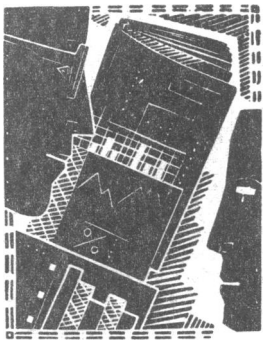


Is necropsy a valid monitor of clinical diagnosis performance?

Rodolfo Saracci

Abstract

***Objective*—To improve the validity of comparisons between clinical and postmortem diagnoses when postmortem diagnosis is used to monitor clinical diagnosis performance.**

***Design*—Analysis of elementary examples.**

***Main outcome measures*—Sensitivity and specificity of clinical and postmortem diagnoses and confirmation and agreement rates. Sensitivity and specificity permit valid comparisons of clinical and postmortem diagnoses among different procedures, sites, or times whereas agreement and confirmation rates may be misleading. Estimates of sensitivity and specificity, however, can be severely distorted by factors such as non-random selection of cases for necropsy or by unrecognised errors in postmortem diagnosis. Such distortion may be minimised by (*a*) estimating the likely magnitude of errors in postmortem diagnosis, (*b*) specifying standard conditions for performing necropsies, and (*c*) ensuring an unbiased sample of moderate size rather than a large biased sample.**

***Conclusion*—Sensitivity and specificity should be used as measures of agreement between clinical and postmortem diagnoses.**

***Implication*—Monitoring of clinical diagnosis performance by necropsy surveys requires ensuring accuracy of pathological examinations and validity of study design and analysis.**

Introduction

The role of necropsy in today's medicine is being recurrently discussed in view of the sharp decline in its prevalence in most countries[1-6] and evidence showing persistently unsatisfactory "agreement" percentages (derived in various arithmetical ways) between clinical diagnosis or diagnosis on the death certificate and postmortem diagnosis, often ranging from as low as 30% to 80% and rarely reaching 90%.[7-11] However, the requirements that the postmortem procedures should themselves satisfy in order to constitute a valid monitoring instrument have received little attention. This paper recalls, at an elementary level, some of these requirements, as they pertain to the design, analysis, and interpretation of comparisons between clinical and postmortem diagnoses.

Comparing clinical and postmortem diagnoses: sensitivity and specificity

The objective of surveys (occasional, periodical, or permanent) comparing clinical and postmortem diagnoses should be to measure the two fundamental properties of sensitivity and specificity of the clinical diagnosis process with necropsy ("true" diagnosis) taken as the standard (table I). The table also shows two other indexes, the overall "agreement rate" and the "confirmation rate" of the clinical diagnosis, once that has been examined by a necropsy. Both these indexes have the drawback of mixing together sensitivity and specificity: it is unfortunate that their widespread use in reporting results may obscure and make it impossible to judge what is going on at the level of the performance of the diagnostic process, which is the objective of the comparison between postmortem and clinical diagnoses. Table II, for instance, displays a comparison between the hospitals of city A and city B (this could equally apply to the comparison between two different years at the same hospitals). In terms of both agreement rate and confirmation rate the hospitals in city B seem only slightly better than those in city A. However, in terms of sensitivity and specificity the picture is much more illuminating as in the hospitals in city B the sensitivity is as high as 100% (not a single case of myocardial infarction being missed) whereas in the other hospitals it is only 80% (as many as 20 cases out of 100 are being missed), the specificity being 95% in both cities. Table III shows a further illustration of how the use of the postmortem "confirmation rate," seemingly the most direct indicator of clinical diagnosis success, may be wholly misleading. In a given geographical area between 1955 and 1985 the confirmation rate of the clinical diagnoses of cancers seemed to have increased

TABLE I—*Sensitivity, specificity, agreement rate, and confirmation rate for clinical versus postmortem diagnosis*

	Postmortem diagnosis (all deaths)		
	Myocardial infarction	Other	Total
Clinical diagnosis:			
Myocardial infarction	24	14	38
Other	6	256	262
Total	30	270	300
Agreement rate	((24+256)/300)×100=93% (Disagreement rate=7%)		
Confirmation rate	(24/38)×100=63%		
Sensitivity	(24/30)×100=80% (false negatives=20%)		
Specificity	(256/270)×100=95% (false positives=5%)		

