UDC 681.3 : 616 (047)

Some Sources and Potential Consequences of Errors in Medical Data Recording*)

(From the Health Law Project, University of Pennsylvania, and the Department of Anatomic Pathology, Philadelphia General Hospital, Philadelphia)

L. Hendrickson, J. Myers

One year's patient data from an urban teaching hospital was retrieved from the computer tapes of a large abstracting service. A variable but significant error rate was found in the recording of each entry studied. Such errors caused the data to be unsuitable for a quality of care study since the error rate was either very high or there was evidence of a recording bias. In either case, chart review would have been required and thus the data would have been, at best, only a guide to selecting parameters to study in a chart review.

The »information« obtained was so obviously inaccurate that it is unlikely that anyone would consider using it. However, in-hospital monitoring of data summaries received from the abstracting service might well lead to the elimination of obvious errors while retaining plausible but still inaccurate data. The acceptance of such could lead to serious false conclusions. For example, our »data« suggested that ward patients and non-ward patients received a different quality of care, a result not confirmed by chart review.

Given the system used, it was concluded that possible errors in data recording could be recognized most efficiently through »check« programs at the abstracting service and that by using such programs, the abstracting service could and should initiate improvements in hospital practices and coder training to minimize errors.

Key-Words: Medical Data Recording, Computer-stored Patient Abstracts, Error Sources

FEHLERQUELLEN BEI DER MEDIZINISCHEN BEFUNDDOKUMENTATION

Aus den Magnetbändern eines großen Krankenblatt-Bearbeitungsdienstes wurde ein Jahrgang von Patientendaten eines städtischen Lehrkrankenhauses näher analysiert. Bei allen untersuchten Sachverhalten fand sich eine variable, aber erhebliche Fehlerquote. Diese Fehler ließen die Daten für ein Studium der Behandlungsqualität ungeeignet erscheinen, da entweder die Fehlerquote recht hoch war oder ein offensichtlicher Bias bei der Datenerfassung mitspielte. In jedem Falle wäre eine Überprüfung der Ablochbelege erforderlich gewesen; die Daten hätten daher bestenfalls als Hinweis dafür dienen können, welche Auswahlparameter bei einer solchen Nachprüfung zu untersuchen wären.

Die erhaltene »Information« war augenscheinlich so ungenau, daß wohl niemand davon Gebrauch machen würde. Die klinikinterne Überprüfung der von dem Bearbeitungsdienst gelieferten Datenübersichten könnte allerdings zur Ausmerzung augenfälliger Fehler führen, wobei zwar plausibel erscheinende, aber immer noch ungenaue Daten resultieren würden. Die Annahme solcher Daten könnte zu gefährlichen Fehlschlüssen führen. Die eigenen Daten der Autoren ließen beispielsweise vermuten, daß bezüglich der Güte der Behandlung ein deutlicher Unterschied bei ambulanten und stationären Patienten bestand, wofür sich aber bei einer Nachprüfung kein Anhalt fand.

Mögliche Fehler der Datenverarbeitung könnten in dem vorgegebenen System durch Fehlerprogramme erkannt werden. Bei Durchführung solcher Programme könnte und sollte der Verarbeitungsdienst Verbesserungen bei der Datenerfassung im Krankenhaus und der Verschlüsselungsroutine im Interesse einer Fehlerreduktion anregen.

Schlüssel-Wörter: Medizinische Befunddokumentation, Computer-gespeicherte Patienten-Daten, Fehlerquellen

Introduction

Despite the voluminous literature on computerized medical data processing and hospital information systems, there has been insufficient stress on the problem of the reliability of the data. Much of the extensive literature on the unreliability of medical and other data is in a context separate from computer applications. A list of the relevant literature was compiled in 1964 [5] as an aid and reminder to workers in the field but »it is almost not perceived by most of our colleagues« [7]. Furthermore, it is our opinion that too little attention has been paid to what Wagner [7] calls »formal errors« as opposed to »factual errors«. By this he means errors that could be detected without any knowledge of the true state of affairs, such as a non-existent date or an incompatibility between entries — e.g., a missed abortion in a male.

Most people are aware of the high frequency of factual errors, and a physician on receiving an unlikely laboratory result would probably request a repeat examination before taking any other action. In addition, the significance of factual errors is probably not much greater in computerized information systems than it was before their advent. Formal errors, on the other hand, are more significant in computerized information systems for several reasons. Converting a record into computer-readable form (coding) adds another step in which an error can be introduced. The ease with which the computer can perform complex analyses encourages such use and increases the possibility that the results will be accepted as reflections of the truth. The acceptance of the results is furthered by the ability of the computer to analyze the data without anyone seeing the individual data elements, including those with obvious

*) Reprint requests to Dr. L. Hendrickson, 133 S. 36th St., Room 310, Philadelphia 19104.

formal errors, and to print out the results in impressive fashion. Such errors are not only more dangerous in computerized applications but are less excusable since the checking of formal errors is done better by the computer than by man [7]. Computers can also detect invalid patterns of data where the individual elements are plausible [2], and possible but implausible data aggregates. It must not be assumed that errors will not lead to false statistical conclusions when they are either random or infrequent. Large random errors and small non-random errors can lead to false conclusions [1, 3, 6].

