

# Overestimation of clinical diagnostic performance caused by low necropsy rates

K G Shojania, E C Burton, K M McDonald, L Goldman

*Qual Saf Health Care* 2005;14:408–413. doi: 10.1136/qshc.2004.011973

**Background:** Diagnostic sensitivity is calculated as the number of correct diagnoses divided by the sum of correct diagnoses plus the number of missed or false negative diagnoses. Because missed diagnoses are generally detected during clinical follow up or at necropsy, the low necropsy rates seen in current practice may result in overestimates of diagnostic performance. Using three target conditions (aortic dissection, pulmonary embolism, and active tuberculosis), the prevalence of clinically missed cases among necropsied and non-necropsied deaths was estimated and the impact of low necropsy rates on the apparent sensitivity of antemortem diagnosis determined.

**Methods:** After reviewing case series for each target condition, the most recent study that included cases first detected at necropsy was selected and the reported sensitivity of clinical diagnosis adjusted by estimating the total number of cases that would have been detected had all decedents undergone necropsy. These estimates were based on available data for necropsy rates, time period, country (US v non-US), and case mix.

**Results:** For all three target diagnoses, adjusting for the estimated prevalence of clinically missed cases among non-necropsied deaths produced sensitivity values outside the 95% confidence interval for the originally reported values, and well below sensitivities reported for the diagnostic tests that are usually used to detect these conditions. For active tuberculosis the sensitivity of antemortem diagnosis decreased from an apparent value of 96% to a corrected value of 83%, with a plausible range of 42–91%; for aortic dissection the sensitivity decreased from 86% to 74%; and for pulmonary embolism the reduction fell only modestly from 97% to 91% but was still lower than generally reported values of 98% or more.

**Conclusions:** Failure to adjust for the prevalence of missed cases among non-necropsied deaths may substantially overstate the performance of diagnostic tests and antemortem diagnosis in general, especially for conditions with high early case fatality.

See end of article for authors' affiliations

Correspondence to:  
Dr K G Shojania, The  
Ottawa Hospital - Civic  
Campus, 1053 Carling  
Avenue, Room C403, Box  
693, Ottawa, ON,  
Canada K1Y 4E9;  
kshojania@ohri.ca

Accepted for publication  
31 July 2005

Diagnostic errors have received surprisingly little attention in the literature on medical errors and patient safety,<sup>1</sup> apart from specific issues such as errors in radiograph interpretation<sup>2</sup> and discussions of the psychological underpinnings of common pitfalls in diagnostic reasoning.<sup>3</sup> One area with a long history of studying errors in clinical diagnosis is the literature on necropsies. Numerous studies document substantial rates of major clinically unsuspected diagnoses detected at necropsy, including missed diagnoses that probably affected outcome.<sup>4–10</sup>

Clinicians have generally attributed these persistent and roughly unchanged<sup>6–8,9</sup> discrepancies between antemortem and postmortem diagnoses to selection bias, arguing that cases sent for necropsy are precisely those in which there is diagnostic uncertainty. Despite its plausibility, this view is not supported by the available evidence.<sup>6–11,12</sup> If rates of clinically important diagnoses first detected at necropsy largely reflected case selection by clinicians, one would expect studies with high necropsy rates to report substantially lower error rates. However, the literature bears out this expectation only weakly. For instance, among recent studies in intensive care settings, a British study in which only 8% of decedents underwent necropsy reported clinically important missed diagnoses in 39% of cases<sup>13</sup>; a French study from an intensive care unit with a necropsy rate of 53%,<sup>10</sup> over six times higher than in the British study, reported clinically important missed diagnoses first detected at necropsy at a slightly lower rate of 32%; and a study from a Belgian intensive care unit reported detecting such diagnoses in 26% of necropsies, despite having a necropsy rate of 93%.<sup>14</sup> Thus, even large increases in necropsy rates were associated with

only modest decreases in the rates at which necropsy revealed clinically important missed diagnoses.

