

The Analysis of Family Practice Workloads by Seriousness

R.C. Westbury, MA, MB, BChir (Cantab)
Calgary, Alberta

There is a need for a measure of the overall seriousness of a given family practice workload. In the past, such measurements have been attempted in various ways.

In this work, the rubrics of the Canuck Book Classification (a classification of health problems in family practice) were rated for seriousness. There was sufficient agreement on 96 rubrics for them to be used as indicators of the seriousness of workloads in general.

Some examples of the uses of the system are shown. Several difficulties were encountered; these are not insuperable, and the method deserves to be developed further.

Any effort to understand the mechanics of family medicine must focus on the doctor-patient contact. It was logical that attempts to quantify the workload of family doctors should begin by counting these encounters. However, it soon became apparent that not all encounters were equal; in order to quantify practice workloads and be able to compare one practice to another, it was necessary to subdivide doctor-patient contacts according to various criteria. The location of the encounter (office, hospital, house call, etc) is clearly important. The age and sex distribution of the patients seen makes a considerable difference when comparing workloads.

It is evident also that the content of each individual encounter must influ-

ence workload. European studies have tended to be based on the "one problem per visit" model. By contrast, there has been a tendency in North American work to emphasize the universal truth that patients often bring several quite separate problems to the family physician. From this model has arisen the concept of the doctor-problem contact; each doctor-patient contact must contain at least one and may involve many doctor-problem contacts. (In our practice, seven was the maximum number of doctor-problem contacts per doctor-patient contact and the average was 1.24; the corresponding average given by Bentsen¹ was 2.54 problems per encounter.)

But not all doctor-problem contacts are equal either; they differ in terms of the amount of time they take, and in the amount of stress and worry they bring to the patient and his family doctor.

It is to the latter consideration that this paper is directed; doctors think of some problems as being more serious and some as being less serious. The more serious ones require more diligence and more effort from the doctor. Coping with these generally demands more emotionally and intellectually. How can we measure the level of seriousness in our daily workload, so that we can give statistical consideration to the quality as well as the quantity of our patient contacts?

Background

The problem can be approached in two ways.

(1) *On-the-spot judgment.* Here the doctor, as he records the visit, assigns it to a certain category of seriousness. Or, he can classify each separate problem with which he copes at each visit according to its perceived seriousness. A differential count is then made to assess the seriousness of his total workload.

(2) *Rubric rating.* Here the rubrics of a diagnostic classification are rated according to their seriousness. The doctor simply records the problems with which he copes at each patient encounter, using the rated classification. The seriousness of the workload is then calculated, using an agreed upon formula, from the seriousness of the individual doctor-problem contacts which it comprises.

Each of these methods has advan-

Requests for reprints should be addressed to Dr. R.C. Westbury, Cambrian Medical Clinic, 41 Chelsea Street, Calgary, Alberta, T2K 1P1.

tages and disadvantages. On-the-spot evaluation is highly specific, whereas the rubric rating approach involves considerable generalization. The first method maintains its accuracy even when applied to small numbers of contacts. On the other hand, the rubric rating technique depends on the diagnosis, which is a relatively unemotional judgment; assigning seriousness directly to a visit depends on how the recorder felt in general that day, his state of mind at the moment of recording, and, in particular, how he felt subjectively about the patient whom he had just seen. Finally, the first method involves having recorders trained to record seriousness according to a uniform and clearly defined protocol; the second method allows a retrospective assessment to be made of the seriousness of any workload which has been recorded according to any accepted classification of disease problems. Such workload records are becoming increasingly common all over the world; many opportunities for interesting comparisons are beginning to appear. A tool to compare seriousness is urgently needed.

There have been many surveys which have used the first approach: the results of eight of these have been tabulated for easy comparison in *Present State and Future Needs of General Practice*.² The researchers cited therein classified problems as "major," "minor," or "chronic." Minor problems accounted for 51 to 75 percent of the workloads. Major problems showed less variation; five of the eight recorders reported from 16 to 18 percent of new diagnoses as major. The definitions of major and minor varied from study to study.

