The problem of diagnostic variability in general practice

D L Crombie, K W Cross, D M Fleming

Abstract

Study objective—The aim was to examine the scale, source, and relevance of variation between general practices in respect of the rates with which patients consulted with illnesses falling in each of several diagnostic groups.

Design—This study involved a general practice morbidity survey conducted over two years, 1970–72. All patients who consulted their general practitioners were identified and the number of these who consulted with diagnoses attributable to each of the 18 main chapters of the *International classification of diseases* were counted. Patients who consulted for more than one diagnosis within a chapter were counted once only; those who consulted for one or more diagnoses in each of several chapters were counted once for each chapter.

Setting—This was a national survey involving general practitioners in England and Wales.

Subjects—The study involved 214 524 patients from 53 selected general practices (115 doctors) who were registered with their general practitioners for the whole of the year 1970–71 and for whom their morbidity data had been linked with their social data from the 1971 census.

Measurements and main results—Using the numbers of patients on the practice lists as denominators, practice patient consulting rates (PPCR) were calculated for each practice and for each ICD chapter. Variability in chapter PPCR was examined by calculating coefficients of variation and, after allowance for random variation, coefficients of residual variation. There were large interpractice (doctor) variations in all chapter rates. These variations were only marginally attributable to: chance; different age, sex and social class mixes of practice populations; geographical locations; and practice organisation. The rates were, however, consistent from one year to the next for any one practice. Approximately half of the interpractice (doctor) diagnostic variability was associated with overall patient consulting behaviour. When the effects of this behaviour were discounted, any major residual diagnostic variability was confined largely to ICD chapters I–V, XVI, and XVII, ie, those chapters where aetiology forms the basis of classification.

Conclusion—Variations in recorded diagnostic rates are mainly due to the consistent

but idiosyncratic and selective exclusion by practitioners of some components from the total set which often coexist in a new diagnosis. Because of the scale of interpractice diagnostic variability, the use of algorithms and information technology is largely precluded from outcome studies, auditing procedures, and studies of practice work loads in general. However, (1) the consistency of any individual doctor's pattern of diagnostic recording from one year to another permits studies of trends; and (2) given a reasonable number of recording practices, the population mean practice consulting rates can be estimated with sufficient accuracy for many epidemiological research and administrative uses.

J Epidemiol Community Health 1992; 46: 447–454

Persons bring problems to general practitioners who interpret them in the form of one or more diagnostic labels. For example, a female aged 45 years may consult a general practitioner with menorrhagia and, on detailed questioning, she may have other menopausal symptoms and may also appear depressed. Examination findings might include a fibroid and a haemoglobin of 10.5 g/dl. In total, a description of this patient's problem may include from 1–5 labels (fibroids, menorrhagia, reactive depression, microcytic anaemia, menopausal symptoms). The labels categorise the illness into various components which doctors enter into a recording system as episodes. These may extend across the boundaries of the chapters of the *International classification of diseases* (ICD) as in this example which encompasses three chapters. The general practitioner, using his clinical judgement and personal assessment of the priority for management, will select the one (or more if necessary) episode label best fitted to the problem for the record. His choice of label may not include any conventional diagnosis, especially where problems are more closely related to social issues than to morbid processes. Many labels have little diagnostic content¹ and they can be applied as an aide memoire in a holding operation pending the exclusion of some particular diagnosis.

In this paper we shall examine the scale, source, and relevance of the resulting diagnostic interpractice variability.

Methods

The data for this study are taken from the recording years 1970–71² and 1971–72 of the Second National Morbidity Survey.³ This survey

Birmingham
Research Unit, Royal
College of General
Practitioners,
Lordswood House,
54 Lordswood Road,
Harborne,
Birmingham B17 9DB,
United Kingdom
D L Crombie
D M Fleming
Department of Social
Medicine, Medical
School, Edgbaston,
Birmingham B15 2TJ,
United Kingdom
K W Cross

Correspondence to:
Dr Crombie

Accepted for publication
January 1992

was used because it is the only national morbidity survey of general practice which (1) spanned two consecutive years, and (2) provided a linkage with social data collected from a national census. In 1970-71, 115 doctors in a total of 53 practices looking after 300 000 patients recorded all face to face contacts with their patients in modified disease indices. This population at risk was representative of the population for England and Wales by age and sex.² Of the original 53 practices, 43 continued to record for a second year. The basic entry in the index was the episode of illness and a diagnostic classification was used based on the 7th revision of the ICD adapted for general practice. Linkage of morbidity data with selected social data from the 1971 census was achieved for 214 524 patients present for the whole year; these form the population studied here. All the data used in this study are person based for each practice separately.

