

World perspectives: The OMGE database for inflammatory bowel disease

F TIMOTHY DE DOMBAL, MA, MD, FRCS

FT DE DOMBAL. World perspectives: The OMGE database for inflammatory bowel disease. *Can J Gastroenterol* 1993;7(7):550-556. This paper discusses several aspects of the creation, establishment and maintenance of a multinational database concerning patients with inflammatory bowel disease (IBD). The discussion is based on experience gained with the World Organisation of Gastroenterology (OMGE) Research Committee's multinational survey of patients with IBD, comprising (in 1992) 4612 cases entered from 50 centres in 27 countries. The purposes of such a database are described, as are a number of practical problems in its creation and maintenance, such as deciding upon a (minimum) dataset, measured to maximize objectivity, and other difficulties in implementation. Finally, the benefits of this type of research are discussed – and illustrated by highlighting some of the recent findings from the OMGE Research Committee's survey.

Key Words: *Computers, Inflammatory bowel disease, Multinational survey, Natural history*

Perspectives mondiales: Base de données OMGE pour les maladies intestinales inflammatoires

RÉSUMÉ: Cet article aborde divers aspects de la création, du fonctionnement et de la mise à jour continue d'une base de données internationale pour les patients atteints de maladie intestinale inflammatoire. La présentation se base sur l'expérience tirée d'une enquête internationale du comité de recherche de l'Organisation mondiale de gastroentérologie (OMGE) ayant porté sur des patients atteints de maladie intestinale inflammatoire, et qui comprenait, en 1992, 4 612 cas inscrits dans 50 centres répartis dans 27 pays. Les buts de cette base de données sont décrits, ainsi qu'un certain nombre de problèmes pratiques relatifs à sa mise au point et à son maintien, comme par exemple, les décisions qui portent sur les séries de données minimum utilisées pour rehausser le degré d'objectivité et autres difficultés de réalisation. Finalement, les avantages de ce type de recherche sont présentés et illustrés à l'aide de certains résultats récents tirés de l'enquête menée par le comité de recherche de l'OMGE.

Clinical Information Science Unit, University of Leeds, Leeds, United Kingdom

Correspondence: Professor FT de Dombal, Director, Clinical Information Science Unit, 22 Hyde Terrace, Leeds, LS2 9LN United Kingdom. Telephone +44 532 334961, Fax +44 532 429078

UNTIL THE 1960s, MOST 'RESEARCH' in clinical medicine was concerned primarily with the annotation of experience. Since then, the pendulum has swung in a different direction, so that nowadays any research which does not include a randomized controlled trial or deal with some highly complex aspect of serology or immunology is greeted with the gravest suspicion.

Yet it is possible to argue that perhaps the pendulum has swung too far, and two developments have lent this argument some force. The first is the increasing ease of communication between centres, countries and continents, so that 'shared experience' in real time becomes a practical proposition. The second concerns the development of information technology – to the point where data storage is limited solely by our ability to collect the data concerned. These reasons account for a revived interest in the multicentre 'survey' but there is a clear lesson to be learned from earlier single-centre studies. This concerns the *objectivity* of the data collected; it is not sensible (nor scientifically permissible) to correlate data between different centres when these data are collected under different circumstances and using different criteria.

The twin keys, therefore, to the establishment and maintenance of a use-

ful database in any clinical area are data collection from a wide number of representative centres, and the strictest possible rigour and quality control over the collected data. It is in these two areas that the World Organisation of Gastroenterology (OMGE) inflammatory bowel disease (IBD) survey has perhaps broken new ground, and why this OMGE survey forms the template for discussion in this paper.

SURVEY DESCRIPTION

The multinational survey of IBD which has given rise to the database discussed in this presentation has been well documented elsewhere (1-5).

