

Post-vagotomy insulin test: improved predictability of ulcer recurrence after corrections for height and collection errors

N. K. MAYBURY, R. G. FABER, AND M. HOBSLEY

From the Department of Surgical Studies, The Middlesex Hospital and Medical School, London

SUMMARY Insulin stimulated gastric secretion was studied in 74 unoperated duodenal ulcer patients (DUs), (20 women and 54 men). Three indices of secretion were studied—observed volume, acid output, and volume of gastric juice corrected for pyloric loss and duodenal reflux (V_G). These three measurements were expressed both as peak secretion and as secretion during the $\frac{1}{2}$ to two hour period after insulin, and also both before and after standardisation for height, making 12 different indices in all. From the data a significant correlation between insulin-stimulated secretion and height in DUs was found. A method of standardising each patient's secretion for height is described. We confirm a significantly higher insulin-stimulated secretion in men than in women and show that this difference can be explained by their difference in height. For each of the 12 indices of secretion, the range of secretion for the unoperated subjects was obtained. The same indices were measured in 155 postvagotomy patients, including 33 patients with recurrent DUs, and compared with the ranges of secretion established in the unoperated patients. Responses above the lower 95% tolerance limit of the preoperative range were designated positive and those below negative. The Hollander status was determined. It was found that the least satisfactory criterion was Hollander's (7% false negative and 69% false positive). The best was $\frac{1}{2}$ -2 V_G standardised for height (3% false negatives and 43% false positives). The improvement in predictability was significant at the 0.0005 level.

The postvagotomy insulin test was introduced by Hollander (1946) for detecting the presence of any intact vagal fibres after operation. Two years later he defined criteria for interpreting the test (Hollander, 1948), but even then was extremely cautious about placing any reliance on the test for predicting the individual patient's chances of developing a recurrent ulcer. In 1950 he clearly stated: 'The insulin test cannot be used to prognosticate clinical results of vagotomy' (Weinstein *et al.*, 1950). Despite this, the original Hollander criteria are still widely applied to the postvagotomy insulin test and conclusions concerning the ability of a particular surgeon or the clinical course of the patient are unjustifiably drawn from the results of these criteria.

In 1975 a 'New interpretation of the insulin test' in men was described from this department (Faber *et al.*, 1975). These new criteria were based on the comparison of the insulin-stimulated secretion of an

individual after operation with the range of insulin-stimulated secretion of a group of patients with duodenal ulcer tested before operation: they completely ignored basal secretion. Good discrimination was obtained by dividing the postvagotomy subjects into two groups: those with secretion within the preoperative range had a 50% liability to recurrent ulcer; those with secretion less than the lower 95% tolerance limit of the preoperative range had a zero liability to recurrent ulcer.

These new criteria applied only to men; they gave unsatisfactory results for women. The possibility that this difference was related to the smaller stature of women seemed attractive for several reasons. Work from this department has shown that the difference in maximally-stimulated gastric secretion between normal men and women can be explained in terms of stature (Hassan and Hobsley, 1971). Also maximal gastric secretion is positively correlated with height in both normal controls (Baron, 1969; Hassan and Hobsley, 1971) and in patients

with duodenal ulcer before operation (Hassan and Hobsley, 1971; Hobsley *et al.*, 1975). Moreover, it has been shown (Spencer *et al.*, 1969) that, at least in terms of acid output, insulin-stimulated secretion is greater in male than female DUs.

The objectives of the present study have therefore been: (1) To determine whether there is a significant correlation between insulin-stimulated secretion and height in DUs; (2) if such a correlation is confirmed, to establish a method of standardisation for height so that the secretions of individuals of different heights become comparable; (3) to confirm the difference in insulin-stimulated secretion between men and women and to test whether the difference can be explained in terms of height; (4) to establish ranges for various indices of secretion in DU patients before operation; and finally (5) to study insulin-stimulated secretion in patients after vagotomy for duodenal ulcer in the light of this information.

Methods

PATIENTS AND EXPERIMENTAL DESIGN

Studies were performed on (1) 74 unoperated duodenal ulcer patients (54 men and 20 women) and (2) 155 patients after vagotomy (124 men and 31 women), of whom 60 had undergone truncal vagotomy and drainage, 40 selective vagotomy and drainage, and 55 proximal gastric vagotomy. The follow-up was between four months and six years.