TABLE II—*Sensitivity, specificity, agreement rate, and confirmation rate for clinical versus postmortem diagnosis in hospitals in two cities*

	Postmortem diagnosis (all deaths)			
	City A		City B	
	Myocardial infarction	Other	Myocardial infarction	Other
Clinical diagnosis:				
Myocardial infarction	24	14	30	14
Other	6	256		256
Total	30	270	30	270
Agreement rate (%)	93		95	
Confirmation rate (%)	63		68	
Sensitivity (%)	80		100	
Specificity (%)	95		95	

Unit of Analytical Epidemiology, International Agency for Research on Cancer, 69372 Lyons Cedex 08, France
Rodolfo Saracci, MD, *chief*

BMJ 1991;303:898-900

TABLE III—*Confirmation rate for clinical versus postmortem diagnoses 1955 and 1985*

	Postmortem diagnosis (all deaths)			
	1955		1985	
	Cancer	Other	Cancer	Other
Clinical diagnosis:				
Cancer	27	14	81	10
Other	3	256	9	200
Total	30	270	90	210
Confirmation rate	(27/(27+14))×100=66%		(81/(81+10))×100=89%	
Sensitivity	(27/30) ×100=90%		(81/90) ×100=90%	
Specificity	(256/270)×100=95%		(200/210)×100=95%	

TABLE IV—*Sensitivity and specificity for clinical versus postmortem diagnoses according to selection of deaths for necropsy (1985 data)*

	Postmortem diagnosis			
	All deaths		Selected deaths*	
	Cancer	Other	Cancer	Other
Clinical diagnosis:				
Cancer	81	10	81	10
Other	9	200	3	67
Total	90	210	84	77
Sensitivity	(81/90) ×100=90%		(81/84)×100=96%	
Specificity	(200/210)×100=95%		(67/77)×100=87%	

*All clinical diagnoses of cancer selectively brought to necropsy; one in three of "other" diagnoses brought to necropsy.

appreciably (from 66% to 89%), consistent with all the diagnostic improvements since 1955. Unfortunately, as the table also shows, the performance of clinical diagnosis had not changed, as the sensitivity and specificity were exactly the same in both years, the increase in the confirmation rate reflecting the increase in deaths from cancer, from 10% of total deaths in 1955 to 30% in 1985. (A sensitivity of 90% and a specificity of 95% as used in these constructed examples are in the range of realistic values in everyday medical practice; for simplicity sampling errors are ignored here as in the other examples.)

Sensitivity and specificity and selection of deaths for necropsy

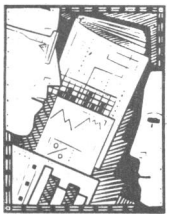
In the previous examples it was assumed for simplicity that all deaths in a geographical area came to necropsy (and also that there were no differences in sensitivity and specificity by age and sex) whereas in actual practice only a proportion, often a small proportion, of deaths are so investigated. Table IV shows the sensitivity and specificity of diagnosis for all deaths in 1985 and the results when a selection takes place, based on the clinical diagnosis, whereby all cases of clinically diagnosed cancers go to necropsy but only, say, one in three of the other diagnoses. The estimates of sensitivity and specificity are then distorted and seem to be 96% and 87% instead of 90% and 95%. (Should the selection be even stronger, with only one in five of the other diagnoses coming to necropsy, sensitivity and specificity would seem to be 98% and 80%.) When this type of selection takes place, prompted by desire to investigate certain kinds of suspected diagnoses necroscopically (probably the most common situation in practice), it may become impossible to measure specificity and sensitivity with any accuracy. Quantitative comparisons between postmortem and clinical diagnoses in different circumstances, from which implications about clinical diagnosis performance are drawn, may then end with equivocal answers, largely defeating the very purpose of the exercise. For example, in a thorough investigation carried out on postmortem and clinical material at one hospital in the United States out of 100 necropsies in each of the three years 1960, 1970, and 1980 the percentages of major diagnoses of any kind missed by the clinician (false negatives) were respectively 8%, 12%, and 11%, indicating, if anything, a deteriorating clinical diagnosis performance with time. Missed major diagnoses were defined as those in which their detection before death would probably have led to a change in management that might have resulted in a cure or prolonged survival. However, during the observation period the proportion of inpatient deaths resulting in a necropsy fell from 75% in 1960 and 71% in 1970 to 38% in 1980. By taking these figures into account and on the assumption that no major missed diagnoses would have been included among the cases that did not result in a necropsy (probably being those for which the diagnosis was clear enough without necropsy) the percentages of false negatives would become 6%, 8·5%, and 4·2% respectively, pointing to a long term improvement of the clinical diagnosis, contrary to the previous conclusion. Moreover, this time trend could be accentuated or, instead, flattened or reversed, depending on the extent to which the assumption of no false negative diagnoses in the group for which no necropsy was performed applies equally, or differently, or not at all, in the three years 1960, 1970, and 1980.