Analysis has been based on practice person consulting rates (PPCR) for the aggregated data for each of the 17 ICD chapters concerned with morbidity, excepting chapter XI (Deliveries and complications of pregnancy, childbirth and puerperium) and chapters XIV and XV (Congenital anomalies, and Certain causes of perinatal morbidity and mortality), in which numbers were particularly small and often zero. The number of persons consulting over a year per 1000 at risk (ie, the PPCR), unlike measures such as consultation or episode rates, is a true proportion with the person only counted once regardless of the number of consultations undertaken; thus a PPCR is clearly a binomial variate.

In order to contrast variation between practices for different diagnostic groups, we need a measure of variability that takes account of differing mean rates. The obvious choice is the coefficient of variation, but without adjustments this does not take account of the random element of variability

which depends upon the number of practices in the subset, the values of the proportions considered, and the variable denominators, ie, practice populations. A method which takes these influences into account and leads to the calculation of coefficients of residual variation and variance measured in absolute terms is outlined in appendix A.

The influence of age and social class upon practice variability is examined by calculating, for each sex, standardised rates by the indirect method and hence adjusted coefficients of variation. These are contrasted with the coefficients of variation for the unstandardised rates to highlight any differences which these two characteristics may make to the variability.

Results

The mean PPCR values (unstandardised) for each sex for each chapter of the disease classification are given in the first two columns of table I. The interquartile range is specified under each mean to give some idea of the variation inherent in the PPCR values. For illustrative purposes the full distributions of male PPCR values for chapters II, V, and VI are displayed in fig 1; these have been

Table I Mean practice person consulting rates (PPCR) by sex and diagnostic group (interquartile ranges in brackets) and coefficients of variation and residual variation

ICD chapter	Mean PPCR per 100 patients at risk		Coefficients of variation (CV)		Coefficients of residual variation (CRV)	
	M	F	M	F	M	F
I Infective and parasitic diseases	7.5 (5.8-9.1)	7.5 (5.9-9.2)	31	32	30	31
II Neoplasms	1.0 (0.7-1.2)	1.4 (0.9-1.7)	50	44	42	38
III Endocrine, nutritional and metabolic disease	1.7 (1.1-2.1)	3.9 (2.6-4.8)	42	40	36	37
IV Diseases of blood and blood-forming organs	0.5 (0.3-0.7)	2.0 (1.2-2.5)	60	62	46	60
V Mental disorders	7.5 (5.7-8.7)	15.3 (12.0-18.0)	44	36	43	35
VI Diseases of nervous systems and sense organs	11.6 (10.4-12.8)	12.0 (10.7-13.6)	20	20	18	19
VII Diseases of circulatory system	5.9 (4.6-6.8)	8.0 (6.5-9.8)	28	26	26	25
VIII Diseases of respiratory system	27.5 (23.4-32.1)	27.9 (23.4-31.3)	21	22	21	21
IX Diseases of digestive system	9.2 (7.3-10.9)	9.4 (7.7-11.0)	32	28	30	27
X Diseases of genitourinary system	2.5 (2.1-2.9)	12.6 (11.6-14.4)	23	18	16	17
XII Diseases of skin and subcutaneous tissue	11.2 (9.4-13.0)	12.7 (10.9-14.6)	21	21	19	20
XIII Diseases of musculoskeletal system	8.7 (7.8-10.0)	10.9 (9.5-12.5)	22	22	21	21
XVI Symptoms, signs and ill defined conditions	10.4 (7.6-13.4)	13.3 (9.0-18.0)	40	40	39	39
XVII Accidents, poisoning and violence	11.9 (8.6-13.5)	9.1 (7.0-10.2)	43	35	42	34
All diseases and conditions	63.4 (61.3-66.4)	70.4 (67.7-73.4)	9	7	9	7

Figure 1 Frequency distribution of practice person consulting rate (PPCR) values

chosen to show the variation around a very low mean (chapter II), and to show large (chapter V) and small (chapter VI) degrees of variation around "average" mean values.

The coefficients of variation are also given in table I (columns 4 and 5). For all diseases combined the values were small (coefficient of variation for males = 9, for females = 7). Broadly speaking, chapter coefficients of variation fall into three groups: chapters VI, VIII, X, XII, and XIII with values of the order of 20; chapters I, VII, and IX with values around 30; and the remainder with values of about 40 (or above). For most chapters, the PPCR values were approximately normally distributed (see also fig 1), and the interpretation of the coefficients of variation is facilitated by reference to appendix B, which gives 95th/5th centile ratios of PPCR values equivalent to different coefficients of variation. For example, the coefficient of variation for chapter VI—"Diseases of the nervous system"—equals 20 so that for each sex the PPCR at the 95th centile is double that at the 5th centile.

