Diagnostic errors in the new millennium: a follow-up autopsy study

Stefanie Schwanda-Burger[1], Holger Moch[2], Jörg Muntwyler[1] and Franco Salomon[1]

[1]Medical Clinic, Department of Internal Medicine, Zurich, Switzerland and [2]Institute of Surgical Pathology, Department of Pathology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

A systematic review of the second half of the last century suggested that diagnostic errors have decreased over time. Our previous study covering the years 1972–1992 was then the only time series showing a significant reduction of diagnostic errors from a single institution. We report here the results of a follow-up study a decade later. We analyzed discrepancies between clinical and autoptic diagnoses in 100 randomly selected medical patients who died in the wards and in the medical intensive care unit at a tertiary-care teaching hospital in Switzerland in the year 2002. Autopsy rate declined from around 90% in the years from 1972 to 1992 to 54% in the present study. Major diagnostic errors (class I and II) declined significantly from 30 to 7% ($P < 0.001$) over the last 30 years. Class I errors decreased from 16 to 2% ($P < 0.001$) in the year 2002. Sensitivity for cardiovascular diseases increased from 69 to 92% ($P = 0.006$), for infectious diseases from 25 to 90% ($P = 0.013$) and for neoplastic diseases from 89 to 100% ($P = 0.053$). Specificity for cardiovascular diseases increased from 85 to 98% ($P < 0.001$) but was unchanged at a high level for infectious diseases and neoplastic diseases. The number of diagnostic procedures increased from 144 to 281 ($P < 0.001$) with an increase in the number of computer tomography investigations and of tissue sampling in the last decade. The frequency of major diagnostic errors has been further reduced at the beginning of the new millennium probably due in large part to new diagnostic tools.

The armamentarium of diagnostic procedures in clinical medicine has become vast and sophisticated over the past decades. A systematic review of autopsy-detected diagnostic errors over the last 40 years of the 20th century disclosed a relative decrease of about 20% for major errors and of one third for class I errors per decade.[1] Time series from single institutions included in the systematic review showed no decrease in diagnostic discrepancies over time[2–6] with one exception,[7] reflecting probably a combination of inadequate power and decreasing autopsy rates leading to selection bias.[1] We report here the 10 year follow-up of the previously mentioned study[7] to assess the impact of new diagnostic tools as spiral computed tomography (CT) angiography,[8,9] an array of biomarkers[10] and the use of electronic medical records on the frequency of diagnostic errors detected by autopsy.

Materials and methods

Selection of Cases

We analyzed retrospectively the medical and necropsy records of 100 randomly selected adult patients who died at the medical clinic A and B and the medical intensive care unit at the Department of Internal Medicine at the University Hospital Zurich, Zurich, Switzerland, in the year 2002. Data from 1972, 1982 and 1992 were published in a previous report.[7] Random sampling was performed with a random-number table[11] from the list of patients who died in the medical clinic and underwent necropsy. In the late nineties, Zurich switched from tacit consent to informed consent for autopsy-permission. In all patients who died in the medical clinic in 2002, informed consent for necropsy was sought.

Correspondence: Dr H Moch, MD, Institute for Surgical Pathology, University Hospital Zurich, Schmelzbergstrasse 12, CH-8902 Zurich, Switzerland.
E-mail: holger.moch@usz.ch

Box 1 Description of discrepancy classes according to Goldman and colleagues,[3] modified by Battle and colleagues[13]

Major discrepancies
Class I
Discrepancies in major diagnoses. Knowledge of diagnosis before death would have led to changes in management that could have prolonged survival or cured the patient (e.g., pulmonary infarction treated as pneumonia, fungal pneumonia treated as bacterial infection).
Class II
Discrepancies in major diagnoses whose detection before death would not have changed survival even with correct treatment (e.g., biventricular cardiac insufficiency due to severe aortic stenosis with missed pulmonary emboli, correctly treated bacterial sepsis with multiorgan failure because of unrecognised postoperative cervical osteomyelitis in a patient with rheumatoid arthritis). No treatment available at the time (eg cytomegalovirus infection up to the early eighties).

Minor discrepancies
Class III
Discrepancies in minor diagnoses not directly related to cause of death, but with symptoms that should have been treated or would have eventually affected prognosis (e.g., pulmonary carcinoma in a patient with ruptured infrarenal aortic aneurysm).
Class IV
Discrepancies in minor occult diagnoses (non-diagnosable) but with possible epidemiological or genetic importance (e.g., symptomless gallstones, goitre).

