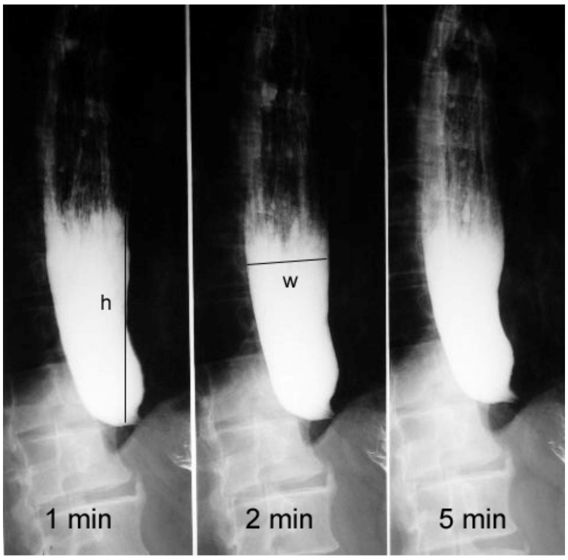

Figure 2. The height of the barium column (h) is measured from the top of the barium column to the 'birds's beak' at the level of the cardia. The maximum diameter (w) is measured at the widest point perpendicular to the long axis of the esophagus.

In **study I, II** and **III**, all subsequent radiological assessments of patients and controls were done blindly in random order without knowledge of the history, diagnosis or treatment of the respective subjects. The distance (cm) from the distal esophagus to the top of the barium column (height) and the maximum diameter (width) were measured on the films (Figure 2). The distal extent of the barium column was measured at level of lower esophageal sphincter, identified by the "bird´s beak" appearance. The top of the barium column was measured from the level at which the barium-foam interface was best defined. If it was impossible to distinguish such an interface, the top of the barium column was assumed to be situated half way between the top of the foam layer and the level at which the foam first was observed to mix with contrast. Due to the potential difficulties in differentiating a small barium column in the distal esophagus from mere coating of contrast material on the mucosa, residual barium with a height of 2 cm or less was considered as complete emptying, unless a distinct fluid level was observed. The differentiation between coating of the mucosa and the existence of a true barium column may also be complicated on a too lightly exposed film. For this reason, if the observer considered that the film was inadequately exposed, the examinations were excluded from analysis. The maximum width was measured perpendicular to the approximated long-axis of the barium column, at its widest point. Since the maximum width may not represent the mean width of the total barium column, leading to an overestimation of the total amount of barium in the esophagus in some instances, we added the mean width to our measurements. The mean width was measured by drawing two lines parallel to the respective outer margins of the barium column from the top to the bottom. The lines were drawn in such a way that the estimated contrast-containing area outside the line equaled the area not containing contrast inside the line (Figure 3). The mean width was recorded as the distance between the two lines.

Figure 3. The mean width (d) is measured by drawing two lines parallell to the respective outer margins of the barium column. The lines are drawn in such a way that the estimated area containing contrast material outside the line equals the area not containing contrast material inside the line. The mean width is the distance between the parallell lines.

We considered the product height times mean width to be a rough estimate of the area of the esophageal barium column. The percentage of change in this area as well as the change in barium column height from the 1 min to the 5 min film formed the basis for the analysis of the degree of esophageal emptying in study I. In study II and III, emptying was also evaluated by calculating the volume of barium in the esophagus at the respective time-point. After correction for magnification by dividing the measurements by the magnification factor of 1.35, the volume of barium in the esophagus was calculated according to the formula: (mean radius)^2 x 3.14 x height. Subtraction of the calculated volume from the actually ingested volume of barium in each individual resulted in an estimation of the emptied volume, expressed in milliliters.

In **study IV**, swallowing function was assessed on the video recordings and the morphology primarily on the spot films, and the findings were recorded on a data sheet by two reviewers in consensus. Another data sheet was completed by a third, independent reviewer, to allow for calculation of inter-observer variability. The oral and pharyngeal phases of swallowing were analyzed in slow motion. Functional parameters of the oral phase included: premature spill of bolus from the oral cavity, delay in initiation of swallowing, impairment of tongue motion, and residuals in the mouth after swallowing. Aspects of the pharyngeal phase analyzed were: delayed triggering of the pharyngeal phase, degree of soft-palate elevation, posterior tongue thrust and contact between the tongue base and the posterior wall of the pharynx, effectiveness of pharyngeal peristalsis, and bolus-clearance. A composite evaluation of these aspects was done and oral and pharyngeal dysfunction was graded as none, mild, moderate or severe. The occurrence of oral, pharyngeal, or nasopharyngeal regurgitations and the amount of after-swallowing was likewise graded on a four-point scale ranging from none to very frequent (more than 4-5). In order to quantify the temporal aspects of swallowing, a frame-by-frame analysis of the pharyngeal phase of swallowing was performed. The pharyngeal transit time was defined as the time from when the head of the bolus crossed the anterior border of the vertical ramus of the mandible until the tail of the bolus passed through the upper anastomosis. The pharyngo-jejunal connection was assessed regarding width and location and the presence of any morphological abnormalities, such as a pseudopouch. In case of a pouch, the volume was calculated according to a standardized formula for ellipsoid lesions (width x height x length / 2) (Figure 4) (65). Retention of bolus in the pouch was graded as none, mild, moderate or severe. The length of the jejunal interponate as well as that of the remaining, native esophagus was recorded. The function of the jejunal graft was assessed in relation to the degree of retention and delay in bolus transit. None was scored if the graft emptied completely with no retention of barium observed, mild if transit was somewhat slow but without bolus-retention, moderate if transit was somewhat slow with some retention of contrast material and poor if transit was slow and most of the bolus was retained in or above the graft. The degree of intrinsic activity in the graft (- = absent, + = weak, ++ = moderate, +++ = lively), as well as any localized delay or hold-up in transit of bolus, was noted. The motility in the remaining native esophagus was evaluated with regard to the presence of non-propulsive, tertiary contractions, delayed esophageal emptying and impaired lower esophageal sphincter (LES) relaxation. If considered to be abnormal, the abnormality was graded as mild, moderate or severe.

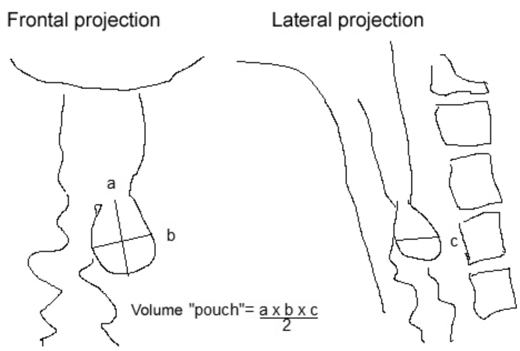

Figure 4. Calculation of the volume of the pouch according to a standardized formula.

Manometry

Manometry was carried out according to a predefined, standardized protocol. The manometric catheter was water-perfused and equipped with side holes located 5 cm apart to record the presence of primary peristaltic waves and the amplitude of esophageal contractions in the upper, middle and lower parts, respectively. To accurately record the intraluminal pressure from within the LES, a 6-cm-long sleeve sensor was used. This device contains a silicone membrane under which water is perfused, so that when pressure is applied anywhere along the length of the membrane, the resistance to flow beneath the membrane increases (66). The distal end of the sleeve device contained a side hole to record the intragastric pressure. Each catheter also had a pharyngeal port, which allowed swallowing signals to be recorded. The catheter assembly was connected to a pressure transducer and continuously perfused with degassed water through a low compliance perfusion pump (Arndorfer System, Arndorfer Medical Specialities, Greendale, Wisc, USA). The signals were transmitted to a computerized recording system (Medtronic, Stockholm, Sweden).

The patients were investigated after an overnight fast and, during the investigation, kept recumbent in the right lateral position. Before introduction of the catheter, each pressure port had been calibrated to standard pressure levels at room temperature. Following 10 swallows of 5 ml of room-temperature water, the esophageal peristalsis was evaluated. The mean contraction amplitude in respective segment (upper, middle and lower) of the esophageal body was calculated. Simultaneous onset of contractions in the segments was categorized as functional aperistalsis. Peristalsis was classified as failed, if the peristaltic wave disappeared during the aboral transmission through the esophagus or if the contraction amplitude never reached higher than 10 mm Hg. The basal LES pressure was recorded at 1-minute intervals, in periods of stable pressure levels with no interference from swallows. The intraluminal end-expiratory gastric pressure served as reference. The nadir pressure was defined as the lowest pressure level achieved following a standard 5 ml water swallow. Maximal LES relaxation was instituted by allowing the patient to drink 100 ml of water through a straw and simultaneously recording LES pressure. Accordingly, the lowest pressure plateau hereby reached was registered.

Clinical assessment

All patients in **study II** were asked to complete a self-assessment questionnaire, which evaluated symptoms by a previously described scoring system (49, 67). Symptoms assessed were dysphagia for solids, dysphagia for liquids, acid regurgitation, chest pain and heartburn. The frequency of each symptom was graded on a scale from 0 to 5 (0 = none, 1 = rare, once per month or less; 2 = occasional, once a week, up to 3 to 4 times a month; 3 = frequent, 2 to 4 times a week; 4 = often, once a day; 5 = severe, several times a day). In addition the more specific Watson dysphagia score was applied (68) and included in the self-assessment questionnaire. This instrument enables a comprehensive description of the character of the swallowing difficulties, combining information about difficulty in swallowing nine types of liquids and solids:

- Water
- Milk (or thin soup)
- Custard (or yoghurt or pureed fruit)
- Jelly
- Scrambled egg (or baked beans or mashed potatoes)
- Baked fish (or steamed potatoes or cooked carrot)
- Bread (or pastries)
- Apple (or raw carrot)
- Steak (or pork or lamb chop)

If dysphagia for the item always is present, the score is 1 point, sometimes present = ½ point, never present = 0 points. The score for each substance is multiplied with the respective line number (first line equals 9, last line equals 1), and then all lines are summarized, resulting in a score that increases with the severity of dysphagia, ranging from 0-45. The intensity of dysphagia was also evaluated by the dysphagia item of the disease-specific gastrointestinal symptom rating scale (GSRS) (69, 70). The GSRS uses a seven-point scale, and higher scores mean more pronounced symptoms. Patients also assessed the duration of the disease manifestations and the delay, if any, in the diagnosis due to the physician. Doctor's delay was defined as the interval from the time the patient first sought medical advice until the diagnosis of achalasia was established.

In **study III**, the same self-assessment questionnaire as in study II was mailed to the patients at 1, 3, 6, 12, 24 and 36 months after treatment. During follow-up, the occurrence of treatment failures was recorded. A treatment failure was defined as incomplete symptom control or symptom relapse that required more than three additional dilatations, or switchover to the alternative treatment required due to serious complications or requested by the patient because of dissatisfaction with the allocated therapy.

In **study IV**, all patients were clinically assessed by use of the Karnofsky Performance Status Scale Index (KPSSI) (71) (100 = no evidence of disease, 0 = dead) and, if established, regarding their speech valve function. The latter was assessed by a surgeon and a speech pathologist in concordance with regard to the number of syllables per breath, intelligibility and voice use and rated according to Ahmad et al.(72) as good, average or poor.

Dysphagia was graded according to Ogilvie et al. (73) (0 = no dysphagia; 1 = some dysphagia, but no dietary limitations; 2 = can drink, but only eat semisolid food; 3 = can only drink; 4 = total dysphagia). A more detailed description of potential swallowing difficulties was captured by the Watson Dysphagia Score (68).

Patients were asked to complete the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) version 3.0 (74), to which was added the Esophageal Module (EORTC QLQ-OES 18) (75). The former consists of 30 tumor-specific questions, and in this study, only the global health status/QoL score is presented. A high score on this scale represents a high quality of life. The latter questionnaire consists of questions focusing on problems due to the specific tumor location and treatment, and contains 4 scales and 6 single items. A high score on these scales and items represents a high level of symptoms.

Treatment procedures

In **study III**, in patients allocated to pneumatic dilatation, the dilatation was performed as an outpatient procedure using intravenous sedation (n= 21) or a short intubation anaesthesia (n= 5). A predefined, graded dilatation protocol was followed, starting with a 30 mm balloon in women and a 35 mm balloon in men (Rigiflex ABD, Boston Scientific, Boston, MA, USA) (32). The clinical response was evaluated after 7-10 days and in patients with persistent symptoms (n= 8) the procedure was repeated with a 35 mm balloon in women and a 40 mm balloon in men. In patients allocated to surgical myotomy, a laparoscopic complete anterior cardiomyotomy was carried out. The myotomy extended well above what was considered to be the upper margin of the LES, and distally the sling fibers of the gastric portion of the sphincter were divided. To prevent post-operative reflux, a partial posterior fundoplication according to Toupet was added.

Randomization between treatment groups was performed in a 1:1 fashion by a computer-based algorithm stratifying by a so called minimization technique for age, gender, and previous medical treatment. A study nurse conducted the randomization centrally.

In **study IV**, the surgery was performed as a joint venture between upper gastrointestinal- , ENT- and plastic reconstructive surgeons. In addition to lymphadenoidectomy, the larynx, hypopharynx and proximal esophagus were resected en-bloc with the intention to get a tumor free margin of ≥ 2 cm. A jejunal segment, 15-20 cm of length with a suitable long mesenteric pedicle was harvested via midline abdominal incision and subsequently used as an interposition. The proximal end of the jejunal segment was closed by staples and the pharyngo-jejunostomy was constructed either end-to-side or end-to-end by use of interrupted invaginated absorbable sutures. The distal jejuno-esophagostomy was sutured accordingly end-to-end, again with absorbable suture material. Micro-vascular end-to-end and/or end-to-side anastomoses were performed to recipient vessels in the neck. Approximately three months after initial surgery, a secondary tracheo-jejunal puncture using a speech valve (Provox I) was established.