Briefly, the survey began in 1976 with fairly modest aims: namely to establish the initial presentation and management of patients with IBD; and to study the ulcerative colitis: Crohn's disease (UC: CD) ratio in various centres. The survey now encompasses 50 centres in 27 countries and, as of 1992, some 4612 patients had been entered into the survey. Of these patients, 2779 are categorized as having UC and 1712 as having CD. A further 75 patients (1.6% of the total group) are classified as having indeterminate colitis. The remaining 46 patients have been rejected from the survey because of incomplete data collection or the establishment of an alternative diagnosis. The original aims of the survey have been much expanded. In recent years, attention has moved from initial presentation to follow-up of these patients. A total of 9000 patient-years of observed follow-up had been recorded by 1990 – possibly the largest multinational follow-up ever attempted in relation to IBD.

The data from this survey have thus given rise to an international database of considerable proportions. For example, the presentation data alone amount to 627,000 datapoints, with a missing data rate of just 1 or 2% from the patients entered into the survey. As a result, the survey data present a unique opportunity to: provide new and different insights into the presentation and natural history of IBD; and provide a basis for discussion of the purposes, difficulties and practical imple-

mentation of a large multicentre database in this clinical area. These latter aspects will be discussed.

DEFINITION AND PURPOSE OF A 'DATABASE'

What is a 'database'? A database is defined in the dictionary as 'a collection of information suitable for storing on a computer'. This definition reflects a common process, understandable but erroneous. In this process, a survey is begun (often with vague aims) and, as the number of patients increases and the amount of data expands, analysis becomes difficult and it is decided 'to computerize the survey'.

This process loses sight of the essential purpose behind large-scale survey work and computerized databases. The purpose of a large-scale survey is primarily the *acquisition of information from a representative sample*. This process is different from (though complementary to) the small-scale controlled clinical trial, the strength of which is that the conditions of study can be carefully and scientifically controlled, but the defect of which is that the results may only apply to a small series of patients (since it is virtually impossible – especially in IBD – to guarantee that a handful of patients will be representative).

There are thus many possible reasons why a large-scale representative sample of patients should also be studied. These surveys supplement carefully controlled small-scale studies by allowing the correlation of results between many individual centres or even between countries and continents, establishment and definition of consensus practice, provision of guidelines for educational purposes and so on.

Moreover, once it has been decided to collect a large representative sample, it makes good sense nowadays to minimize the tedium of analysis by establishing a database of information on an appropriate computer system. It remains true, however, that the need for computerization should be driven by an a priori consideration of the purposes of the survey and not (as is customary) by a dataset which has grown beyond the capacity of those managing it to analyze it by hand.

PRACTICAL PROBLEMS CONCERNING DATABASES

Data conformity: Surprisingly, one of the major difficulties (even today) in making comparisons between centres and continents is that different centres collect different information. There is a clear need in establishing a database to maximize conformity by the construction and circulation of agreed pro formata which constitute a minimum dataset of information about the patients concerned. The content of the pro formata (and hence the database) is defined by the purpose of the study, recalling that the dataset collected constitutes a 'trade-off'. The more data collected, the more extensive the potential analysis, but the more data required, the more likely participants are to drop out of the survey because of the cost in time, effort or dollars.

The original OMGE dataset and pro formata have been described elsewhere (1), as has another version of the data collection form devised in collaboration with the National Foundation for Ileitis and Colitis and the International Organisation for the Study of Inflammatory Bowel Disease (6). A more recent version of this second form trialled by OMGE centres and designed to maximize ease of data collection at individual patient visit is shown in Figure 1.

Maximizing objectivity: Any database of information is only as valid as the data contained within it. If the data within it are not capable of reliable and reproducible elicitation, the whole database may be worthless. In the case of IBD, problems in eliciting and recording reproducible data are severe. This applies to clinical symptoms and signs (7), physicians' estimates of patient progress (8), and calculation of various indices of severity and activity (9).