Non discrepancy
Class V
Non-discrepant diagnoses.

Non-classifiable cases
Class VI
Patient died immediately after admission with no diagnostic procedures, or refused any diagnostic procedures or treatment. Necropsy was unsatisfactory with no clear findings and no diagnosis could be established after review of clinical and necropsy data.[14]

The role of the medical clinic at the University Hospital of Zurich as referral center of eastern Switzerland remained unchanged, as did the organization of the medical clinic. All medical inpatients of the University Hospital of Zurich were cared for in the emergency room and then admitted to the medical clinics A and B or to the medical intensive care unit. There were no specialised medical wards. All patients were cared for by full-time hospitalists,[12] who were in charge of the emergency room, of all the wards and the medical intensive care unit. There was a close collaboration with all the medical specialties in the Department of Medicine. An in-house developed computer-based patient records system was stepwise introduced since 1995. In 2002 medical and nurse reports, laboratory results and radiology reports were available at all the working places.

Analysis of Reports

We recorded age, sex, number of admissions in the 12 months before the index hospital stay, and length of stay. Clinical diagnoses were those listed by the clinician on the necropsy request and all diagnoses that were established or assumed and led to specific treatment. Autopsy diagnoses were the diagnoses listed on the final autopsy report. All included patients underwent a complete necropsy, including histological assessment of each organ, which was performed by a junior pathologist. A staff pathologist reviewed macroscopic and microscopic findings. Each week the results of autopsies were presented and discussed with the medical clinic's clinicians. The main diagnoses were listed separately on the autopsy report, but were not classified

and no cause of death was given. Clinical and necropsy diagnoses were grouped into seven classifications according to the International Classification of Disease, 9th edition (classification 1979–1983): infectious diseases (ICD-9 1–139); neoplastic diseases (ICD-9 140–239); cardiovascular diseases (ICD-9 390–459); pulmonary diseases (ICD-9 460–519); gastrointestinal diseases (ICD-9 520–579); renal diseases (ICD-9 580–629); and miscellaneous (remaining diagnoses).

We classified discrepancies between the clinical and autopsy diagnoses according to the method of Goldman *et al*[3] the modification of Battle *et al*[13] and as non-classifiable cases[14] (Box 1). Major diagnoses were those involving the principal underlying cause of death and major contributors to it.[13] Minor diagnoses were antecedent disorders, related diagnoses, contributing causes or other important disorders.[14] If the decision to limit or stop the diagnostic or therapeutic process was made during the hospital stay, we assessed and classified the clinical diagnostic process up to the point at which the process was stopped. For example, a patient was diagnosed as having a diffuse metastasizing carcinoma and it was agreed and stated in the notes that no further investigation and treatment should be undertaken apart from analgesic treatment and supportive care. The patient died 10 days later. Pneumonia and deep-vein thrombosis seen at necropsy were not taken into account for the classification of discrepancies. If, however, the carcinoma had not been confirmed, the diagnosis would have been classified as a major discrepancy.

A single class of discrepancy was assigned to each case. For cases with more than one class, the most severe was chosen. Discrepancies were classified by agreement of two clinicians (SSB and FS) and one

Table 1 General data of study patients

	1972 (n = 100)	1982 (n = 100)	1992 (n = 100)	2002 (n = 100)	P-value
Male/female	60/40	59/41	62/38	60/40	0.984
Median age (years)	65	67.5	64	68	0.304
Median length of stay (days)	10.5	5.5	8.0	7.0	0.002
Patients with no previous hospital admissions	49%	56%	52%	40%	0.137
Patients with one previous hospital admissions	29%	33%	26%	21%	0.282
Patients with ≥ 2 previous hospital admissions	22%	11%	22%	39%	0.001

pathologist (HM). All class I and II errors were assessed a second time, but in no case was a reclassification necessary.

We calculated accuracy, sensitivity, and specificity for the three most frequent clinical categories of diagnoses—cardiovascular, neoplastic and infectious diseases. Accuracy was calculated as the sum of true-positive and true-negative diagnoses in each category divided by all cases. We calculated sensitivity as the proportion of true positives divided by the sum of true positives and false negatives. Specificity resulted from the proportion of true negative divided by the sum of true negatives and false positives. We counted class I and II discrepancies to be false-negative diagnoses in the case in which the necropsy diagnosis was not in the same diagnostic group as the clinical diagnosis is. False-positive diagnoses were cases with class I and II discrepancies, in which the clinical diagnosis was not in the same diagnostic group as the necropsy diagnosis. When the diagnoses were in the same diagnostic group but a major discrepancy was present, we took them to be false positive and false negative within the assessed diagnostic group. If, for example, aortic dissection was wrongly diagnosed as myocardial infarction, the diagnosis was counted as false negative (missed aortic dissection) and as false positive (myocardial infarction not present).