We next consider the corresponding coefficients of residual variation (table I, columns 6 and 7). The harmonic means of the practice list sizes which were used in their calculation (see appendix A) were 1378 for males and 1482 for females. Random (binomial) variation based on these denominators was very small in comparison with the systematic variation (vide infra); thus the residual variation coefficients differ only marginally from the corresponding variation coefficients, except for chapters II and III (both sexes) and chapters IV and X (males only).

Values of coefficients of residual variation for each chapter range from a maximum for chapter IV—Diseases of blood and blood-forming organs—to a minimum for chapter X—Disorders of the genitourinary system. Broadly speaking, greater systematic variation in PPCR values is

evident for chapters where diseases are grouped together on an aetiological basis (I–V and XVII) than for those grouped by site. The high values for chapter XVI—Symptoms, signs etc—were not unexpected and tell us more about the handling of clinical uncertainty by the various practices. For most chapters the male and female coefficients of residual variation were very similar; exceptions were chapters II, IV, V, and XVII.

INFLUENCE OF AGE, SEX, AND SOCIAL CLASS

Although the age, sex, and social class structures of individual practice populations differ, allowances for such differences had little effect upon interpractice variability in PPCR values. The coefficients of variation for age standardised rates (by the indirect method) as given in table II (columns 2 and 3) were virtually the same as those of the unstandardised rates (table I, columns 4 and 5). The only exception was chapter VII (Diseases of the circulatory system) for females with coefficients of variation of 22 and 26. For this chapter, age specific mean patient consulting rates for all practices combined varied between 2.6 per 1000 patients at risk for the under 5s to 273 for those aged 75 years and over. No other chapter of the ICD has more than a 10-fold difference between age specific rates. The correlation coefficient for males and females combined for unstandardised versus age standardised PPCR values for this chapter was 0.77. The next lowest was 0.94 for chapter XIII.

Comparison of the coefficients of variation for age standardised rates with those for age and social class standardised rates (columns 4 and 5) shows that social class also had no appreciable influence on practice variability, with the exception of males for chapter XVII—Accidents, poisoning and violence (coefficient of variation of 42 reduced to 37).

Finally, the coefficients of residual variation for the rates standardised by age and social class are given in columns 6 and 7 of table II. These are virtually the same as the corresponding coefficients of variation because of the small random components of variation (except for those chapters with very small PPCR values). Interestingly, the ratios of systematic (non-random) to random variation calculated by the method in appendix A as applied to the standardised rates can be shown to agree closely with the corresponding ratios using McPherson's method⁴ (which assumes Poisson variation) of estimating separately the random and systematic components of variation; the only notable exception was the comparison of ratios for chapter VIII—"Diseases of the respiratory system"—where the mean PPCR (for either sex) was relatively large (28 per 100).

INFLUENCE OF PATIENT CONSULTING AND DOCTOR DIAGNOSTIC BEHAVIOUR

Apart from the random variation due to sampling, and to age, sex, and social class differences, interpractice variation has two other main components: (1) that due to overall patient consulting behaviour (measured by person consulting rates for all diseases combined); (2) that due to the doctor's diagnostic behaviour (measured in major part by person consulting rates by ICD chapter).

Table II Coefficients of variation referable to distribution of standardised PPCR values by sex and diagnostic group

ICD chapter	Coefficients of variation of age standardised PPCR values		Coefficients of variation of age and social class standardised PPCR values		Coefficients of residual variation of age and social class standardised PPCR values	
	M	F	M	F	M	F
I Infective and parasitic diseases	29	30	29	29	28	28
II Neoplasms	50	44	50	45	42	39
III Endocrine nutritional and metabolic disease	43	40	43	39	38	36
IV Diseases of blood and blood-forming organs	60	63	59	63	46	60
V Mental disorders	43	35	43	35	42	34
VI Diseases of nervous systems and sense organs	20	20	20	20	18	18
VII Diseases of circulatory system	26	22	26	21	26	19
VIII Diseases of respiratory system	20	21	20	20	19	20
IX Diseases of digestive system	32	28	30	28	29	26
X Diseases of genitourinary system	24	17	24	16	17	15
XII Diseases of skin and subcutaneous tissue	20	20	20	20	18	19
XIII Diseases of musculoskeletal system	24	23	24	23	22	22
XVI Symptoms, signs and ill defined conditions	40	41	40	40	39	39
XVII Accidents, poisoning and violence	42	35	37	33	37	32
All diseases and conditions	9.1	6.8	8.7	6.7	8.5	6.5

PPCR = practice person consulting rate

The proportions of total variances in chapter PPCR values standardised for age and social class (columns 2 and 3, table III) have been obtained by regressing these chapter rates in turn against the all disease PPCR values, and then using the R^2 values. Bivariate scatter plots of male PPCR values for chapters II, V, and VI against the overall PPCR values are given in fig 2 as examples.