We recently encountered an example of false conclusions from erroneous data in a most striking manner since it seemed to confirm the rather widely held impression that ward patients receive an inferior quality of medical care. As part of an evaluation program of a large urban teaching hospital performed by the Health Law Project of the University of Pennsylvania, we tried to determine whether there were significant differences in the quality of care received by ward and non-ward patients. The hospital subscribed to a data abstracting service and thus, abstracts of patient charts were computer retrievable. Unfortunately, our retrievals yielded »data« incompatible with the actual practices of the hospital and which, on statistical analysis, implied a difference between ward and non-ward practices. We investigated all of the routine laboratory studies, categories where the indicated frequency of entry was much lower than expected, and categories where the frequency of the entry was significantly different for ward and non-ward patients. The sources of the errors were determined where possible and the incidence of error estimated. We found a high correlation between categories with a high error rate and categories with a significant difference in the ward and non-ward frequency, suggesting that the differences were not real but a result of the high error rate.

The Data Processing Service

The Professional Activity Study (PAS) is the oldest, largest, and most comprehensive of all hospital abstracting services. It was developed by the Commission on Professional and Hospital Activities (CPHA) and is used in over 1700 hospitals in the United States and Canada. A standard abstract for each patient chart is prepared after discharge and these abstract forms are sent to CPHA where they are placed on computer tapes. Over seventy million abstracts have been sent to CPHA thus making it the largest single repository of patient information in the United States.

PAS, in common with other abstracting and data retrieval services, collects a minimum basic data set which includes patient characteristics such as age, sex, race, length of stay, diagnosis, attending physician, operation, and surgeon. In addition PAS routinely lists over 150 other variables, including routine admission and follow-up laboratory tests and x-rays, blood pressure, temperature, drugs and other types of therapy. Once the data are entered in computer readable form they are available to the hospital on magnetic tapes. CPHA summarizes the information and provides participating hospitals with monthly and semi-annual reports. CPHA also uses the data in the *Hospital Records Study**) and the *PAS Reporter***) in which there is considerable description and analysis of the pooled data.

*) Published quarterly by Lea Inc., Ambler, Pennsylvania.

**) Published weekly by the Commission on Professional and Hospital Activities, Ann Arbor, Michigan.

Materials and Methods

We obtained the PAS data tape containing abstracts of over 16,000 in-patient charts for the year 1969. We had the computer print out the over-all incidence of each entry (test, treatment, examination, etc. — see Figure 1) and the incidence of each entry for all ward and all non-ward (private and semi-private) patients. We obtained a similar listing for entries that we considered relevant in each of 46 disease categories. When it became apparent that the error rate was high we examined charts in order to estimate the difference between the actual and recorded incidence of the entry. Because of the volume, we restricted our investigation to routine laboratory tests, entries whose incidence was obviously in error, and entries with a significant ward, non-ward difference. We then tried to determine why certain entries were misrecorded so frequently. Since we were investigating the accuracy of recording charted data on abstract forms almost all of the errors could be ascribed to errors on the part of the people charged with coding (coders). However, investigation of charting practices, analysis of PAS instructions and examination of the PAS system in conjunction with coder interviews revealed numerous areas in which the error rate could have been minimized. The analysis also indicated how »check« programs could be incorporated into the PAS data processing to detect patterns of misrecording by coders. Thus responsible people in participating hospitals could be alerted to areas of weakness and abstracts could be returned for recoding before entering the data pool.

Results

Analysis of Routine Laboratory Studies (Table 1)

The hemoglobin determinations as recorded appear reasonable in view of the number of admissions and the number of determinations reported by the laboratory. The minimum number of determinations is close to the number done in the laboratory; the remainder could well represent repeat determinations.

The hematocrit figures suggest that the .1% rate is an accurate overall rate of hematocrit recognition, the admission figures resulting almost entirely from recognition of hemoglobin determinations under the category of »admission hemoglobin or hematocrit«. Differential white counts also appear to go unrecorded although total white blood cell counts are recorded with substantial frequency.

The bacteriology and stool results suggested that stool for blood and stool analysis were almost always checked in tandem as were bacterial smear and culture. We were virtually certain this was so since the figures by patient classification in most disease categories were identical.

The low count of antibiograms (antibiotic sensitivity testing) would appear to result from a defect in planning the system. Anyone familiar with a bacteriology laboratory would expect the number of antibiograms performed to be almost equal to the number of cultures. On the other hand, a laboratory slip cannot indicate sensitivities without bacterial growth. We concluded that this category may give some information if it is regarded as the number of cultures in which at least one organism grew. Differential white blood cell counts were not recorded because the lab slip did not contain the word »differential«.