Study designs

The design in diagnostic studies is by definition observational in nature, as apposed to the experimental study design often used in treatment research. Observational studies can be of an analytical or descriptive type. If the study has a comparison or control group, the study is termed analytical. If not, it is a descriptive study (76).

In **study I**, we did have a small control group of healthy volunteers. However, the purpose of this design was to ascertain that TBE results in patients with achalasia differ from those in normal persons ("phase I question") (77). As the control group did not consist of patients suspected of having achalasia, this phase of evaluation of TBE cannot be translated into diagnostic action, and it would not be appropriate to use the term analytical for this type of study design. The patients were prospectively, but not consecutively, recruited among patients managed for achalasia at the Department of Surgery, Sahlgrenska University Hospital, and the purpose of the study was to evaluate the reproducibility and observer agreement of TBE in this spectrum of patients.

Study II is also a descriptive study of a cross-sectional type (78). We compared our test method TBE with the present gold standard method, esophageal manometry, and with cinical evaluation, in 46 consecutive patients with newly diagnosed achalasia.

In **study III**, we applied TBE in the setting of a prospective, randomized clinical trial comparing pneumatic dilatation and laparoscopic myotomy in patients with newly diagnosed achalasia. The outcome of interest was esophageal emptying assessed by TBE after respective treatment. As the patients were allocated to different treatment groups and then followed for the outcome, it may be appropriate to characterize this part of the study as an analytical study. All of the patients were also followed as a cohort for the development of symptomatic response of up to 36 months after treatment, but as there was no control group in this part of the study, it should preferably be characterized as a descriptive study, rather than a true cohort study.

Study IV is a case series evaluation of 10 survivors after pharyngo-laryngo-esophagectomy (PLE). The evaluation included assessment with a modified TBE-technique, and instruments for evaluation of dysphagia, health related quality of life (HRQL) and voice quality.

Comments

Historically, much of the radiology literature has been descriptive in nature. In such study design, bias may be more difficult to control, particularly in retrospective studies (79). A potential pitfall is to misuse the data and draw causal or temporal inferences, which are not possible in studies without a comparison group (80). In the hierarchy of research designs, the results of randomized controlled trials are considered to be evidence of the highest grade, whereas observational studies are viewed as having less validity because the risk of overestimating effects (81). A major advantage of a randomized controlled clinical trial is the control over unknown confounders, i.e. factors that cannot be adjusted for since they are unknown. However, randomized controlled trials can also produce heterogeneous results and are difficult to perform in rare diseases such as achalasia, where only a limited number of patients are available for inclusion. Descriptive studies may be useful for the exploration of new concepts and are important for generation of hypotheses, that later can be tested in more rigorous studies with comparison groups (78). Judging the validity of a study is therefore not as simple as just categorizing the study according to the research design applied (81).

Statistics and ethics

Statistical advice in study I was obtained from Gunnar Ekeroth, biostatistician, Statistiska Konsultgruppen, Göteborg. The studies were approved by the local ethics committees and informed consent was obtained from each participating patient before inclusion.

In general, data were presented as median and interquartile ranges (25% to 75%).

The non-parametric two-sided Mann-Whitney U-test was used for analysis of significant differences between medians between groups (study I, II and III). In study I, the mean values of the TBE variables measured at the test and retest were compared by paired samples *t*-test. Wilcoxon signed rank test was used to compare pre-and post-treatment TBE variables in the same patients in study III.

To assess whether two continuous variables were associated, Pearson's correlation analysis was used (study I, II and III). In addition to the Pearson's correlation coefficient, the analysis of agreement was done according to methods proposed by Bland and Altman (82). The differences between repeated measurements were plotted against the respective means, to obtain Bland-Altman plots. The limits of agreement (LoA) were calculated as the 95% range of agreement for individuals. The coefficient of variation (CV) was defined as the standard deviation of the differences relative to the mean value, expressed in per cent.

To elucidate the association between continuous TBE- and manometric data and categorical symptom scores, Spearman rank correlation analysis was applied (study III).

Inter-observer agreement of the radiological findings in study IV was assessed by calculation of the weighted kappa-value (83)

A p-value of less than 0.05 was considered as significant. All variables were stored in an Excel database on which the SPSS statistical programme (Chicago, Ill. USA) was utilized.

Results and comments

Study I

Eleven of the 21 patients examined with repeated TBE were able to drink the allocated volume of 250 ml at both test occasions. On average, the patients ingested slightly less barium at the second test occasion (203 ml) compared to the amount ingested at the first (211 ml), but the difference was not statistically significant. All but one examination performed in the achalasia patients showed a column of barium in the esophagus after 1 minute and an esophagus that tapered into a 'bird-beak' at the gastro-esophageal junction. The measurements of TBE-parameters in the patient group in the test and re-test situation are listed in Table 1. None of the means of the measured parameters showed any significant difference between the test occasions. The correlation coefficients of the static TBE variables ranged from 0.72 to 0.86. The relation of the corresponding barium heights at 1, 2 and 5 minutes are depicted in Figure 5 a-c. The differences between repeated measurements of barium heights were also plotted against the respective means, to obtain Bland-Altman plots, indicating the limits of agreement (i.e. 95% range of agreement for individuals) (82) (Figure 6 a-c). When assessing the day-to-day variability of the dynamic measures of esophageal emptying, poorer correlations were found. The percentage emptying calculated by subtracting the area of the barium column (height times mean width) at 5 min from the same area at 1 min attained an *r*-value of 0.50 ($p < 0.05$), whereas calculating emptying by using only the relative change in barium column height reached a correlation coefficient of 0.28 ($p = 0.24$).

Variable	Measurements at first examination (cm)	Measurements at second examination (cm)	p Value	Correlation	LoA (cm)
Barium height 1 min	13.5 (7.6)	12.8 (9.0)	0.49	*r*= 0.86 p< 0.01	± 9.2
Barium height 2 min	12.2 (5.7)	11.0 (5.7)	0.16	*r*= 0.73 p< 0.01	± 10.6
Barium height 5 min	9.1 (5.7)	8.3 (7.6)	0.49	*r*= 0.79 p< 0.01	± 9.2
Maximum width 1 min	4.1 (1.6)	4.1 (1.5)	0.98	*r*= 0.76 p< 0.01	± 2.2
Maximum width 2 min	4.1 (1.9)	3.9 (1.6)	0.34	*r*= 0.79 p< 0.01	± 2.4
Maximum width 5 min	3.6 (1.8)	4.1 (1.6)	0.51	*r*= 0.84 p< 0.01	± 2.0
Mean width 1 min	3.1 (1.1)	3.1 (1.2)	0.75	*r*= 0.85 p< 0.01	± 1.2
Mean width 2 min	3.1 (1.5)	3.0 (1.2)	0.56	*r*= 0.82 p< 0.01	± 1.8
Mean width 5 min	2.7 (1.5)	2.6 (1.1)	0.39	*r*= 0.72 p< 0.01	± 2.2
Percentage emtying 1-5 min (area method)	37.8 (31.7)	49.1 (36.3)	0.14	*r*= 0.50 p< 0.05	± 68 %
Percentage emptying 1-5 min (height method)	28.3 (29.7)	42.6 (40.1)	0.09	*r*= 0.28 p= 0.24	± 84 %

Table 1. Measurements of TBE variables on two occasions in 21 patients. Measurements are presented as mean (SD). The means were compared by paired samples t-test and no significant difference between the respective means were found (column marked "p-value"). All measurements performed, except percentage emptying 1-5 min (height method), correlated significantly between test and retest and the degree of correlation is given by the Pearson correlation coefficient *r*. LoA = 95% limits of agreement of differences in individual measurements (cm).

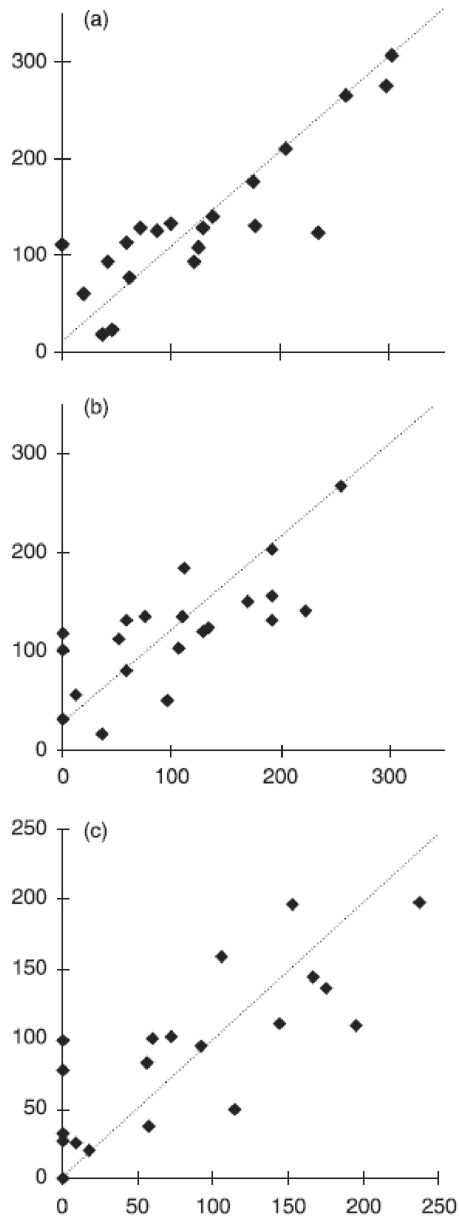

Figure 5 a-c. Relationship of barium column height at repeated examinations in patients with achalasia. Numbers on the axes represent barium column height in mm. The line of equality is shown. Examples show barium column height at 1 min (r= 0.86) (a), at 2 min (r= 0.73) (b) and at 5 min (r= 0.79) (c).

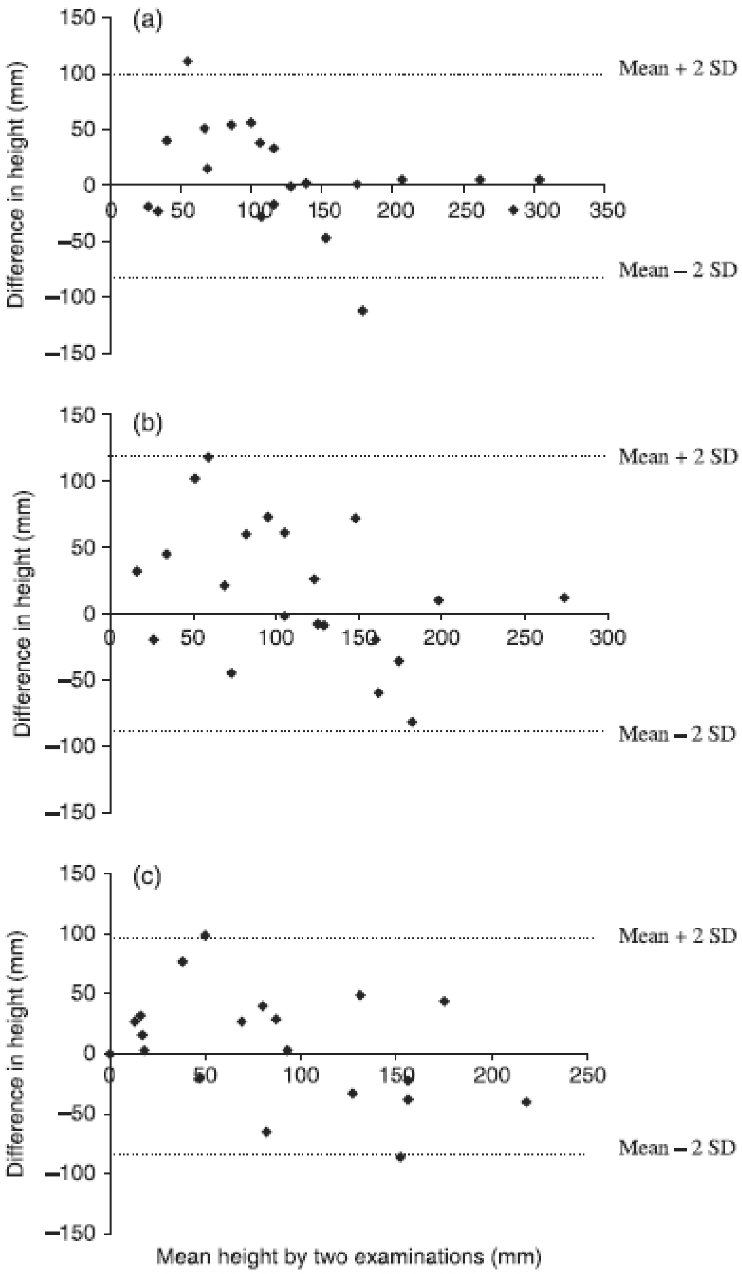

Figure 6 a-c. Bland-Altman plots showing the differences in the measurements (in mm) at repeated examinations (Y-axis) plotted against their respective means (X-axis). Examples show measurements of height at 1 min (a), at 2 min (b) and at 5 min (c).

Intra-and inter-observer agreement are summarized in Table 2. All measurements correlated significantly. 2.1% of the films had to be excluded from the analysis due to inadequate exposure. The inter-observer agreement was of the same magnitude as the intra-observer agreement. There was a tendency for somewhat higher variation of measurements at the 5-min time point compared with the 1-min time point. The limits of agreement (LoA) define the limits within which 95% of the differences between measurements lies.

All TBE variables in the healthy controls differed significantly compared to the patients. Some of the healthy individuals retained small amounts of barium in the esophagus after 1 minute, but all had emptied their esophagi after 2 minutes with no significant amount of contrast remaining in the lumen. No healthy subject showed signs of dilatation of the esophagus (i.e. had maximal width over 2 cm).