Each patient fasted and abstained from medication known to affect gastric secretion from midnight. At 9.00 a.m. the patient lay semirecumbent on a couch. A special two-lumen nasogastric tube was passed (Thomson *et al.*, 1973), and its position checked by the water-recovery test of Hassan and Hobsley (1970). In all subjects phenol red was instilled into the stomach by the small lumen. After a basal collection of at least half an hour, soluble insulin 0.2 units/kg was injected intravenously and gastric juice was collected by continuous aspiration in 10-minute samples for a further two hours. Blood sugar was measured, and only tests where hypoglycaemia of less than 2.2 mmol/l was achieved were included.

TECHNIQUES AND CALCULATIONS

The volume of each sample was measured. Then after filtration through Whatman No. 1 filter paper the following measurements were made: titratable acidity (Radiometer automatic titrator, titration to pH7), sodium, potassium, and chloride concentrations (EEL flame photometer and chloride meter), and phenol red concentration (Unicam SP spectrophotometer) as described by Faber *et al.* (1975).

The observed volume of each 10-minute sample

(V_{obs} ml/10 min) was corrected, firstly, for pyloric losses by means of the marker phenol red and, secondly, for duodenal reflux by means of a formula based on the output of sodium ions (Faber *et al.*, 1974; Hobsley, 1974). The volume of secretion after these corrections was designated V_G ml/10 min.

For each subject, whether tested before or after operation, the insulin-stimulated secretion was expressed in terms of V_{obs} , V_G and also acid output (AO), both as peak and half-to-two hour ($\frac{1}{2}$ -2) secretions. The peak secretion was defined as the sum of the three highest consecutive 10-minute samples multiplied by 2 and expressed as ml/hour or mmol/hour; $\frac{1}{2}$ -2 secretion was defined as the sum of the nine 10-minute samples during the period from half to two hours after the injection of insulin and divided by 1.5, thus expressing the result as ml/hour or mmol/hour.

Hollander status was determined for 149 of the 155 post-vagotomy studies; in the remaining six studies basal secretion was discarded. The detailed assessment of Hollander status was in general according to the exact criteria laid down by Hollander (1948), but with the exception that patients falling into Hollander's 'intermediate' group were classified by us, in common with most other workers, as negative.

Gastric secretion, whether expressed as volume or acid output, is not distributed on a normal scale. However, after logarithmic transformation the distribution is not significantly different from normality (Faber *et al.*, 1975; Faber and Hobsley, 1976). For this reason, all gastric secretion data were transformed to the natural logarithm before determining the mean and standard deviation and before performing correlations against height.

It is possible to express a logarithmic mean as its arithmetical equivalent by simply taking its anti-logarithm and this is the figure quoted for the means in Table 2. The same procedure cannot be used with a logarithmic standard deviation: two logarithmic standard deviations below the logarithmic mean is not numerically the same interval as two logarithmic standard deviations above the logarithmic mean, because the addition and subtraction of logs corresponds to multiplication and division in a numerical scale. There is therefore no meaning in the anti-logarithm of the logarithmic standard deviation, and so no value for standard deviation is quoted for a logarithmic mean in these results. However, the lower 95% tolerance limits of the mean *can* be expressed numerically by taking the antilog of the expression (log mean—2 log standard deviations).

Differences between means for unpaired data were tested by Student's *t* test (after logarithmic trans-

formation). The success of the various criteria in predicting recurrence was tested by the χ^2 distribution. All statistical tests were applied as described by Cooper (1969).

Results¹

1. RELATIONSHIP OF SECRETION WITH HEIGHT

For the 74 unoperated patients, each of the three indices of gastric secretion, V_{obs} , AO and V_G , during both peak and $\frac{1}{2}$ -2 hour periods, were correlated against the height of the patients. By all indices of secretion, both peak and $\frac{1}{2}$ -2 hour, there was a highly significant ($P < 0.001$) positive correlation (Table 1). An example of this correlation with $\frac{1}{2}$ -2 V_G ($r = 0.509$) is shown in Fig. 1.