Diagnostic tests (except blood tests and microbiological investigations) were separated into the following groups:[3,4] standard non-contrast radiological procedures, contrast radiological procedures, endoscopies (gastrointestinal and respiratory tract), biopsies and surgical explorations; scintigraphy; ultrasonography; echocardiography; CT; and magnetic resonance imaging. As defined by Goldman et al[3] the number of different types rather than the number of procedures were counted.

Statistical Analysis

We calculated means and proportions of baseline characteristics. We compared the changes in the proportion of errors across the study years with the exact Cochran–Armitage test for trend. Analysis was carried out with SAS (version 9.1) and SPSS (version 12.0.1). All reported P-values are two sided and $P \leq 0.05$ was taken to be significant.

Table 2 Distribution of clinical diagnoses

Disorder	1972	1982	1992	2002
Infectious diseases	0	4	14	10
Neoplastic diseases	33	29	23	36
Cardiovascular diseases	33	45	30	36
Pulmonary diseases	4	2	7	4
Gastrointestinal diseases	5	7	12	9
Renal diseases	9	1	1	1
Miscellaneous	16	12	13	4
Total	100	100	100	100

Results

The main characteristics of patients were similar in each year (Table 1). The necropsy rate was 94.0% in 1972, 89.2% in 1982 and 1992 and declined to 53.6% in 2002. Cardiovascular and neoplastic diseases were the largest diagnostic groups in 2002 followed by infectious diseases (Table 2).

The changes in the six discrepancy classes (Box 1) over time are shown in Figure 1. Major diagnostic errors (class I and II) declined significantly from 30 to 7% ($P < 0.001$). In the last 10 years major diagnostic errors declined from 14 to 7% with class I errors being reduced from 7 to 2% and class II errors from 7 to 5%. The major diagnostic errors in 2002 were pneumonia (two cases), myocarditis (two cases) and one case each of pulmonary embolism, intestinal ischemia and metastatic subdural empyema due to pneumococcal septicemia. Minor diagnostic errors (class III and IV) increased from 23 to 53% ($P < 0.001$). Class III errors decreased from 25 to 16% in the last decade. Minor occult diagnoses (class IV discrepancies) increased from 21 to 37% in the same period.

Sensitivity for cardiovascular diseases increased from 69 to 92% ($P = 0.006$), for infectious diseases from 25 to 90% ($P = 0.013$) and for neoplastic diseases from 89 to 100% ($P = 0.053$) (Table 3). Specificity for cardiovascular diseases increased from 85 to 98% ($P < 0.001$) but was unchanged for infectious diseases (100–99%, $P = 0.245$) and for neoplastic diseases (92–99%, $P = 0.125$).

The number of diagnostic procedures increased from 144 to 281 ($P < 0.001$) with a higher number of CT investigations and of biopsies and fine-needle aspirations in the last decade (Table 4).
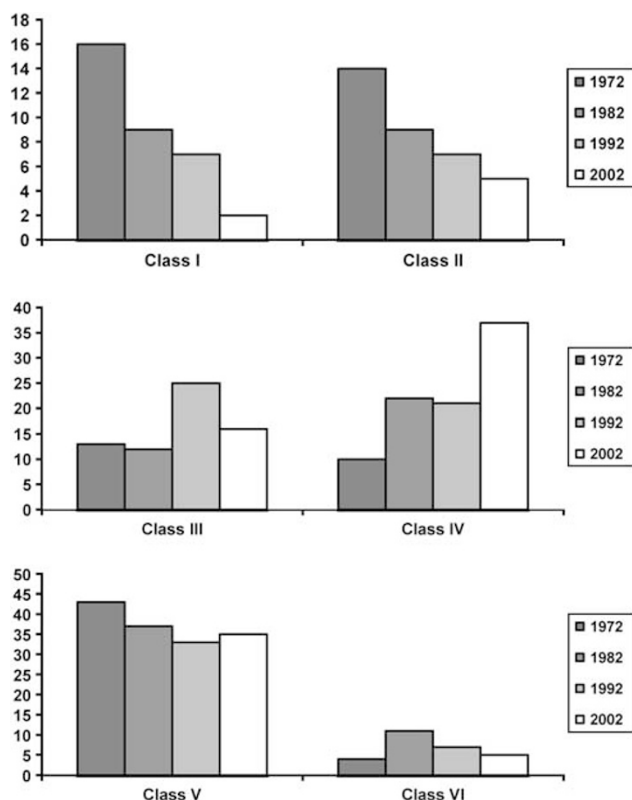
Figure 1 Changes in discrepancy classes I–VI over the study years. Class I decreased from 16 to 2% ($P < 0.001$) and class II decreased from 14 to 5% ($P = 0.026$). Major diagnostic errors (class I and II) decreased from 30 to 7% ($P < 0.001$). Class III increased from 13 to 16% ($P = 0.207$) and class IV from 10 to 37% ($P < 0.001$). Minor diagnostic errors increased from 23 to 53% ($P < 0.001$). Class V decreased from 43 to 35% ($P = 0.212$) and class VI from 4 to 5% ($P = 1.000$).