Sensitivity and specificity as affected by errors in postmortem diagnosis

It would be surprising if the complex processes of macroscopic and microscopic observations and data interpretation leading to the postmortem diagnosis were to be error free, though it is also surprising that no studies specifically addressing this issue seem to have been published. Reports are available, however, on variations between and within observers in assessing histopathological and macroscopic necropsy lesions—for example, atherosclerotic lesions[12]—and this gives some support to the a priori notion that the postmortem diagnosis itself must be subject to a margin of variation and uncertainty. In table V postmortem diagnosis has been granted with an error of 2%: the clinical categorisation of all cases remaining fixed, 2% of the "other" cases are erroneously attributed at necropsy to primary lung cancer (bronchial carcinoma), raising the total from 200 to 236. The results in the table clearly indicate that in such a case the unrecognised error in postmortem diagnosis would show up as an appreciable underestimate of the sensitivity of the clinical diagnosis, the true sensitivity being 90% and the observed sensitivity only 77%, unduly underrating the clinical diagnosis performance. Other patterns and sizes of errors could occur and, as a safe rule of thumb, a correct estimation of sensitivity and specificity would be ensured only if the frequency of errors in necropsy were one to two orders of magnitude smaller than the

TABLE V—*Sensitivity and specificity for clinical versus true and observed postmortem diagnosis*

	Postmortem diagnosis			
	True		Observed	
	Bronchial carcinoma	Other	Bronchial carcinoma	Other
Clinical diagnosis:				
Bronchial carcinoma	180	90	182	88
Other	20	1710	54	1676
Total	200	1800	236	1764
Sensitivity	(180/200) ×100=90%		(182/236) ×100=77%	
Specificity	(1710/1800)×100=95%		(1676/1764)×100=95%	



frequency of the conditions to be diagnosed (that is, (1% to 0·1% in the example).

Some suggestions

From these elementary arguments one may attempt some suggestions. Sensitivity and specificity of clinical diagnosis, rather than agreement or confirmation rates, ought to be measured validly for necropsy to have an effective role in the systematic monitoring of quality of clinical diagnosis (in addition to its continuing role in research and teaching and as a final diagnostic tool in individual diagnostically or therapeutically problematic cases). This involves in turn three conditions. Firstly, the likely size of the errors in the postmortem diagnosis itself must be estimated; of course, as there is no further standard than necropsy itself the ultimate "truth" is destined to remain unknown. It is possible, however, to obtain empirical estimates of the degree of reproducibility of postmortem diagnosis under different circumstances (which also provide an estimate of the maximum degree of correlation of the postmortem diagnosis with the unknown "true" diagnosis). For example, this could be done by having one pathologist performing the necropsy and two observing pathologists independently noting the findings and coming up with their own diagnoses, or by having two pathologists independently interpreting the written record provided by a third, or by assigning at random to two pathologists consecutive cases with a given clinical diagnosis. Secondly, the conditions under which the necropsy is performed must be clearly specified for the whole period of the survey, which may stretch over many years, with respect to procedure, degree of completeness (extent of organ sampling and macroscopic and microscopic examination), pathologist(s), and any clinical information available to the pathologist when formulating the postmortem diagnosis. Thirdly, an appropriate sampling procedure should be adopted for deaths submitted to necropsy. This is a sine qua non to make the results interpretable. As subjects move from a healthy to a diseased state, outside hospital and in hospital, and to death and to necropsy non-random selections take place—for instance, in who is admitted to hospital compared with who is not and in which out of all hospital deaths is a necropsy performed. To minimise the distorting effect that such non-random selections have on the estimates of sensitivity and specificity the best sampling plan in a given geographical area consists of taking for necropsy all deaths or randomly sampling a proportion of them. This basic design can be usefully and flexibly modified in various ways, particularly in the sense of classification according to criteria such as sex, age, and inpatient or outpatient and sampling at random, possibly with different sampling fractions, within the classes. Also, a record needs to be kept of the individual characteristics (demographic and clinical) of those subjects included in the sample for whom a necropsy was not performed notwithstanding all efforts by the pathologists, and reasons (for example, refusal by relatives). This sampling scheme may prove less complicated and more feasible than it seems at first, as suggested by the experience in, for example, Trieste or Malmö, where a high proportion of deaths result in necropsy.[5,13] At any rate, efforts in securing an unbiased representative sample of reasonable size are bound to pay a much higher return in terms of exploitable information than efforts to obtain very large but biased series of cases.