We formally confirmed this weak relationship between clinical selection, as measured by the percentage of decedents undergoing necropsy, and the rates at which necropsy detects clinically important missed diagnoses in a systematic review of the necropsy literature over a 40 year period.<sup>11</sup> In addition to our analysis, several prospective necropsy studies have found clinicians to have little<sup>4–5,15–19</sup> or no<sup>7</sup> ability to predict cases in which necropsy will reveal important diagnoses that escaped clinical detection.

Despite these findings, there exists a general perception that necropsies are no longer necessary as antemortem diagnosis identifies the principal cause of death and other clinically significant diagnoses in the vast majority of cases. This perception has undoubtedly contributed to the progressive decline in necropsy rates over the past 30–40 years. In most jurisdictions, including the UK,<sup>20</sup> Australia,<sup>21</sup> France,<sup>22</sup> and the United States,<sup>23</sup> 10% or fewer of all natural deaths undergo necropsy. These average rates reflect a wide range of institutional necropsy rates, with a small number continuing to performing relatively frequent necropsies but many performing almost none. A survey of hospitals in Louisiana revealed that 63% of hospitals performed no necropsies in a single year.<sup>24</sup> Consequently, many clinicians, radiologists, and others involved in the antemortem diagnostic process never have the opportunity to learn of major missed diagnoses among patients who died under their care, further contributing to the perception that necropsies are no longer necessary.

In addition to affecting perceptions about the performance of antemortem diagnostic performance, low necropsy rates

affect formal assessments of diagnostic performance. Even assuming that necropsied deaths are more likely to include missed diagnoses than non-necropsied deaths (an assumption only weakly borne out by the available evidence), the non-necropsied deaths outnumber necropsied deaths by a factor of 10–20. Consequently, the number of important missed diagnoses among non-necropsied deaths could approximate or even exceed the number observed at necropsy. In conditions where misdiagnosis confers substantial short term mortality, conventional estimates of diagnostic performance—for individual tests or for the entirety of antemortem diagnosis—may substantially overstate diagnostic performance because they do not take into account the possibility of missed cases among non-necropsied deaths. We explore this possibility using the results of our previous analysis<sup>11</sup> and published estimates of the sensitivity of antemortem diagnosis for acute aortic dissection, pulmonary embolism, and active tuberculosis—three conditions of major clinical significance known to escape clinical detection until necropsy.<sup>6–9 12 18 25–37</sup>

## METHODS

### Definitions

The necropsy literature defines “major errors” as clinically missed diagnoses involving a principal underlying disease or primary cause of death.<sup>6 12</sup> Class I errors are defined as clinically missed diagnoses which, if detected during life, “would”, “could”, possibly” or “might” have affected patient prognosis or outcome—at a minimum, discharge from the hospital alive. The studies included in the previous systematic review<sup>11</sup> all used this classification system or reported data in such a manner that we could reclassify cases in this way. However, we avoid referring to these cases as “diagnostic errors” in the present paper as we do not know the extent to which they reflect atypical presentations or acceptable limitations to current diagnosis. We refer instead to rates of diagnostic discrepancies, meaning discrepancies between antemortem and postmortem diagnoses. Removing judgments about errors and referring only to diagnostic discrepancies also reflects the recognition that, like any diagnostic test, the necropsy has an error rate of its own.<sup>38</sup>

