

The influence of age on gastro-oesophageal reflux: a re-appraisal of the DeMeester scoring system[☆]

Anjum Jalal*, Helena R.J. Payne, K. Jeyasingham

Frenchay Hospital, Bristol, UK

Received 6 September 1999; received in revised form 30 March 2000; accepted 7 August 2000

Abstract

Objective: The statistical evaluation of the influence of age on the pattern of gastro-oesophageal reflux (GOR) identified by prolonged pH monitoring in asymptomatic subjects. Re-appraisal of the DeMeester scoring system for GOR. **Method:** Prolonged pH monitoring was performed on 45 asymptomatic elderly adults with normal contrast oesophagogram, manometry and endoscopy. They included 36 males and nine females. The mean age was 66.6 years. The monitoring time ranged between 20 and 24 h, including one complete daily feeding cycle. GOR was defined as a reflux event with a pH of <4. The mean, standard deviation, kurtosis and skew were calculated for six parameters listed in the results. The mean values were compared with those of normal values determined by DeMeester et al. (in: Read NW, editor. Gastrointestinal motility: which test? 1989, pp. 43–52) from their study of 50 young healthy adults, and the *t*-test was applied to determine the statistical significance of differences. The ‘null hypothesis’ for each parameter was defined as ‘the mean values of the elderly population are not statistically different from DeMeester’s normal values’. **Results:** The means (\pm SD) of six parameters studied in pH monitoring were as follows: supine reflux time as a percentage of total study time, $2.94 \pm 5.18\%$; upright reflux time as a percentage of total study time, $4.14 \pm 5.71\%$; total reflux time as a percentage of total study time, $3.5 \pm 4.38\%$; duration of longest reflux episode, 14.98 ± 24.92 min; number of reflux episodes lasting >5 min, 1.76 ± 2.75 ; total number of reflux episodes during study, 13.49 ± 11.31 . These results were significantly different from the normal values reported by DeMeester. In addition, the data for each individual parameter was grossly skewed, as well as kurtotic, which implied that the data did not represent a normally distributed population. Moreover, we believe that the equation used for calculation of the DeMeester score, is inappropriate. **Conclusions:** The null hypothesis is rejected as the mean values of these parameters in our group are significantly higher than those used as normal. This implies that the normal values defined by DeMeester would over-diagnose gastro-oesophageal reflux disease (GORD). Moreover, we have found that the formula used to calculate the DeMeester’s score is not according to the principle it is based on. DeMeester’s system scoring is therefore inappropriate. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: DeMeester score; Gastro-oesophageal reflux; Prolonged pH monitoring

1. Introduction

The DeMeester score is widely used for diagnosing and quantifying gastro-oesophageal reflux disease (GORD). It assigns a score to the values of six parameters used in prolonged pH monitoring on the basis of their difference from normal mean values. The normal values have been determined by Johnson and DeMeester [1] in their original study of 50 healthy young adults. In practice, we observed that even completely asymptomatic subjects tend to have higher DeMeester scores in the elderly population. This

study was done to perform a detailed statistical evaluation of the difference in the pattern of gastro-oesophageal reflux (GOR). Moreover, after a detailed re-appraisal of this scoring system, we discovered that the method of calculation of the DeMeester score was based on an inappropriate principle and an incorrect statistical formula.

2. Subjects and methods

2.1. Study design

A group of 100 elderly subjects, who attended our thoracic unit for non-oesophageal diagnostic or therapeutic procedures, was included in this study. All these patients were completely asymptomatic of reflux, were not taking any antacids and did not have any previous history of gastro-

[☆] Presented at the 13th Annual Meeting of the European Association for Cardio-thoracic Surgery, Glasgow, Scotland, UK, September 5–8, 1999.

* Corresponding author. The Cardio-thoracic Centre, Thomas Drive, Liverpool L14 3PE, UK. Tel.: +44-370-600-705; fax: +44-151-293-2256.

E-mail address: anjumalal1@hotmail.com (A. Jalal).

oesophageal disorder. They were not on any other medicines which could provoke GOR or affect oesophageal motility. After detailed clinical evaluation, they underwent the preliminary barium swallow test to exclude hiatal herniae and any gross abnormality of oesophagus. Clinically normal subjects underwent static manometric evaluation of the oesophagus, especially regarding the location and function of the lower oesophageal sphincter (LOS). This was followed by prolonged pH monitoring. Later on, they underwent endoscopic evaluation of oesophagus and stomach before surgery for pulmonary lesions.