TBE-variable	Inter-observer variability			Intra-observer variability		
	Mean difference ± SD (cm)	*r*	LoA (cm)	Mean difference ± SD (cm)	*r*	LoA (cm)
Barium height 1 min	0.4 ± 1.2	0.99	± 2.3	0.1 ± 0.8	0.99	± 1.6
Barium height 2 min	0.2 ± 0.8	0.99	± 1.5	0.1 ± 0.5	0.99	± 1.0
Barium height 5 min	0.2 ± 1.2	0.98	± 2.4	0.3 ± 1.2	0.98	± 2.4
Maximum width 1 min	-0.1 ± 0.3	0.99	± 0.6	0.0 ± 0.1	0.99	± 0.2
Maximum width 2 min	-0.1 ± 0.3	0.99	± 0.6	0.1 ± 0.2	0.99	± 0.4
Maximum width 5 min	-0.1 ± 0.6	0.95	± 1.2	-0.1 ± 0.6	0.93	± 1.2
Mean width 1 min	-0.2 ± 0.3	0.98	± 0.6	0.2 ± 0.2	0.97	± 0.4
Mean width 2 min	-0.1 ± 0.4	0.97	± 0.8	0.0 ± 0.2	0.98	± 0.4
Mean width 5 min	0.0 ± 0.4	0.96	± 0.8	0.3 ± 0.5	0.91	± 1.0

Table 2. Inter- and intra-observer variability. *r* = correlation coefficient. LoA = 95% limits of agreement of differences in individual measurements (cm).

Comments

When analyzing the repeatability of a single measurement method or when comparing measurements by two observers, the same methods may be used as when assessing agreement between two measurement methods (84). Although frequently used, the Pearson correlation coefficient actually does not measure agreement, but rather determines the strength of linear association between two variables. This can be misleading in the case of systematic bias between the measurements. For example, if the second measurement is exactly twice the size of the first, this would result in perfect correlation, but the agreement would be very poor (85). If there is no consistent bias between repeated measurements by the same method, as seems to be the case in our study, correlation can be used in the analysis (84). However, even when it is appropriate, the correlation coefficient does not help us to interpret a clinical measurement on a given patient. The best approach for this purpose is to analyse the differences between the measurements on each subject (82). In the so called Bland-Altman plot, the differences between the measurements are plotted against the average of the two measurements. This makes it easy to visualize the size of the differences and their distribution around the zero. To estimate how well the measurements are likely to agree for an individual, the standard deviation of the differences (SD_{diff}) are used. The 95 % limits of agreement (LoA) are defined as the mean ± $2SD_{diff}$, and it follows that about 95% of the observations are included within these limits. For repeated measurements by the same method, when the average difference is zero, the term repeatability coefficient has been recommended for the absolute value of the 95% limits of agreement (84, 86). In our data, the mean difference between repeated measurements of the barium column height is negligible and there is no obvious tendency that the differences are related to the size of the measurements (Figure 6 a-c). Nineteen of the 21 differences (90 %) in the measured height of the barium column after 1 minute lie within approximately 5 cm. There are two outliers with differences of about 10 cm, but we retained these measurements for the analysis of limits of agreement. It must be remembered that the interpretation of the limits of agreement must depend on the clinical circumstances – it is not possible to use statistics to define acceptable agreement (87).

When evaluating the repeatability of a test, even a small sample is valuable, provided that it is representative, and the duplicate tests are genuinely independent, the observer being unable to identify the pairs (88).

The very first step in the study of a new, potential diagnostic test is to examine whether patients with the target disorder have different test results from normal individuals (89). In the evidenced based architecture of diagnostic research this step is the equivalent to the phase I study in a clinical trial (77). This so called phase I question can be answered with a minimum of effort, expense and time. A negative answer to the question indicates that the test is without value and removes the need to answer the more time-consuming, and costlier questions of the following validation phases. We found large differences between the median values of all TBE-parameters between patients and healthy controls. All of the control subjects had emptied their esophagi completely after 2 minutes. This finding was found in only 2 of 42 examinations performed in patients (i.e. overlapping results in 5 % of the examinations performed in achalasia patients).

Study II

All 46 patients examined with TBE showed signs of impaired LES opening, manifested by a tapered bird's beak appearance adjacent to the gastroesophageal junction. The height of the barium column after 1 minute varied from 6.0 to 30.7 cm. Only two patients had a non-dilated esophagus with a width of 2 cm or less. The static parameters; median height, maximum and mean width of the barium column were 16.0, 4.4 and 3.3

cm at the 1 minute time-point after contrast ingestion, and 13.0, 3.8 and 2.7 cm, respectively, at 5 minutes. Complete subjective evaluation of achalasia symptoms were available from 45 of the 46 patients, and 39 (85%) of the patients provided the Watson dysphagia score. The median Watson score in these patients was 33.3, indicating profound swallowing difficulties. Manometry was performed in 42 of the patients and the contraction pattern of the esophageal body could be evaluated in all of these patients. The manometric characteristics of the esophageal body as well as the LES were typical for achalasia. In two cases it was not possible to pass the gastro-esophageal junction (GEJ) with the catheter assembly. However, other sources of technical difficulties during the subsequent examination precluded the recording of the LES tone in some of the remaining patients. Thus, adequate recordings of the LES resting pressure were available in 36 patients, Nadir pressure (=the lowest pressure level achieved following a 5 ml water swallow) in 34 patients, and maximal relaxation pressure in 26 of the patients. None of the patients revealed complete LES relaxation.

Esophageal emptying by TBE, calculated by subtracting the volume of barium in the esophagus from the actually ingested volume of barium in each individual, was significantly larger during the first minute (median 143 ml) than from 1 to 5 minutes (median 16 ml). Emptying during the first minute showed significant inverse correlation with the resting and the maximal relaxing pressure of the LES (r= -0.34 and r= -0.54, respectively). When we analyzed the potential associations between emptying and clinical variables, we found an inverse correlation between the volume of emptied barium after 5 minutes at TBE and the duration of symptoms, i.e. decreasing emptying with longer duration (r= -0.36, p< 0.05). There was also a weak correlation between width of the barium column at 1 minute and the duration of symptoms (increasing width with longer duration), but this correlation did not reach statistical significance (r= 0.27, p= 0.09). There was, however, a significant inverse relationship between barium column width and the symptom score for post-prandial chest pain (r= -0.44, p< 0.01). We were unable to reveal any association between esophageal emptying, expressed as the relative changes in barium column height or area between the 1- and 5-minute time-points, and manometric findings. Likewise no relationship was revealed between those TBE parameters and the duration or severity of symptoms.

Comments

Although our method of estimating the volume of barium in the esophagus has not been validated, the data are not compatible with esophageal emptying being linear in time, i.e. constant volume of barium emptied into the stomach in a given time. The finding of larger emptying during the first minute after ingestion than during the subsequent 4 minutes is instead consistent with emptying being exponential, or at least non-linear. Some support for this pattern can be found from radionuclide studies by analyzing esophageal emptying curves in achalasia patients. These curves demonstrate an initial higher velocity of emptying to the stomach after intake of the isotope labelled bolus or test meal, followed by slower emptying during the time of the registration up to 10 to 20 minutes after the ingestion (90-93). Peak retention of isotope and retention in the esophagus at 5, 10, and 20 minutes have been shown to discriminate well between normal controls, untreated and treated achalasia patents (54, 90-92). However, after the initial phase of emptying, the slopes of the emptying curves seem to have similar inclination in studied subjects irrespective of previous treatment (54, 91, 92). In one study, the emptying rate (velocity) between 2 and 10 minutes was even shown to be faster in untreated compared with treated patients (92). When assessing emptying, which is the objective of TBE, it may therefore be suboptimal to use the relative changes in esophageal retention between 1 and 5 minutes as a measure of emptying, as proposed in the original description by de Oliveira et al (47). It is possible that the repeated swallowing actions, that is required to ingest the rather large volume of fluid in TBE, provides good stimulation to LES relaxation, thereby explaining the observed faster "initial" emptying. As emptying during this initial phase may reflect the functional impairment of the LES more accurately (and thus the severity of the disease process in achalasia) it is reasonable to conclude that a primary aim of TBE should be to quantitate this emptying. This conclusion is supported by our observation of inverse correlation between emptying during the first minute and the resting and maximal relaxing pressures of the LES at manometry, and the duration of symptoms, respectively. As many patients have difficulties in ingesting the allocated volume in TBE (illustrated by the 41% of the patients in our study that were unable to do so), it is important to subtract the *estimated* volume of barium in the esophagus after 1 minute from the *actually* ingested volume to assess "initial" emptying.

Following the initial passage of barium after deglutition, it is likely that esophageal emptying is passive and probably the result of gravity (92). It is known that esophageal function varies with the body posture through the effect of gravity, and that emptying is delayed in the supine position compared with an upright position in both healthy subjects as well as in patients with achalasia (94, 95). It is interesting to note that the calculated hydrostatic pressure from the mean height of the barium column after 1 min at TBE is in the same order as the mean resting pressure of the LES measured at manometry (16.0 mmHg vs 17.7 mmHg).

There are scanty scientific evidence available of the actual pros and cons of radiology and manometry, respectively, for the diagnosis of achalasia. In a prospective study in 88 symptomatic patients examined with videofluoroscopy, with manometry as the reference standard, the barium study had a sensitivity of 87% in the detection of achalasia (13 of 15 cases) (96). In another prospective study of achalasia patients diagnosed by manometry, Howard and co-workers found that barium esophagography was suggestive of achalasia in only 64% of study participants (39). In a recent study of 38 patients also diagnosed manometrically, achalasia was stated as a diagnostic possibility in the radiologist report in only 58% (97). On the other hand, in a retrospective study in 21 patients with typical radiographic findings of achalasia, one-third of the patients had normal LES relaxation at manometry (98). However, all 21 patients had excellent resolution of symptoms after treatment, suggesting that they in fact had achalasia. This led the authors to conclude that in patients with typical radiographic findings of achalasia, the barium study can be used to guide treatment without a need for manometry. To further elucidate the respective roles of radiology and manometry in the diagnosis of achalasia, a large prospective study in patients suspected of having achalasia, with both examinations performed in all cases, would be of value.

Study III

Of the 43 post-treatment TBE examinations, two had to be excluded for technical reasons as the technique used deviated from the prescribed protocol. In two patients (one in each treatment group) treatment failed early and they crossed over to the alternative treatment before the post-treatment TBE. Complete data from both the pre- and post-treatment TBE-examination were available in 35 of the 51 patients (69 %) originally randomized in the treatment study. Eighteen of the patients were allocated to treatment with dilatation and 17 to treatment with surgery, and the results of the TBE-examinations were subsequently used for comparison of esophageal emptying after respective treatment. The demographic, pre-treatment manometric, and TBE characteristics were well balanced in the two study groups (Table 3).

The Watson dysphagia score in patients examined with TBE decreased from median 28.5 (22.1-42.0) pre-treatment to median 18 (4.3-30.0) post-treatment, with no significant difference between the treatment groups. Pooling the results from both treatment groups, a significant improvement was found for all TBE-parameters in response to treatment. The median height of the barium column at 1 minute decreased from 16.5 cm to 7.0 cm ($p< 0.001$) and the volume of retained barium at 1 minute decreased from median 81.0 ml to 16.0 ml ($p< 0.001$). There was, however, no significant inter-group difference between the postoperative TBE-parameters (Table 4). The improvement in TBE-emptying in the respective treatment groups was not affected by the patients' gender or age. However, in patients treated with pneumatic dilatation, those with a wider esophagus at baseline showed inferior improvement in the height of the barium column in response to therapy ($p< 0.05$, Pearson's r= -0.53 at 1 minute). No such correlation between a wider esophagus and inferior improvement in emptying could be found among patients treated with surgery.

	Dilatation (n = 18)	Surgery (n = 17)	*p* value
Age (years)	40.5 (28.0-60.0)	44.0 (31.5-55.5)	0.86
Gender (M/F)	8/10	11/6	
Resting LES pressure (mm Hg)	21.2 (12.3-36.3) (n=16)	18.6 (12.0-27.1) (n=14)	0.45
Nadir pressure LES (mm Hg)	9.8 (5.1-25.0) (n=11)	6.4 (4.4-7.8) (n=11)	0.06
Time from pre-op TBE to treatment (days)	12.0 (5.0-27.5)	75.0 (25.0-116.0)	0.003
Height of barium column at 1 min pre-op TBE (cm)	16.1 (13.5-22.1)	17.4 (11.2-20.7)	0.99
Maximum width of barium column at 1 min pre-op TBE (cm)	4.3 (4.0-5.1)	4.7 (3.6-5.9)	0.53
Emptying at 1 min pre-op TBE (ml)	139.0 (57.8-172.8)	139.0 (30.5-174.0)	0.93
Time from treatment to post-op TBE (months)	6.5 (3.0-17.0)	6.0 (3.0-25.5)	0.77

Table 3. Demographic, manometric and TBE-characteristics (median and interquartile range) of patients with newly diagnosed achalasia subsequently randomized to dilatation or surgery. The difference in the variables between treatment groups was analyzed using the Mann-Whitney test for unpaired, non-parametric data.
M= males. F= females. LES= Lower esophageal sphincter. TBE= Timed barium esophagogram.