Using these proportions of the variance for each individual chapter, the coefficients of variation referable to interpractice patient consulting behaviour (columns 4 and 5) and to (doctor) diagnostic variability (columns 6 and 7) can then be estimated. For example, the coefficient of variation for males with mental disorders (chapter V) was 43 (column 4, table II). The proportion of total variance attributable to overall patient consulting behaviour was 21% (column 2, table III) so that the coefficient of variation corresponding to this component of the variance was 20 (column 4, table III). The residual variance proportion of 79% in turn gave a coefficient of variation for the doctor diagnostic behaviour of 38 (column 6).

Examination of these estimates of the coefficients of variation referable to doctor diagnostic variation (columns 6 and 7) shows that the interpractice variability was still considerable (coefficients of variation > 20) for chapters II-V and XVI-XVII; and the coefficients for chapter I—Infective and parasitic diseases—were also relatively high (24 and 26). However, the coefficients for the remaining seven chapters were of the order of 20 (or even less) indicating only a twofold (or less) ratio of the adjusted PPCR at the 95th centile over that at the 5th centile. Noteworthy differences between the sexes in respect of these adjusted coefficients of variation (columns 6 and 7) were evident for chapter IV, where the female coefficient exceeded the male, and for chapters III, V, and X, where the converse was true. When the same approach was applied to the unstandardised PPCR values, substantially higher (a difference > 3) coefficients of variation refer-

able to doctor diagnostic variation were found only for chapter VII (Diseases of the circulatory system).

CONSISTENCY OF VARIABILITY

Finally, table IV gives the correlation coefficients obtained when the unstandardised PPCR values for the first year of the Second National Morbidity Survey (1970–71) were correlated with those of the second year (1971–72) for the 43 practices which participated in both years. All the coefficients were very highly significant ($p < 0.001$) and in general the values for the sexes were similar. From this comparison we can say that the morbidity presented to the practices was similar in quantity in both years and that the recording habits within the practices were consistent. By implication, 80% (0.9 squared) of the variance was thus contained within the practice recording habit and prevailing morbidity among those patients.

Table III Percentage of total variances of distribution of chapter PPCR values for all diseases (standardised for age and social class) and contribution of component variances to chapter coefficients of variation (CV)

ICD chapter	Percentage of total variance		CVs referable to patient consulting behaviour		CVs referable to doctor diagnostic behaviour	
	M	F	M	F	M	F
I Infective and parasitic diseases	30.9	20.6	16	13	24	26
II Neoplasms	16.6	4.5	20	10	46	44
III Endocrine, nutritional and metabolic disease	23.0	23.5	21	19	38	34
IV Diseases of blood and blood-forming organs	1.7	1.1	8	7	58	63
V Mental disorders	21.0	19.2	20	15	38	31
VI Diseases of nervous systems and sense organs	47.1	49.5	14	14	15	14
VII Diseases of circulatory system	39.9	17.5	16	9	20	19
VIII Diseases of respiratory system	58.5	43.1	15	13	13	15
IX Diseases of digestive system	55.3	38.0	22	17	20	22
X Diseases of genitourinary system	29.5	49.8	13	11	20	11
XII Diseases of skin and subcutaneous tissue	45.4	49.1	13	14	15	14
XIII Diseases of musculoskeletal system	46.3	42.8	16	15	18	17
XVI Symptoms, signs and ill defined conditions	8.8	12.5	12	14	38	37
XVII Accidents, poisoning and violence	38.3	15.4	23	13	29	30

PPCR = practice person consulting rate

Figure 2 Examples of bivariate plots of chapter practice person consulting rate (PPCR) values against overall PPCR values

It is of particular interest that among the highest correlation coefficients were the values of 0.94 for each of the sexes with mental disorders. The whole mechanism of diagnostic labelling is weakest for mental disorders (*vide infra*) yet individual practices were most consistent in recording such conditions.