2.2. Exclusion criteria

Fifty-five subjects were excluded due to any one or more of the following criteria:

- Abnormal barium test.
- Abnormal profile of static manometry demonstrating abnormal motility or incompetent LOS.
- Lack of compliance, dislike of procedure or excessive anxiety during manometry or pH monitoring.
- Development of any GI symptoms, like nausea, vomiting or excessive swallowing, during the pH monitoring.
- Total pH monitoring time of <20 h or missing any meals during the study period which would have been taken during a usual 24 h circadian cycle.
- Reflux oesophagitis or Barrett's oesophagus noticed on subsequent endoscopy.

2.3. Subjects

Finally, we had 45 subjects in the study. They included 36 males and nine females. The mean age was 66.6 years with a standard deviation of 5.6 years. The age ranged between 54.5 and 76.25 years, and the median age was 65 years. The 5th and 95th percentiles were 57.6 and 76 years, respectively which demonstrates that the majority of subjects (90%) were more than 60 years of age.

2.4. Methods

A preliminary oesophageal manometry was performed using the station pull-through technique. The LOS was located using a Gaeltec solid state triple probe manometric catheter connected to a Roche amplifier-recorder trolley system. A Synectics pH probe was positioned 5 cm above the recorded level of the LOS and connected to a Synectics Mark II (Gold) Digitrapper for continuous pH monitoring over 23 h. Although the initial manometry and insertion of the pH probe were performed on an empty stomach, the subjects were permitted to carry on with normal day to day activities once the pH monitoring was commenced. On completion of the 24 h period, the data in the digitrapper were downloaded onto an IBM compatible desktop personal

computer for analysis of the results using the Synectics Gastrosoft package.

2.5. Statistical analysis

Summary statistics (e.g. mean, median, standard deviation, etc.) for each of the six parameters of prolonged pH monitoring, as well as for various parameters of manometry, were calculated. We compared the results of pH monitoring of our group of people with the normal values described by DeMeester et al. The unpaired two tailed *t*-test was used and *P* values were determined to find the statistical significance of the differences. The null hypothesis (H_0) was defined as 'the mean values of study group (X_1) were not higher than DeMeester's group (X_2)', i.e. Eq. (1):

$$H_0 = X_1 = X_2 \quad (1)$$

The alternate hypothesis (H_α) was therefore defined as in Eq. (2):

$$H_\alpha = X_1 \neq X_2 \quad (2)$$

Differences were considered statistically significant if the *P* values were ≤ 0.05 . Since most of the values in our study were higher than DeMeester's normal values, we also determined *P* values for the single tailed *t*-test to prove whether there was any statistical significance in this finding.

One limitation of this analysis was the unavailability of raw data from the original study of DeMeester and colleagues. The only available information were 'means and standard deviations'. Although the *Z*-test showed a significant difference between both groups, it is not very reliable due to a lack of normal distribution in our data. We therefore had to rely on the *t*-test for comparison of the means and standard deviations of both groups.

3. Results

3.1. Manometry and endoscopy

As mentioned earlier, all subjects included in the study were free of symptoms and had normal barium study. All of the 45 subjects selected for study had normal LOS function on manometry. Table 1 gives the means, medians, standard deviations, maximum and minimum values of the individual parameters used in static manometry. All these values are consistent with normal LOS function, and represent no abnormality that could lead to GOR. All these subjects had normal endoscopic findings. None had Barrett's oesophagus.