The OMGE survey has attempted to address this problem in a number of ways. First, observer variation studies concerning the data to be recorded have been carried out prior to any widespread data collection. Second, preliminary versions of the data collection form have been piloted prior to widespread distribution. Third, and crucial

CROHN'S DISEASE ACTIVITY PROFILE

SEND

PATIENT: Name _____

SURVEY NO: _____

SEEN BY DOCTOR: _____

TIME FRAME FOR THIS PROFILE: _____

FROM DATE: ____/____/____

TO DATE: ____/____/____

PATIENT WELL-BEING: WELL : SLIGHTLY UNWELL : CONSIDERABLY UNWELL : TERRIBLE

SYMPTOM CARD TOTALS: BOWELS:____ PAIN:____ WELLBEING:____

SYMPTOMS	BOWELS:						COMPLICATIONS	SYMPTOM CARD TOTALS:		
	Never	1/mth or less	↔	1/week	↔	Every day		Never Present	Current/Active	Known/Inactive
Excess frequency 4+/day							REMOTE:			
Liquidity							Arthritis			
Blood/pus							Aphthous Stomatitis			
Disrupts activity/sleep							E. Nodosum			
Constipation							Pyoderma			
Incontinence							Episcleritis/keratitis			
PAIN:							PERIANAL:			
Tolerable							Fissure			
Terrible							Tags			
FEVER:							Abscess			
Temp 38°C or more							Fistula			
Drenching sweats							OTHER: (Specify)			
Highest Recorded Temp: _____ °C										

EXAMINATION	MASS:			CURRENT THERAPY
	None	Dubious	Definite (Diam.)	
	Not Tender	Mildly Tender	Very Tender	** Tick box if being prescribed and state preprn. and dose *
TENDERNESS:	None	Diffuse	Local	Analgesics
	Dubious	Mild	Severe	Syst. Steroids
ENTERIC FISTULA:	None	Dubious	Definite	Local Steroids
	Inflamed	Draining	Neither	5 ASA
NUTRITIONAL STATUS:				Sulfasalazine
Height _____ cm/in. Weight _____ kg/lb.				Azathioprine
Usual premorb. wt _____ Wasting / Emaciated				Metronidazole
				Anti-diarrheals
				Restricted Diet
				T.P.N.

INVESTS	** Attach separately copies of X-ray endoscopy reports **					
	Hb. _____ WBC. _____ ESR. _____ Creatinine. _____ SGOT. _____ Alk. Phosphatase. _____					
Na. _____ K. _____ Cl. _____ Calcium. _____ Alb. _____ Proteins Total. _____						

** Add CRP, orosomucoid platelets if available **

PHYSICIAN'S OVERALL JUDGEMENT: BETTER : SAME : WORSE

WELL : SLIGHTLY UNWELL : CONSIDERABLY UNWELL : TERRIBLE EXCELLENT |-----| TERRIBLE

NEXT APPOINTMENT DATE: ____/____/____ TREATMENT MEANWHILE: _____

Figure 1) Crohn's disease activity profile. Reproduced with permission from de Dombal. Measures of disease activity. In: Bayless TM, Ed. Current Management of Inflammatory Bowel Disease. Toronto: BC Decker, 1988:1-7

TABLE 1
Stability of match between computer-aided prediction and actual centre diagnosis (at the patient's first referral) in 4066 cases over 12 years

Year	Cases in survey	Diagnostic match*	Percentage
1978	1056	988	93.6%
1982	2657	2476	93.1%
1986	3175	2924	92.1%
1990	4066	3761	92.5%

*Match between on-site initial diagnosis in centre concerned and computer-aided diagnostic prediction based on data from other survey cases

TABLE 2
Stability of diagnosis between referral and subsequent review after mean follow-up of nine years in 1445 patients

Initial diagnosis	Follow-up diagnosis*		
	CD	UC	Indeterminate
CD	459	10	10
UC	24	894	18
Indeterminate	6	6	18

*After mean of 9.0 years; adjustment levels over that period were ulcerative colitis (UC) to Crohn's disease (CD) 2.6%, CD to UC 2.1%

TABLE 3
Attack rates in current World Organisation of Gastroenterology (OMGE) series (with percentage of severe attacks) in 1855 observed patient-years of follow-up

	Current OMGE Series		1960s data*
	SASP	No SASP	
Patient-years observed	964	890	1323
Attack rate/year	44%	69%	55.7%
% severe attacks**	8%	11%	14.6%