The protocols of several of these studies showed a surprising tendency to make ex cathedra generalizations about the seriousness of various groups of diseases, thus sacrificing the flexibility which is the great advantage of the on-the-spot approach. The work of Backett et al,³ was an example of this: "Such conditions as pulmonary tuberculosis, cancer, established peptic ulcer, and pneumonia were always classified as serious." Many family physicians would find it hard to consider a positive Mantoux test, a carcinoma-in-situ, duodenal ulcer, or viral pneumonia as *always*, or even generally, especially serious.

In Canada, Steele, Kraus, and

Smith⁴ divided 891 encounters into "Trivial," "Non-Serious," and "Serious," representing 9.2, 61.8, and 28.9 percent of the workload, respectively.

Several of the classic surveys of the National Health Service ran into difficulties over the evaluation of seriousness in general practice workloads. The reports by Hadfield⁵ and Gemmil⁶ specifically mentioned the wide variations in what general practitioners classed as serious or trivial. The General Practitioners Association (GPA) Study Group⁷ pointed out that symptoms, the stalwarts of general practice diagnosis, are especially hard to assess for seriousness.

The major work to study the rubric rating approach was that of Wyler, Masuda, and Holmes.⁸ They selected, by a rather complicated series of choices and rejections, a list of 126 problems. The items were ranked for seriousness by comparing them to the seriousness of peptic ulcer on a scale of approximately 1,000 points. The ranking was performed by 141 lay people and 117 doctors. The final rank order was calculated from the geometric means of the individual ranks. This study is a model of beautiful design and careful method. It emphasizes the degree of agreement between the doctors and laymen in assessing disease seriousness. Unfortunately, it has several severe defects from the point of view of its practical application in the front line of medicine: (1) the terms are not from a classification which is used in general practice, (2) many terms could not be readily used by family physicians as they had been rendered into "laymen's language," and are, therefore, rather too vague and diffuse, and (3) the rating system is too broadly based and fails to give sufficient weight to the ranking by the physician respondents, especially family physicians who are exposed to the widest range of human problems. It is important to bear in mind the distinction between severity, how the patient subjectively and functionally considers his illness, and seriousness, how outsiders (especially doctors) assess the problem. Attempts are being made to unite the two concepts.⁹

Several Canadian studies have struggled with the rubric rating approach. The Economic Planning Branch of the Department of Treasury and Economics of the Province of Ontario (1971 - Private Communication)

prepared a similar ranking of 38 morbidity states, and has plans to expand the range of its "Index of Health." The Committee on Educational Objectives of the Canadian College of Family Physicians¹⁰ devised an ingenious system for weighting diagnostic terms according to their incidence, seriousness, and treatability. They point out that, initially, different scores for each of these three variables were assigned to each diagnostic term by each of the nine members of the Committee. However, they were able to resolve these differences of opinion by discussion.

Method

Of the two methods available for assessing seriousness, I chose the rubric rating approach because of its greater reliability and its ease of application.

Thirty-five family physicians (including my partner, Tarrant, and myself) were invited to participate in the study. All had an interest in teaching and/or research. We were presented with 274 rubrics of the 389 rubrics in the "Canuck Book."¹¹ This was an interim classification, created for use in Canadian general practice until an internationally acceptable classification could be agreed upon. From 1971 to 1975, it was found to be acceptable and was widely used by GP recorders in Canada. It shows a high level of concordance with the *International Classification of Health Problems in Primary Care (ICHPPC)*,¹² which is now coming into general use. One hundred and fifteen rubrics (29.6 percent) were eliminated because they embraced so many variegated conditions that they could not possibly be assigned a degree of seriousness.