	Dilatation (n = 18)	Surgery (n = 17)	*p* value
Height of barium column 1 min (cm)	8.4 (3.9-10.2)	6.7 (0.0-9.0)	0.08
Improvement in height pre-postop (percent)	60.5 (21.8-75.0)	70.0 (50.5-100.0)	0.14
Maximum width of barium column at 1 min (cm)	2.6 (2.2-3.9)	2.8 (2.1-4.2)	0.70
Improvement in maximum width pre-postop (percent)	62.5 (50.5-77.5)	60.0 (52.5-100.0)	0.82
Volume of barium column 1 min (ml)	14.5 (2.0-54.3)	24.0 (0.0-37.5)	0.77
Improvement in retained volume of barium at 1 min pre-postop (percent)	78.5 (48.5-95.5)	77.0 (48.0-100.0)	0.68
Height of barium column 5 min (cm)	2.3 (0.0-8.2)	2.6 (0.0-8.6)	0.80
Improvement in height at 5 min pre-postop (percent)	66.5 (29.4-100.0)	72.0 (48.0-100.0)	1.00
Volume of barium column 5 min (ml)	1.5 (0.0-28.8)	12.0 (0.0-27.0)	0.60
Improvement in retained volume of barium at 5 min pre-postop (percent)	94.5 (28.3-100.0)	46.0 (-9.0-100.0)	0.40

Table 4. Post-treatment TBE-characteristics (median and interquartile range) of patients with newly diagnosed achalasia treated with dilatation and surgery, respectively. The difference in the variables between the treatment groups was analyzed using the Mann-Whitney test for unpaired, non-parametric data.

Concerning post-therapy TBE variables and symptom outcomes during postoperative follow-up, complete data sets were obtained from 32 patients, with a median time interval between treatment and the TBE of 6 months (interquartile range 3-16 months) and a median follow-up time after the TBE of 18 months (interquartile range 9-27 months). Five of the 41 technically adequate TBE-examinations had to be excluded due to a considerable delay between treatment and the radiological investigation (range 26-48 months). In four patients complete clinical data were not captured.
Correlations between the TBE variables and the symptomatic outcome are detailed in Table 5. There were significant correlations between the height of the barium column at 1 minute after barium ingestion and the symptom scores for “dysphagia for liquids” ($p< 0.05$, rho= 0.47), “chest pain” ($p< 0.05$, rho= 0.42) and “Watson score” ($p< 0.05$, rho= 0.46), respectively. Moreover, the estimated emptied volume of barium from 1 to 5 minutes related to the scores for “dysphagia for solids” ($p< 0.05$, rho= 0.39), “dysphagia for liquids” ($p< 0.05$, rho= 0.40), “chest pain” ($p< 0.05$, rho= 0.41) and “Watson score” ($p< 0.05$, rho= 0.40).

	Dysphagia for solids	Dysphagia for liquids	Heart burn	Chest pain	Acid regurgi-tations	Watson score
Height of barium column at 1 min (cm)	N.S. rho= 0.36	P< 0.05 rho= 0.47	N.S. rho= -0.04	P< 0.05 rho= 0.42	N.S. rho= 0.08	P< 0.05 rho= 0.46
Volume of barium column at 1 min (ml)	N.S. rho= 0.36	N.S. rho =0.31	N.S. rho= -0.08	N.S. rho= 0.13	N.S. rho= 0.00	N.S. rho= 0.26
Height of barium column at 5 min (cm)	N.S. rho= 0.30	N.S. rho= 0.17	N.S. rho= -0.14	N.S. rho= 0.07	N.S. rho= -0.18	N.S. rho= 0.15
Volume of barium column at 5 min (ml)	N.S. rho= 0.33	N.S. rho= 0.22	N.S. rho= -0.08	N.S. rho= 0.00	N.S. rho= -0.07	N.S. rho= 0.09
Emptied volume of barium from 1 to 5 min (ml)	P< 0.05 rho= 0.39	P< 0.05 rho= 0.40	N.S. rho= 0.03	P< 0.05 rho= 0.41	N.S. rho= 0.13	P< 0.05 rho= 0.40

Table 5. Results of two-tailed Spearman correlation analysis between post-treatment TBE-variables and symptom scores at the end of follow-up (median 18 months from the TBE-examination). N.S. = non-significant.

In four patients, treatment failure, i.e. recurrence of symptoms or a complication necessitating additional treatment, occurred early post-operatively prior to the scheduled TBE-examination. In 28 patients, it was possible to relate the degree of improvement in TBE-emptying after treatment to the number of treatment failures during the subsequent follow-up (mean 29 months). Ten patients showed less than 50 % improvement in the barium column height at 1 minute, and four of these patients failed treatment during follow-up (40 %). Of the 18 patients with more than 50 % improvement in this variable, none displayed treatment failure during the follow-up. The improvement in the width of the barium column, the height of the barium column at 5 minutes or in the estimated emptied volume of barium, did not show any predictive value for the frequency of subsequent treatment failures. Sixteen of the patients showed signs of barium retention at 5 minutes while 13 patients had emptied their esophagus completely at 5 minutes. One of the patients with complete emptying showed treatment failure during follow-up. The median Watson score did not differ significantly between patients with complete emptying and patients with signs of remaining barium in the esophagus after 5 minutes (Watson score 10.5 vs 21.8).

Only 17 patients agreed to undergo manometry during follow-up. Thus, it was not meaningful to statistically compare the results of manometry between the treatment groups. In patients treated with surgery and examined with TBE (n= 10), the median resting LES pressure was 5.2 (2.5-13.4) mm Hg post-treatment and the median relaxing LES Nadir pressure was 1.5 (0.7-4.1) mm Hg. In patients treated with pneumatic dilatation and also examined with TBE (n= 6), the corresponding values were 3.7 (2.0-6.5) mm Hg and 2.9 (0.6-5.5) mm Hg, respectively. In all patients treated, the resting LES pressure decreased from median 20.8 (14.4-27.0) mm Hg pre-treatment to median 5.0 (2.4-7.5) mm Hg post-treatment ($p < 0.01$). The median LES Nadir pressure decreased from 6.4 (4.8-10.5) mm Hg to 1.9 (0.7-4.1) mm Hg ($p < 0.05$). No significant correlations were found between the post-treatment manometric recordings and TBE parameters.

Comments

There are several therapeutic options for achalasia. However, response to medication is poor (31) and the effect of intrasphincteric botulinum toxin injection is transient (99). The principal therapeutic decision is between pneumatic dilatation and surgical myotomy (45). Although the body of literature indicates that the long-term results may be more favourable after surgery than after dilatation (31, 100-102), the small number of controlled therapeutic trials comparing the effectiveness of surgical myotomy and pneumatic dilatation makes it difficult to objectively assess the outcomes after respective treatment. A retrospective cohort study by Vela et al demonstrated no difference in the early outcome of PD and HM, respectively, and showed that the success rate of both methods decreases over time (90% vs. 89% at 6 months, to 44% vs. 56% at 6 years) (103). However, most studies on the outcomes of achalasia treatment describe the effectiveness of only one treatment method. The interpretation of such studies is difficult because of different patient groups, different interventions and varying outcome measures. Using the appraisal methods of evidence-based practice (EBP), the methodological quality of such case series is weak. According to the levels of evidence developed by the Oxford Centre for Evidence-Based Medicine, they represent level 4 evidence (104).
Several studies have used radiography to assess the *response to treatment in achalasia*. It has been shown that the radiological diameter of the esophagus decreases significantly after treatment with myotomy as well as after dilatation (34, 50, 105). Gockel at al did not find any significant difference between the percentage reductions of the diameter after myotomy and dilatation, respectively (33). However, radiographic as well as scintigraphic studies have shown that patients do not regain normal esophageal emptying after these treatments (54). In a retrospective study of previously untreated achalasia patients, Vela et al found similar degrees of improvement in TBE emptying in patients treated with Heller myotomy (n= 72) and pneumatic dilatation (n= 111) (56). In the myotomy group, 44% of the patients showed more than 80% improvement in barium column height after treatment, compared with 54% in the dilatation group (non significant). Using TBE to evaluate the response to graded dilatation with the Rigiflex ballon, Vaezi et al found that the barium column height at 5 minutes decreased mean 58.4% six months post-therapy compared to the pre-therapy height (106).

Vaezi et al in another study found TBE to have predictive value for the *long-term outcome after treatment* for achalasia. They concluded that patients treated with pneumatic dilatation and showing less than 50 % improvement in barium column height at 5 minutes had a very high rate of treatment failures within one year (9 out of 10 patients = 90%) (107). Only 2 of the 22 patients with improvement in barium column height failed therapy within 1 year. Vela et al found that lack of improvement in postdilatation barium height at 5 minutes was associated with an early risk for needing a second dilatation (103). In our study we found that patients with less than 50% improvement of the height of the barium column post-therapy had 40% risk of treatment failure during follow-up. Although the cutoff point of 50% improvement has been shown to have predictive value for treatment failure, it was not pre-specified in our study, thereby increasing the risk of overestimating the performance of the test (108).

To our knowledge, our study is the first to report an association between results of TBE and the symptomatic outcome in patients treated with pneumatic dilatation as well as with myotomy during a medium to long-term follow-up. Interestingly, this association was only found between symptoms and the height of the barium column at 1 minute. As almost all of the patients were able to drink all of the allocated 250 ml barium after treatment, utilizing the emptied volume of barium did not, unlike in study II, result in a closer correlation with symptom scores. In this setting, a seemingly contradictory finding was that emptying calculated as the difference in barium retention between 1 and 5 minutes also correlated positively with dysphagia scores (i.e. the larger the emptying the more dysphagia symptoms). However, this association is likely explained by the strong correlation between this measure and the height of the barium column at 1 minute ($p< 0.001$, $r= 0.80$). That is, a larger barium column will result in greater emptying due to the increased hydrostatic pressure exerted on the LES compared to a smaller column. In any way, this positive correlation strengthens the notion that this is not a valid method of estimating esophageal emptying, despite the fact that it was proposed as a measure of emptying in the original description of TBE (47). Further support for the invalidity of this method can be found in studies I and II. In study I we found that the percentage change in the area or the height of the barium column from the 1 min to the 5 min film showed poor reproducibility in the test-retest situation and in study II we found no correlation between these measures and manometric values of LES tone.

Study IV

The radiological examinations revealed adequate function (mild or no dysfunction) of the oral and pharyngeal phases of the swallowing in five of seven patients. In these patients, the bolus pharyngeal transit time was also within the upper limit reported in normal subjects (1.2 sec) (109). However, swallowing disturbances in the form of oral and/or pharyngeal regurgitations were seen in three of these patients and all patients demonstrated a mild to severe amount of after-swallowing (Table 6). The closure of the soft palate was effective in all patients, and consequently, no case of naso-pharyngeal regurgitation was found.

Finding	Pt. 1	Pt. 4	Pt. 6	Pt. 7	Pt. 8	Pt. 9	Pt. 10	Mean
Oral dysfunction	+	+	−	++	++	−	+	
Oral regurgitation	+	−	+	++	−	+++	−	
Nasopharyngeal regurgitation	−	−	−	−	−	−	−	
Pharyngeal dysfunction	−	+	−	−	−	−	+	
PTT (seconds)	0.88	0.84	0.72	1.76	1.80	0.52	1.00	1.07
Pharyngeal regurgitations	−	−	+	+	−	+++	−	
After-swallowing	++	+	++	++	+	+++	++	
Pouch (ml)	6.5	−	2.3	20.5	−	10.1	−	
Impaired transit due to pouch	++	−	+	++	−	+++	−	
Length of graft (cm)	22	10	18	9	17	14	7.5	13.9
Intrinsic activity in graft	++	+	++	++	++	++	++	
Delay in transit through graft	−	+++	+++	−	++	−	++	
Location of delay	Pouch + native esophagus	Lower third of graft	Middle third of graft + native esophagus	Pouch	Upper third of graft + native esophagus	Pouch	Middle third of graft	
Length of native esophagus (cm)	12.5	21	13.5	20	12.5	19.5	19	16.9
Abnormal motility pattern in esophagus	+++	+	++	−	++	+	−	

PTT: pharyngeal transit time—head of the bolus at the anterior border of the vertical ramus of the mandible until the tail of bolus reaches the upper (pharyngojejunal) anastomosis

Assessment: –, none; + mild; ++ moderate; +++ severe

Table 6. Results of radiologic assessments (two reviewers in consensus). For definitions of qualitative swallowing measures, see definition in text.

Dysfunction of the graft was observed in all patients, but of varying degree. In three patients (nos. 1, 7 and 9), a pouch at the proximal anastomosis seemed to be the principal reason for the impaired bolus passage into and through the graft (Figure 7). One patient (no. 6) had a small pouch that did not seem to affect the transit time through the graft at all. Another patient (no. 4) had a stricture at the distal anastomosis causing retention of bolus transit in the lower third; and in three patients (nos. 6, 8 and 10), retention of barium in the graft was found without signs of any organic blockage. Intrinsic motility of the graft, that was largely independent of swallowing, was registered in all subjects. The length of the jejunal graft did not seem to affect the function with regard to peristalsis or radiographic bolus retention. However, in patients with a relatively shorter remaining native esophagus (nos. 1, 6, and 8), a moderate to severe abnormality in the motility pattern of the esophagus with delayed emptying was observed.

Figure 7. Barium examination in the lateral projection showing retention in a pouch.

Interobserver agreement between the consensus evaluation of the video-fluoroscopic findings and the assessment of the same findings by the third, independent reviewer was good (weighted kappa = 0.81) (83).

The patients' grades of dysphagia, assessed according to Ogilvie et al., were mild ranging from 0 to 1 (Table 7). The Watson Dysphagia Score varied between 0.5 and 45, with a mean score of 16.2. For most patients, solid food generated more dysphagia compared to liquids.

HRQL questionnaires were returned from all patients. The global health status/QoL score varied between 25 and 100 (mean 74), and the dysphagia scale score varied between 0 and 100 (mean 36). Other scores from the scales and items of the EORTC QLQ OES-18 questionnaire were all within the lower range, except for trouble with swallowing saliva, problem with taste and trouble with speech which had mean scores of 46, 43 and 43 respectively.