3.2. The data of individual parameters is not normally distributed

The statistical values of kurtosis and skewness of different parameters given in Table 2 clearly demonstrate that these values are not normally distributed. All these values

Table 1
Results of static manometry^a

Parameter	N	Mean	Median	SD	Min.	Max.
LOS pressure (mmHg)	45	12.1	12.0	04.0	05.0	20.0
LOS contraction amplitude (mmHg)	45	32.1	27.0	14.5	17.0	63.0
LOS contraction duration (s)	45	08.8	07.8	03.9	04.2	21.6
LOS relaxation amplitude (mmHg)	45	15.2	15.0	05.4	05.0	26.0
LOS relaxation duration (s)	45	05.7	06.0	02.0	03.0	08.7
UOS pressure (mmHg)	17	26.4	27.0	07.6	15.0	40.0
UOS contraction amplitude (mmHg)	17	54.2	48.0	18.6	27.0	85.0
UOS contraction duration (s)	17	03.3	02.3	02.7	01.0	10.0
UOS relaxation amplitude (mmHg)	17	28.7	26.0	09.3	17.0	50.0
UOS relaxation duration (s)	17	00.8	00.8	00.3	00.5	01.6
BO contraction amplitude (mmHg)	17	32.5	32.0	06.7	21.0	46.0
BO contraction duration (s)	17	03.0	03.0	00.5	02.1	03.8
Pharyngeal contraction amplitude (mmHg)	17	47.0	50.0	09.7	20.0	60.0
Pharyngeal contraction duration (s)	17	00.4	00.4	00.2	00.2	00.8

^a UOS, upper oesophageal sphincter; BO, body of oesophagus; N, number of cases.

are skewed towards the right hand side which implies that the values are clustered on the higher side. We believe this is an extremely important observation as it directly questions the validity of a scoring system which is based on the concept of standard deviations from the means of a normally distributed population. The logic and procedure of determining the DeMeester score is further elaborated in Section 4.

3.3. Prolonged pH monitoring and pattern of the oesophageal reflux in old age

The means and standard deviations of various parameters in our study are generally higher than DeMeester's normal values (Table 3). This confirms that oesophageal reflux is higher in old age. In addition to this, we noticed that the difference was more marked in cases of some parameters than others. For instance, the supine reflux was nearly 4.6 times higher than the normal value, whereas the upright reflux was only 1.8 times higher than normal. More inter-

Table 2
Results of 24 h pH monitoring in old age subjects^a

Parameter	Mean ± SD	Kurtosis	Skewness
Supine reflux	2.94 ± 5.18	3.59	2.05
Upright reflux	4.14 ± 5.71	5.58	2.25
Total reflux	3.50 ± 4.38	1.44	1.49
Number of episodes lasting >5 min	1.76 ± 2.75	3.82	1.98
Duration of the longest episode	14.98 ± 24.92	17.92	3.74
Total number of episodes	13.49 ± 11.31	2.66	1.30

^a N = 45.

estingly, in this study, the older people had less frequent episodes of reflux as the total number of reflux episodes was 71% of the normal. The duration of the longest reflux episode was, in general, much longer than normal. The mean duration of the longest episode in the present study was found to be 14.98 min (DeMeester's mean, 6.74 min), and the highest value noted in our study was 147 min. The pH of that particular reflux episode was just below 4. This observation suggests that mechanisms for the clearance of reflux become inefficient in old age. All of the above-mentioned observations suggest that the pattern of reflux in old age is different from that in the younger age.

4. Discussion

GOR is a normal physiological phenomenon and occurs in every individual. It is called GORD when it is associated with symptoms or histological changes in the lower oesophagus. Within physiological limits, GOR is influenced by a variety of factors, e.g. type of food, variations in the oesophageal motility, medications, psychological stress and posture, etc. Large epidemiological surveys have demonstrated that old age, male sex and white ethnicity are risk factors in the development of much more severe GORD [2]. However, little is known about the relevance of changed pattern of GOR in producing disease and its complications in old age.

Up until the early 1990s, there were not many large studies available in the literature to evaluate the influence of age on the pattern of oesophageal reflux. Spence et al. [3] and Fass et al. [4] have reported that there are no differences in the pattern of GOR studied in young and old subjects. Both of these studies have very limited value, as the number of subjects in each study is too small to make any clinically significant generalization. However, studies on the physiology of the ageing oesophagus have provided evidence that, in old age, the production of saliva is reduced and secondary oesophageal peristalsis is less frequent, as well as less consistent [5]. Since these two factors are major mechanisms of lower oesophageal clearance, this can explain why reflux episodes tend to last much longer in elderly people. This fact has been observed by Smout et al. [6], as well as in the current study. In addition to this, the clinical presentation of reflux is also altered, which is either due to the relative insensitivity of the oesophagus to acid exposure or the decreased content of acid in relatively longer reflux episodes.