In the preface, the theoretical background and objectives were briefly explained. The instructions were: "*Go through the list and opposite each rubric indicate your opinion as to the seriousness of that condition by circling the numbers according to this code: 0 - usually not serious, 2 - usually serious, and 1 - somewhere in between.*" (*Opposite the diagnostic description of each category, the numbers 0 . . . 1 . . . 2 had been printed to allow one of these to be circled.*)

The philosophy of Wyler, Masuda, and Holmes⁸ was used in the instructions for the respondents. Two points should be noted: (1) the respondents

Table 1. Distribution of Doctor-Patient Contacts by Seriousness*

	Office		Home		Acute Hospital		Chronic Hospital	
	#	%	#	%	#	%	#	%
Dr. T.								
Non-serious	3,720	17.7	482	24.3	30	3.1	1	1.4
Grey zone	16,391	78.2	1,361	68.7	727	75.8	57	80.3
Serious	853	4.1	139	7.0	202	21.1	13	18.3
Index of Seriousness	0.229		0.288		6.733		(13.000)	
Overall Index	0.285							
Dr. W.								
Non-serious	3,777	14.8	870	23.8	34	3.6	5	2.4
Grey zone	20,154	79.2	2,592	71.0	685	72.7	178	84.4
Serious	1,520	6.0	187	5.1	223	23.7	28	13.3
Index of Seriousness	0.402		0.215		6.559		(5.600)	
Overall Index	0.418							

*For Drs. T. and W., the distribution of doctor-patient contacts by seriousness in the five-year period for various locations, and the Index of Seriousness for each location, as well as the Overall Index of Seriousness for each doctor.

were asked to *generalize* according to their experience in their practice. Many said that this was no easy task for physicians dedicated to considering each patient as an individual. Two respondents returned blank questionnaires because of this objection. (2) Deliberately, no exact instructions were offered for the *basis* on which seriousness was to be considered. I was forcefully attacked in both the written and spoken word for this deficiency. These attacks had been clearly anticipated, and were suffered with carefully rehearsed stoicism. To understand the reason for this, the reader is asked to momentarily adopt each of the two extreme positions: if one were to consider only the viewpoint of the individual patient, every rubric could be called serious. Few patients would consciously waste their time, or that of their doctor, with problems which they consider to be of no importance. At the other extreme, if the definition of "serious" were "the potential

killers" (Fry¹³ used this definition), there would be no need to go further than the nearest pathology book, but the concept of seriousness would lose many of the finer nuances of family medicine. For the instrument to be useful, it must represent the way that first-contact physicians amalgamate these extreme viewpoints in their daily professional lives. The importance of not offering directions to the respondents was emphasized by the interesting observation that there was *only one* rubric which *every* respondent classified as "non-serious": that was "Wax in the Ear." This problem was used in the preamble, as an example of how, under special circumstances, even the most trivial problems could be serious.

Of the 35 forms distributed, 20 forms were returned. Five were eliminated (two were too late for analysis, one was incomplete, and two others were returned blank), leaving 15 for analysis.

Results

The Choice of Indicator Rubrics

For each rubric, a note was made of the number of doctors who assigned it to each of the three categories of seriousness. If ten, or more, of the 15 doctors had assigned a rubric to either Code 0 or Code 2, then that rubric was categorized as "non-serious" or "serious," respectively. Calculations based on the Multinomial Distribution show that the chances of this level of agreement occurring by a purely random process are only 1.7 percent.

In this way, I developed a list of 53 "serious" and 43 "non-serious" indicator rubrics. These represent 96 (35 percent) of the 274 rubrics presented for rating, and 14 and 11 percent, respectively, of the 389 rubrics in the "Canuck Book."