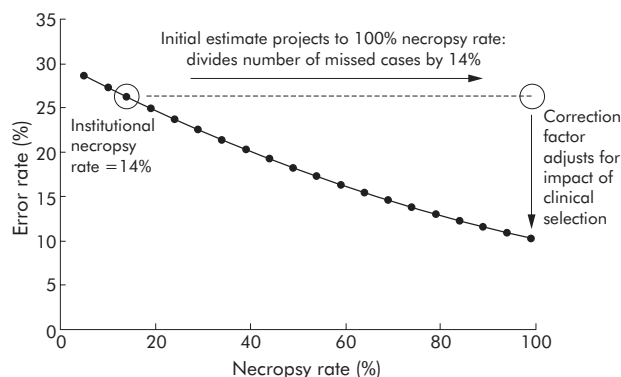
### Search strategy and study selection

We supplemented our previous systematic review<sup>11</sup> with an extensive search for studies reporting data relevant to calculating the sensitivity of antemortem diagnosis. Eligible studies reported case series of patients with acute aortic dissection, pulmonary embolism, or active tuberculosis in which investigators explicitly identified clinically missed cases detected at necropsy. For each target condition we selected the most recent eligible study to emphasize assessments of contemporary diagnostic performance.

### Estimating clinical diagnostic sensitivity

The sensitivity of antemortem diagnosis for a particular target condition equals the ratio of true positive diagnoses to the sum of true positives and false negatives. Most studies estimate false negatives using subsequent clinical follow up and/or cases first detected at necropsy, ignoring the possibility of clinically missed cases among non-necropsied deaths.

Although the prevalence of undetected diagnoses among non-necropsied deaths cannot be calculated with certainty, our previous analysis<sup>11</sup> permits estimation of the total number of clinically missed diagnoses among all deaths—not just necropsied deaths—if the necropsy rate, time period, and case mix are known. Thus, a conservative estimate of the sensitivity of clinical diagnosis depends only on adding to the above denominator a term for clinically missed cases among



**Figure 1** Graph showing major diagnostic errors as a function of necropsy rate, adjusting for case mix, study period, and country.<sup>13</sup> The parameters of the curve correspond to those in the larger of the two aortic dissection studies discussed in the paper<sup>41</sup> (case mix=general hospital necropsies, year=1990, country=US). The calculation to estimate the corrected sensitivity of antemortem diagnosis is illustrated below using the data from this study.<sup>41</sup> The study reported 77 cases of aortic dissection detected antemortem and six cases first diagnosed at necropsy<sup>41</sup> giving an apparent sensitivity of  $77/(77 + 6) = 93\%$ . The denominator in this estimate should include all clinically missed cases (i.e. necropsied and non-necropsied cases). Given the institutional necropsy rate of 14% (indicated on the figure), a rough estimate of the total number of clinically missed cases would be 6 divided by 0.14 = 43 cases. As indicated by the horizontal dashed line, this result overestimates clinical false negatives because it assumes that the same proportion of errors would have been observed if the necropsy rate were 100%. Because the relationship between necropsy rates and major error rates is roughly linear, the correction for the initial estimate of 43 missed cases can be approximated by the ratio of the diagnostic error rate expected if the institution had a necropsy rate of 100% (in this case 10%) to the error rate observed with its actual necropsy rate (in this case 26%). Thus, the correction for this case series would be  $0.10/0.26 = 0.38$ , making the estimated total number of clinically missed cases  $0.38 \times 43 = 16$ . The corrected estimate of the sensitivity of clinical diagnosis in this example would therefore be  $77/(77+16) = 82\%$ .

non-necropsied deaths. We deem this estimate conservative because it assumes all clinical diagnoses are true positives and ignores clinically missed diagnoses that have been lost to follow up.

### Estimating missed cases among non-necropsied deaths

If necropsies occurred independently of the chance of detecting clinically missed diagnoses, the total number of clinically missed cases among patients who died would equal the number of clinically missed cases detected at necropsy divided by the necropsy rate. This result would overestimate clinical false negatives as it assumes that missed cases are distributed evenly between necropsied and non-necropsied deaths—that is, the absence of any selection by clinicians. Because the relationship between necropsy rates and detection rates for missed diagnoses is roughly linear (fig 1), we can correct for this overestimate using the ratio of the diagnostic error rate expected if the institution had a necropsy rate of 100% to the error rate observed with its actual necropsy rate. This correction assumes that the analysis of diagnostic errors as a group applies to errors involving specific target conditions such as aortic dissection, pulmonary embolism, and tuberculosis. This assumption seems reasonable because these diagnoses commonly appear in the necropsy studies that generated these estimates in the first place.<sup>39</sup>