Discussion

In this longitudinal study we observed a further significant reduction in major diagnostic errors at the beginning of the new millennium. A similar reduction from 15.0 to 6.1% was reported recently from an other single center study of 970 autopsies continuously analyzed over a period of 10 years from 1997 to 2006, with an average autopsy rate of 50%.[15] Such studies reflect in the first instance the local efforts to improve the diagnostic and therapeutic process. But by analyzing the possible contributing factors to the improvement there could be some indications for a more general trend. The changes in diagnostic accuracy, sensitivity and specificity can give some indications.[16]

In the present study the largest changes in accuracy, sensitivity and specificity for cardiovascular diseases occurred between 1972 and 1992. In the study of Thurnheer *et al*,[15] the significant increase in accuracy, sensitivity and specificity for cardiovascular diseases was seen between 1997 and 2002 coinciding with the introduction of d-dimers, troponin and spiral CT angiography for the diagnosis of cardiovascular diseases. The levels of

accuracy, sensitivity and specificity for cardiovascular diseases in the present study are remarkably similar to Thurnheer *et al*,[15] but they were reached in the year 1992 before the introduction of the abovementioned new diagnostic tools. This example shows how the same result can be achieved by different means depending on the local circumstances regarding patient characteristics, diagnostic tools and diagnostic know-how. The further increase in sensitivity for cardiovascular diseases observed in this study between 1992 and 2002 could well be the result of the new diagnostic tools.

In most autopsy studies the number of missed neoplastic diseases was low,[3,4,15] with exceptions even in recent times.[17] In contrast with our previous study,[7] we now found a significant increase in diagnostic sensitivity and accuracy for tumors over the last 30 years. This improvement goes along with an increased use of diagnostic procedures such as CT and tissue sampling most pronounced in the last decade (Table 4).

Having analyzed the changes in sensitivity and specificity for cardiovascular and neoplastic diseases it seems likely that increased use of diagnostic tools have contributed to the reduction of diagnostic errors. But as argued above, other factors could come into play. In trying to find such factors, speculations are inevitable in a retrospective study on an extremely complex task such as the diagnostic process. Graber *et al*[18] found that cognitive factors were responsible in 90% of the cases with diagnostic errors, whereas in cases with delayed diagnosis, system-related factors were the main cause.

Faulty information synthesis was the most frequent cause of cognitive-based diagnostic errors and premature closure the single-most frequent mechanism.[18] Premature closure can occur at any stage of the diagnostic process.[7] The tendency to stop considering other diagnosis is independent of clinical experience[19] and is associated with overconfidence in the already available findings leading to a false-positive diagnosis. Overconfidence is therefore a main factor leading to diagnostic errors and by the same mechanism autoptic verification of diagnoses is considered unnecessary by clinicians,[20] as reflected by the very low autopsy rates.[21–23] This fact is rarely stated openly but often disguised as complacency.[20,24] Post-mortem case review without autopsy is often used as a substitute,[25,26] but this approach has been shown to leave 85% of main diagnostic errors undetected.[27] The invaluable advantage of feedback and learning through autopsies is that uncertainty is almost eliminated and confidence in the diagnostic workup is strengthened in cases with no discrepancies, which are still the largest group. In cases with major or minor discrepancies clinicians become aware of their fallibility as a prerequisite to further improve the diagnostic process[28] by correcting overconfidence.[20,29] Autopsy should therefore be an integral part of any effort to reduce diagnostic errors.[30]