*For comparison, data from typical 1960s (Leeds - collected 1956-1964) series included. Note attack rate reduced in patients receiving maintenance therapy, but severe attack rate reduced in all current OMGE patients; **As judged by criteria of Truelove & Witts (1954) (16). SASP Receiving sulphasalazine maintenance therapy

to the successful prosecution of any multicentre survey, the OMGE survey team have instituted a system of computer-aided quality control during and throughout the data collection process. **Computer-aided quality control:** Computer-aided quality control is the hallmark of the difference between early surveys (with all their inherent faults) and the more rigorous data collection currently possible. The process has been described in more detail elsewhere (1). Essentially, in this process, each participating centre's data are entered into a computer-based quality

control system which constantly checks the individual centre's data against the remaining data pool from the survey as a whole. Differences which exceed limits of tolerance are highlighted, and are drawn to the attention of the survey team and the participating centre.

Data analysis: Another problem which overcomes many surveys and databases constitutes what is commonly known as 'the data graveyard'. In this situation, the amount of data eventually overwhelms the investigator's ability to analyze it, so that increasing volumes of data lie unanalyzed.

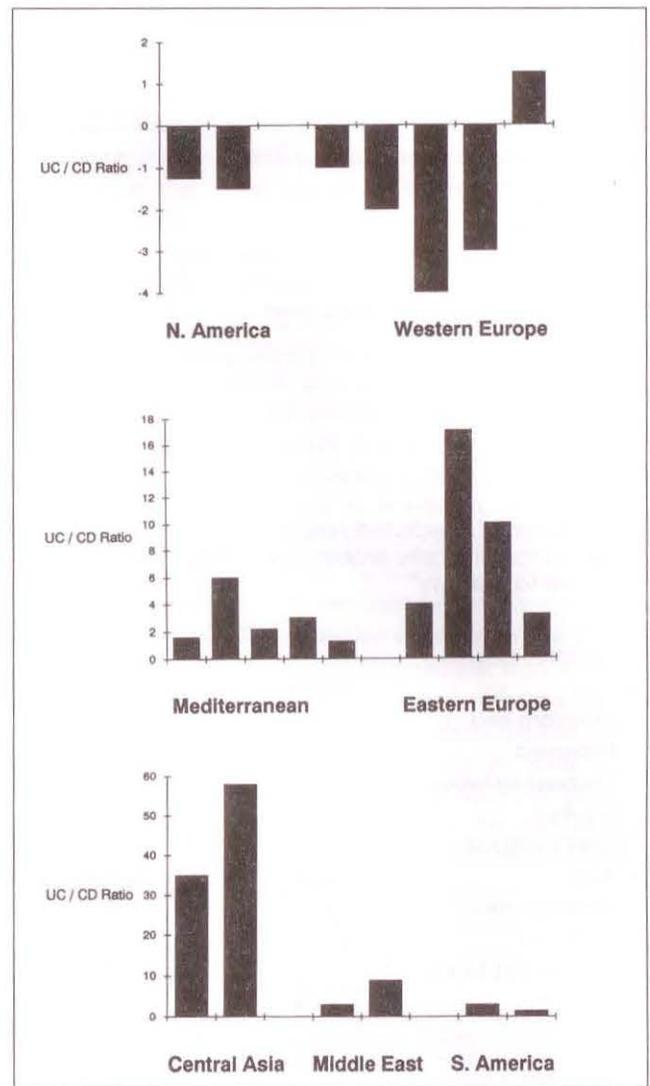


Figure 2) Regional and continental differences in the ulcerative colitis (UC):Crohn's disease (CD) ratio in 4000 patients with inflammatory bowel disease entered into the World Organisation of Gastroenterology (OMGE) survey (based on work by Softley [11]). Positive ratio favours UC, negative favours CD. The OMGE Inflammatory Bowel Disease Survey 1976-90: data presented to the 9th World Congress of Gastroenterology, Sydney, Australia, 1990

The remedy for this situation is, in principle, simple and twofold: computerization and the creation of a dedicated analysis team. Of these, the second is the more important. In the case of the OMGE IBD survey, the problem was addressed by prior creation of a central dedicated analysis team to ensure that analyses agreed prior to the commencement of the survey were carried out, further queries and suggestions from survey participants might be responded to so that the best possible value might be obtained from the data collected.