To the extent that the requirements discussed here are not met, numerical estimates of sensitivity and specificity are helpful as rough orientations, but they can hardly be of use to compare quantitatively the relative merits of different diagnostic methods or diagnostic services in different places or times, which are key objectives in evaluating diagnostic procedures and diagnoses, as used in epidemiological studies.

1 Cameron HM, McGoogan E, Clarke J, Wilson BA. Trends in hospital necropsy rates. Scotland 1961-74. *BMJ* 1977;i:1577-80.
2 Roberts WC. The autopsy: its decline and a suggestion for its revival. *N Engl J Med* 1978;**299**:332-8.
3 Ahronheim JC, Bernholc AS, Clark WD. Age trends in autopsy rates. Striking decline in late life. *JAMA* 1983;**250**:1182-6.
4 Geller SA. Autopsy. *Scientific American* 1983;**248**:110-21.
5 Silvestri F, Bussani R, Giarelli L. High autopsy rate in Trieste, 1901-1985: age associated increase in necropsy practice. *Pathologica* 1988;**80**:523-32.
6 Peacock SJ, Machin D, Duboulay CEH, Kirkham N. The autopsy: a useful tool or an old relic? *J Pathol* 1988;**156**:9-14.
7 Britton M. Diagnostic errors discovered at autopsy. *Acta Med Scand* 1974;**196**:203-10.
8 Cameron HM, McGoogan E, Watson H. Necropsy: a yardstick for clinical diagnoses. *BMJ* 1980;**281**:985-8.
9 Burrows S. The post-mortem examination: scientific necessity or folly? *JAMA* 1975;**233**:441-3.
10 Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three medical eras. *N Engl J Med* 1983;**308**:1000-5.
11 Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *N Engl J Med* 1985;**313**:1263-9.
12 Uemura K, Sternby N, Vanecek R, Vihert A, Kagan A. Grading atherosclerosis in aorta and coronary arteries obtained at autopsy. *Bull World Health Organ* 1964;**31**:297-320.
13 Sternby NH. The role of autopsy in cancer registration in Sweden, with particular reference to findings in Malmö. In: Riboli E, Delendi M, eds. *Autopsy in epidemiology and medical research*. Lyons: International Agency for Research on Cancer, 1991:217-22. (IARC scientific publications No 112.)

(Accepted 20 August 1991)

International Quality Assurance

Quality and health care in Sweden

Peter Reizenstein

This is the first in a series of commissioned articles describing the nature of health care quality in Europe and elsewhere.

Haematology Laboratory and Institute for Radiation Protection, Karolinska Hospital, PO Box 60500, S-104 01 Stockholm, Sweden
Peter Reizenstein, MD, *professor*

BMJ 1991;**303**:900-2

Mahatma Gandhi was once asked what he thought about Western civilisation. He replied: that it would be a good idea. The same is true of medical audit in Sweden.

Swedish health care system

Swedish hospitals and district medical offices are mostly owned by county councils, communities with an average of 300 000 inhabitants and their own elected local parliament, which have the right to tax inhabitants about 20%-25% of their income, about three quarters of which is spent on health. Government owned compulsory health insurance contributes a fixed, age adjusted sum per inhabitant to the councils but no fees for service.

Private medical offices receive fees for service, which, however, are deducted from the sum given to the county councils. Private offices can be established only with the councils' permission, which is given restrictively, particularly in big cities. The largest private sector is company owned or union owned industrial medicine costing about a tenth of the total expenditure on health care.