Some Practical Applications

The many uses of the technique can be appreciated from an examination of

Table 2. Relation of Age and Sex to Seriousness*

Age-sex group		Serious		Non-serious		Index of Seriousness
		#	%	#	%	
<1	Male	9	1.3	162	23.2	(.055)
	Female	8	1.4	105	18.2	(.076)
	Total	17	1.3	267	20.9	(.064)
1-11	Male	149	2.1	1,590	22.3	.094
	Female	96	1.6	1,540	25.8	.062
	Total	245	1.9	3,130	23.9	.078
12-20	Male	195	3.9	929	18.7	.210
	Female	176	3.3	1,081	20.5	.163
	Total	371	3.6	2,010	19.6	.184
21-64	Male	643	6.4	1,406	14.0	.457
	Female	1,243	7.4	1,963	11.6	.633
	Total	1,886	7.0	3,369	12.5	.560
65+	Male	278	24.2	74	6.4	3.757
	Female	368	23.6	69	4.4	5.333
	Total	646	23.9	143	5.3	4.517
Total	Male	1,274	5.3	4,161	17.3	.306
	Female	1,891	6.2	4,758	15.7	.397
	Total	3,165	5.8	8,919	16.4	.355

*The seriousness of doctor-patient contacts by age and sex, showing (for each age and sex group) (1) the number of "serious" contacts (2) this figure expressed as a percentage of all doctor-patient contacts in that group (3) the number of "non-serious" contacts (4) this figure expressed as a percentage of all doctor-patient contacts in that group (5) the Index of Seriousness. The figures are for all doctor-patient contacts made by Drs. T. & W. over a five-year period.

Tables 1, 2, and 3 which are derived from records kept of each doctor-patient contact (and each doctor-problem contact) in the practice of Tarrant and myself, for the five years from 1967 to 1971. The details of each contact were written onto a card; the data was later key-punched into machine-readable form and analyzed by computer.

The first Table shows differences in the seriousness of the doctor-patient contacts made by my partner and me in various locations (a doctor-patient contact is considered to be as serious as the most serious of the doctor-problem contacts which it includes). The numbers of "serious" and "non-serious" problems can be compared readily in each cell of Table 1.

However, it is often useful to have a "shorthand" summary of the seriousness of workloads: for this purpose, an "index of seriousness" can be calculated by dividing the number of "serious" contacts by the number of "non-serious" contacts.

Table 2 demonstrates clearly the differences in seriousness of doctor-patient contacts grouped by the age and sex of the patient. The increasing seriousness of morbidity with increasing age is intuitively appreciated by all family doctors, and, indeed, is tangibly recognized by the age-weighting of the capitation fees paid to British family physicians. The tendency for adult females to present with more serious problems than adult males is not so well known.

Finally, the third Table demonstrates some small, but interesting and significant, differences in the seriousness of our workload for various days of the week. The workload for Wednesday and Thursday tends to be more serious, mainly because serious problems are relatively over-represented. Friday, a busier day generally, tends to have a less serious workload with serious problems under-represented and non-serious problems over-represented. The contacts at weekends are not strictly comparable because a large proportion of the contacts are house calls.

Discussion

While the technique described has proved to be useful for various analyses within our practice (such as those just referred to), its potential for inter-practice comparisons has not been achieved because it relates to an obsolete classification. It would be most useful if a similar weighting for seriousness were to be developed for ICHPPC, a classification which is likely to be used very widely. Before this is done, it would be worthwhile to consider some difficulties which have been brought into focus by this work.

The number of physicians used to rate seriousness in this study was too small to be representative; furthermore, they were a selected group. Ideally, as ICHPPC is an international classification, opinions about seriousness should be obtained under identical test conditions from many doctors in many countries. This would be an interesting (but difficult) exercise in itself.

The small number of the raters was to some extent offset by the strictness of the criteria for agreement: only those rubrics on which two thirds of the doctors could agree were used. The level of agreement was thus comfortably high, but the number of indicator rubrics so obtained was correspondingly low.