### Sources of relevant necropsy rates

Because none of the clinical studies (table 1) reported necropsy rates, we drew relevant data from other

**Table 1** Studies and data sources for target diagnoses

Target diagnosis	Patients	Setting	Clinically detected cases	Clinically missed cases detected at necropsy
Aortic dissection <sup>40</sup>	Patients admitted through emergency department and diagnosed with aortic dissection at any point during hospitalization or at necropsy	Three hospitals affiliated with University of Pittsburgh; 1992–6	37	6
Aortic dissection <sup>41</sup>	Patients diagnosed at any point during hospitalization or at necropsy	Henry Ford Health System, Detroit; 1985–94	77	6
Pulmonary embolism <sup>42</sup>	International registry of patients with suspected or confirmed pulmonary embolism	52 teaching hospitals in seven countries; 1995–6	2393	61
Active tuberculosis <sup>43</sup>	Cases of active tuberculosis, including those diagnosed after death*, reported to a regional department of public health	San Francisco; 1986–95	2982	120*

\*Tuberculosis first detected after death were those cases identified at necropsy or using antemortem cultures that became positive after death in patients receiving fewer than two anti-tuberculosis medications during the last 24 hours of life, as defined in other epidemiological studies.<sup>46</sup>

sources.<sup>23 37 44 45</sup> For the most recent study of aortic dissection<sup>40</sup> a study from the same institution reported a necropsy rate for 1994, the mid point of the aortic dissection study period. For the second aortic dissection study<sup>41</sup> (included because of the small size of the most recent study) we contacted the study's first author to ascertain the necropsy rate during the study period. We used 14%, the rate from the study mid point, as the best estimate of the relevant necropsy rate and conducted a sensitivity analysis using the rates from the beginning and end of the 10 year study period (table 2). Necropsy rates for the pulmonary embolism study<sup>42</sup> were neither reported nor available from the investigators due to the large number of participating institutions. We therefore used median necropsy rates obtained from a survey of 410 institutions in the US and Canada<sup>44</sup> during a comparable time period.

For the study of tuberculosis detected after death<sup>43</sup> we approximated necropsy rates for hospitals in the San Francisco area using national data from the same period. In 1990, the mid point of the study's observation period, 7% of natural deaths underwent necropsy.<sup>23</sup> In choosing a lower bound we considered the degree to which average necropsy rates tend to be skewed by the high rates achieved by a minority of institutions. For instance, the previously mentioned survey of 410 necropsy departments in the US and Canada<sup>44</sup> reported a very skewed distribution with a small number of institutions reporting very high necropsy rates.

Even more dramatically, a recent survey of Louisiana hospitals reported that 63% of hospitals performed no necropsies in a single year.<sup>24</sup> This study reported a mean necropsy rate of 5% but a median rate of 2%. Using this same ratio of mean to median necropsy rates, we conducted our analysis of the San Francisco tuberculosis study using 3% as the lower bound of the relevant necropsy rate (table 2). For the upper bound of the range of relevant necropsy rates we used the necropsy rate at San Francisco General Hospital which contributed the majority of detected cases (Kathryn DeRiemer, personal communication). In 1990, the mid point of the study period, San Francisco General Hospital had a necropsy rate of 17% (Walter Finkbeiner, personal communication). This value was counted as an upper bound because other hospitals in the San Francisco area would be unlikely to have a necropsy rate higher than a large teaching hospital that is also the only level 1 trauma center in the city.