Table 3 Accuracy, sensitivity and specificity for (a) cardiovascular diseases; (b) neoplastic diseases; (c) infectious diseases

	1972	*1982*	*1992*	*2002*	P-*value*
(a)					
Accuracy					
Value (%)	83	87	95	97	<0.001
(True pos+true neg)/all	(33+50/100)	(45+42)/100	(30+65)/100	(35+62)/100	
Sensitivity					
Value (%)	69	82	86	92	0.006
True pos/(true pos+false neg)	(33/48)	(45/55)	(30/35)	(35/38)	
Specificity					
Value (%)	85	82	97	98	<0.001
True neg/(true neg+false pos)	(50/59)	(42/51)	(65/67)	(62/63)	
(b)					
Accuracy					
Value (%)	93	94	96	100	0.014
(True pos+true neg)/all	(32+61/100)	(29+65)/100	(22+74)/100	(34+66)/100	
Sensitivity					
Value (%)	89	88	96	100	0.053
True pos/(true pos+false neg)	(32/36)	(29/33)	(22/23)	(34/34)	
Specificity					
Value (%)	92	97	96	99	0.125
True neg/(true neg+false pos)	(61/66)	(65/67)	(74/77)	(66/67)	
(c)					
Accuracy					
Value (%)	97	98	97	99	0.552
(True pos+true neg)/all	(1+96/100)	(4+94)/100	(12+85)/100	(9+90)/100	
Sensitivity					
Value (%)	25	67	86	90	0.013
True pos/(true pos+false neg)	(1/4)	(4/6)	(12/14)	(9/10)	
Specificity					
Value (%)	100	100	99	99	0.245
True neg/(true neg+false pos)	(96/96)	(94/94)	(85/86)	(90/91)	

Abbreviations: neg, negative; pos, positive.

Table 4 Number of patients with diagnostic procedures in study in years 1972, 1982, 1992 and 2002

Procedure	*1972*	*1982*	*1992*	*2002*
Non-contrast radiological procedures	92	87	81	80
Contrast radiological procedures	13	2	11	7
Endoscopies	14	18	32	36
Biopsies, fine-needle aspirations and surgical explorations	13	10	15	34
Scintigraphy	12	2	3	5
Ultrasound	0	15	49	42
Echocardiography	0	2	24	26
Computed tomography	0	11	17	45
Magnetic resonance imaging	0	0	3	6

Timely information retrieval is a fundamental part of making a correct diagnosis. Graber *et al*[18] found that non-availability of data contributed to diagnostic errors at the system-related level. In our Department of Internal Medicine, a clinically intuitive computer-based patient record system was developed and was used by doctors and nurses since 1995. The all-around availability of patient records may have contributed to reduce diagnostic errors in the present study. It is interesting to note that the introduction of a computer-based patient record system can change knowledge organization and reasoning pattern in medical decision making.[31]

There are several limitations in the interpretation of the present study. Most importantly there has been a drop in autopsy rate in the last decade from

around 90 to 54%. This was in part due to the change in legislation in the county of Zürich from tacid to informed consent regarding autopsy request as observed by others.[22] The lower autopsy rate makes the interpretation of the reduction of diagnostic errors difficult. The distribution of main diagnostic groups showed no evidence of selection bias (Table 2). We compared 100 patients without autopsy who died in 2002 with the study population and found no difference in age, gender, length of stay, previous hospitalizations or in the number of diagnostic procedures (unpublished data). A systematic review of the relationship between autopsy-detected diagnostic errors and autopsy rates found that lower autopsy rates were associated with higher rates of major diagnostic errors.[1] We have been very careful to use the same criteria to assign the discrepancy classes as in the previous study but subtle unintended shifts cannot be excluded. Contrary to the previous study this time a pathologist was part of the team to classify the cases. He was not involved in performing and analyzing the autopsies.

The apparent increase in minor diagnostic errors is due to the classification system used, where only the most severe discrepancy was counted. With the reduction of major diagnostic errors more class III and IV errors emerged as discrepancies, with a predominance of class IV discrepancies in the last decade.

In summary we observed a further improvement of diagnostic performance assessed by autopsy from unselected patients who died in the wards and in the intensive care unit of an academic Department of Internal Medicine. This reduction of diagnostic errors is likely to be the result of new diagnostic methods, of continuous feedback and learning through autopsy and improved availability of patient information.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Shojania KG, Burton EC, McDonald KM, *et al.* Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. JAMA 2003;289:2849–2856.