2. STANDARDISATION FOR HEIGHT

The regression equations established under (1) above

¹Full data are available on request to the authors.

Table 1 Correlation coefficients (r) comparing V_{obs} , AO and V_G with height

Index of secretion	Correlation coefficient (r)	
	Peak	$\frac{1}{2}$ -2 h
V_{obs}	0.415	0.431
AO	0.503	0.503
V_G	0.529	0.509

In all cases $P < 0.001$.

were then used to construct formulae for standardising the secretion of any individual to the average height of the whole group. Details of these formulae are given in the Appendix.

3. DIFFERENCES BETWEEN MEN AND WOMEN

The secretion data in the unoperated patients yielded 12 different indices of gastric secretion: V_{obs} , AO, and V_G , each expressed both as peak and $\frac{1}{2}$ -2 hour values, and these six in turn before and after standardisation for height. The means were calculated after logarithmic transformation for all indices for the 20 women and 54 men separately (Table 2), and the difference between their means compared by Student's t test for unpaired data. For all indices not standardised for height the difference in secretion between the women and men was highly significant ($P < 0.001$). After standardisation for height there was still a significant difference between the men and women in terms of V_{obs} and AO (P values in Table 2). However, in terms of V_G , both peak and $\frac{1}{2}$ -2, no significant difference could be demonstrated between men and women ($P > 0.1$).

4. RANGES OF SECRETION BEFORE OPERATION

The ranges of secretion of all 74 unoperated patients (20 women, 54 men) were examined in terms of all 12 indices described above. The means and lower 95% tolerance limits are given in Table 3. These data provide a background against which the secretion of patients after vagotomy can be assessed. Figure 2 illustrates the range of $\frac{1}{2}$ -2 V_G in unoperated

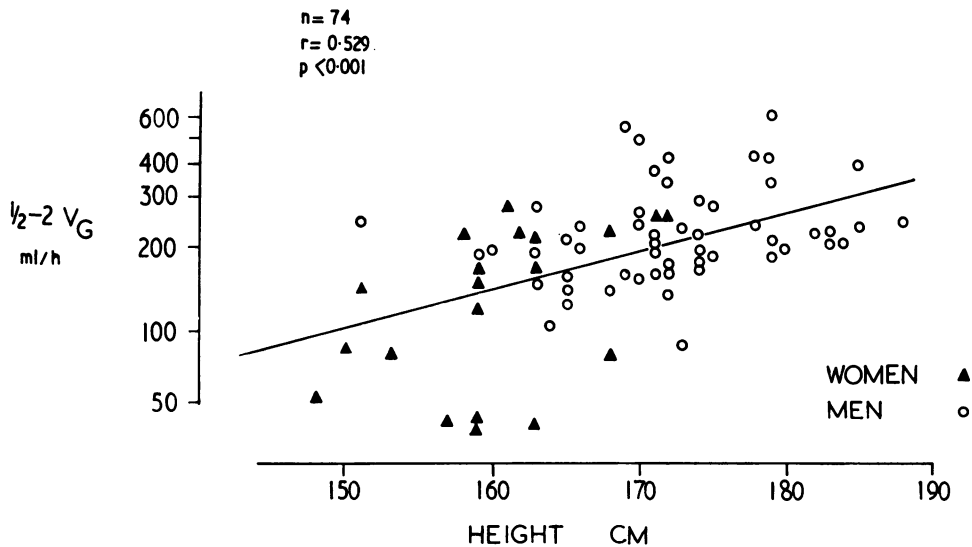


Fig. 1 The correlation between $\frac{1}{2}$ -2 V_G , plotted on a logarithmic scale, height measured in centimetres for 74 unoperated DUs is shown. That most women are shorter and secrete less gastric juice than men is clearly shown.