## RESULTS

### Identified studies of clinical diagnostic performance

Table 1 presents the most recent study meeting our inclusion criteria for each of the three target diagnoses.<sup>40–43</sup> The most recent included aortic dissection study involved 43 patients initially evaluated in the emergency departments of three University of Pittsburgh hospitals from 1992–6 with six cases first detected at necropsy.<sup>40</sup> Because the most recent study of aortic dissection was relatively small, we also included the

**Table 2** Relevant necropsy rates and correction factors for estimating clinically missed diagnoses among non-necropsied deaths

Target diagnosis and source of data for sensitivity of clinical diagnosis	Best estimate for relevant necropsy rate (range of values used to test analysis)	Source of estimate
Aortic dissection (University of Pittsburgh) <sup>40</sup>	19% (14–24%)	Study <sup>37</sup> from same institution in midpoint of aortic dissection study period; $\pm 5\%$ used as potential range to test analysis
Aortic dissection (Henry Ford Hospital, Detroit) <sup>41</sup>	14% (10–18%)	Best estimate and boundaries of range for sensitivity analysis correspond to necropsy rates from mid point, beginning, and end of study period, based on personal communication with study's first author <sup>41</sup> and data presented in another study <sup>45</sup> from same institution
Pulmonary embolism (seven countries in Europe and North America) <sup>42</sup>	15% (8–19%)	Median necropsy rates for 410 institutions in the US and Canada <sup>44</sup> during a comparable time period; lower bound corresponds to non-teaching hospitals and upper bound to hospitals with pathology training programs
Active tuberculosis (all hospitals reporting to San Francisco Department of Health) <sup>43</sup>	7% (3–17%)	Best estimate and lower bound based on national <sup>23</sup> and state <sup>24</sup> level data; upper bound corresponds to necropsy rate at hospital providing majority of cases

next most recent eligible study.<sup>41</sup> This study of aortic dissection also reported six cases first detected at necropsy, but with a total of 77 cases diagnosed before death. The International Cooperative Pulmonary Embolism Registry<sup>42</sup> included 2454 patients with suspected or confirmed acute pulmonary embolism from 52 hospitals in seven countries. In 61 cases detection first occurred at necropsy. We confirmed the completeness of the registry with the study's first author who estimated that the Registry had missed no more than 5% of eligible patients at participating institutions (Samuel Z Goldhaber, personal communication). The most recent study of clinically missed active tuberculosis reviewed all cases reported to the San Francisco Department of Health from 1986 to 1995.<sup>43</sup> Among 3102 reported cases of tuberculosis, 120 (3.9%) met the definition for diagnosis after death. We considered discarding this study because of the concern that the prevalence of AIDS in San Francisco might limit generalisability. However, a study at the national level reported approximately the same result, with 5.1% of cases meeting the same definition for diagnosis after death.<sup>46</sup> Given this close agreement and because we had access to key data from the investigators in the San Francisco study,<sup>43</sup> we chose to keep this study instead of the national one.<sup>46</sup>

### Corrected estimates of clinical sensitivity

The contribution of missed cases among non-necropsied deaths substantially lowered the sensitivity of clinical diagnosis for all three target conditions (table 3). With the exception of the smaller study of aortic dissection,<sup>40</sup> the corrected sensitivities were outside the 95% confidence intervals associated with the apparent sensitivities implied by the original studies.<sup>41–43</sup> For tuberculosis the adjusted sensitivity associated with the lower bound of our estimate for the underlying necropsy rate is strikingly low at 42%. Even using 5% (instead of 3%) as the lower bound for the underlying necropsy rate, the adjusted sensitivity for detecting tuberculosis before death would be just 68%.

In addition to the question of the appropriate range for the underlying necropsy rate, the analysis of clinically missed cases of tuberculosis involved two issues that did not apply to the studies of aortic dissection and pulmonary embolism. Firstly, case identification relied on Public Health Department records which may underestimate clinically detected cases by almost 20%.<sup>47</sup> We repeated our analysis correcting for this potential underestimation of clinically detected cases, but the correction exerted minimal impact given the difference in magnitude between the numbers of clinically detected cases and cases detected after death. Secondly, some of the clinically missed cases may have been identified by antemortem culture results that became positive after death, independent of a necropsy. The first author of the study confirmed that 13% of cases detected after death occurred on the basis of antemortem culture results (Kathryn DeRiemer, personal communication). Including cases

detected by antemortem culture results among the cases detected after death can be regarded as increasing the effective necropsy rate—that is, with respect to diagnosing tuberculosis, sending cultures is equivalent to performing more necropsies. Since such a small percentage of cases was diagnosed in this manner, the range of values we considered for the underlying necropsy rate (3–17%) captured this effect.