Despite seemingly plausible figures for hemoglobin and white blood cell counts, we suspected recording inaccuracies when we obtained the incidence of these tests for the ward and non-ward patients (see Table 4, below).

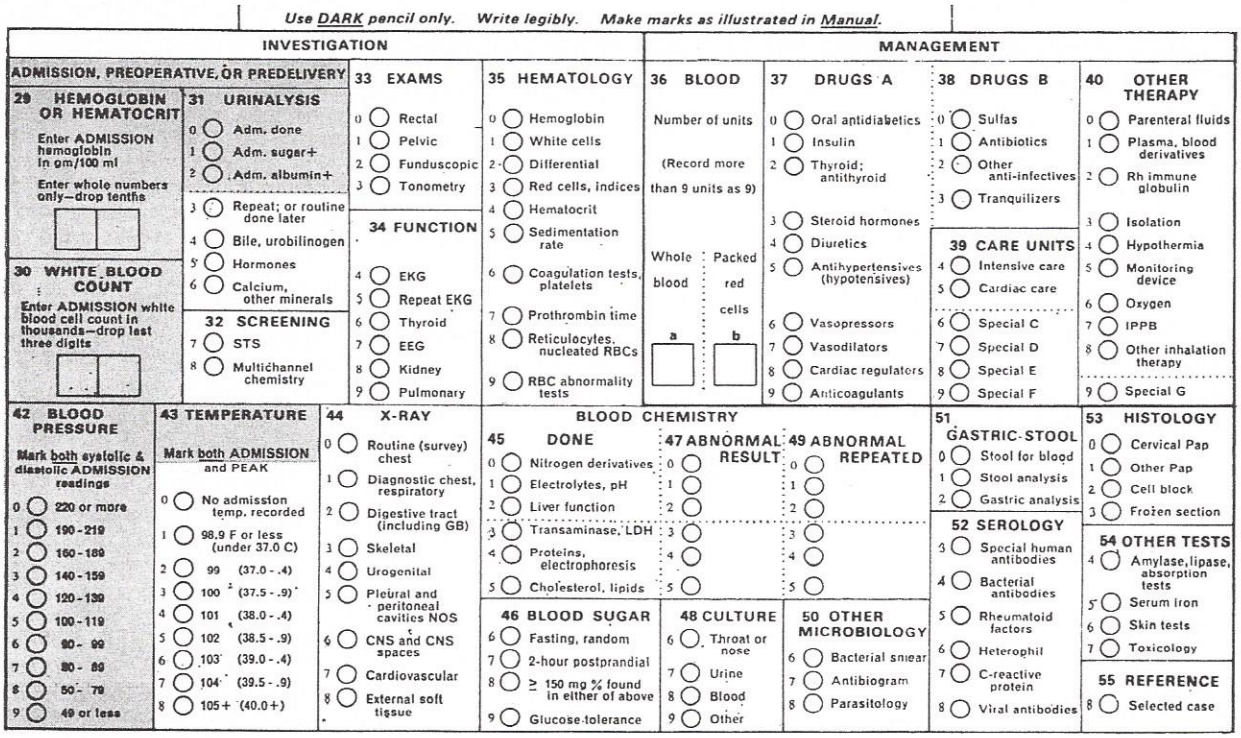

Figure 1: Part of a form on which a single case is abstracted. Our analysis did not distinguish between numeric data (squares) and »yes-no« data (circles). Any value for an admission hemoglobin or hematocrit (entry 29) would indicate that the test was done. Where numeric data are indicated by »yes-no« checking (entry 42-0 to 42-9), we used the information in the analysis

Since we would be concerned with whether a test had been done rather than with when, we ran another retrieval by ward and non-ward for hemoglobin, white blood cell count and urinalysis in the following manner: »no test«, »admission only«, »done later only«, »admission and repeat« (Table 2). This breakdown did not explain the apparent anti-ward bias which, if anything, appears more marked than before. However, it does suggest the possibility of a recording bias rather than a quality of care bias since the pattern is so unlikely. We investigated further to determine if recording practices influenced these data and what might have directed such recording practices. We tried to account for all of the following:

1.) The number of hemoglobin tests recorded in the PAS file was probably in error because of the high percentage of »no tests« (22%).

2.) There was a zero incidence of »admission only« hemoglobin tests and white blood cell counts recorded.

3.) There was an improbably low incidence of white blood cell counts (0.4%) and urinalysis (1%) done after but not on admission.

4.) There was a consistent anti-ward bias.