In the younger population, GORD can be diagnosed with confidence on careful clinical evaluation of the symptoms. On the contrary, in older age, it frequently presents as a part of a differential diagnosis of rather more serious conditions, and therefore, requires objective evidence for confirmation of diagnosis. A variety of investigations, with their own sensitivities and specificities (Table 4), can be performed to establish the diagnosis. These investigations, which include lower oesophageal manometry, endoscopy and

Table 3

Comparison of prolonged pH monitoring results of present study vs. DeMeester's normal values^a

Parameter	Present study (mean ± SD; n = 45)	DeMeester's (mean ± SD; n = 50)	Z	t	P (two tailed)	P (one tailed)
Supine reflux time (as percentage of total study time)	2.94 ± 5.18	0.63 ± 1.0	2.96	3.20	< 0.002	< 0.001
Upright reflux time (as percentage of total study time)	4.14 ± 5.71	2.34 ± 2.34	1.76	1.90	> 0.05	< 0.05
Total reflux time (as percentage of total study time)	3.5 ± 4.38	1.51 ± 1.36	2.92	3.10	< 0.002	< 0.001
Number of episodes lasting >5 min	1.76 ± 2.75	0.84 ± 1.18	2.00	2.14	< 0.05	< 0.025
Duration of the longest episode (min)	14.98 ± 24.92	6.74 ± 7.85	2.12	2.25	< 0.05	< 0.025
Total number of episodes	13.49 ± 11.31	19.00 ± 12.76	2.23	2.26	< 0.05	< 0.025

^a P values are determined by the unpaired Student's *t*-test. Values below 0.05 are considered significant. $|Z| > 1.96$ is significant.

biopsy, contrast oesophagogram, gastro-oesophageal scintiscanning, Bernstein's acid perfusion test and oesophageal pH monitoring, can provide this evidence.

Estimation of oesophageal reflux by measuring oesophageal pH was described by Skinner and Booth in 1970 [7]. In 1974, Johnson and DeMeester modified the methodology of oesophageal pH monitoring. They increased the monitoring period to 24 h [1] to make a thorough evaluation of the reflux pattern during the day and night, and in the upright as well as supine positions. For interpretation of pH data acquired over 24 h, Johnson and DeMeester identified six important parameters which are listed in Tables 2 and 3. The normal values (means and standard deviations) were determined by the same group, by initially studying 15 healthy adult volunteers, and later on, by including data from 35 more healthy young adults. The 95th percentile values of this group were accepted as the upper normal limit. Since their pioneer study, these results have been used as standard.

On the basis of their extensive experience, DeMeester and colleagues [8] developed a scoring system for the diagnosis of GORD. In this system, two standard deviations below the mean of normal subjects was defined as the zero point, and an increment of each standard deviation above this was given one extra point. The formula used for calculating the component score of each parameter is given as follows in Eq. (3):

$$((\text{Patient's value} - \text{mean values of normal})/\text{SD of normal}) + 1 \quad (3)$$

Table 4

Sensitivity and specificity of tests for the diagnosis of oesophageal reflux^a

Diagnostic test	Sensitivity (%)	Specificity (%)
LOS manometry	58	84
Endoscopy	68	96
Mucosal biopsy	77	91
Scintiscanning	61	95
Barium	40	85
Bernstein test	79	82
Prolonged pH monitoring	88	98

^a Reproduced from [9].

A sum of component scores of the six parameters is called DeMeester's score. The upper normal limit of this score is 14, and any score above 14 represents GORD. Fig. 1 shows the steps used to derive the above-mentioned formula (Eq. (3)) as described in one of the articles by DeMeester [9]. This formula is, in fact, a modification of a statistical formula used for the standardization of individual values taken from a normally distributed population. The actual statistical expression is as follows (Eq. (4)):

$$Z = (\text{individual value} - \text{mean})/\text{SD} \quad (4)$$

This Z-score [10] tells how many standard deviations a particular value will be away from the mean value, and is an extremely useful method for comparison of the individual values, as well finding where a particular value will be roughly located in a normally distributed population. However, one of the basic requirements for using the Z statistic is that the reference population should have a normal distribution, which means 50% of the data should be below the mean and 50% above. In other words, it should be neither skewed nor kurtotic in any particular direction [10].