Discussion

The main aim of this paper has been to establish the scale, source, and relevance of diagnostic variability. This initial study has been based upon the ICD chapter structure which is sufficient for this purpose. Simple univariate regression techniques, together with the new method for estimating the random component of variability and hence systematic variance in absolute terms, suffice to fulfil the limited aims. The relationship, if any, of diagnosis to the other three pillars of clinical management—referrals, investigations, and drug therapy—will be examined in a subsequent contribution.

SCALE OF VARIABILITY

In this paper we have shown that at the level of ICD chapters, the morbidity recorded by general practitioners during a year (as measured by person consulting rates) varies considerably from practice to practice. This was particularly true of morbidity subsumed under those chapters where aetiology forms the basis of classification; about one third of all episodes of illness seen by general practitioners fall into these chapters. Even for those chapters based upon the site of the disease, a minimum twofold difference between rates referable to the 5th and 95th centile practices was found.

SOURCE OF VARIABILITY

We first considered the possible contribution of random variation to the interpractice variability in PPCR values and showed that it was negligible, except for those chapters with very small PPCR values: chapter II (Neoplasms), chapter III (Endocrine and metabolic disorders), and (for males only) chapter IV (Diseases of blood and

blood-forming organs). These three chapters account for only 5.1% of all patients consulting. It is, of course, not surprising that the components of random variation for all other chapters were so trivial in view of the larger numbers of patients upon which the PPCR values are based. Furthermore, this was also true when standardised PPCR values were considered.

Although sex, age, and social class have well documented influences upon patient consulting rates, the effects of these characteristics upon the interpractice variability in the rates have been found to be minimal. The only notable exceptions were some (small) reductions in variability in the PPCR values for (1) chapter VII—Diseases of the circulatory system—after standardisation by age, and (2) chapter XVII—Accidents and poisoning, etc—after standardisation by social class. As regards the latter, there are established differences between social classes in the use of hospital accident and emergency departments without first seeing the general practitioner.

One major contribution to interpractice (doctor) variability is the varying propensity of different practice populations to consult. Person consulting rates for all diseases combined will include doctor availability and accessibility, in addition to overall morbidity, as specific influences on overall patient consulting behaviour. The variation in practice patient consulting rates for all diseases combined is relatively small, but it has a substantial influence on the chapter PPCR values, especially for chapters VI–XIII.

Doctor diagnostic behaviour

It has been consistently found from many different studies across the world^{5–9} that on average each new clinical problem presented to general practitioners contains some 2.5 separate episode elements, and that doctors at the first consultation record only 1.2–2.1⁷ of these separate episode elements in their records. In the first year of the second National Morbidity Survey (on average) only 1.1 components were recorded and entered as separate episodes. The variability between practices results from the idiosyncratic choice of the episodes actually recorded, and their chapter allocations, from the total of 2.5 (on average) present in each new problem.

This idiosyncratic, but as we shall see, consistent pattern of choice of a diagnostic label in many clinical problems has two main sources: the inherent diagnostic uncertainty in clinical problem solving; and the way in which each doctor restricts his choice from the possible episode labels.

Diagnostic uncertainty

Many of the problems managed by general practitioners resolve spontaneously and often only one consultation is required. There is, inevitably, considerable diagnostic imprecision about such illness episodes and that uncertainty is manifested in the way diagnostic labels are used. This is also the reason why many episodes of illness are labelled only in symptomatic terms. Variability was particularly great in chapter XVI—Symptoms, signs and ill-defined conditions—and minimal in chapters VI–XIII, which account for two thirds of all episodes.

Table IV Correlation coefficients (Pearson) for the PPCR values obtained in the year 1970–71 with those in the year 1971–72

ICD chapter	Male	Female
I Infective and parasitic diseases	0.66	0.66
II Neoplasms	0.67	0.75
III Endocrine, nutritional and metabolic disease	0.82	0.91
IV Diseases of blood and blood-forming organs	0.68	0.91
V Mental disorders	0.94	0.94
VI Diseases of nervous systems and sense organs	0.89	0.84
VII Diseases of circulatory system	0.91	0.92
VIII Diseases of respiratory system	0.92	0.90
IX Diseases of digestive system	0.74	0.68
X Diseases of genitourinary system	0.69	0.87
XII Diseases of skin and subcutaneous tissue	0.85	0.88
XIII Diseases of musculoskeletal system	0.85	0.90
XVI Symptoms, signs and ill defined conditions	0.81	0.85
XVII Accidents, poisoning and violence	0.95	0.92
All diseases and conditions	0.91	0.87

PPCR = practice person consulting rate

Consultation threshold

McPherson¹⁰ has made the point that the true incidence and prevalence rates for morbidity are the only satisfactory basis for any estimates of quality of outcome but are exceedingly difficult to establish. Jones and Lydeard have shown¹¹ that there was remarkably little difference in the true prevalence of dyspeptic symptoms in the populations of eight practices (estimated coefficient of residual variation = 7, using the method described in appendix A), though the rates at which the patients consulted their doctors varied considerably.