Table 2 Comparison of gastric secretion between women (n = 24) and men (n = 54) for each of 12 indices of secretion

Index of secretion	$\frac{1}{2}$ -2				Peak			
	Women Mean	Men Mean	t	P	Women Mean	Men Mean	t	P
Not standardised for height								
V _{obs} ml/h	105	199	4.82	<0.001	131	243	4.96	<0.001
AO mmol/h	7.06	19.86	5.22	<0.001	9.12	26.36	5.62	<0.001
V _G ml/h	123	222	4.59	<0.001	151	280	5.12	<0.001
Standardised for height								
V _{obs} ml/h	135	182	2.29	<0.025	170	222	2.12	<0.025
AO mmol/h	10.9	16.9	2.23	<0.025	13.87	20.2	1.81	<0.05
V _G ml/h	176	201	1.59	>0.1	207	247	1.53	>0.1

The means have all been obtained by a logarithmic transformation. SD are not shown (see techniques and calculations).

Table 3 Range of secretion of all unoperated DUs, both men and women combined for each of 12 indices

Index of secretion	$\frac{1}{2}$ -2		Peak	
	Mean	Lower 95% tol. limit	Mean	Lower 95% tol. limit
Not standardised for height				
V _{obs} ml/h	168	53	197	60
AO mmol/h	15.02	2.55	19.76	3.58
V _G ml/h	189	63	236	82
Standardised for height				
V _{obs} ml/h	168	59	207	78
AO mmol/h	15.11	3.22	18.46	5.07
V _G ml/h	191	73	236	101

DUs both before and after standardising for height.

5. POSTVAGOTOMY SECRETION

The subjects after vagotomy were assessed from two separate viewpoints: (a) their clinical course, with reference to recurrence of peptic ulceration; and (b) their insulin-stimulated secretion, in comparison with the preoperative ranges (Table 3).

a. Of the 155 patients, 33 had a recurrent ulcer (RU) proven by endoscopy and/or reoperation, 10 patients had symptoms suggestive of recurrent ulceration but without firm evidence, and the remaining 112 patients had no symptoms suggestive of recurrence. The distribution of recurrent ulcers between the men and women was similar. Of 31 women, six (19%) had RU, of 124 men, 27 (22%) had RU $P > 0.9$ by the χ^2 distribution.

b. The insulin-stimulated secretion of each of the 155 postvagotomy subjects was compared with ranges of secretion established for the unoperated DUs for each of the 12 indices of secretion defined above. Figure 2 shows this comparison for $\frac{1}{2}$ -2 V_G. Those whose secretion by any one index fell above the lower 95% tolerance limit (Table 3) were called positive by that index, and those below negative. Each of these two groups (positives and negatives) were divided into those who had a proven RU and those who had not. In the positive group those who did not have a RU were designated 'false positives'

and their percentage of the whole positive group calculated (% false positives). In the negative group those patients who had a RU were designated 'false negatives' and their percentage of the whole negative group calculated (% false negatives; Table 4).

The Hollander status of 149 of the postvagotomy cases was available. Of the six remaining patients, in whom the basal secretion had been discarded, none had a RU. The 149 patients were divided into Hollander positive and Hollander negative. As with the other criteria, the number of false positive patients was expressed as a percentage of the total Hollander-positive group, and the number of false negative patients was expressed as a percentage of the Hollander-negative group.

By any index of secretion, the numbers of false negatives were small: the percentage of false negatives for the 12 new indices of secretion varied from 1 to 4%; for the Hollander test it was 7% (Table 4). These numbers are too small to allow a statistical comparison to be made.

By all indices of secretion the incidence of false positives was considerably higher than the incidence of false negatives. The false positives for the 12 new indices varied from 43 to 68%, while for the Hollander test the figure was 69% (Table 4). Figure 3 illustrates the graduation of the incidence of false positives over the whole range of 13 indices. The best prediction of RU was given by height standardised $\frac{1}{2}$ -2 V_G; the worst by Hollander's criteria. Height-standardised $\frac{1}{2}$ -2 V_G is the index that makes allowance for height, pyloric losses, duodenal reflux, and, in contrast with peak values, removes the difficulty of defining a peak. The Hollander criteria make no allowance for any of these factors, and also rely on basal secretion. Examination of Fig. 3 shows that there is improvement in predictive accuracy as, firstly, basal secretion is discarded; secondly, corrections are made for pyloric losses and duodenal reflux; and, thirdly, the data are standardised for height.