Since, in the DeMeester scoring system, a Z value of negative 2 is fixed as the zero point, it therefore required the addition of the factor '2' to the standard score (Fig. 1). However, as mentioned in the figure, they ignored the justification of adding '2', and simplified the formula by adding only '1' instead of '2' (Fig. 1; last equation). In their view, '2' was a constant value for all parameters, and could therefore be omitted altogether. However, they used '1' as a constant, with the reasoning that some of the values had a possibility of being negative, and hence, the addition of '1' helped in converting them to a positive value [9]. In actual fact, '2' had its place in the formula to set two standard deviations below the mean as the zero point. The replacement of '2' with '1' sets this zero point at one standard deviation below the mean value, and hence, is against the principle on which the whole scoring system is based. For that very reason, the whole scoring system is based on erroneous formula.

$$\begin{aligned}
 \text{Component Score} &= (\text{Patient Value} / \text{SD}) + (\text{SD}/\text{SD}) - [(\text{Mean}-\text{SD}) / \text{SD}] \\
 &= (\text{Patient Value} - \text{Mean} + 2\text{SD}) / \text{SD} \\
 &= [(\text{Patient Value} - \text{Mean}) / \text{SD}] + 2
 \end{aligned}$$

The constant 2 is replaced by 1 to get the final formula to calculate component score as following :

$$\text{Component Score} = [(\text{Patient Value} - \text{Mean}) / \text{SD}] + 1$$

Note : SD= Standard Deviation

† Reproduced from :

DeMeester TR
 Prolonged Oesophageal pH Monitoring
 In : NW Read (Ed.) *Gastrointestinal motility: Which test?*
 Wrightson Biomedical Publishing Ltd 1989; p 43-52

Fig. 1. Formula for the calculation of DeMeester's score.

We further question the justification of using the value of the 95th percentile from a small group of 50 healthy young American adults as the upper normal value and two standard deviations below the mean as the zero point. We know the group of 50 people studied by DeMeester et al. were all free of disease, and hence, using the 95th percentile directly regards 5% of them as abnormal. This narrows down the limits of normality to 95% of 50 normal young American adults only. We therefore believe that, for a better scoring system, either a much bigger population with a wider range of ages should be used to determine normal values or different values should be used for paediatric, adult and older age groups after studying the correlation of scores with the actual presence of disease in each group.

The word 'score' very specifically denotes a 'mark' or 'grade' of severity of a problem. Mean and standard deviations have no ability to quantify a pathology as they represent central tendency and dispersion in a normally distributed population, respectively. The most appropriate

approach to develop a scoring system should therefore take into account the role of individual parameters for the development of a particular pathology. It appears that initially, DeMeester and colleagues also attempted to develop a scoring system on this basic principle. In their earlier studies [11], they tried to develop a scoring system by assigning points to individual parameters according to their contribution in producing pathological effects, e.g. if supine reflux is more important than upright reflux, then the presence of supine reflux should get more points than the presence of upright reflux. However, DeMeester's group abandoned this approach because they thought that the points assigned to any abnormal parameter were determined arbitrarily and were not based on any objective reason derived from the available data [8]. More recently, the much easier availability of modern computers and advanced statistical software has made it possible to study the correlation between the presence of disease and the degree of severity of contributory factors. This kind of multivariate analysis can provide

very objective weighing factors for the individual variables. For that very reason, in the last 10 years, investigators have used the techniques of multivariate analysis for evaluation of the role of individual parameters in producing pathological effects of GOR [12,13]. We believe more detailed and larger studies with such analyses can provide better information to design a scoring system. Until then, we recommend not to entirely rely on DeMeester's score. It is much more appropriate to evaluate the whole pattern of pH changes in 24 h.

5. Conclusions

The null hypothesis defined at the beginning is rejected as the mean values of parameters recorded in our study are different from those used as normal. This implies that conventional normal values defined by DeMeester et al. can over-diagnose the majority of normal elderly subjects as suffering from GORD. If on clinico-pathological grounds, the different reflux pattern, noted in this study, can be accepted as normal for relatively older people, then the normal values for the elderly population deserve modification. Moreover, the formula used for the calculation of DeMeester's score does not correspond to the principle it is based on. At the same time, the basic principle is, itself, questionable, and we therefore recommend that the scoring system should not be used for quantification of reflux.