Priorities in clinical interpretations

Variability in these priorities is most evident in attitudes to the psychological/emotional component of problems. In the study cited earlier, based on the clinical assessment of video recordings of consultations,⁵ the proportion of all patients consulting with any psychological/emotional element, however clinically insignificant, was 40%. In this study the equivalent proportions for chapter V—Mental disorders—varied from 5% to 43% with a mean of 16.5%. We can be certain, therefore, that only the practice with the highest recording rate may contain an element of misdiagnosis or overdiagnosis. The degree of non-perception and/or non-recording is related to the differing degrees of clinical importance attributed to the psychological/emotional components of morbidity by individual doctors.

This example suggests that a large part of the interpractice (doctor) diagnostic variability is due to differences in clinical priorities arising from (1) selective non-recording, if not non-perception, among multiple episodes in patient problems; (2) selective use of one or more rubrics among possible alternatives; or (3) misdiagnosis.

Secular influences

These can be implied from data in table IV, which deal with recording consistency. Changes in rates from one year to another, however, are due in part to the summation of all other influences leading to change. These will include changes in staff (in particular doctors) and in recording methods, criteria, definitions, and habits. On the whole these latter influences will be similar for all chapters. For Infectious and communicable diseases (chapter I), the correlation coefficients for one year on the other in both sexes were comparatively low. This is, of course, to be expected since the epidemic quality of disorders in this chapter does mean that at least for some conditions there is variability from one year to the next. This also explains the relatively high coefficients of residual variation for this chapter.

RELEVANCE OF DIAGNOSTIC VARIABILITY

Diagnostic groupings

Cluster analysis techniques¹² have been used to bypass the difficulty of interpreting clinical data in the presence of diagnostic variability. However, if, as we have shown here, aggregation of individual diagnostic terms into clusters as large as ICD chapters still leaves very large inter-recorder variability, it is not surprising that ad hoc clustering such as the use of diagnostically related

groupings has not solved the basic problem. The best that has been achieved is that the rank ordering of the frequencies of certain ad hoc clusters is similar when measured in different practices.

Prior probabilities

Against this background, we have to acknowledge that no universal "prior probabilities" can be generated from data where the contributing sets give rise to coefficients of variation of 20–50%. This rules out diagnostic approaches based on Bayes' theorem and any logic branching techniques based on prior probabilities. Such "artificial intelligence" systems would need to be designed for each doctor (practice) separately, since each practice has its own unique set of rates, including patient consulting rates, and therefore also prior probabilities.

Outcome studies

Until we have proper measures of the outcome of care and of the true incidence and prevalence of morbidity, it is impossible to be certain whether the practice (doctor) with the highest recording and/or perception rate of morbidity is more effective than the one with the lowest rate.

Auditing

Grave doubts must exist about the validity of auditing procedures which are based on the comparison of actual performance against some idealistic norms in the face of the diagnostic variability established here. We must concentrate on the "process" of diagnosis to gain the insights necessary to develop hypotheses about the underlying reasons for interdoctor diagnostic variability which can then be the subject of true experiments.¹³ However, the actual person consulting rate must be known if the practice (doctor) contribution to total diagnostic variability is to be estimated accurately. The measurement of the proportion of patients consulting in a given length of time is extremely difficult and requires a very high level of discipline in recording which has to include all contacts.

Trends over time

Against this background of interpractice variability, trends and changes in rates over time can still be measured accurately because of the consistency of recording over time by each individual practice; hence the need to use at least some of the same practices in consecutive surveys.

Epidemiology, administration, and research

Paradoxically, the large scale interpractice (doctor) variability which has been described above presents less serious problems than might be expected when the objective is the estimation of *population* person consulting rates for epidemiological or administrative studies.¹⁴ We can illustrate this paradox using as examples Chapter V (Mental illness)^{15 16} and chapter IV (Diseases of blood and blood forming organs).

It has been shown that the annual mean practice person consulting rate for males with mental illness was 7.5 per 100 persons at risk (ie, large numbers of patients consulted) and the interpractice variability was particularly high for

reasons already discussed. However, the standard error of the distribution of the sample mean for the 53 PPCR values was only 0.38, so that the 95% confidence limits for the mean for all practices were approximately 6.7 to 8.3%. Very similar results are obtained for the estimate of the standard error of the population mean person consulting rate assuming the practice populations (list sizes) are considered as random cluster samples.