### DISCUSSION

Adjusting for clinically missed cases among non-necropsied deaths substantially lowered the sensitivity of clinical diagnosis for tuberculosis and aortic dissection. While not as dramatic, the decreased sensitivity for acute pulmonary embolism from 97% to 94% (with a plausible range extending as low as 90%) is still noteworthy because the corrected value lies outside the generally accepted miss rate of 4%, based on the reported false negative rate for normal lung scintigraphy.<sup>48</sup> Certainly, a 90–94% sensitivity is well below the approximately 99% sensitivity reported for a clinical algorithm for the investigation of pulmonary embolism<sup>49</sup> and for helical computed tomography.<sup>50</sup>

The major limitation of the analysis concerns the number of assumptions required to generate the corrected estimates for the performance of antemortem diagnosis. Given the need for these assumptions, the conservative nature of our analysis deserves special attention. Firstly, we ignored false positive clinical diagnoses so that cases of aortic dissection, pulmonary embolism, and tuberculosis were counted as “true positives” even though clinical overdiagnosis clearly occurs for pulmonary embolism and tuberculosis.<sup>48–51</sup> Secondly, we ignored clinically missed cases lost to follow up such as those resulting in outpatient death or admission to another institution. Thirdly, we counted as “clinically detected” all cases diagnosed any time before death, thus ignoring the problem of delayed diagnosis.<sup>52–55</sup> We also counted serendipitous diagnoses as clinically detected cases. Studies of aortic dissection specifically noted that a substantial proportion of cases were detected as a result of investigations intended to diagnose other conditions.<sup>40–41–56</sup> Thus, while the details of the calculations used to generate the corrected estimates of antemortem diagnostic performance involve some speculative assumptions, the overall framework of the analysis heavily favours antemortem diagnosis. Our results are therefore unlikely to represent an overstatement of the problem.

One limitation of the calculations we employed concerns the instability of our estimates of corrected diagnostic performance with very low necropsy rates. For example, in the study of tuberculosis from San Francisco area hospitals, changing the lower bound for the underlying necropsy rate from 3% to 5% increased the corrected sensitivity from 68% to 77%. We would suggest, however, that this instability provides an argument for maintaining necropsy rates higher than seen in contemporary hospitals. In other words,

**Table 3** Reported diagnostic sensitivity and estimates corrected for missed cases among non-necropsied deaths

Target diagnosis	Reported sensitivity of clinical diagnosis* (95% CI)	Sensitivity of clinical diagnosis adjusted for missed cases among non-necropsied deaths (possible range)†
Aortic dissection (Pittsburgh) <sup>40</sup>	86% (71 to 94)	74% (70–78%)
Aortic dissection (Detroit) <sup>41</sup>	93% (84 to 97)	82% (77–85%)
Pulmonary embolism <sup>42</sup>	97% (96 to 98)	94% (90–95%)
Active tuberculosis <sup>43</sup>	96% (95 to 97)	83% (68–91%)

\*Reported sensitivity equals clinically detected cases divided by sum of clinically detected cases and those first detected at necropsy (derived from table 1).

†Ranges represent likely ranges of values for each estimate, with justifications explained in text and in table 2.



necropsy rates in the 0–10% range simply do not permit reliable assessment of the performance of clinical diagnosis.