References

- [1] Johnson LF, DeMeester TR. Twenty-four hour pH monitoring of distal esophagus: a quantitative measure of gastro-esophageal reflux. *Am J Gastroenterol* 1974;62:325–332.
- [2] El-Serag HB, Sonnenberg A. Association between different forms of gastro-oesophageal reflux disease. *Gut* 1997;41(5):594–599.
- [3] Spence RA, Collins BJ, Parks TG, Love AH. Does age influence normal gastro-oesophageal reflux? *Gut* 1985;26(8):799–801.
- [4] Fass R, Sampliner RE, Mackel C, McGee D, Rappaport W. Age and gender related differences in 24 hour esophageal pH monitoring of normal subjects. *Dig Dis Sci* 1993;38(10):1926–1928.
- [5] Sonnenberg A, Steinkamp U, Weise A, Berges W, Wienback M, Rohner HG, Peter P. Salivary secretion in reflux esophagitis. *Gastroenterology* 1982;83:889–895.
- [6] Smout AJPM, Breedijk M, Van der Zouw C, Akkermans LMA. Physiologic gastro-esophageal reflux and esophageal motor activity studied with a new system for 24-hour reading and automated analysis. *Dig Dis Sci* 1989;34:372–378.
- [7] Skinner DB, Booth DJ. Assessment of distal oesophageal function in patients with hiatus hernia and/or gastroesophageal reflux. *Ann Surg* 1970;172:627.
- [8] Johnson LF, DeMeester TR. Development of the 24 hour intraoesophageal pH monitoring composite scoring system. *J Clin Gastroenterol* 1986;8(Suppl 1):52–58.
- [9] DeMeester TR. Prolonged oesophageal pH monitoring. In: Read NW, editor. *Gastrointestinal motility: which test?* Petersfield, UK: Wrightson Biomedical Publishing Ltd, 1989. pp. 43–52.
- [10] Zar JH. *Biostatistical analysis*. Englewood Cliffs, NJ: Prentice Hall, 1996.
- [11] DeMeester TR, Wang C-I, Wernly JA, Pellegrini CA, Little AG, Klementsitsch P, Bermudez G, Johnson LF, Skinner DB. Technique, indications, and clinical use of 24 hour esophageal pH monitoring. *J Thorac Cardiovasc Surg* 1980;79:656–670.
- [12] Cadiot G, Bruhat A, Rigaud D, Coste T, Vuagnat A, Benyedder Y, Vallot T, Le Guludec D, Mignon M. Multivariate analysis of pathophysiological factors in reflux oesophagitis. *Gut* 1997;40:167–174.
- [13] Ghillebert G, Demeyere AM, Janssens J, Vantrappen G. How well can quantitative 24-hour intraoesophageal pH monitoring distinguish various degrees of reflux disease? *Dig Dis Sci* 1995;40:1317–1324.

Appendix A. Conference discussion

Dr M. Migliore (Catania, Italy): You know that in 4 h, you can have several number of reflux, and you are saying that your pH value is being performed between 20 and 24 h. So, I believe that the mean values that you have presented are higher than those you published because the DeMeester score is on 24 h pH monitoring and your pH monitoring performance is between 20 and 24 h. So I believe that the number is not what we can find pH monitoring engagement in 24 h.

The second question is this. In elderly patients, you have absence or decrease of motility in the oesophagus. Some of your patients can have oesophagitis without symptoms. I would like to know how many of your patients have endoscopy? And if so, what are the findings?

Mr Jalal: The answer to the first question is that, as mentioned earlier, we had very strict exclusion criteria on the basis of which 55 out of 100 subjects were dropped from the study. In fact, the majority of patients completed 24 h, and all were studied for more than 20 h. Only a few had less than 24 h due to various reasons not affecting the pattern of reflux. After carefully examining the pH recordings of individual subjects, we excluded all those whose result would have changed significantly in the remaining few hours. Moreover, even if we presume that there was a theoretical chance of more reflux episodes in these few subjects, this would have slightly increased the values of individual parameters and further consolidated our conclusion that in old age, the normal values of the individual parameters are higher than the younger age adults.