Radiographic grading of disturbed swallowing, manifested in the form of oro-pharyngeal dysfunction, regurgitations or frequent after-swallowing, or as impaired bolus passage through the graft, was not able to separate between patients with various degrees of dysphagia or HRQL scores.

Parameter	Pt. 1	Pt. 2	Pt. 3	Pt. 4	Pt. 5	Pt. 6	Pt. 7	Pt. 8	Pt. 9	Pt. 10	Mean
Karnofsky index (0–100)	90	60	90	90	70	90	80	80	90	80	82
Speech valve function	–	+	+	–	+	+	–	+	+	+	
Degree of dysphagia (0–4)	1	4	0	0	2	1	1	1	1	1	1
Watson dysphagia score (0–45)	6.5	45.0	0.5	4.5	33.0	12.0	22.5	4.5	8.0	25.5	16.2
EORTC QLQ C-30 (0–100)											
Global health status/QoL	83	33	92	67	67	100	83	92	100	25	74
EORTC QLQ OES-18 (0–100)											
Dysphagia scale	11	100	0	11	100	22	33	11	11	67	36
Eating scale	8	42	0	25	42	33	8	8	33	42	25
Reflux scale	0	0	0	33	0	0	17	0	0	0	5
Local pain scale	0	33	0	0	11	0	0	0	0	0	4
Trouble swallowing saliva	33	100	33	67	100	33	33	0	33	33	46
Problems with choking	33	33	0	0	67	0	0	0	67	0	20
Dry mouth	0	0	33	0	0	0	0	33	0	0	7
Problem with taste	33	33	33	33	33	100	33	33	33	67	43
Trouble with coughing	0	33	33	67	100	0	0	0	33	33	30
Trouble with speech	33	33	33	67	33	67	0	33	33	100	43

–: speech valve not established; +: poor speech valve function

Table 7. Clinical assessment and results from HRQL questionnaires.

Comments

Although none of our patients had any significant problems with dysphagia, post-operative dysphagia has been reported to be a frequent problem after reconstruction with free jejunal grafts of the pharynx (110). Dysphagia in these patients may be related to functional or anatomic factors, or a combination thereof. Williford et al followed 22 patients with jejunal autotransplants with radiography and observed acute and delayed complications in the form of anastomotic leaks, fistulae, and strictures in 11 patients

(111). Kerlin et al evaluated another 12 patients 2-40 months after surgery (112). The oro-pharyngeal phase of swallowing was not studied in detail by radiography, but it was noted that the barium rapidly passed to the stomach in 7 of the patients. At manometry it was found that swallowing generally failed to induce peristaltic contractions in the graft, explaining the delay in barium transit without any associated stricture, which was observed in three of the patients.

For evaluation of the oral and pharyngeal phases of swallowing we used a video-fluorographic technique, slightly modified compared to the technique described in an article by Dodds, Stewart and Logemann (113). However, as patients surgically treated with reconstruction using a free jejunal transplant lack several anatomic structures (such as the hyoid bone and the larynx), only some of the criteria used to define the oral and pharyngeal phases of swallowing in normal subjects could be used. With this limitation in mind, most patients examined demonstrated seemingly adequate oro-pharyngeal swallowing. When assessing the bolus transit through the jejunal graft, it would have been preferable to use an objective method in the same way that esophageal emptying is evaluated at TBE. Unfortunately, no suitable approach for reliable objective quantification of barium retention could be conceived. The high interobserver agreement in the estimation of impaired transit through the graft found in our study can, however, be regarded as indirect evidence of the reliability of the subjective estimation used. Some further support for the validity of this subjective estimation can be found in the original study describing TBE, where a high agreement between subjective and objective estimation of barium emptying was demonstrated (47).

Signs of various degrees of disturbed bolus passage through the jejunal graft were observed in all patients. It is likely that the frequent findings of oral and/or pharyngeal regurgitations and of after-swallowing were caused by impaired bolus-transit through the graft. Based on the clinical assessments and the results from the HRQL questionnaires, dysphagia was, however, relatively mild in those patients who were free of tumour recurrence, with a high rate of full oral nutrition. Furthermore, the mean global health status/QoL score was 74, which compares well to the general Swedish population (114). The cause of this discrepancy between impaired transit and symptoms of dysphagia is unknown, but it has been noted that the correlation between objective and subjective measures of outcome is imperfect at best in most disorders of gastrointestinal motility, perhaps more than in other disease areas (115). As examples, poor relationships between symptoms and manometry have been observed in esophageal motility disorders (16, 116) as well as between symptoms and scintigraphically measured delayed gastric emptying in diabetes mellitus (117). Impairment of perception due to damage to afferent pathways of the autonomic nervous system in diabetes has been proposed as the cause of the poor correlation between reported symptoms and esophageal dysmotility in this condition (116). In the case of our patients with a free jejunal transplant, one may speculate that lack of perception in a denervated tissue is responsible for the lack of dysphagia in these patients. Correlations between sensory disturbances in the form of hypersensitivity and the severity of symptoms have been demonstrated in motor disorders of the esophagus as well as in Irritable Bowel Syndrome (IBS) (118, 119). It must also be remembered that the pre-surgical expectations on the swallowing ability after the operation might have influenced the self-assessed scoring, contributing to a response shift in the HRQL parameters. Thus, low expectations may have contributed to a relatively good score in a patient with a jejunal transplant, while an otherwise healthy person would have scored the same swallowing difficulties as more severe.

Discussion

Quantitative assessment of treatment response in achalasia

The aims of treatment in achalasia are to decrease the LES pressure, improving esophageal emptying, and most importantly, relieve the symptoms of achalasia. The usual method for assessing the results of treatment of achalasia is clinical evaluation (120). However, it can be argued that objective methods are important in order to estimate the results reliably.

Several authors have observed a disparity between symptomatic response and objective improvement in achalasia (47, 48, 51, 58, 59, 121). It seems reasonable to assume that the placebo effect is responsible for the fact that, in some instances, patients report improvement of symptoms after treatment, despite the continuation of poor esophageal emptying demonstrated by objective methods. If left untreated, persistent poor esophageal emptying is believed to lead to progressive dilatation and ultimately to end-stage megaesophagus, which may require esophagectomy (35). Additionally, as regurgitations, especially nocturnal regurgitations, becomes more frequent with impaired emptying and increasing esophageal dilatation, patients with increased risk of pulmonary complications due to chronic aspiration may go unrecognized if objective methods are not utilized (31, 32, 122).

It also has to be remembered that treatment in achalasia is palliative and the rate of therapeutic success decreases steadily over time irrespective of the type of treatment given (100, 103, 123, 124). Careful questioning of treated patients has disclosed that many patients remain severely symptomatic and have a decreased HRQL compared to healthy controls (125). Additionally, many patients fail to seek help when symptoms recur, possibly due to an erroneous assumption that no further improvement can be expected (126). The chronic character of the disease and the appearance of late recurrences makes life-long follow-up of the patient necessary, and emphasises the need for objective tests that can easily be applied (124). On the other hand, a normal emptying test in the presence of persistent or recurrent symptoms would argue strongly against repeated treatment and for the need to investigate other potential sources for the symptoms, such as GERD. Finally, objective tests are useful for comparison of different treatment modalities for achalasia.

Diminished visceral sensation in achalasia patients compared with controls has been demonstrated and may explain poor perception of esophageal distension and retention (53). Taken together with presented evidence of diminished visceral sensitivity with aging (52), these alterations in esophageal sensation may well be important factors for the subjective perception of symptoms in achalasia.

For the objective evaluation of achalasia patients post-therapy, three methods are commonly used: esophageal manometry, radionuclide studies and barium studies. Each method has pros and cons:

Manometry is the most commonly used objective method. However, manometry does not quantitate the functional disturbance present in achalasia, namely impaired emptying of food (54). A few studies have found a correlation between the post-therapy LES pressure and long-term symptom improvement (51, 120, 127). Eckardt et al studied patients post-dilatation and found that all patients with a LES pressure of 10 mmHg or less was in clinical remission at two years (51). Although this sign had high predictive value, the sensitivity was poorer, as only 8 out of 28 patients (29 %) in remission at two years achieved this degree of LES pressure post-therapy. Ponce and co-workers studied 117 achalasia patients treated with pneumatic dilatation and also found that patients with a post-dilatation LES pressure equal or lower than 10 mm Hg presented a better outcome than the rest (127). Patients with a LES pressure <10 mm Hg had a probability of being in clinical remission one and eight years after the dilatation of 63% and 53%, respectively. The corresponding remission rates in patients with LES pressure >10 mm Hg was 40% and 23%, respectively. Alonso et al found that a decrease of LES pressure to 17 mm Hg or less predicted successful outcome of treatment, evaluated after four to 12 months, with a sensitivity of 89% and a specificity of 93% (120).

Other studies have not been able to find any manometric predictors of symptom response (33, 48, 63, 121, 128, 129). If the resting LES pressure is within the normal range before treatment, which is the case in approximately 40-50% of patients, this value may be less useful in the prediction of the outcome (16). Taking these facts into consideration, it can be questioned if the association between the LES pressure and a prolonged remission is strong enough to be of clinical value. According to guidelines from the American Gastroenterological Association, manometry is *possibly* indicated to assess symptoms in patients who have undergone treatment for achalasia (130). There are some other drawbacks of using esophageal manometry. This test is not well tolerated by patients (59). Parkman et al used a discomfort score (0: no discomfort to 5: extreme discomfort) to judge patient tolerance and found that the discomfort score for esophageal manometry (3.8 ± 1.4; mean ± SD) was significantly higher than that for videoesophagography (2.0 ± 1.4; mean ± SD) (18). The compliance of the patients in completing this test may be suboptimal, as demonstrated in our study III, in which only 17 of the patients scheduled for post-treatment manometry agreed to be examined compared with the 43 patients that completed post-treatment TBE.

Adequacy of esophageal manometry is highly dependent on the technical expertise of the person performing the test. It has been shown that diagnoses of esophageal motility disturbances, including determination of the LES pressure, when interpreted from manometry tracings, have high interobserver variability (131). No study of the reproducibility of esophageal manometry when performed on the same patient in repeated sessions can be found by searching the PubMed. The cost of esophageal manometry is higher than that of a barium study (132) and manometry may not be available in small hospitals. Furthermore, difficulty in the placement of the manometric catheter in a dilated esophagus may sometimes require endoscopy. However, new technologies such as multichannel intraluminal impedance (MII) may improve and complement the manometric diagnosis in patients treated for achalasia, but this remains to be shown (133, 134).

Radionuclide studies have also been used in the objective assessment of treatment response in achalasia. Although there are significant uncertainties involved in estimating effective radiation doses, the typical effective dose from a radionuclide GI-emptying study has been reported to be 0.4 mSv (135). There are no data concerning the radiation exposure in TBE, but the typical effective dose of a "barium swallow" is estimated to be 1.5 mSv (136). As only two or three radiographs are exposed in a TBE study, the effective dose is likely to be less than 1.5 mSv. These doses can be compared with the average annual effective dose from background radiation of about 3 mSv (135). Several investigators have found correlation between short-term clinical response and scintigraphy (48, 54, 95, 137). However, scintigraphy has not been shown to predict the long-term effects of treatment (51, 63, 128). Other limitations of the test include a high degree of technical variability, low availability, and high cost. Therefore, scintigraphy is currently not recommended for the long-term objective assessment after treatment for achalasia (59).

Although a standard barium esophagogram can be used in the objective evaluation of patients post-therapy, there are several draw-backs of this examination. Studies have shown that the radiologic diameter of the esophagus consistently decreases after treatment (34, 50, 105). However, in the case of a greatly dilated esophagus, loss of functional reserve of the esophageal muscular layer may make it impossible for the esophageal diameter to return to normal (120). Also, if the esophagus is not yet dilated before treatment, it is difficult to evaluate the outcome through a decrease in diameter. On the other hand, using only the height of the barium column may also be problematic, as a worsening of the disease with associated increase in diameter leads to a decrease in height (138). Studies that have assessed esophageal emptying post-therapy by the width or the height of the barium column alone, have not found these radiological findings to correlate with the long-term clinical course (51, 55, 63, 64, 128).

Better means of evaluating the disease severity and esophageal emptying after treatment may be to include measurements of both the height and the width of the barium column, the surface area or the volume of the barium column (138). A standardized approach, like that of TBE, may further clarify the role of barium esophagogram in the objective post-therapy assessment of patients (59).

In summary, a very large number of studies have been performed to evaluate the ability of different objective methods to predict the clinical outcome after treatment for achalasia. It seems reasonable to conclude that the correlation between these tests and the symptomatic outcome generally is weak. Further, it has not been possible to agree on a single cut-off value that consistently discriminates between patients with good and poor outcome after treatment with both a high sensitivity and specificity.

TBE has several potential advantages. It is non-invasive, easy to perform, causes relatively mild discomfort for the patient, is inexpensive, and is available outside the specialized institutions. However, TBE as a diagnostic test has prior to this dissertation not been fully validated.

Reliability and validity

The term *validity* refers to truthfulness; does the test measure what it intends to measure (139)? On the variable level and in the field of statistics, validity implies absence of a systematic error, or bias (140). The term *reliability* is the extent to which the measurements of a test remain consistent over repeated tests. This can either be measurements by the same method in a test-retest situation, or in the case of subjective instruments, whether two independent assessors give similar scores. Reliability is inversely related to random error in measurement (140). Reliability is sometimes stratified into reproducibility and repeatability. According to the International Standards Organisation (ISO: 5725) the term reproducibility refers to the strength of agreement between repeated measurements obtained with the same method on the same items under different conditions (different laboratories, operators and/or equipment). The term repeatability refers to the strength of agreement obtained in the same laboratory by the same operator using the same equipment within a "short" interval of time (86). This implies that it may have been more appropriate to use the term repeatability in our study I, but as the time interval may have played an important role in the study, this could motivate the use of the term reproducibility.