After coder interviews, re-reading PAS instructions and examining charts from cases without recorded hemoglobin tests and urinalyses we concluded that non-random recording errors would best account for the findings. The first test spotted by the coders on reviewing the chart would be entered as an admission test. Hemoglobin determinations and white blood cell counts thus entered would be entered again under »repeat or done later«. This was technically admissable since the PAS instructions state that it is »not necessary« to check these items under hematology if an admission test is checked. This option

Table 1: Laboratory Tests Performed — All In-Patients

Test	Number in PAS File
Admission hemoglobin or hematocrit	9884 (61%)
Hemoglobin, repeat or done later	12508 (78%)
Minimum number of hemoglobin determinations *)	22392
Total in-patient hemoglobins (laboratory records)	25417
Hematocrit, repeat or done later	198 (0.1%)
Total in-patient hematocrits (laboratory records)	25082
Admission white blood cell count	9426 (58%)
White blood cell count, repeat or done later	9489 (59%)
Differential white blood cell count	6 (0%)
Admission urinalysis **)	10776 (67%)
Urinalysis, repeat or done later	4632 (29%)
Stool for blood	779 (5%)
Stool analysis	785 (5%)
Bacterial smear	3884 (24%)
Bacterial culture	3882 (24%)
Antibiogram	2396 (15%)

*) Assuming hemoglobins are recognized and hematocrits are not (see text).

**) Includes normal admission urine and admission albuminuria (see text).

Table 2: PAS Recording of Selected Laboratory Tests

	Ward	Non-ward	Private
Hemoglobin			
None recorded	2488 (29.5%)	1116 (14.5%)	
Admission only	0 (0.0%)	1 (0.0%)	
Done later only	1757 (20.9%)	868 (11.3%)	
Admission and repeat	4178 (49.6%)	5705 (74.2%)	
White Blood Cell Count			
None recorded	4470 (53.1%)	2153 (28.0%)	
Admission only	0 (0.0%)	1 (0.0%)	
Done later only	37 (0.4%)	27 (0.4%)	
Admission and repeat	3916 (46.5%)	5509 (71.6%)	
Urinalysis			
No normal admission, No later test	4575 (54.3%)	2225 (28.9%)	
No normal admission, Done later	180 (2.1%)	58 (0.8%)	
Normal admission No later test	1276 (15.1%)	3405 (44.3%)	
Normal admission Repeat test	2392 (28.4%)	2002 (26.0%)	
No normal admission-Total	4755	2283	310 *)
Admission albuminuria	1698	2278	315 *)

*) The number of patients with albuminuria cannot, of course, exceed the number of patients not having normal urines. These figures resulted from checking both normal and albuminuria for the same urine.

(where apparently none is intended) thus virtually eliminated any recording of only an admission test. Significantly, instructions for urinalyses did not contain a »not necessary« clause and a high proportion of charts are coded as admission urinalysis only. We could not account for the 16% of charts coded as hemoglobin »done later« only, except to guess that the lumping together of admission hemoglobins and hematocrits on the PAS forms had some influence on the recording. We dropped the subject when the coders became defensive since we tried throughout the study to avoid anything in the nature of the questions and in our manner of asking them which might appear critical.

The instructions for entering admission urinalyses were extremely complicated. If done and negative, »admission urinalysis done« was to be checked, whereas, if the sugar or albumin were positive, one or both were to be checked instead. Thus an unchecked admission urinalysis could mean either »not done« or »positive«. The coders sometimes marked both »done« and »positive« accounting for the apparent low number of no-tests in the non-ward group. We did not check the actual numbers of such duplications because of the computer costs. We were certain they occurred since among the private patients there were more positive albumins recorded than there were unchecked admission tests. The total positives recorded were further increased by inclusion of all trace readings as positives. PAS specifically states that this is optional. While the »illegal« coding is technically a coder error, the labelling of the forms and the instructions leave much to be desired. If, instead of »urinalysis-admission done«, the same space would have been labelled »admission urine negative for sugar and albumin« there would have been no way to misinterpret the instructions.

Evidence supporting a recording bias rather than a quality of care bias to account for the 2:1 ratio of »no tests« in the ward patients was obtained when we reviewed 30 charts selected at random from the group without recorded hemoglobin tests or urinalyses. Every chart was that of a ward obstetrical patient and each chart contained the laboratory tests*) in question. The tests were missed because they were done on the delivery floor upon arrival of the patient and recorded only in the notes and not on routine laboratory slips. Since there were approximately 2000 ward deliveries and 1000 non-ward deliveries, this accounted for a considerable number of missed entries. We could not exclude the possibility of a quality of care differences as well since the absence of non-ward charts from the randomly selected group might indicate a higher number of regular laboratory follow-up tests but, as in previous analyses, we concluded that any quality of care study would necessitate chart review. Interestingly, two categories, »urinalysis-repeat or done later« and »hemoglobin-repeat or done later«, which showed a significant pro-ward bias (Table 3) appear to be unlikely chance occurrences rather than quality of care differences or recording bias. This serves to demonstrate another potential danger in purely statistical analysis.