2 Carvalho FM, Widmer MR, Cruz M, *et al.* Clinical diagnosis *versus* autopsy. Bull Pan Am Health Organ 1991;25:41–46.

3 Goldman L, Sayson R, Robbins S, *et al.* The value of the autopsy in three medical eras. N Engl J Med 1983;308:1000–1005.

4 Kirch W, Schafii C. Misdiagnosis at a university hospital in 4 medical eras. Medicine (Baltimore) 1996;75:29–40.

5 Poli L, Pich A, Zanocchi M, *et al.* Autopsy and multiple pathology in the elderly. Gerontology 1993;39:55–63.

6 Veress B, Alafuzoff I. A retrospective analysis of clinical diagnoses and autopsy findings in 3,042 cases during two different time periods. Hum Pathol 1994;25:140–145.

7 Sonderegger-Iseli K, Burger S, Muntwyler J, *et al.* Diagnostic errors in three medical eras: a necropsy study. Lancet 2000;355:2027–2031.

8 Richman PB, Courtney DM, Friese J, *et al.* Prevalence and significance of nonthromboembolic findings on chest computed tomography angiography performed to rule out pulmonary embolism: a multicenter study of 1025 emergency department patients. Acad Emerg Med 2004;11:642–647.

9 van Strijen MJ, Bloem JL, de Monye W, *et al.* Helical computed tomography and alternative diagnosis in patients with excluded pulmonary embolism. J Thromb Haemost 2005;3:2449–2456.

10 Mueller C, Muller B, Perruchoud AP. Biomarkers: past, present, and future. Swiss Med Wkly 2008;138: 225–229.

11 Wissenschaftliche Tabellen, Geigy Basel. 1980; **Teilband Statistik**..

12 Flanders SA, Wachter RM. Hospitalists: the new model of inpatient medical care in the United States. Eur J Intern Med 2003;14:65–70.

13 Battle RM, Pathak D, Humble CG, *et al.* Factors influencing discrepancies between premortem and postmortem diagnoses. JAMA 1987;258:339–344.

14 Bellwald M. [Autopsies with unsatisfactory results]. Schweiz Med Wochenschr 1982;112:75–82.

15 Thurnheer R, Hoess C, Doenecke C, *et al.* Diagnostic performance in a primary referral hospital assessed by autopsy: Evolution over a ten-year period. Eur J Intern Med 2009;20:784–787.

16 Saracci R. Is necropsy a valid monitor of clinical diagnosis performance? BMJ 1991;303:898–900.

17 Burton EC, Troxclair DA, Newman 3rd WP. Autopsy diagnoses of malignant neoplasms: how often are clinical diagnoses incorrect? JAMA 1998;280: 1245–1248.

18 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med 2005;165: 1493–1499.

19 Voytovich AE, Rippey RM, Suffredini A. Premature conclusions in diagnostic reasoning. J Med Educ 1985;60:302–307.

20 Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. Am J Med 2008;121: S2–23.

21 Burton JL, Underwood J. Clinical, educational, and epidemiological value of autopsy. Lancet 2007;369: 1471–1480.

22 Lundberg GD. Low-tech autopsies in the era of high-tech medicine: continued value for quality assurance and patient safety. JAMA 1998;280: 1273–1274.

23 Shojania KG, Burton EC. The vanishing nonforensic autopsy. N Engl J Med 2008;358:873–875.

24 Zijlstra JG. The value of autopsy, believe it or not. Lancet 2007;370:27.

25 Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. JAMA 2001;286:415–420.

26 Hayward RA, McMahon Jr LF, Bernard AM. Evaluating the care of general medicine inpatients: how good is implicit review? Ann Intern Med 1993; 118:550–556.

27 Pelletier Jr LL, Klutzow F, Lancaster H. The autopsy: its role in the evaluation of patient care. J Gen Intern Med 1989;4:300–303.

28 Leape LL. Error in medicine. JAMA 1994;272:1851–1857.

29 Lowry F. Failure to perform autopsies means some Mds 'walking in a fog of misplaced optimism'. CMAJ 1995;153:811–814.

30 Newman-Toker DE, Pronovost PJ. Diagnostic errors—the next frontier for patient safety. JAMA 2009;301:1060–1062.

31 Patel VL, Kushniruk AW, Yang S, *et al.* Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning. J Am Med Inform Assoc 2000;7:569–585.