The variability between test-and retest measurements in study I can be divided into measurement errors (variability in the implementation of the test and in observer variability) and biologic variability within subjects. Although the variability in the way the examinations were performed as well as the observer variability seems to be small, it is not possible to separate the relative influence of these sources of variability from the biological variability. By repeating the test within a very short time interval, it would have been possible to minimize the effect of fluctuations in physiology within subjects. However, this approach seems to be of limited interest, since the most important aim of the study was to examine the day-to-day variability in conditions as realistic as possible.

The reliability of a test has implications for its validity. The degree of reliability of a measurement method affects the probabilities to detect relevant changes in the variable of interest. The smaller the random variability, the better are the chances to detect significant changes. However, neither validity nor reliability is an either/or dichotomy; there are degrees of each. If the reproducibility of a test is mediocre, and yet the test discriminates well between those with and without the target condition, it still is very useful.

To evaluate the discriminatory power of a new diagnostic test, its results should be compared with an independently established "gold standard" in an appropriate spectrum of patients (89). This step in the validation process concerns the "correlative" or "concurrent" validity of the test (139). However, a perfect gold standard test (with 100% sensitivity and specificity) rarely exists. Therefore, the term "reference standard" is nowadays considered better than "gold standard" (89). In study II we compared TBE with the reference standard manometry, and found significant correlations between certain measures of esophageal emptying at TBE and measures of LES pressure at manometry. It is reasonable to infer that manometry is a suitable reference standard for diagnostic purposes in achalasia (141) and that the "content" validity of manometry in this setting is likely to be high (142). It is also reasonable to assume that symptoms in achalasia primarily are related to impairment in esophageal emptying, a variable that is not measured at manometry. As the destruction of the myenteric plexus is irreversible even after treatment, the functionality of the esphagus is more important for management decision than the neuro-anatomic status. The "content" validity of manometry in the evaluation of response to treatment, could therefore be lower.

Another important step in the validation process of a new, potential diagnostic test is evaluation of its "predictive" validity (139). To be useful for management decisions, the results of the diagnostic test must give information about the clinical outcome or the prognosis of the patients. In study III, we found that the height of the barium column at 1 minute at the post-therapy TBE correlated significantly with dysphagia scores 18 months later. We also found that patients with less than 50% improvement of the height of the barium column post-therapy had a 40% risk of treatment failure during follow-up.

Evidence based diagnostics

Quality standards for diagnostic research are not yet as high as those for treatment research (143). Diagnostic tests are often much less rigorously evaluated than new drugs. Unlike the situation with regard to clinical trials, there are no formal requirements that a diagnostic test must meet in order to be accepted or retained as a routine part of health care. As a consequence of the introduction of evidence-based medicine into radiology, more research efforts are now being devoted into issues addressing to what extent diagnostic testing affects the effectiveness of clinical practice (144). Once the specificity and sensitivity of a test have been established, the final question is whether tested patients fare better than similar untested patients. Architectures for research into diagnostic tests that parallels the established phases in drug research have been proposed (77, 145). A hierarchic model based on six levels of efficacy outcomes in diagnostic imaging has also been described and is now commonly accepted (146) (Table 8). The efficacy outcomes of the two first levels in this model are the traditional goals of diagnostic imaging, technical quality and diagnostic accuracy.

Even if diagnostic imaging is important in modern medicine, clinical diagnosis is often not based on the result of just one test. Setting a diagnosis is a multivariate process of estimating the diagnostic probability of disease presence given available information on patient history, physical findings and combinations of test results (147). On level 3, other sources of information, such as the dysphagia score in the case of achalasia, may dominate, so even a highly accurate test may have little or no effect on the diagnosis. Until recently few authors have investigated diagnostic effect, that is, has the radiologic investigation altered the clinician's diagnosis or rendered other investigations unnecessary (144)? However, calculation of to what extent a test result increase or decrease the probability of disease presence, as estimated from the previous results in an individual patient, involves multivariable prediction models using complex statistics such as logistic

Level of efficacy	Outcome
1. Technical	Imaging parameters Image quality assessment Resolution Noise
2. Diagnostic accuracy	Sensitivity Specificity Accuracy Receiver operating characteristics
3. Diagnostic thinking	Change in pre- and post-test probability Degree of which imaging aids in or changes diagnosis
4. Therapeutic	Degree of which imaging aids in or changes therapy or management
5. Patient outcome	Change in outcome with vs without imaging Morbidity prevented as result of imaging Change in quality-adjusted life expectancy
6. Societal	Cost-effectiveness from societal perspective Cost benefit from societal perspective
Source – from references (149) and (150)	

Table 8. Hierarchical model of efficacy: Typical measures of analysis.

regression or a repeated Bayesian approach (148). As a consequence, not surprisingly, dissociation exists between academic proposals concerning how clinical diagnoses should be made and what doctors actually do when weighting a diagnosis. In particular, it has been found that the actual use of the indices and quantitative procedures proposed by academic circles is the exception rather than the rule (151).

Decision analysis is an objective technique which incorporates probability of events as well as the valuations of possible outcomes of these events (152). Methods in decision analysis include decision trees and threshold probabilities (153). These methods can enable the decision maker to understand the decision problem and choose the management alternative that is most likely to help the patient. If the probability of disease is above the so called treatment threshold probability, treatment should be given. However, in the case of patients treated for achalasia in study III, as the net benefits of treating patients with poor esophageal emptying and the net costs of treating patients with adequate emptying is unknown, it is hard to define this treatment threshold probability. As often in clinical practice this threshold has to be chosen intuitively. However, the physician not only has the options to treat or not treat, it is also possible to choose to obtain additional diagnostic information. By for example performing manometry in all patients that show less than 50% improvement in barium column height at TBE, and selecting those patients with a positive manometry for additional treatment, a better selection of patients for additional treatment may be possible. In terms of decision analysis, the probability of treatment failure in patients with a positive manometry test is then assumed to be lifted above the treatment threshold probability (139). In any case, intuitively the 40% probability of treatment failure in patients with poor improvement in TBE-emptying in study III seems to be above the so called "investigation threshold" (probability above which additional testing is motivated).

To further evaluate the efficacy of a diagnostic test on level 4 and 5 (impact on choice of treatment and on health outcomes), it has been proposed to use randomized study designs (77, 139, 145). However, randomized controlled studies are hard to conduct as they demand large study populations and a long observation time. As a consequence, very few randomized trials deal with diagnostic tests (136, 148). It has also been claimed that a randomized trial may give invalid results and is not always the best approach for many types of diagnostic research (154). Due to the rarity of achalasia it seems unrealistic to verify the usefulness of TBE by means of randomizing the participants to undergo the test or not in such a study.

Due to the difficulties in performing direct experimental studies, assessments of imaging technologies at level 5 and 6 of the efficacy hierarchy are often conducted by using decision analysis (149). Cost-effectiveness analysis is a form of decision analysis that involves evaluation of the costs of health care, as well as the outcomes (155). It involves the creation of algorithms, usually displayed as decision trees, which incorporate probabilities of events and, in case of a cost-utility analysis, the valuations of possible outcomes of these events. Even if such a cost-effectiveness analysis is outside the scope of this thesis, some conclusions from available data seem reasonable. In particular, if it can be assumed that the difference in diagnostic accuracy between TBE and manometry in the evaluation of treatment response is modest, the lower cost (128 € vs. 347 €) and the better patient tolerance would favour the former diagnostic modality.

In conclusion, we have in this thesis evaluated a new diagnostic test, TBE, primarily by obtaining information on the levels of technical efficacy and diagnostic accuracy. At present, there are insufficient data available to determine whether patients tested with TBE fare better in the long run than similar untested patients. Awaiting further studies, considering the degree of diagnostic accuracy, the non-invasiveness, the availability, and the low cost, TBE seems to be the test that has the highest clinical utility for the evaluation of response to treatment for achalasia among available objective tests.

Conclusions

- When examining the day-to-day variability of the TBE-variables, we found moderate to good agreement between measurements of the static variables (height and width of the barium column). However, estimating esophageal emptying by measuring the percentage change of the height or area of the barium column from 1 to 5 minutes showed poor agreement between test and retest.

- The intra- and inter-observer variability of the static TBE-variables was small and well within clinically acceptable limits.

- All patients with newly diagnosed idiopathic achalasia , confirmed by manometry, showed poor esophageal emptying when evaluated by TBE.

- Emptying, expressed as volume of barium at TBE, showed significant inverse correlation with the resting and the maximal relaxing pressure of the LES at manometry (r= -0.34 and r= -0.54, respectively) and with the duration of symptoms (r= -0.36).

- TBE variables measured at the 1 minute time-point showed the closest correlation with the degree of functional impairment, using the manometric resting and relaxing pressure of the LES as the reference. Using the change in barium column height or area from 1 to 5 minutes does not seem to be a valid way to estimate emptying.

- Pneumatic balloon dilatation and laparoscopic myotomy similarly affected esophageal function as assessed by TBE.

- The height of the barium column after 1 minute at the TBE examination performed 6 months post-therapy correlated significantly with symptom scores evaluated median 18 months later. Lack of improvement in barium column height post-treatment was associated with an increased risk of treatment failure.

- Radiological signs of disturbed bolus passage were common in patients treated with radical resection of cancer of the PEJ with reconstruction with a free jejunal transplant, but the clinical impact on dysphagia symptoms and HRQL seemed to be low.

- The clinical utility of TBE and the impact of routinely performing TBE on the long-term outcome of achalasia patients need to be studied in further prospective trials.

Acknowledgements

I wish to express my gratitude to all co-workers, colleagues and friends who have helped and encouraged me and made this thesis possible. In particular, I would like to thank:

- Mikael Hellström, my main supervisor, for your friendship since many years, for your never ending support and skilful guidance. At times, your remarks may be perceived as being very meticulous indeed, but they are always justified and constructive. Thank you, Mikael!
- Lars Lundell, my co-supervisor, for initiating the whole project, for encouraging me, for indispensable help with writing the scientific papers, and for adding a touch of eminence in clinical research to the work.
- Magnus Ruth, my co-supervisor, for valuable guidance and support, and for skilful assistance in completing the manuscripts. Magnus, you are a very nice person!
- Srdjan Kostic, co-author, for managing the achalasia patients and for referring them for radiological examination in a timely manner.
- Henrik Bergquist, co-author, for sometimes slightly bewildered but always enjoyable discussions when assessing the radiological findings in the "dark dungeons" of the radiology department.
- Hans Lönroth, Ann Kjellin and Hasse Ejnell, co-authors, for all valuable contributions.
- My Engström, study nurse at the "GI-lab", for cheerful support and invaluable help with the data collection and clinical assessment of the achalasia patients.
- Ann-Christine Mjörnheim at the "Esophagus-lab" for skilful performance of the manometry studies.
- Liselotte Robertsson for all help with "practical stuff".
- Kjell Geterud and Farida Hashimi for generously allocating time for the research work despite many empty spaces in the daily work schedule. Mats Asztély for friendship and support.
- The staff at the "Uro-gastro-kärl" Unit for helps with conducting the radiological studies.
- All other colleagues and friends at the Department of Radiology for their good humour and friendly support throughout the years.

Sammanfattning på svenska

Akalasi (achalasia cardiae) är en ovanlig sjukdom som ger svåra sväljningsbesvär (dysfagi). Vid akalasi föreligger total aperistaltik och ofullständig relaxation av den nedre esofagussfinktern vid sväljning. Målet för all behandling vid akalasi är att reducera motståndet i den nedre esofagussfinktern. De viktigaste behandlingsalternativen är laparoskopisk myotomi och endoskopisk dilatationsbehandling. Objektiva metoder är viktiga för evaluering av behandlingsresultatet, bl.a. för att tidigt kunna upptäcka progredierande funktions-nedsättning. 1997 beskrevs en standardiserad röntgen-undersökning för kvantifiering av esofugus-tömningen, kallad ”Timed Barium Esophagogram” (TBE).

Syftet med delarbete I var att undersöka graden av reproducerbarhet vid TBE. Tjugoen patienter med akalasi undersöktes med upprepad TBE efter ca en vecka. Röntgenbilder av esofagus togs 1, 2 och 5 minuter efter att patienten intagit upp till 250 ml bariumkontrast. Höjden, vidden och ytan av kontrastpelaren samt deras procentuella ändring från 1 till 5 minuter mättes. De statiska måtten visade god repeterbarhet mellan undersökningarna, medan deras procentuella ändring från 1 till 5 minuter visade dålig korrelation med korrelations-koefficient på 0,50. Intra- respektive interobservatörsvariabiliteten vid mätning av kontrastpelarens höjd och vidd testades i 21 respektive 30 slumpvis utvalda undersökningar. Överensstämmelsen var mycket god med korrelationskoefficienter mellan 0,91-0,99. 14 TBE-undersökningar utfördes även på 8 friska försökspersoner. Alla TBE-parametrar skilde sig signifikant mellan försökspersonerna och akalasipatienterna.

Målet för delarbete II var att beskriva TBE-fynden hos konsekutiva patienter med nydiagnosticerad idiopatisk akalasi (n=46) och att korrelera dessa med manometri och med patienternas symptom. Alla patienterna visade retention av bariumkontrast i esofagus. Tömningen av kontrast uttryckt i ml korrelerade inverst med vilotrycket och det ”maximala relaxationstrycket” i den nedre esofagussfinktern samt med symptomdurationen.