The sources of error in the recording of routine laboratory tests are legion. The coders were inadequately

*) Micro hematocrits done on the floor should have appeared under »admission hemoglobin or hematocrit«.

trained, did not follow PAS instructions and were careless in search techniques. The periodic summaries received by the hospital most likely were not reviewed by medically informed personnel. PAS instructions were needlessly ambiguous and the PAS computer programs lacked the simplest monitoring techniques to check the data received, accepting such data as 1 patient out of over 16,000 having only an admission hemoglobin and white count.

any items done within 48 hours of admission and indicated on the chart*). The remaining 6 were attributed to careless search. Coders stated that radiology reports reached charts after coding but in each case x-ray findings were included in notes. The probable incidence of skeletal x-rays in fractures is 100%.

The patients listed with eye diagnoses and no funduscopic examination provided rather odd findings.

Table 3: Selected Data Entries by Disease Category

Entry	Disease Category	Recorded Incidence	Incidence *) Estimated
Skeletal x-ray	Fracture	75% (145/192)	100%
Pelvic examination	Gynecologic	95% (960/1006)	
	Pregnancy complications	7% (43/648)	
	Abortions	80% (377/473)	
	Normal delivery	1% (17/1497)	
	Complicated delivery	3% (35/1374)	
	Puerperal complications	76% (22/29)	
	All obstetrical	12% (494/4021)	
	All OB-GYN	29% (1454/5027)	100%
	All female patients	21% (2258/10884)	> 46%
	All patients	14% (2258/16113)	> 31%
Funduscopic examination	Eye	10% (8/82)	59%
	Hypertension	56% (75/143)	97%
	All patients	12% (1953/16113)	
Antihypertensive drugs	Hypertension	16% (23/143)	
Diuretic drugs	Hypertension	25% (36/143)	
Hypotensive or diuretic drugs	Hypertension	37% (53/143)	57%
Frozen section	Breast	18% (23/126)	> 90%

*) See text

Analysis of Unusual Results by Disease Category

We expected the entries chosen for analysis in these disease categories to approach 100%. We reviewed charts only from the group not containing the entry in question, assuming that where an entry was both expected and present that it was accurate. The sample size varied with the number. A probable incidence was calculated by adding the number obtained from extrapolation of the sample error to the number of entries present. Coder interviews and chart analysis usually allowed us to determine why errors occurred.

The pelvic examinations were encoded when there was a gynecology consult or the word »pelvic« was written in the margin. The examination was missed when the findings were merely recorded in notes. Thus, the high incidence of missed pelvics. There were 5027 admissions to the OB-GYN Service and we used this figure to project an overall incidence of greater than 31% for pelvic examinations.

We reviewed 10 charts from the fracture-no skeletal x-ray group. Four were presumably missed because the x-rays were taken immediately prior to admission. The x-ray findings in each case were present in the notes and should have been marked (PAS instructions are to include

Thirteen charts were examined. One chart had »toxic goiter« and one had »Grave's disease« as the primary diagnosis. We could only guess that the charts were coded before the front sheets were completed and that the coders found the words »exophthalmic goiter« on the chart and fastened on the wrong word. Of the remaining 11, two were admissions for blepharoplastics. These did not have funduscopic examinations recorded, nor did one other chart. A single chart had the funduscopic results prior to admission and hence was not recorded because of the coder's misunderstanding of this point. The wording in the charts of 2 patients with cataracts made it impossible to state for certain whether or not an ophthalmoscope was employed. The remaining 5 charts were of patients with cataracts in which an examination was attempted and recorded as unsatisfactory because of the cataracts. Despite our reservations as to coder search accuracy we attributed these errors to inadequate coder training. We discarded the toxic goiters from our count, and retained the blepheroplastics and unclear charts as legitimate »no examinations« to project a probable incidence of 59%.

*) The coders misunderstood this point without exception. The practice was to include only post admission items.

Twenty-four charts from the hypertensive no funduscopic examination group were reviewed. Of these, 2 had no funduscopic examination listed. Two of the misses were attributed to inadequate coder training (patient had bilateral cataracts) and twenty to carelessness in search. The probable incidence of funduscopic examination is 97% in this category.

Five charts from the breast »no frozen section« category all had frozen sections. Frozen sections were routinely performed on specimens when requested. One of the charts had 5 separate notations of a frozen section. These misses were attributed to carelessness in search and a probable frozen section rate of at least 90% was assigned.