The mean male PPCR in this study for chapter IV (the chapter with the smallest mean rate and highest interpractice variability) was 0.50 per 100 patients and the 95% confidence limits of the mean PPCR for all practices were approximately 0.42 to 0.58. Thus for this condition, as well as for mental illness with which much larger numbers of patients consult, estimates of the population mean values for persons consulting are probably sufficiently accurate for purposes of epidemiological study.

We must accept that the decision to record at all will always be dictated by the priorities of the clinical situation as seen by each individual doctor. The confidence limits will, of course, always be sensitive to the number of practices involved.

CONCLUSIONS

A minimal twofold difference between chapter practice person consulting rates for the 5th and 95th centile practices cannot be ignored. The problem of variability is central to the issues of quality of care. Variability must be expressed in absolute terms if it is to be understood and compared in different contexts. This has dictated the use here of coefficients of variability and of residual variance measured in absolute terms. The mean PPCR values, coefficients of variation, and coefficients of residual variation based on these practice populations for which census data were available are in fact almost identical to the corresponding values based upon all practices and all patients in the second National Morbidity Survey, with a total survey population of 300 000 patient years.¹⁴ For some chapters, patient consulting behaviour shows nearly as much variability as doctor diagnostic variability; we have treated them here as though they were independent, although this is obviously not the case. While doctors may well influence the consulting habits of their patients via accessibility and availability, it is less likely that patients influence the diagnostic behaviour of their doctors to the same extent.

This diagnostic variability (which is particularly marked for chapters I-V based on aetiology rather than site) has its roots in the most fundamental aspects of professional expertise and training. Whatever we may believe about medical education, it does not appear to inculcate any consistent view of illness in practitioners, even for the basic activity of describing the clinical problems on which all management logically should follow. Diagnostic variability clearly has important implications and the profession may well have to consider how best it should be dealt with.

Any reappraisal will have to reconcile two presently irreconcilable antagonistic requirements: (1) the belief that there should be an agreed set of definitions for the terms used in clinical problem solving, in particular diagnostic termin-

ology; and (2) the need for maximising economy in a clinical problem solving system which has to be efficient as well as effective. Diagnoses are signposts to action.¹⁷ A logical framework is needed for appropriate selection of clinically important episode elements in each problem from the total, and guidance on how far the efficient search for final diagnostic accuracy and effectiveness should be prosecuted in each ad hoc problem.

Appendix A: Coefficients of residual variation

Practice patient consulting rates (PPCR) are true proportions and each will be subject to binomial variation. A method is required which will take account of such variation in the examination of the variation in PPCR values for each ICD chapter. In essence, an analysis of variance needs to be carried out which separates the binomial from the true extraneous variation in a situation where the proportions are based upon different denominators (ie, practice list sizes).

Cochrane¹⁸ addressed the problem of analysing the variance for percentages based upon unequal numbers and drew attention to the need for appropriate weighting of the observed proportions, depending upon the relative amounts of extraneous and binomial variation. Inspection of the distribution of PPCR values for any diagnostic group here considered suggests that most of the variation is extraneous. In fact, most PPCR values are less than 200 per 1000, the denominators (practice list sizes by sex) exceed 1000, and the ratio of the binomial to the total variance is less than the 30% which Cochrane considered to be the level below which weighted proportions are most appropriate.

Consider the i 'th practice with a list size of n_i , and observed PPCR (as a decimal fraction) of p_i , and assume that this is a sample from a population with a true mean PPCR of P_i .

The variance of p_i about P_i is given by the usual expression for the binomial distribution:

$$P_i(1 - P_i) \div n_i$$

The true proportions, such as p_i , will vary from practice to practice round a mean of P , and for the total variance of an observed PPCR we may therefore write:

$$[P_i(1 - P_i) \div n_i] + \sigma_i^2$$

Since σ_i^2 , which represents the extraneous variance, is the variance of the true proportion it will not in general depend upon n_i .

Summing over k practices (independent samples) and dividing by k , the expected error mean square of the unweighted observed PPCR values will be:

$$\frac{1}{k} \sum_{i=1}^k [P_i(1 - P_i) \div n_i] + \frac{1}{k} \sum_{i=1}^k \sigma_i^2$$

The values of the true proportions are unknown, although we may assume from the distributions of observed proportions that they will be approximately normally distributed. We may reasonably substitute p , the weighted average of the observed p_i values, for these unknowns so that the above expression becomes:

$$[p(1 - p) \div n_h] + \sigma_e^2$$

where n_h is the harmonic mean of the practice list sizes and σ_e^2 is the average extraneous variance.