The differences between the Hollander criteria and

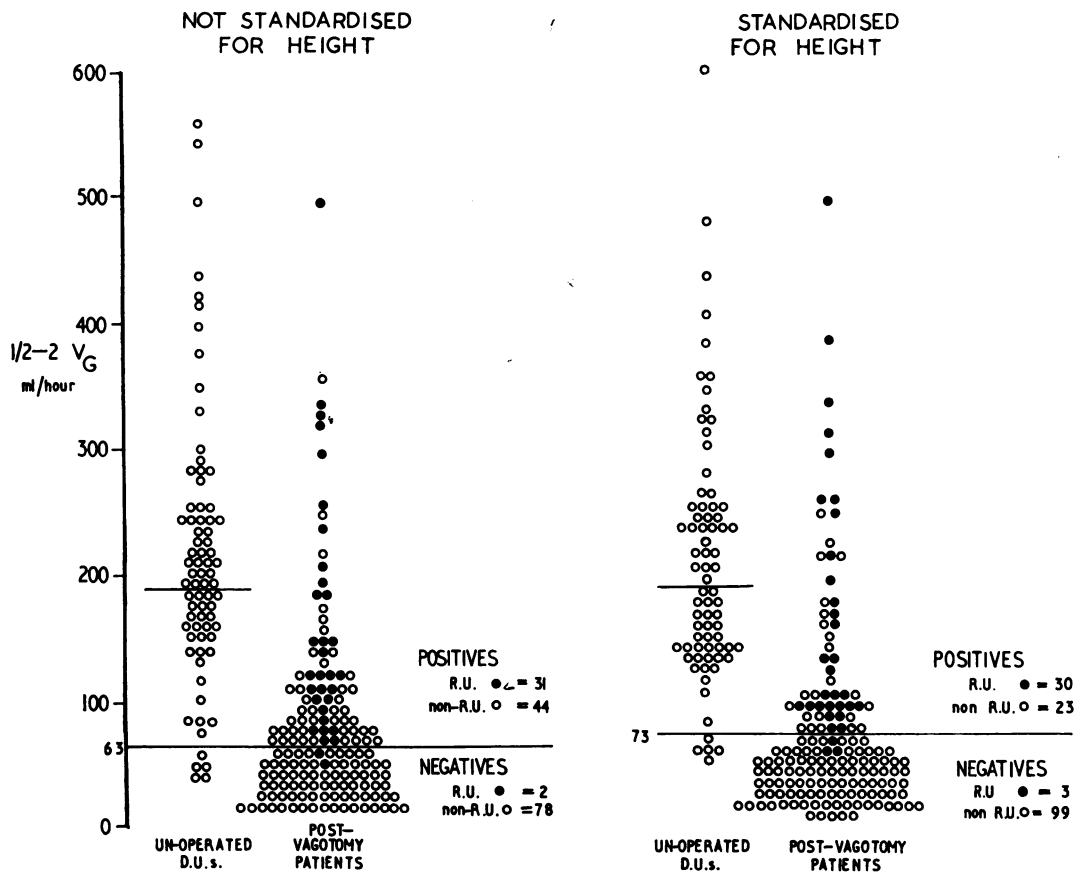


Fig. 2 The secretory response, $\frac{1}{2}$ -2 V_G ml/hour, to insulin 0.2 units/kg IV in 74 unoperated DUs is shown both before and after standardisation for height. Against this range is compared the post-vagotomy response to insulin of 155 patients again before and after standardisation for height. Those whose post-vagotomy response to insulin is greater than the lower 95% tolerance limit of unoperated subjects are deemed positive, and those below negative. Means and lower 95% tolerance limits have been calculated after logarithmic transformation.

Table 4 Distribution of patients by reference to their clinical status and index of secretion

	H	Not standardised for height						Standardised for height					
		$\frac{1}{2}$ -2 h			Peak			$\frac{1}{2}$ -2 h			Peak		
		V_{obs}	AO	V_G	V_{obs}	AO	V_G	V_{obs}	AO	V_G	V_{obs}	AO	V_G
Insulin positive results													
Recurrent ulcers	29	31	31	31	32	31	31	31	29	30	32	29	32
Non-recurrent ulcers (false positives)	64	51	41	44	67	38	40	36	30	23	42	31	35
% False positive	69	62	68	59	68	55	56	54	51	43	57	48	52
Insulin negative results													
Recurrent ulcers (false negatives)	4	2	2	2	1	2	2	2	4	3	1	4	1
Non-recurrent ulcers	52	71	81	78	55	84	82	86	92	99	80	91	87
% False negative	7	3	2	3	2	2	2	2	4	3	1	4	1