The analysis of drugs in hypertensive patients proved quite complex. Drugs with both diuretic and antihypertensive properties were coded as diuretics only (contrary to PAS instructions) because diuretics were listed first. We assumed that all diuretics in this group should have also been marked as antihypertensive agents and attributed the 30 charts listing a diuretic only to a failure to follow PAS instructions. We then investigated 30 of the 90 charts of patients who were recorded as having received neither a diuretic nor a hypotensive agent. Of 30 patients, 9 had received a drug which should have been coded as an antihypertensive agent according to PAS and as either antihypertensive or diuretic by the system admittedly employed by the coders (these misses were attributed to carelessness of search). Of the remaining 21, 4 should not have had hypertension as the primary diagnosis. We retained 2 of these in calculating a probable incidence of drug treatment since elevated blood pressures were recorded and eliminated two charts from normo-tensive patients from our calculations. Of the remaining 17 patients, 16 were mild hypertensives who were in a special program for treating mildly hypertensive, obese patients with diet alone. Hospitalization was only required for diet enforcement. These patients either responded to diet or were discharged on medication. The untreated patient died within a month. The probable incidence of antihypertensive drugs in this category is 57%.

Analysis of Ward — Non-Ward Differences

There were 54 PAS items in which the difference between all ward and all non-ward patients was significant (99% confidence limits). Thirty were numerically in favor of non-ward and 24 in favor of ward patients. By disease category (only relevant retrievals were obtained), there were 35 PAS items with significant differences, 6 numerically favoring non-ward and 29 favoring ward patients. We did not analyze most of these differences, since the differences fit in with our general impression and would not be useful for a quality of care study without extensive chart analysis in support. For example, more tranquilizers in the non-ward group and more bacterial cultures in the ward group probably reflect patient demands and house staff disregard of costs rather than real care differences. However, certain findings might well imply a difference. Nineteen such items showing a significant difference are listed in Table 4.

In analyzing these apparent differences we noted that 14 of the 19 categories showing significant differences are those in which we had previously found a high rate of incorrect entries. The corrected figures previously obtained lend themselves to all sorts of statistical manipulations which do not answer the question of real ward — non-ward care differences. For instance, a ward and non-ward frequency of pelvic examinations could be recalculated using a correction based on the difference in the error for each group in the category of gynecologic diseases. The new rate would then show a significant difference in favor of ward patients, reversing the bias but leading us no closer to the truth. We concluded that these differences could not be used in a quality of care study and that factors other than a difference in the actual frequencies could well be causing these unusual results.

We did not investigate the accuracy of the remaining items showing a ward — non-ward difference. Three of these favored ward and 2, non-ward patients. This and the nature of the studies suggested that even if accurate such figures would not be useful for a quality of care study. Furthermore, the error rate was so high in the data already investigated that it was apparent that no data could be used without checking it against the charts thereby negating the value of computerized data.

Comment

We found an extremely high rate of incorrect entries on investigation of one hospital's data for the year 1969. In addition we found the incidence and the nature of the errors to be such that the data were useless for the type of analysis for which they should be most useful. Since we may have investigated a hospital situation in which the accuracy of the data is worse than most, we should attempt to determine whether similar errors are likely to appear in other institutions and to propose remedies. Since coder deficiencies are the leading source or error, it might be assumed that the coders were uniquely unfit for the job and that a better choice of personnel would

Table 4: Ward — Non-Ward Differences

Disease Category	Entry	Percent with Entry	
		Non-ward	Ward
All Patients	White Blood Cell Count	72	47
	Urinalysis, done later	27	30
	Admission Urine	70	43
	Differential WBC	.03	.05
	Skeletal X-ray	11	8
	Funduscopic Examin.	14	11
	Antihypertensive Drugs *)	.8	.3
	Pelvic Examination	17	11
	Hemoglobin, done later	11	21
	Admission Hemoglobin	74	50
Malignant Neoplasms	Chest X-ray	24	43
	Gastrointestinal X-ray	15	26
Nutritional and Metabolic	Chest X-ray	5	58
	Gastrointestinal X-ray	67	17
	Radioactive Tracers	67	29
Eye	Funduscopic Examination	7	60
Miscellaneous Gastrointestinal	Hematocrit	0	10
Gynecological	Pelvic Examination	99	90
Obstetrical	Pelvic Examination	15	10

*) Ward had higher pressures and more in hypertensive range

have produced more accurate data. Theoretically, this is so, but the nature of the work is such that it is unlikely that people with the qualifications to approach theoretical accuracy would agree to serve as coders.

The coders we interviewed were mature and apparently trying hard but limited by an almost total ignorance of hospital procedures. Quite probably, they are typical of those who would be found in such work. Furthermore, adequate knowledge in itself would not guarantee accuracy without check programs. In a recent review of pathologic diagnosis codes entered by physicians at another institution, we found a 10 percent error rate.