Another limitation relates to the age of the studies, which reported data collected approximately 10–15 years ago. Antemortem diagnosis may have improved since then, especially in the case of pulmonary embolism, given the technological advances with helical computed tomography. On the other hand, of the three conditions, antemortem diagnosis of pulmonary embolism may be the one most affected by false positive clinical diagnoses. With the advent of helical computed tomography as the primary mode of evaluation for patients with suspected pulmonary embolism, clinicians tend to treat virtually all positive results without further confirmatory testing.<sup>57</sup> Preliminary results from Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) II confirm the problem of false positive diagnoses among patients who did not have a high pretest probability of disease, which represents the majority of patients referred for testing.<sup>58</sup> Thus, while technological advances may have occurred, the high yield of antemortem diagnosis for pulmonary embolism probably includes a substantial number of false positive cases, inflating the apparent sensitivity of antemortem performance.

An additional limitation of our analysis relates to the necropsy itself. Like most complex procedures involving multiple observational and cognitive elements, the necropsy almost certainly has an error rate of its own, although this issue has received relatively little attention.<sup>58</sup> Few data address the properties of the necropsy as a diagnostic test, such as inter-rater reliability.<sup>59</sup> Future research addressing this issue would be very valuable. However, the three conditions analysed in the present work can be diagnosed at necropsy with little ambiguity. Pulmonary embolism offers some room for error as clotting can occur after death. However, specific macroscopic and microscopic characteristics reliably distinguish antemortem and postmortem thromboemboli.<sup>60–63</sup>

As mentioned earlier, the necropsy literature frequently refers to clinically missed diagnoses as “diagnostic errors”. We have avoided this term because many so-called “errors” undoubtedly represent atypical presentations or acceptable limitations to current diagnosis. On the other hand, increased awareness of the full spectrum of clinical presentations associated with a given condition can prompt revision of what constitutes an atypical presentation. For example, abrupt onset of pain, the presence of any pulse deficit, or the presence of a murmur of aortic regurgitation—three classic signs of acute aortic dissection—are each significantly less likely to occur in patients over 70 years of age with aortic dissection than in younger patients presenting with acute dissection.<sup>64</sup>

Part of the reduction in sensitivity we have illustrated probably reflects discrepancies between effectiveness and efficacy. Just as therapeutic outcomes obtained in routine practice can diverge from the results reported in clinical trials,<sup>65–67</sup> diagnostic tests may exhibit important discrepancies between efficacy and effectiveness. For example, despite the impressive results in formal evaluations of the sensitivity and specificity of computed tomography in diagnosing appendicitis,<sup>68</sup> a population level analysis indicated no change in relevant outcomes such as finding a normal or ruptured appendix at laparotomy.<sup>69</sup> Suboptimal diagnostic performance, whether for a particular test or the entirety of antemortem diagnosis, may reflect a variety of process failures from the adequacy of clinical examination<sup>41</sup> to the appropriateness of test ordering, adequacy of test interpretation, and even eroded performance of diagnostic tests in patient populations typically excluded from formal evaluations of diagnostic tests.<sup>70 71</sup>

Regardless of their causes, clinically significant diagnoses escape detection at a much greater rate than generally appreciated, and clinicians must entertain the possibility of missed causes of death when considering whether or not to request a necropsy. Equally important, investigators who assess diagnostic tests by including clinically missed cases detected only at necropsy overestimate diagnostic sensitivity, because low necropsy rates lead to underestimation of the false negative rate. We have shown that, given the low necropsy rates in both teaching and non-teaching hospitals, including necropsy detected cases by itself is insufficient. The assessment of sensitivity for any diagnostic process or test must adjust for the substantial prevalence of clinically missed cases among non-necropsied deaths. Interventions to increase necropsy rates will be required if accurate measures of diagnostic sensitivity are to be calculated for key clinical diagnoses.

#### Authors' affiliations

**K G Shojania, I Goldman**, Department of Medicine, University of California San Francisco, CA, USA

**E C Burton**, Department of Pathology and Laboratory Medicine, Baylor Health Care System, USA

**K M McDonald**, Center for Primary Care and Outcomes Research, Stanford University, CA, USA

This article is based in part on work performed by the UCSF-Stanford Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (Contract No. 290-970013), Rockville, MD. The authors are responsible for the contents of this article. No statement in this article should be construed as an official position of the Agency for Healthcare Research and Quality or of the US Department of Health and Human Services.