I delarbete III studerades TBE-metoden inom en prospektiv, randomiserad behandlings-studie mellan laparoskopisk myotomi (n=25) och endoskopisk ballongdilatation (n=26). Bedömning gjordes med TBE före (n=46) och efter behandling (n=43). Förmågan hos TBE att prediktera patienterna symptom efter en genomsnittlig uppföljningstid på 18 månader undersöktes även. Det var ingen signifikant skillnad i graden av esofagustömning mätt med TBE efter respektive behandling. Det förelåg signifikanta samband mellan höjden av kontrast-pelaren efter 1 minut och graden av dysfagi för vätska, bröstsmärta och ”Watsons dysfagi-score” efter 18 månaders uppföljning. Patienter med mindre än 50 % förbättring av kontrast-pelarens höjd efter i minut hade 40 % risk för ”treatment failure” under uppföljningen.

I delarbete IV undersöktes sju patienter, opererade p.g.a. hypofarynx- eller proximal esofaguscancer och rekonstruerade med jenunum-interponat, med en modifierad TBE-teknik. Fynden jämfördes med resultatet av strukturerad symptom- och livskvalitets-evaluering. Trots att de flesta patienterna endast rapporterade milda dysfagibesvär så visade röntgen-undersökningen försvårad kontrastpassage i anslutning till interponatet i samtliga fall.

Sammanfattningsvis så fann vi att TBE är en enkel och reproducerbar metod för att objektivt bedöma esofagus tömningsförmåga före och efter behandling av akalasi. Hos patienter med nydiagnosticerad akalasi så medförde beräkning av den tömda volymen kontrast bättre korrelation med det manometriska trycket i den nedre esofagussfinktern än hittills använda mått. Esofagusfunktionen mätt med TBE var likartad efter behandling med laparoskopisk myotomi respektive ballongdilatation. Utebliven minskning av kontrastpelarens höjd efter behandling var förenad med en ökad risk för behov av kompletterande behandling. Trots att undersökning med TBE-liknande teknik av patienter opererade med faryngo-laryngo-esofagektomi visade nedsatt bolustransport genom jejunuminterponatet hos samtliga uppgav patienterna endast lindriga sväljningsbesvär.

References

1. Tibbling L, Gustafsson B. Dysphagia and its consequencies in the elderly. Dysphagia 1991; 6:200-202.
2. Chen P, Golub J, Hapner E, Johns M. Prevalence of perceived dysphagia and qualtity-of-life impairment in a geriatric population. Dysphagia 2008; (Epub ahead of print) Mar 27.
3. Layne K, Losinski D, Zenner P, Ament J. Using the Fleming Index of dysphagia to establish prevalence. Dysphagia 1989; 4:39-42.
4. Spieker M. Evaluating dysphagia. Am Fam Physician 2000; 61:3639-3648.
5. Levine M, Rubesin S. Radiologic investigation of dysphagia. AJR 1990; 154:1157-1163.
6. Lee G, Craig P, Freiman J, de Carle D, Cook I. Intermittent dysphagia for solids associated with a multiringed esophagus: clinical features and response to dilatation. Dysphagia 2007; 22:55-62.
7. Cooper G. Indications and contraindications for upper gastrointestinal endoscopy. Gastrointest Endosc Clin N Am 1994; 4:439-454.
8. Feussner H, Kauer W, Siewert J. The place of esophageal manometry in the diagnosis of dysphagia. Dysphagia 1993; 8:98-104.
9. Panebianco V, Habib F, Tomei E, et al. Initial experience with magnetic resonance fluoroscopy in the evaluation of oesophageal motility disorders. Comparison with manometry and barium fluoroscopy. Eur Radiol 2006; 16:1926-1933.
10. Langmore S, Schatz K, Olsen N. Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. Ann Otol Rhinol Laryngol 1991; 100:678-681.
11. Logemann J. Evaluation and treatment of swallowing disorders. Austin, Tex: Pro-Ed, 1998.
12. Kuo B, Urma D. Esophagus-anatomy and development. http://www.nature.com/gimo/contents/pt1/full/gimo6.html Accessed July 2008.
13. Yamato S, Spechler S, Goyal R. Role of nitric oxide in esophageal peristalsis in the opossum. Gastroenterology 1992; 103:197-204.
14. Crist J, Gidda J, Goyal R. Intramural mechanism of esophageal peristalsis: Roles of cholinergic and noncholinergic nerves. Proc Natl Acad Sci USA 1984; 81:3595-3599.
15. Pandolfino J, Kahrilas P. AGA technical review on the clinical use of esophageal manometry. Gastroenterology 2005; 128:209-224.
16. Richter JE. Oesophageal motility disorders. Lancet 2001; 358:823-828.
17. Chen Y, Ott D, Hewson E, et al. Diffuse esophageal spasm: Radiographic and manometric correlation. Radiology 1989; 170:807-810.
18. Parkman H. Optimal evaluation of patients with nonobstructive esophageal dysphagia. Manometry, scintigraphy, or videoesophagography? Dig Dis Sci 1996; 41:1355-1368.
19. Leite L, Johnston B, Barret J, et a. Ineffective esophageal motility (IEM). The primary finding in patients with nonspecific esophageal motility disorder. Dig Dis Sci 1997; 42:1859-1865.
20. Hollis J, Castell D. Esophageal function in elderly men: a new look at presbyesophagus. Ann Intern Med 1974; 91:897-904.
21. Podas T. Achalasia: a critical review of epidemiological studies. Am J Gastroenterol 1998; 93:2345-2347.

22. Willis T. Pharmaceutice rationalis sive diatribe de medicamentorum operationibus in humano corpore. London, 1672.
23. Hurst A, Rake G. Achalasia of the cardia. Q J Med 1930; 23:491-507.
24. Mayberry J. Epidemiology and demographics of achalasia. Gastrointest Endosc Clin N Am 2001; 11:235-248.
25. Payne S, Ellis FH, Olsen AM. Treatment of cardiospasm (achalasia of the esophagus) in children. Surgery 1961; 50:731-735.
26. Birgisson S, Richter JE. Achalasia: What's new in diagnosis and treatment? Dig Dis 1997; 15:1-27.
27. Goldblum J, Whyte R, Orringer M, al. e. Achalasia: a morphologic study of 42 resected specimens. Am J Surg Pathol 1994; 18:327-337.
28. Holloway R, Dodds W, Helms J, et al. Integrity of cholinergic innervation of the lower esophageal sphincter in achalasia. Gastroenterology 1986; 90:924-929.
29. Park W, Vaezi MF. Etiology and pathogenesis of achalasia: The current understanding. Am J Gastroenterol 2005; 100:1404-1414.
30. Csendes P, Csendes A, Cortez C, Burgos A. Evolutive radiological changes of the esophagus in patients with achalasia who did not receive treatment. Surgery Today 2007; 37:183-186.
31. Spiess A, Kahrilas P. Treating achalasia: From whalebone to laparoscope. JAMA 1998; 280:638-642.
32. Vaezi MF, Richter JE. Diagnosis and management of achalasia. Am J Gastroenterol 1999; 94:3406-3412.
33. Gockel I, Junginger T, Eckardt V. Effects of pneumatic dilation and myotomy on esophageal function and morphology in patients with achalasia. Am Surg 2005; 71:128-131.
34. Csendes A, Bragghetto I, Henriques A, Cortés C. Late results of a prospective randomized study comparing forceful dilatation and oesophagomyotomy in patients with achalasia. Gut 1989; 30:299-304.
35. Pohl D, Tutuian R. Achalasia: an overview of diagnosis and treatment. J Gastrointest Liver Dis 2007; 16:297-303.
36. Blam M, Delfyett W, Levine M, Metz D, Katzka D. Achalasia: A disease of varied and subtle symptoms that do not correlate with radiographic findings. Am J Gastroenterol 2002; 97:1916-1923.
37. Garofalo J, Pofahl W. Achalasia: A brief review of treatment options and efficacy. Current surgery 2002; 59:549-553.
38. Prior A. Oesophageal achalasia mistaken for anorexia nervosa. BMJ 1992; 305:833-834.
39. Howard P, Mahler L, Pryde A, Cameron E, Heading R. Five year prospective study of the incidence, clinical features, and diagnosis of achalasia in Edinburgh. Gut 1992; 33:1011-1015.
40. Debas H, Payne W, Cameron A, et al. Physiopathology of the lower esophageal diverticulum and its complications. Surg Gynecol Obstet 1980; 151:593-600.
41. Ergun GA, Kahrilas PJ. Clinical applications of esophageal manometry and pH monitoring. Am J Gastroenterol 1996; 91:1077-1089.
42. Katz PO, Richter JE, Cowan R, et al. Apparent complete lower esophageal sphincter relaxation in achalasia. Gastroenterology 1986; 90:978-983.

43. Mearin F, Malegelada J. Complete lower esophageal sphincter relaxation observed in achalasia patients is functionally inadequate. Am J Physiol Gastrointest Liver Physiol 2000; 278:376-383.
44. Kahrilas P, Kishk S, Helm J, Dodds W, Harig J, Hogan W. Comparison of pseudoachalasia and achalasia. Am J Med 1987; 82:439-446.
45. Boeckxstaens GE. Achalasia. Best Pract Res Clin Gastroenterol 2007; 21:595-608.
46. Kostic S, Kjellin A, Ruth M, et al. Pneumatic dilatation or laparoscopic cardiomyotomy in the management of newly diagnosed idiopathic achalasia. Results of a randomized controlled trial. World J Surg 2007; 31:470-478.
47. de Oliveira J, Birgisson S, Doinoff C, et al. Timed Barium Swallow: A simple technique for evaluating esophageal emptying in patients with achalasia. AJR 1997; 169:473-479.
48. Robertsson CS, Hardy JG, Atkinson M. Quantitative assessment of the response to therapy in achalasia of the cardia. Gut 1989; 30:768-773.
49. Vaezi M, Baker M, Richter J. Assessment of the esophageal emptying post-pneumatic dilation: use of the timed barium esophagogram. Am J Gastroenterol 1999; 94:1802-1807.
50. Vantrappen G, Hellemans J, Deloof W, Valembois P, Vandenbroucke J. Treatment of achalasia with pneumatic dilatations. Gut 1971; 12:268-275.
51. Eckardt VF, Aignherr C, Bernhard G. Predictors of outcome in patients with achalasia treated by pneumatic dilation. Gastroenterology 1992; 103:1732-1738.
52. Lasch H, Castell D, Castell J. Evidence of diminished visceral pain with aging: Studies using graded intraoesophageal balloon distension. Am J Physiol 1997; 272:G1-G3.
53. Brackbill S, Shi G, Hirano I. Diminished mechanosensitivity and chemosensitivity in patients with achalasia. Am J Physiol Gastrointest Liver Physiol 2003; 285:G1198-G1203.
54. Holloway R, Krosin G, Lange R, Baue A, McCallum R. Radionuclide esophageal emptying of a solid meal to quantitate results of therapy in achalasia. Gastroenterology 1983; 84:771-776.
55. Lee J, Cecil B, Brown P, Wright R. The Cohen test does not predict outcome in achalasia after pneumatic dilation. Gastrointest Endosc 1993; 39:157-160.
56. Vela M, Richter J, Wachsberger D, Connor J, Rice T. Complexities of managing achalasia at a tertiary referral center: use of pneumatic dilatation, Heller myotomy, and botulinum toxin injection. Am J Gastroenterol 2004; 99:1029-1036.
57. West R, Hirsch D, Batelsman J, et al. Long term results of pneumatic dilatation in achalasia followed for more than 5 years. Am J Gastroenterol 2002; 97:1346-1351.
58. Vantrappen G, Hellemans J. Treatment of achalasia and related motor disorders. Gastrenterology 1980; 79:144-154.
59. Vaezi M. Quantitative methods to determine efficacy of treatment in achalasia. Gastrointest Endosc Clin North Am 2001; 11:409-424.
60. Gockel I, Junginger T. The value of scoring achalasia: A comparison of current systems and the impact on treatment-the surgeon's viewpoint. Am Surg 2007; 73:327-331.
61. Urbach D, Tomlinson G, Harnish J, Martino R, Diamant N. A measure of disease-specific health-related quality of life for achalasia. Am J Gastroenterol 2005; 100:1668-1676.
62. Cohen N. An endpoint for pneumatic dilation of achalasia. Gastrointest Endosc 1975; 22:29.