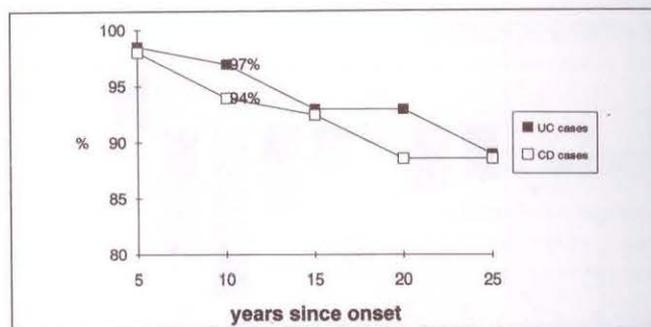


Figure 3) Cumulative mortality of inflammatory bowel disease patients over a 10-year period from disease onset; based on work by Softley (11). CD Crohn's disease; UC Ulcerative colitis

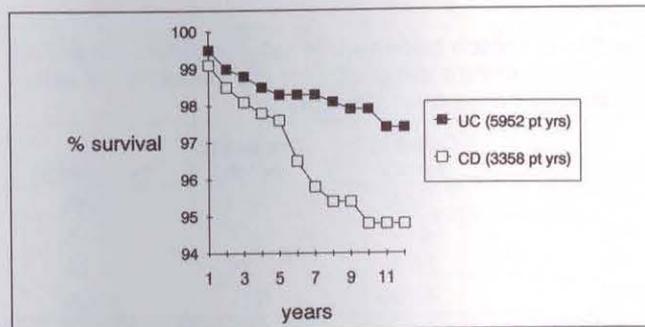


Figure 4) Cumulative mortality over a 10-year period from patients entered into the World Organisation of Gastroenterology survey in 9310 observed years of follow-up, based on work by Softley (11). CD Crohn's disease; pt yrs Patient-years; UC Ulcerative colitis

TABLE 4
Comparison of predicted response to therapy at admission (from the World Organisation of Gastroenterology computer prediction system) versus actual response to therapy*

	Predicted		
ULCERATIVE COLITIS			
<i>Actual</i>	Good	Intermediate	Poor
Symptom-free	31	9	25
Improved	35	21	40
Unchanged/worse	7	10	37
% failed	9.5%	25.0%	36.2%
CROHN'S DISEASE			
<i>Actual</i>	Good	Intermediate	Poor
Symptom-free	24	10	5
Improved	21	14	21
Unchanged/worse	3	12	44
% failed	6.3%	33.3%	62.8%

*Actual response to therapy (assessed one month later by clinicians) in 215 patients with ulcerative colitis and 154 patients with Crohn's disease (preliminary results; for more detailed results in expanded series, see reference 16)

Computerization is by no means a sinecure. There are various disadvantages (including costs and resources). It is all too easy to select the wrong computer package for a database survey if the purposes of the survey are not agreed prior to data collection. Finally, it is worth remembering that a computer system which stores patient data must comply with legislation in force in the various participating countries concerning data processing and data protection. This legislation has, of course, to be respected in relation to OMGE or any other surveys.

BENEFITS OF DATABASES

General benefits: In general terms, the benefits of database collection are several. They include practical benefits; it has been widely shown (10) that the

mere presence of a database and its impact on information flow has beneficial effects upon clinical care, both in terms of physician performance and patient outcome. In the specific case under discussion, it is perhaps best to illustrate some of the benefits of the database by highlighting some of the more interesting results presented at the recent World Congress in Sydney, Australia in 1990 (11).