Dr Turina: The second question was have you verified the findings that the people are really not ill? Did you perform oesophagoscopy?

Mr Jalal: As I defined earlier, the oesophageal reflux disease should have symptoms and/or histological findings. By this definition, all of our 61 subjects were normal old age people. These subjects attended our thoracic surgery unit for diagnostic or therapeutic procedures of their lung tumours. None of these had symptoms or histories suggestive of reflux, and nobody was taking any medicines which could provoke GOR. In clinical practice, subjects without symptoms are not subjected to endoscopic evaluation. That is the reason we did not perform endoscopic evaluation at this initial stage. However, later on, these subjects had oesophagoscopy evaluation during their operations for lung cancer by our senior author, Mr Jeyasingham, and were found to be free of any endoscopic findings diagnostic of reflux disease.

In earlier studies on normal young adults, endoscopic evaluation was not used and only pH monitoring was done to find out normal values of the six parameters. However, later on in one of the studies by DeMeester and colleagues reported in 1980 in the *Journal of Thoracic and Cardiovascular Surgery* [11], endoscopy was also done. In this particular study, they reported that 54% of endoscopically normal subjects had abnormal pH values. From this finding, they inferred that endoscopy is not a very sensitive method of detecting oesophageal reflux. On the contrary, we believe that the normal values reported by Johnson and DeMeester in 1974 in the *American Journal of Gastroenterology* [7] tend to over-diagnose the oesophageal reflux disease because they were based on an entirely inappropriate principle. This is the whole gist of our study and endoscopy did not change our point of view.

Mr A. Mearns (Bradford, UK): I've been doing pH monitoring now for 18 years, and I have a lot of problems with the DeMeester score. And I'm particularly interested in the age group in your study here, it fascinates me. The major problem is that any reflux at all, during the night, when a patient

is recumbent, is abnormal. And the DeMeester score does not cope with this.

But the most important thing is that the majority of people who are elderly get up at night to go to the toilet. And they have 'getting up and going to the toilet' reflux. They have the nuisance of getting up out of bed. So, all the elderly patients have a dip in their pH at night associated with getting out of bed to urinate. The DeMeester scoring system misses this completely. The most important thing that you do, after the event, is to discuss their nightly behaviour with them, because they do not record this in their diary. This is one of the major causes of the failure of the DeMeester system to take into account the higher scores in the elderly. Those young adults that DeMeester used were fit young men who hadn't had too much beer and did not get up for a pee at night.

Appendix B. Editorial comment

This paper summarized the results of manometry and pH recordings in a group of elderly asymptomatic patients and reassesses the value of the DeMeester scoring system. The paper is of considerable interest as it highlights the complexities and difficulties in the interpretation of data, as well as the appropriate application of statistical methods of analysis.

It is of note that 55 asymptomatic patients were excluded from the study due to abnormal studies. This highlights the fact that occult reflux and motility disorders are not uncommon in the elderly. Comparing historic data may be reasonable, but interlaboratory variation is not uncommon, and a normal control group from the Bristol Laboratory would have

strengthened the results in this paper. Many laboratories have adopted the DeMeester scoring system, and in most centres, there is good correlation with positive scores and pathological changes.

In the original publication, the DeMeester scoring system was purposely simplified to facilitate interpretation, and analysis presumed a normal Gaussian distribution. The results in this paper are analyzed by the Student's *t*-test. This, however, also assumes normal distribution. Non-parametric methods of analysis should be adopted and this can be done by discriminant analysis or with receiver operating characteristics analysis [8]. There is controversial evidence of the effects of age and GOR, but most authors agree that there is an increased incidence of GOR with advancing years. There may also be a sex-difference in the incidence of GOR, and a predominantly male population may skew this paper.

This paper is timely in re-appraising the DeMeester scoring system. I do not feel there is enough evidence to reject the 'null hypothesis'. The 'null hypothesis' is the statistical equivalent of 'you are innocent until proven guilty'. In this case, the jury finds insufficient evidence that the DeMeester scoring system is inappropriate, and besides, it has been a solid foundation in the development and understanding of GORD and the standardization of oesophageal reflux.

J.A.C. Thorpe
Leeds General Infirmary,
Leeds,
UK