Thus if we denote the unweighted variance of the observed proportions by S^2 the average extraneous variance after removal of the average binomial variation is approximately:

$$S^2 - [p(1 - p) \div n_h]$$

We finally define a coefficient of residual variation (CRV) as the square root of the above expression divided by p . For large values of n_h as in this study the CRV will clearly correspond closely to the coefficient of variation (CV), although the difference will increase as p decreases.

Appendix B: Equivalence of 95th/5th centile ratios to coefficients of variation (assuming normal distributions).

Coefficients of variation	5	10	15	20	30	40	50
95th/5th centile, assuming normality	1.18	1.40	1.67	2.00	3.01	5.03	11.16

- 1 The Records and Statistical Unit. "Disease labels". *J Coll Gen Pract* 1963; 6: 197-204.
- 2 Office of Population Censuses and Surveys, The Royal College of General Practitioners, Department of Health and Social Security. *Morbidity statistics from general practice—second national study 1970-71*. Studies on Medical and Population Subjects No 26. London: HMSO, 1974.
- 3 Office of Population Censuses and Surveys, Royal College of General Practitioners, Department of Health and Social Security. *Morbidity statistics from general practice 1971-2—second national study*. Studies on Medical and Population Subjects No 36. London: HMSO, 1979.
- 4 McPherson K, Wennberg JE, Hovind O, Clifford P. Small area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. *N Engl J Med* 1982; 307: 1310-14.
- 5 Verhaak PFM. *Variations in the diagnosis of psychosocial disorders: a general practice observation*. Utrecht, The Netherlands: Institute of Primary Health Care, 1986.
- 6 Anderson JE. The utility of routine encounter reporting in primary care research. In: Kilpatrick SJ, Boyle RM, eds. *Encounter records and the denominator problem*. New York: Praeger, 1984.
- 7 Lamberts H, Wood M, eds. *International classification of primary care*. Oxford: Oxford University Press, 1987.
- 8 Bentsen BG. The accuracy of recording patient problems in family practice. *J Med Educ* 1976; 51: 311-6.
- 9 Research Committee of College of General Practitioners. The continuing observation and recording of morbidity. *J Coll Gen Pract* 1958; 1: 107-28.
- 10 McPherson K. Why do variations occur? In: Andersen TF, Mooney G, eds. *The challenge of medical practice variations*. London: Macmillan, 1990.
- 11 Jones R, Lydeard S. Prevalence of symptoms of dyspepsia in the community. *BMJ* 1989; 298: 30-2.
- 12 Metcalfe D, Hodgkin P. Clustering diagnoses: a method of interpreting morbidity data. *Fam Pract* 1984; 1: 228-31.
- 13 Crombie DL, Fleming DM. *Practice activity analysis*. Occasional paper 41. London: Royal College of General Practitioners, 1998.
- 14 Crombie DL. The problem of variability in general practitioner activities. In: *Yearbook of research and development*, 1989. London: HMSO, 1990: 21-4.
- 15 Dunn G. Records of psychiatric morbidity in general practice: the National Morbidity Surveys. *Psychol Med* 1985; 15: 223-6.
- 16 Jenkins R, Smeeton N, Shepherd M. *Classification of mental disorders in primary care*. *Psychol Med* monograph supplement, No 12. Cambridge: Cambridge University Press, 1998.
- 17 Cohen H. The nature, methods and purpose of diagnosis. *Lancet*, 1943; i: 23-7.
- 18 Cochran WG. Analysis of variance for percentages based on unequal numbers. *J Am Stat Assoc* 1943; 38: 287-301.

Cancer Prevention Fellowship Programme

The National Cancer Institute programme provides an opportunity for training in cancer prevention and control. The programme offers Master of Public Health training during the first year at accredited universities. This training is followed by independent research opportunities within the Division of Cancer Prevention and Control, NCI, located in Bethesda and Rockville, Maryland. The programme is for three years with the MPH option; up to three without.

Applications for the programme are due on September 1, 1992. Fellows begin on July 1, 1993.

Further information from Dr Douglas L Weed, Director, or Mrs Barbara Redding, Cancer Prevention Fellowship Program, Division of Cancer Prevention and Control, National Institutes of Health, National Cancer Institute, Bethesda, MD 20892, USA; tel 301 496 8640; fax 301 402 0816