63. Kim C, Cameron A, Hsu L, et al. Achalasia: Prospective evaluation of relationship between lower esophageal sphincter pressure, esophageal transit, and esophageal diameter and symptoms in response to pneumatic dilation. Mayo Clin Proc 1993; 68:1067-1073.
64. Wong R, Madonovitch C. Utility of parameters measured during pneumatic dilatation as predictors of successful dilatation. Am J Gastroenterol 1996; 91:1126-1129.
65. Pavlik E, DePriest P, Gallion H, et al. Ovary volume related to age. Gynocol Oncol 2000; 77:410-412.
66. Dent J, Chir B. A new technique for continuous sphincter pressure measurement. Gastroenterology 1976; 71:263-267.
67. Kostic S, Rice T, Baker M, et al. Timed Barium Esophagogram: A simple physiologic assessment for achalasia. J Thorac Cardiovasc Surg 2000; 120:935-943.
68. Dakkak M, Bennet J. A new dysphagia score with objective validation. J Clin Gastrenterol 1992; 14:99-100.
69. Svedlund J, Sjödin I, Dotevall G. GSRS. A clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. Dig Dis Sci 1988; 33:129-134.
70. Korolilja D, Sauerland S, Wood-Dauphinée S, et al. Evaluation of quality of life after laparoscopic surgery. Surg Endosc 2004; 18:879-897.
71. Schag C, Heinrich R, Ganz P. Karnofsky performance status revisited: reliability, validity, and guidelines. J Clin Oncol 1984; 2:187-193.
72. Ahmad I, Kumar B, Bradford K, et al. Surgical voice restoration following ablative surgery for laryngeal and hypopharyngeal carcinoma. J Laryngol Otol 2000; 114:522-525.
73. Ogilvie A, Dronfield M, Fergusson R, Atkinsson M. Palliative intubation of oesophagogastric neoplasms at fibreoptic endoscopy. Gut 1982; 23:1060-1067.
74. Aronson N, Ahmedzai S, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993; 85:365-376.
75. Blazeby J, Conroy T, Hammerlid E, et al. Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. Eur J Cancer 2003; 39:1384-1394.
76. Grimes D, Schultz K. An overview of clinical research: the lay of the land. Lancet 2002; 359:57-61.
77. Sackett D, Haynes R. The architecture of diagnostic research. BMJ 2002; 324:539-541.
78. Grimes D, Schultz K. Descriptive studies: what they can and cannot do. Lancet 2002; 359:145-149.
79. Blackmore C, Black W, Jarvik J, Langlotz C. A critical synopsis of the diagnostic and screening outcome literature. Acad Radiol 1999; 6:S8-S18.
80. Sica G. Bias in research studies. Radiology 2006; 238:780-789.
81. Concato J, Shah N, Horwitz R. Randomized, controlled trials, observational studies, and the hierarchy of research design. N Engl J Med 2000; 342:1887-1892.
82. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i:307-310.
83. Kundel H, Polansky M. Measurement of observer agreement. Radiology 2003; 228:303-308.
84. Bland J, Altman D. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 2003; 22:85-93.

85. Halligan S. Correspondence: Reproducibility, repeatability, correlation and measurement error. The British Journal of Radiology 2002; 75:193-195.
86. ISO. Accuracy (trueness and precision) of measurement methods and results. In:ISO 5725-1. Geneva, Switzerland, 1994.
87. Altman D. Practical statistics for medical research. London: Chapman & Hall, 1991.
88. Rose G, Barker D. Epidemiology for the uninitiated; observer variation. BMJ 1978; 2:1006-1007.
89. Knottnerus JA. The evidence base of clinical diagnosis: BMJ Books, 2002.
90. Tolin R, Malmud L, Reilley J, Fisher R. Esophageal scintigraphy to quantitate esophageal transit (quantitation of esophageal transit). Gastroenterology 1979; 76:1402-1408.
91. Gross R, Johnson L, Kaminski R. Esophageal emptying in achalasia quantitated by a radioisotope technique. Dig Dis Sci 1979; 24:945-949.
92. Rozen P, Gelfond M, Zaltzman S, Baron J, Gilat T. Dynamic, diagnostic, and pharmacological radionuclide studies of the esophagus in achalasia. Radiology 1982; 144:587-590.
93. Chung J, Park H, Yu S, et al. A comparison of esophagography and esophageal transit scintigraphy in the evaluation of usefulness of endoscopic pneumatic dilatation in achalasia. Acta Radiol 2008; 49:498-505.
94. Mariani G, Boni G, Barreca M, et al. Radionuclide gastroesophageal motor studies. J Nuc Med 2004; 45:1004-1028.
95. Marshall J, Bodnarchuk G, Singh A. Supine and upright radionuclide esophageal transit before and after treatment for achalasia. Clin Nucl Med 1994; 19:683-686.
96. Schima W, Stacher G, Pokieser P, et al. Esophageal motor disorders: videofluoroscopic and manometric evaluation-prospective study in 88 symptomatic patients. Radiology 1992; 185:487-491.
97. El-Takli I, O'Brien P, Paterson W. Clinical diagnosis of achalasia: how reliable is the barium x-ray? Can J Gastroenterol 2006; 20:335-337.
98. Amaravadi R, Levine M, Rubesin S, Laufer I, Redfern R, Katzka D. Achalasia with complete relaxation of lower esophageal sphincter: Radiographic-Manometric correlation. Radiology 2005; 235:886-891.
99. Zaninotto G, Annese V, Costantini M, et al. Randomized controlled trial of botulinum toxin versus laparoscopic Heller myotomy for esophageal achalasia. Ann Surg 2004; 239:364-370.
100. Lopushinsky S, Urbach D. Pneumatic dilatation and surgical myotomy for achalasia. JAMA 2006; 296:2227-2233.
101. Eckardt V, Gockel I, Bernhard G. Pneumatic dilation for achalasia: late results of a prospective follow up investigation. Gut 2004; 53:629-633.
102. Gockel I, Junginger T, Eckardt V. Long-term results of conventional myotomy in patients with achalasia: a prospective 20-year analysis. J Gastrointest Surg 2006; 10:1400-1408.
103. Vela MF, Richter JE, Khandwala F, et al. The long-term efficacy of pneumatic dilatation and Heller myotomy for the treatment of achalasia. Clin Gastroenterol Hepatol 2006; 4:580-587.
104. Levels of evidence. Oxford Centre for Evidence-Based Medicine Web site. http://www.cebm.net/levels_of_evidence.asp. Accessed may 2008.
105. Duranceau A, Lafontine E, Vallieres B. Effects of total fundoplication on function of the esophagus after myotomy for achalasia. Am J Surg 1982; 143:22-28.

106. Vaezi M, Richter J, Wilcox C, et al. Botulinum toxin versus pneumatic dilatation in the treatment of achalasia: a randomized trial. Gut 1999; 44:231-239.
107. Vaezi M, Baker M, Achkar E, Richter J. Timed barium esophagogram: better predictor of long term success after pneumatic dilation in achalasia than symptom assessment. Gut 2002; 50:765-770.
108. Ewal B. Post hoc choice of cut points intoduced bias to diagnostic research. J Clin Epidemiol 2006; 59:798-801.
109. Hamlet SL, Muz J, Patterson R, Jones L. Pharyngeal transit time: assessment with videofluoroscopic and scintigraphic techniques. Dysphagia 1989; 4:4-7.
110. Smith D, Ott D, McGuirt W, Albertson D, Chen M, Gelfand D. Free jejunal grafts of the pharynx: Surgical methods, complications, and radiographic evaluation. Dysphagia 1999; 14:176-182.
111. Williford M, Rice R, Kelvin F, Fisher S, Meyers W, Thompson W. Revascularized jejunal graft replacing the cervical esophagus: Radiographic evaluation. AJR 1985; 145:533-536.
112. Kerlin P, McCafferty G, Robinson D, Thiele D. Function of a free jejunal "conduit" graft in the cervical esophagus. Gastroenterology 1986; 90:1956-1963.
113. Dodds W, Stewart E, Logemann J. Physiology and radiology of the normal oral and pharyngeal phases of swallowing. AJR 1990; 154:953-963.
114. Michelson H, Bolund C, Nilsson B, Brandberg Y. Health-related quality of life measured by the EORTC QLQ-C30--reference values from a large sample of Swedish population. Acta Oncol 2000; 39:477-484.
115. Pasricha P. Commentary: Quality versus quantity: Outcomes after achalasia surgery. Dig Liv Dis 2006; 38:551-553.
116. Faraj J, Melander O, Sundkvist G, et al. Oesophageal dysmotility, delayed gastric emptying and gastrointestinal symptoms in patients with diabetes mellitus. Diabet Med 2007; 24:1235-1239.
117. Horowitz M, O'Donovan D, Jones K, Feinle C, Rayner C, Samsom M. Gastric emptying in diabetes: clinical significance and treatment. Diabet Med 2002; 19:177-194.
118. Rao S, Hayek B, Summers R. Functional chest pain of esophageal origin: hyperalgesia or motor dysfunction. Am J Gastroenterol 2001; 96:2584-2589.
119. Posserud I, Syrous A, Lindström L, Tack J, Abrahamsson H, Simrén M. Altered rectal perception in irritable bowel syndrome is associated with symptom severity. Gastroenterol 2007; 133:1113-1123.
120. Alonso P, Gonzales-Conde B, Macenille R, Pita S, Vazquez-Iglesias J. Achalasia: The usefulness of manometry for evaluation of treatment. Dig Dis Sci 1999; 44:536-541.
121. Gholoum S, Feldman L, Andrew C, et al. Relationship between subjective and objective outcome measures after Heller myotomy and Dor fundoplication for achalasia. Surg Endosc 2006; 20:214-219.
122. Sonnenberg A, Massey B, McCarty D, Jacobsen S. Epidemiology of hospitalization for achalasia in the United States. Dig Dis Sci 1993; 38:233-244.
123. Eckardt V, Gockel I, Bernhard G. Pneumatic dilatation for achalasia: late results of a prospective follow up investigation. Gut 2004; 53:629-633.
124. Ortiz A, Matinez de Haro L, Parilla P, et al. Very long-term objective evaluation of Heller myotomy plus posterior partial fundoplication in patients with achalasia of the cardia. Ann Surg 2008; 247:258-264.

125. Frankuisen R, Herwaarden Mv, Heijkoop R, et al. Persisting symptoms and decreased health-related quality-of-life in a cross-sectional study of treated achalasia patients. Aliment Pharmacol Ther 2007; 26:899-904.
126. Torbey C, Achkar E, Rice T, Baker M, Richter J. Long-term outcome of achalasia treatment:The need for closer follow-up. J Clin Gastrenterol 1999; 28:125-130.
127. Ponce J, Garrigues V, Pertejo V, Sala T, Berenguer J. Individual prediction of response to pneumatic dilation in patients with achalasia. Dig Dis Sci 1996; 41:2135-2141.
128. Kadakia S, Wong R. Graded pneumatic dilation using Rigiflex achalasia dilators in patients with primary esophageal achalasia. Am J Gastroenterol 1993; 88:34-38.
129. Shahi H, Aggarwal R, Misra A, et al. Relationship of manometric findings to symptomatic response after pneumatic dilation in achalasia cardia. Indian J Gastroenterol 1998; 17:19-21.
130. Pandolfino JE, Kahrilas PJ. American Gastroenterological Association medical position statement: Clinical use of esophageal manometry. Gastroenterology 2005; 128:207-208.
131. Nayar D, Khandwala F, Achkar E, et al. Esophageal manometry: Assessment of interpreter consistency. Clin Gastroenterol Hepatol 2005; 3:218-224.
132. Kostic S, Johnsson E, Kjellin A, et al. Health economic evaluation of therapeutic strategies in patients with idiopathic achalasia: results of a randomized trial comparing pneumatic dilatation with laparoscopic cardiomyotomy. Surg Endosc 2007; 21:1184-1189.
133. Agrawal A, Hila A, Tutuian R, Castell D. Manometry and impedance characteristics of achalasia. Facts and myths. J Clin Gastroenterol 2008; 43:266-270.
134. Tatum RP, Wong JA, Figueredo EJ, Martin V, Oelschlager BK. Return of esophageal function after treatment for achalasia as determined by impedance-manometry. J Gastrointest Surg 2007; 11:1403-1409.
135. Mettler F, Huda W, Yoshizumi T, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: A catalog. Radiology 2008; 248:254-263.
136. Referral guidelines for imaging. Adapted by experts representing European radiology and nuclear medicine. In: European commission 2000.
137. Gelfand M, Kozarek R. An experience with polyethylene balloons for pneumatic dilation in achalasia. Am J Gastroenterol 1989; 84:924-927.
138. Montazeri G. Surface area: a better predictor of disease severity than the height and volume of the barium column in patients with primary achalasia. Eur J Gastroenterol Hepatol 2006; 18:1203-1208.
139. Gellerstedt M, Furberg B. Diagnostik - en tolkningsfråga?: MSD-litteratur, 2007.
140. Ranstam J. Methodological note: Accuracy, precision, and validity. Acta Radiol 2008; 49:105-106.
141. Hirano I, Tatum R, Shi G, Sang Q, Joehl R, Kahrilas P. Manometric heterogeneity in patients with idiopathic achalasia. Gastroenterol 2001; 120:789-798.
142. http://allpsych.com/researchmethods/validityreliability.html. In, Accessed july 2008.
143. Buntix F, Knottnerus J. Editorial. Are we at the start of a new era in diagnostic research? J Clin Epidemiol 2006; 59:325-326.
144. The evidence-based radiology working group. Evidence-based radiology: A new approach to the practice of radiology. Radiology 2001; 220:556-575.
145. Gluud C, Gluud L. Evidence based diagnostics. BMJ 2005; 330:724-726.
146. Fryback D, Thornbury J. The efficacy of diagnostic imaging. Med Decis Making 1991; 11:88-94.

147. Moons K, Biesheuvel C, Groebbee D. Test research versus diagnostic research. Clin Chem 2004; 50:473-476.
148. Hernandez-Aguado I. The windling road toward evidence based diagnoses. J Epidemiol Community Health 2002; 56:323-325.
149. Sunshine J, Applegate K. Technology assessment for radiologists. Radiology 2004; 230:309-314.
150. Thornbury J. Clinical efficacy of diagnostic imaging: love it or leave it. AJR 1994; 162:1-8.
151. Reid C, Lane D, Feinstein A. Academic calculations versus clinical judgments: Praticicing physician's use of quantitative measures of test accuracy. Am J Med 1998; 104:374-380.
152. Janne D'Othee B, Black W, Pirard S, Zhuang Z, Bettman M. Decision analysis in radiology. J Radiol 2001; 82:1693-1698.
153. Sox H, Blatt M, Higgins M, Marton K. Medical decision making. Stoneham, MA: Butterworths, 1988.
154. Bossuyt P, Lijmer J, Mol B. Randomised comparisons of medical tests: sometimes invalid, nor always efficient. Lancet 2000; 356:1844-1847.
155. Singer M, Applegate K. Cost-effectiveness analysis in radiology. Radiology 2001; 219:611-620.