Given the same personnel we would have eliminated the most obvious errors. One would expect a marked increase in accuracy if coders were told that »Hct« means »hematocrit«, if the word »differential« were added to laboratory slips, and if the coders were reminded of the instructions. Since review and revision was apparently lacking in the hospital we studied, it is unlikely that the overall error rate of hospitals contributing data to PAS is as high as the rate we found. The amount of significant information that would be added by physician review and corrective procedures is not so obvious. Some errors are easy to find and correct because the information is known. A computer is not needed to tell us that white counts are almost invariably accompanied by differential counts, that ladies with gynecologic diseases get pelvic examinations or that fractured bones are x-rayed. Had the coders been aware of this they would have produced more plausible data. However, plausible data introduce a real danger of obscuring more subtle errors. For example, the low figure for recorded pelvic examinations of gynecology patients led us to investigate and discover, that both gynecologic and non-gynecologic patients were credited with a pelvic examination only if the word »pelvic« was written in the margin of the notes or if there was a gynecologic consult sheet. Considering the difficulty we had in determining the causes of error and the time spent it is unlikely that in-hospital audits will provide anything better than superficially plausible inaccurate data. The recording of hemoglobins (Tables 1 & 2) where inconsistencies only became obvious on deep analysis is an example of the type of error that is unlikely to be recognized in routine in-hospital audits. The in-hospital audits should decrease the overall error rate by eliminating many glaring errors. Awareness of these should presumably lead to a change of procedure in the hospital, e.g., adding the word »differential« to laboratory slips. Under these circumstances, the most marked improvement will be in areas where the incorrect data were most likely to be discounted anyway. On the other hand the resulting improvement may disguise more subtle errors which can give rise to false conclusions. In this study, obvious inconsistencies discredited all of the data as a basis for evaluation of quality of care studies. For example, the anti-ward bias in recording pelvic examinations of obstetrical and gynecologic patients suggests a recording bias as an alternative to a quality of care bias. Without this obvious inconsistency it is quite possible that the difference in the ward — non-ward incidence of recorded pelvic examinations in patients with other than obstetric or gynecologic diagnoses would be taken as evidence of inferior practice on ward patients.

On the basis of our analysis thus far, it appears that, within the confines of the PAS system, major significant increases in accuracy must be initiated by PAS if they are to come about at all. While it is unrealistic to expect an error-free system it is not unrealistic to hope for a low and unbiased error rate providing data from which useful conclusions can be drawn. Beyond ambiguous instructions which are easily corrected much could be done. PAS could insist on giving standardized training to coders. Based on experience such training programs could be modified and hence, become increasingly more effective. This presupposes however, that the PAS procedures will somehow detect errors. One of us (LH) visited CPHA which produces the PAS system. While it appeared that reasonable procedures were used to minimize errors in the intake operations, no procedures had been incorporated into the program to evaluate the validity or plausibility of the information received. Checking for formal errors and to a lesser extent for implausible data is the »domain of the computer« [7]. The computer is the ideal instrument to compare norms based on reasonable practice with the data received and to do it for subsets of the hospital population such as disease category, sex, age, etc. This could be compared with other data based on pooled input to measure progress and as a means to locate persistent areas of weakness. The data from each hospital could be compared with the arbitrary norms and the pooled data from other hospitals. Persistent differences could be investigated and if they proved to be a true reflection of practice in a particular hospital could be handled appropriately. An example based on our findings will clarify this point. Among the norms used by the computer to analyze the data from each hospital we would include treatment with anti-hypertensive agents for every patient with blood pressures above a certain level or with the primary diagnosis of hypertension. Another reasonable instruction would be to tag any abstracts with the diagnosis of primary hypertension as a coding error if there is neither an elevated blood pressure reading nor an antihypertensive drug recorded. Such a program would have detected the abstracts with missed drugs and with miscoded diagnoses that we discovered in the file. These could have been returned for recoding and presumably lead to changes which would decrease the incidence of such errors. However, there would still be marked and persistent differences between the hospital we studied and the pooled data because of the special program for treating obese hypertensive patients with rest and diet. Once this was determined, a special code could be used to indicate these cases which could thenceforth be excluded from the data pool and in the comparison program. Extra spaces on the abstract for such indicators could be made available by eliminating such apparently useless categories as »bacterial smear« and »antibiogram«.

Both hospital data systems and regional medical data systems have great potential applications. Effective realization of this potential requires a valid data base. CPHA, in purporting to contain such a data base, has achieved a position of considerable importance as a source of information. For example, a recent editorial suggesting that data from CPHA or a comparable agency be used to insure high standards of physician performance states: »Strong and vigorous leadership in record room review is currently visible in the activities of only a few small agencies, the largest of which is the Commission on Professional Activities, Ann Arbor, Michigan« [4]. We agree with the intent and purposes of the editorial, but believe that the use of computerized data to improve standards of health care must be accompanied by scrupulous care in the collection, programming, and analysis of the data.

Acknowledgment

The project reported herein was performed pursuant a grant from the Office of Economic Opportunity, the opinions expressed herein are those of the authors and should not be construed as representing the opinions or policy of any agency of the United States Government.