H: Hollander.

the other indices, as assessed by the χ^2 test, are shown in Fig. 3. Firstly, the upper row of figures, comparing the positive results only, show that at the

extremes of the range, the difference is substantial; for height standardised $\frac{1}{2}$ -2 V_G , $P < 0.005$. Secondly, comparison was made, by means of a 2 x 3 con-

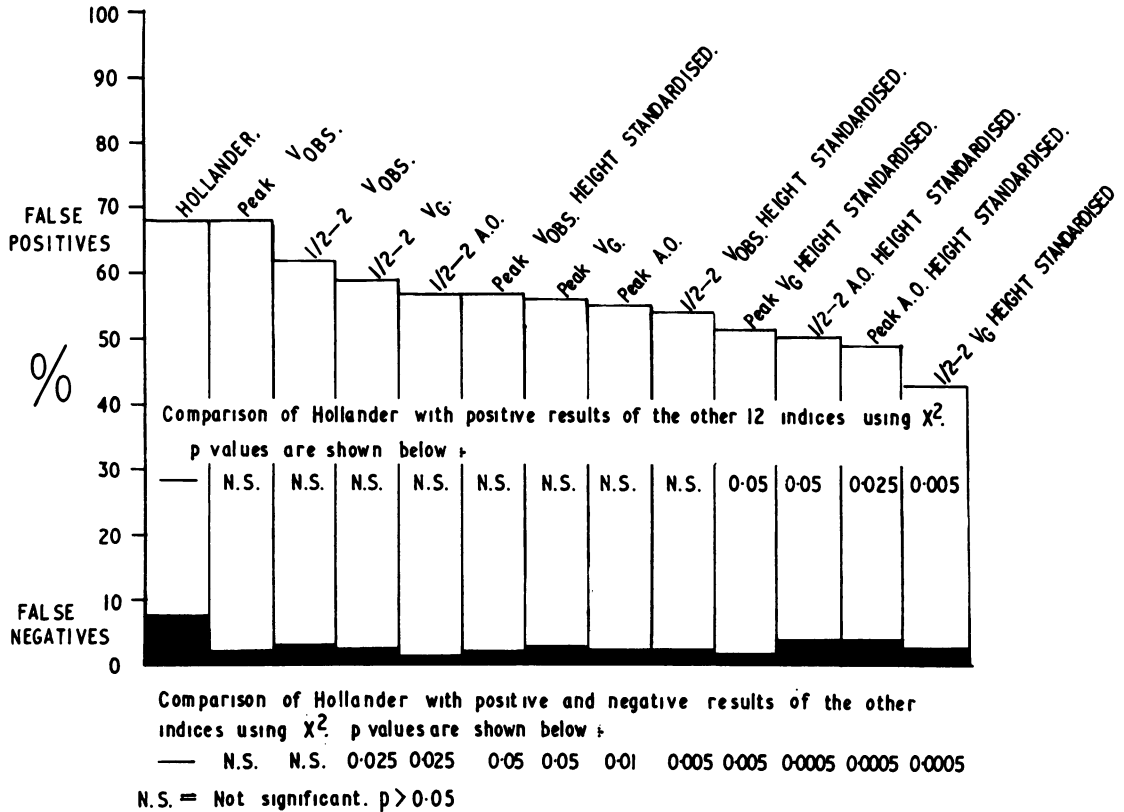


Fig. 3 The comparison of the distribution of the positive tests in terms of the Hollander and the new criteria is shown in the upper row. In the lower row the comparison of the distribution of both positive and negative tests in terms of the Hollander and the new criteria are shown.

tendency table, between the numbers of true and false positives and the sum of the true and false negatives in terms of the Hollander criteria and the new indices. The number of false negatives and true negatives have been combined as the expected number of false negatives is < 5 in all columns and so a 2×4 contingency table is inappropriate. All the new criteria, except V_{obs} not standardised for height, discriminated significantly more accurately than the Hollander criteria, and the probability values are shown in the lower row of Fig. 3.