Diagnosis: By the late 1970s, computer-aided analysis had demonstrated a congruence of diagnostic thought around the world concerning the distinction between UC and CD (1). In the early 1980s, a simple scoring system was devised and tested for this purpose, and proved to be over 90% effective in a series of 510 cases (4). By 1990, repeat studies in 4000 patients showed the

congruence of thought still to exist – the 'match' between computer-aided prediction and actual centre diagnosis being 92.5% in 4066 cases (Table 1). Moreover, the initial diagnosis proved highly stable. Once a diagnosis of CD or UC had been established, after nine years of follow-up, the diagnosis was altered in only 2% of patients (Table 2). **UC:CD ratio:** Early work in the late 1970s revealed considerable variation in the ratio between UC and CD patients in individual centres (varying between 3.4:1 in favour of CD and 9:1 in favour of UC) (1). Computer-aided analysis revealed that the differences were unlikely to be due to differing diagnostic criteria or semantics in the centres concerned (1). More recently, expansion of the series has enabled centres to be grouped by continent and geographical region (5). Interesting differences have emerged; there is a relatively high CD ratio in North America and Western Europe compared with a relatively high UC ratio in Mediterranean countries, eastern Europe and central Asia, with a ratio approximating to 1:1 in the Middle-East and South America (Figure 2).

Natural history: The OMGE survey clearly confirms the suspicion from a number of studies – namely that a considerable change has taken place in the natural history of IBD (particularly of UC) in the past two or three decades.

Studies in the 1960s showed a cumulative mortality for UC of approximately 20% over 10 years, with a lower mortality for CD. By 1990, actuarial analysis of over 9000 patient-years of observed follow-up became possible

from the OMGE survey. This indicated a far lower cumulative mortality: approximately 3 to 4% over 10 years from onset of disease (irrespective of whether the patient had UD or CD) (Figure 3). In terms of survival following referral (Figure 4), similar findings emerged with a cumulative mortality over 10 years of approximately 5% for CD but only 2% UC (in 9310 observed years of follow-up). These same studies also confirmed on a multinational basis, the single-centre data from St Marks Hospital, London, United Kingdom (12) (and other studies) concerning reduction in cancer risk (the OMGE series showed a cumulative risk in patients with total colitis of 9% over 20 years opposed to 50% in earlier 1960s studies [13]).

Attack rates in UC patients: By 1990, 1855 patient-years had become available for detailed study concerning attack rates, both overall and in terms of severity of attack (Table 3). A number of interesting features emerge from this analysis. First, the overall attack rate (between 50 and 60% per annum) was similar to that recorded in other series in the 1960s (14,15). However, in the 1960s series, up to a quarter of all attacks were classified as severe (according to the Truelove and Witts criteria) (16). This figure had fallen to around 10% in the OMGE series, in which data were largely recorded from the 1980s. Finally, the attack rate in UC patients in the OMGE series was significantly affected by maintenance sulphasalazine therapy (44% in those receiving maintenance therapy versus 69% in those who were not). The severity of attacks, however, were unaffected by maintenance therapy.

Activity indices: The observer variation in eliciting data and calculating

indices of 'severity' (particularly in relation to CD) has been explored in conjunction with the International Organisation for the Study of Inflammatory Bowel Disease and is reported elsewhere (9). Further work from the OMGE series has concerned two questions. First, by reference to a target series of 200 OMGE patients, it has been possible to establish that none of the reported indices of severity of activity concerning CD (such as the CD Activity Index [or its modifications] or the Dutch Activity Index) relate to short term prognosis in an individual patient in a close enough way to be used in clinical care or for stratification of patients in clinical trials (2). Second, and more encouragingly, computer-aided analysis has enabled scoring systems to be developed so that individual patient prognostic prediction may be made concerning short term outcome on the basis of data available at presentation to hospital. The results of applying this predictive system to 513 patients with UC and 293 patients with CD (unpublished data) are sufficiently encouraging to warrant intensive further effort to develop accurate short term prognostic predictive systems sufficiently accurate to be of value both in managing individual patients and in stratifying patients for the purposes of conducting clinical trials (Table 4).

DISCUSSION AND CONCLUSIONS

In many ways, the data set out in the preceding tables and figures speak for themselves. For good or ill, what is beyond question is that these data could not have been collected without the benefit of a multicentre, worldwide study and by disciplined effort from the participants in the centres involved,

coupled with rigorous data collection quality control throughout the study.

The results from the study (particularly those concerning the congruence of diagnostic thought, the altered natural history and the accurate prediction of short term prognosis) seem to indicate that there is some practical value in this type of exercise.