References

[1] Diamond, E. L., and Lilienfeld, A. M.: Effects of errors in classification and diagnosis in various types of epidemiological studies. Amer. J. publ. Hlth *52* (1962) 1137—1144.

[2] Lindberg, D. A. B.: Collection, evaluation and transmission of hospital laboratory data. Meth. Inform. Med. *6* (1967) 97—107.

[3] Mainland, D.: Use of case records in study of therapy and other features in chronic disease; planning the survey. Ann. rheum. Dis. *14* (1955) 337—352.

[4] Mueller, C. B.: Continuing assessment of medical performance. New Engl. J. Med. *284* (1971) 1378—1380.

[5] Nacke, O., and Wagner, G.: Bibliographie zum Thema »Die Rolle des Fehlers in der Medizin; Fehlerforschung als Aufgabe der medizinischen Dokumentation«. Meth. Inform. Med. *3* (1964) 133—150.

[6] Rogot, E.: A note on measurement errors and detecting real differences. J. Amer. statist. Ass. *56* (1961) 314—319.

[7] Wagner, G.: Quality control and error checking. In Anderson, J. and Forsythe, J. M. (Eds.): Information Processing of Medical Records, pp. 222—226. (North Holland, Amsterdam 1970).

Addresses of the authors: Leslie Hendrickson, Ph.D., Staff Sociologist, Health Law Project, University of Pennsylvania, 133 S. 36th St., Room 310, Philadelphia/Pennsylvania 19104; Jeffrey Myers, M.D., Ph.D., Assistant Professor of Pathology, University of Pennsylvania School of Medicine, Chief, Division of Surgical Pathology, Philadelphia General Hospital, 700 Civic Center Boulevard, Philadelphia/Pennsylvania 19104.

Computer-Assisted Clinical Decision-Making *)

UDC 681.3 : 616 — 07

(From the Sloan School of Management, Massachusetts Institute of Technology, Cambridge/Mass.)

G. A. Gorry

This paper discusses some research in computer-aided clinical decision-making. Experience with formal (decision theory) models is discussed and the limitations of these models are noted.
It is argued that new approaches are needed to solve complex clinical problems, and one such approach is discussed.

Key-Words: Computer-aided Clinical Decision-making, Artificial Intelligence

COMPUTERUNTERSTÜTZTE ENTSCHEIDUNGSFINDUNG

Die vorliegende Arbeit berichtet über einige Forschungsergebnisse auf dem Gebiet der computerunterstützten klinischen Entscheidungsfindung. Es werden Erfahrungen mit formalen Modellen der Entscheidungstheorie mitgeteilt und die Grenzen solcher Modelle angegeben.
Nach Ansicht des Autors sind zur Lösung komplexer klinischer Probleme neue Ansätze erforderlich; ein solcher Ansatz wird diskutiert.

Schlüssel-Wörter: Computerunterstützte klinische Entscheidungsfindung, künstliche Intelligenz

During the past few years, we have been conducting some preliminary research into the use of computers to augment the decision-making abilities of physicians. A few months ago, we decided to increase our efforts in this area. In what follows, I will outline briefly the motivation for this work, the results obtained to date, and the plan for pursuing this research in the future. The plan for future research is not completely clear, and it undoubtedly can be improved through the active interest and criticism of both computer people and physicians.

I. Motivation for the Research**)

In the past few years, there have appeared in the literature many discussions of the use of computers in the health care system, and the way in which they might improve the efficiency of that system. Such improvements are seen as arising from a wide variety of computer-based activities such as scheduling of hospital admissions, control of laboratories, and the maintenance of medical records. Although these activities (and others as well) can undoubtedly benefit from the introduction of well-designed computer systems, more fundamental problems remain. There is an increasing shortage of physician manpower and a geographical maldistribution because new doctors are reluctant to practice in rural or depressed urban communities. Also these discussions fail to indicate how a high level of physician competence can be maintained in the face of a continued expansion of medical knowledge. The gap between what a doctor should know and what he can retain and utilize is continually widening.

As Schwartz [7] has noted: „The computer thus remains (in the light of conventional projections) as an adjunct to the present [health care] system, serving a palliative function, but not really solving the major problems of that system."

There is, in fact, little reason to believe that any of the current proposals for solving these problems, technologic or other, will do more than mitigate their severity. Despite plans to reorganize patterns of medical care and efforts to enlarge medical school capacity and create new classes of „doctor's assistants", the physician shortage promises to be with us for decades and to pose a serious obstacle to health planning. The problem of maintaining and improving quality appears equally knotty since there is little indication that current programs in postgraduate education will be adequate to the challenge.

*) Paper presented at the 17th Annual Meeting of the Deutsche Gesellschaft für Medizinische Dokumentation und Statistik, München, 8.—11. Oct. 1972.

**) This section is drawn largely from [7].