Discussion

The results of the study have shown that, like maximally stimulated secretion, insulin-stimulated secretion does positively correlate with the height of the patient with a duodenal ulcer. A method for standardising secretion for height is described in the appendix. Furthermore we have confirmed that insulin-stimulated secretion is, indeed, less in women

than men tested before operation. However, when corrections are made for pyloric losses and duodenal reflux and the data are standardised for height, the difference between the sexes disappears.

We have explored the Hollander concept that insulin-stimulated secretion after vagotomy can be used to predict liability to recurrent ulceration. This concept has to be applied in terms of some index of insulin-stimulated secretion, and compared with some standard reference secretion. Hollander's use of basal secretion as a reference suffers from the disadvantage of variability in any individual from one test to another (Baron and Alexander Williams, 1973; Faber and Hobsley, 1976). Also his choice of titratable acidity as an index of secretion is not good because it is particularly prone to errors because of alkaline reflux from the duodenum. We have replaced basal secretion by preoperative insulin-stimulated secretion. The philosophy underlying this choice as a standard is that, if a patient can be shown to have an insulin-stimulated secretion that is

demonstrably and unequivocally depressed, it is likely that the vagotomy has been effective in reducing gastric secretion.

Once this is accepted, the index of secretion chosen has been shown to give increasingly reliable prediction as refinements of measurement were introduced. For example, while a trend was noticeable even in V_{obs} (Fig. 3), a statistically significant improvement required a standardisation for height to be introduced. Height-standardised V_{obs} was significantly better than Hollander, $P < 0.005$. Correction of gastric secretions for pyloric losses and duodenal reflux also gave an improved prediction, and the combination of height-standardisation with the corrections for collection errors gave the best results of all (height-standardised $\frac{1}{2}-2 V_G P < 0.0005$).

We have used both the peak and the $\frac{1}{2}-2$ hour period after an intravenous injection of insulin. We felt that $\frac{1}{2}-2$ hour period might reduce the errors inherent in choosing the peak when the secretion rate is changing rapidly. There is little to choose between the results by these two methods. On the data we have presented the $\frac{1}{2}-2 V_G$ appears to be the most accurate criterion.

Whatever criterion is used, the negative result is more reliable than the positive, for with all indices of secretion the percentage of false negatives is very small (Table 4) and therefore the chances of a patient developing a RU are small.

Of the positive results, Hollander's is so bad that two-thirds of the patients with positive Hollander tests had no recurrence at the time of testing. Even our most accurate criterion ($\frac{1}{2}-2 V_G$) still gives us almost 40% false positives. This figure will probably diminish as time passes: some of the patients have been followed up for only a relatively short period, some already have symptoms, while we have not yet established a diagnosis of RU; it is probable that in at least some of these patients such a diagnosis will ultimately be established. Finally, we draw attention to the fact that one major error affecting gastric secretion remains unchallenged in our studies—salivary contamination. The solution of the problem of correcting for salivary contamination may further improve the accuracy of the secretion studies.

One of the authors, N. K. Maybury, is a Wellcome Research Fellow and gratefully acknowledges the support of the Wellcome Trust. We are grateful to Mr M. Dave for technical assistance.

References

Baron, J. H. (1969). Lean body mass, gastric acid, and peptic ulcer. *Gut*, **10**, 637-642.
 Baron, J. H., and Williams, J. A. (1973). Gastric secretion tests. In *Recent Advances in Surgery*, no. 8, p. 166. Edited