The "grey zone" (eg, rubrics which are neither "serious" nor "non-serious" — Code 1) was extremely large: this is the universal difficulty posed by indicators or indices, as Moroney¹⁴ has stressed. The assumption behind all indices is that the indicator values reflect, in general terms, the unaccessible data in the "grey zone." This, by definition,

Table 3. Relation of Day of Week to Seriousness*

	Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Sunday	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Non-serious	1,630	15.9	1,693	16.1	1,438	15.5	1,430	15.8	1,906	17.0	594	21.4	228	19.6
	8,085	78.8	8,219	78.2	7,251	78.3	7,025	77.8	8,724	77.7	2,016	72.7	825	71.0
Serious	543	5.3	601	5.7	576	6.2	574	6.3	601	5.3	161	5.8	109	9.4
Index of Seriousness	.333		.355		.400		.401		.315		.271		.478	

*The seriousness of the total workload by day of the week for both doctors over the whole five years: (1) the number and percentage of non-serious doctor-patient contacts (DPCs) (2) the number and percentage of serious contacts (3) the Index of Seriousness for each day of the week. (Considering weekdays only and comparing to the distribution of total DPCs on the five days:

for non-serious DPCs — Chi Square Goodness of Fit = 7.89 (4DF), $p < .1$
 for serious DPCs — Chi Square Goodness of Fit = 16.13 (4DF), $p < .005$.)

cannot be proved; it can only be checked empirically.

The technique was not reliable for small workloads: in the "chronic hospital" column of Table 1 and in the cells for infants in Table 2, it can be seen that the addition of one or two serious problems would have radically changed the outcome.

It should be possible to counter many of these difficulties in the future by adopting a ranking technique, and converting the results to graded categories of seriousness. This would expand the number of indicator rubrics and reduce the size of the "grey zone." However, it must be emphasized that it could not eliminate it altogether because of the impossibility of ranking residual categories (eg, "all other diseases of the ... system").

In other words, some of the doubts and difficulties associated with indices will always remain. Before we let this become too discouraging, it is worth

remembering that the Dow Jones Industrial Index represents only 30 percent of the value of shares traded on the New York stock exchange: a computational tool need not be perfect to prove extremely useful.

Acknowledgement

Technical assistance and epidemiologic consultation were provided by Dr. Ed Love.

References

1. Bentsen BG: Direct Medical Audit and Recordkeeping. London, Ontario, University of Western Ontario, 1975
2. Royal College of General Practitioners: Present State and Future Needs of General Practice, ed 3. London, RCGP, 1973
3. Backett EM, Heady JA, Evans JCG: Studies of a general practice (II): The doctor's job in an urban area. Br Med J 1:109-115, 1954

4. Steele R, Kraus AS, Smith PM: Doctor/patient contacts in family practice — an exploratory study. Can Fam Physician 14(9):45-55, 1968

5. Hadfield SJ: A field survey of general practice, 1951-1952. Br Med J 2:683-706, 1953

6. Gemmil PF: Britain's Search for Health. Philadelphia, University of Pennsylvania Press, 1960

7. GPA Study Group: Abuse and Misuse. London, General Practitioners Assoc, 1965

8. Wyler AR, Masuda M, Holmes TH: Seriousness of illness rating scale. J Psychosom Res 11:363-374, 1968

9. Akpom CA, Katz S, Densen PM: Methods of classifying disability and severity of illness in ambulatory care patients. Med Care 11(2):125-131, 1973

10. Committee on Education Objectives: The Canadian Family Doctor, Part II. Toronto, College of Family Physicians of Canada, 1973

11. Classification Committee of CFPC: Canuck Book. Calgary, College of Family Physicians of Canada, 1971

12. Classification Committee of the World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA): The international classification of health problems in primary care. Continuing Education for the Family Physician 3(5):31-37, 1975

13. Fry J: Profiles of Disease. Edinburgh, E & S Livingstone, 1966, p 9

14. Moroney MJ: Facts from Figures, ed 2. London, Penguin Books, 1954, pp 48-54