It needs to be emphasized, however, that before such databases are created, their purpose should be defined and their use should be piloted. Observer variation studies are needed to maximize objectivity, and careful prior consultation is needed to minimize the amount of information which participants will be asked to collect.

All databases need quality. All databases involve different people collecting data and putting it into a central pool. Data collected without regard to quality are worse than useless. Moreover, quality cannot be built in or added on at a late stage. It must be considered prior to the start of data collection, and quality control needs to be instituted throughout the collection period. It is in this role (perhaps above all other) that the computer has a helpful role to play.

Thus, from the present survey some evidence may be adduced that – with attention to the points mentioned – the establishment of a multinational or multicentre database adds considerable value to single-centre studies.

ACKNOWLEDGEMENTS: It is apparent from consideration of these data that this paper could not have been written without the effective and enthusiastic participation of colleagues worldwide. This participation is acknowledged with warm gratitude, and if there is benefit in the resultant analysis, the credit should largely go to those who have given their time and experience to the survey.

REFERENCES

1. Myren J, Bouchier IAD, Watkinson G, de Dombal FT. Inflammatory bowel disease. An OMGE study. *Scand J Gastroenterol* 1979;14:(Suppl 56):1-27.
2. Myren J, Bouchier IAD, Watkinson G, Softley A, Clamp SE, de Dombal FT. The OMGE multinational inflammatory bowel disease survey 1976-1982. *Scand J Gastroenterol* 1984;19:(Suppl 95):1-27.
3. Myren J, Bouchier IAD, Watkinson G, Softley A, Clamp SE, de Dombal FT. OMGE multinational inflammatory bowel disease survey 1976-1988. A further report on 3,175 cases. *Scand J Gastroenterol* 1988;23:(Suppl 144):11-9.
4. Clamp SE, Myren J, Bouchier IAD, Watkinson G, de Dombal FT. Diagnosis of inflammatory bowel disease – A multinational scoring system. *Br Med J* 1982;284:91-5.
5. de Dombal FT, Myren J, Bouchier IAD, Watkinson G, Softley A, eds. *Inflammatory Bowel Disease*, 2nd edn. Some International Data and Reflections. Oxford: Oxford University Press. (In press)
6. de Dombal FT. Measures of disease activity. In: Bayless TM, ed. *Current Management of Inflammatory Bowel Disease*. Toronto: BC Decker Inc, 1988:1-7.
7. Gill PW, Leaper DJ, Guillou PJ,

- Staniland JR, Horrocks JC, de Dombal FT. Observer variation in clinical diagnosis. A computer-aided assessment of the magnitude and importance in 552 patients with abdominal pain. *Methods Inf Med* 1973;12:108-13.
8. Graham NG, de Dombal FT, Goligher JC. Reliability of physical signs in patients with severe attacks of ulcerative colitis. *Br Med J* 1971;2:746-8.
9. de Dombal FT, Softley A. IOIBD Report No 1. Observer variation in calculating indices of severity and activity in CD. *Gut* 1987;28:474-81.
10. Adams ID, Chan M, Clifford PC, et al. Computer-aided diagnosis of abdominal pain: A multi-centre study. *Br Med J* 1986;293:800-4.
11. Softley A. The OMGE inflammatory bowel disease survey 1976-90. Presented at the 9th World Congress of Gastroenterology, Sydney, Australia, August 1990.
12. Lennard-Jones JE, Melville DM, Morson BC, Ritchie JK, Williams CB. Precancer and cancer in extensive ulcerative colitis: Findings amongst 401 patients over 22 years. *Gut* 1990;31:800-6.
13. de Dombal FT, Watts JM, Watkinson G, Goligher JC. Local complications of ulcerative colitis: Stricture, pseudopolyposis and carcinoma of the colon and rectum. *Br Med J* 1966;1:1042.
14. Watts JM, de Dombal FT, Goligher JC. Long term complications and prognosis following major surgery for ulcerative colitis. *Br J Surg* 1966;53:1015.
15. Edwards FC, Truelove SC. Long-term prognosis. Part II. *Gut* 1963;4:309-15.
16. Truelove SC, Witts U. Cortisone in ulcerative colitis. *Br Med J* 1953;2:375-8.
-



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