by Selwyn Taylor. Churchill Livingstone: Edinburgh.
 Cooper, B. E. (1969). *Statistics for Experimentalists*. Pergamon Press: London.
 Faber, R. G., and Hobsley, M. (1976). Basal gastric secretion: Reproducibility and relationship with duodenal ulceration. *Gut*, **18**, 57-63.
 Faber, R. G., Russell, R. C. G., Parkin, J. V., Whitfield, P., and Hobsley, M. (1975). The predictive accuracy of the post-vagotomy insulin test: a new interpretation. *Gut*, **16**, 337-342.
 Faber, R. G., Russell, R. C. G., Royston, C. M. S., Whitfield, P., and Hobsley, M. (1974). Duodenal reflux during insulin stimulated secretion. *Gut*, **15**, 880-884.
 Hassan, M. A., and Hobsley, M. (1970). Positioning of subject and of nasogastric tube during a gastric secretion study. *British Medical Journal*, **1**, 458-460.
 Hassan, M. A., and Hobsley, M. (1971). The accurate assessment of maximal gastric secretion in control subjects and patients with duodenal ulcer. *British Journal of Surgery*, **58**, 171-179.
 Hobsley, M. (1974). Pyloric reflux: a modification of the two component hypothesis of gastric secretion. *Clinical Science and Molecular Medicine*, **47**, 131-141.
 Hobsley, M., Whitfield, P. F., Faber, R. G., and Parkin, J. V. (1975). Hypersecretion and length of history in duodenal ulceration. *Lancet*, **2**, 101-104.
 Hollander, F. (1946). The insulin test for the presence of intact nerve fibers after vagal operations for peptic ulcer. *Gastroenterology*, **7**, 607-614.
 Hollander, F. (1948). Laboratory procedures in the study of vagotomy (with particular reference to the insulin test). *Gastroenterology*, **11**, 419-425.
 Spencer, J., Burns, G. P., Cheng, F. C., Cox, A. G., and Welbourn, R. B. (1969). Differences between males and females in the Hollander insulin test. *Gut*, **10**, 307-310.
 Thomson, J. P. S., Russell, R. C. G., and Hobsley, M. (1973). Double lumen tube for gastric secretion studies. *British Medical Journal*, **3**, 175.
 Weinstein, V. A., Hollander, F., Lauber, F. U., and Colp, R. (1950). Correlation of insulin test studies and clinical results in a series of peptic ulcer cases treated by vagotomy. *Gastroenterology*, **14**, 214-227.

Mathematical appendix

The method of standardising gastric secretion for height is illustrated in Fig. 4 using $\frac{1}{2}-2 V_G$ as the example.

1. V_G is plotted, as its natural logarithm, against height and the linear regression line P-Q determined

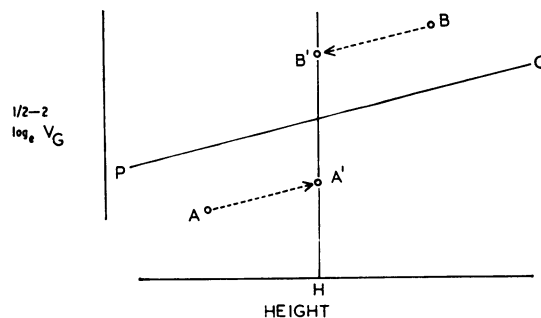


Fig. 4 For explanation see text.

by the method of least squares. The slope of this regression line is b .

2. Any point—for example, A—which represents the secretion value of a patient who is shorter than the mean height (H cm) of the group is standardised for height. This is achieved by moving the point along a line parallel with the regression line to the point A' lying on the vertical through H . Similarly, a point B is moved down the slope to the point B'.

3. The standardised value, A' or B', is read off the left hand scale as the logarithm and the antilogarithm of this value is the height-standardised V_G .

4. These manipulations may be summarised, again using $\frac{1}{2}-2 V_G$ as the example, in a formula.

Height-standardised $\frac{1}{2}-2 V_G = \text{antilog} [\log_e \frac{1}{2}-2 V_G - b(\text{height}-H)]$.

H for our group of patients was 169 cm. The value

of b for $\frac{1}{2}-2 V_G$ was 0.031301². Analogous formulae were derived from our results for the other indices of secretion. H naturally remained 169 cm throughout and the values of b are shown in Table 5.

*When using the value of b in equation A it is important to use all the digits quoted by us. Using fewer digits materially affects the result of the calculations.

Table 5 Value of b for each index of secretion, where b is regression coefficient for each index of secretion when independent variable is height

Index of secretion	Slope = b	
	$\frac{1}{2}-2 h$	Peak
V_{obs}	0.027784	0.25369
AO	0.049691	0.048330
V_G	0.031301	0.035722