Dr Shojania holds a Canada Research Chair in Patient Safety and Quality Improvement.

#### REFERENCES

- 1 **Margo CE**. A pilot study in ophthalmology of inter-rater reliability in classifying diagnostic errors: an underinvestigated area of medical error. *Qual Saf Health Care* 2003;**12**:416–20.
- 2 **Espinosa JA, Nolan TW**. Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study. *BMJ* 2000;**320**:737–40.
- 3 **Graber M**. Metacognitive training to reduce diagnostic errors: ready for prime time? *Acad Med* 2003;**78**:781.
- 4 **Britton M**. Diagnostic errors discovered at autopsy. *Acta Med Scand* 1974;**196**:203–10.
- 5 **Cameron HM, McGoogan E, Watson H**. Necropsy: a yardstick for clinical diagnoses. *BMJ* 1980;**281**:985–8.
- 6 **Goldman L, Sayson R, Robbins S, et al**. The value of the autopsy in three medical eras. *N Engl J Med* 1983;**308**:1000–5.
- 7 **Landefeld CS, Chren MM, Myers A, et al**. Diagnostic yield of the autopsy in a university hospital and a community hospital. *N Engl J Med* 1988;**318**:1249–54.
- 8 **Kirch W, Schafii C**. Misdiagnosis at a university hospital in four medical eras. *Medicine (Baltimore)* 1996;**75**:29–40.
- 9 **Sonderegger-Iseli K, Burger S, Muntwyler J, et al**. Diagnostic errors in three medical eras: a necropsy study. *Lancet* 2000;**355**:2027–31.
- 10 **Combes A, Mokhtari M, Couvelard A, et al**. Clinical and autopsy diagnoses in the intensive care unit: a prospective study. *Arch Intern Med* 2004;**164**:389–92.
- 11 **Shojania KG, Burton EC, McDonald KM, et al**. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA* 2003;**289**:2849–56.
- 12 **Battle RM, Pathak D, Humble CG, et al**. Factors influencing discrepancies between premortem and postmortem diagnoses. *JAMA* 1987;**258**:339–44.
- 13 **Perkins GD, McAuley DF, Davies S, et al**. Discrepancies between clinical and postmortem diagnoses in critically ill patients: an observational study. *Crit Care* 2003;**7**:R129–32.
- 14 **Roosen J, Frans E, Wilmer A, et al**. Comparison of premortem clinical diagnoses in critically ill patients and subsequent autopsy findings. *Mayo Clin Proc* 2000;**75**:562–7.
- 15 **Heasman MA, Lipworth L**. *Accuracy of certification of cause of death*. London: HMSO, 1966.
- 16 **Hartveit F**. Clinical and post-mortem assessment of the cause of death. *J Pathol* 1977;**123**:193–210.
- 17 **Cameron HM, McGoogan E**. A prospective study of 1152 hospital autopsies: I. Inaccuracies in death certification. *J Pathol* 1981;**133**:273–83.

- 18 **Cameron HM**, McGoogan E. A prospective study of 1152 hospital autopsies: II. Analysis of inaccuracies in clinical diagnoses and their significance. *J Pathol* 1981;**133**:285–300.
- 19 **Boers M**, Nieuwenhuysen Kruseman AC, Eulderink F, *et al*. Value of autopsy in internal medicine: a 1-year prospective study of hospital deaths. *Eur J Clin Invest* 1988;**18**:314–20.
- 20 **Burton JL**, Underwood JC. Necropsy practice after the "organ retention scandal": requests, performance, and tissue retention. *J Clin Pathol* 2003;**56**:537–41.
- 21 **Royal College of Pathologists of Australasia Autopsy Working Party**. The decline of the hospital autopsy: a safety and quality issue for healthcare in Australia. *Med J Aust* 2004;**180**:281–5.
- 22 **Chariot P**, Witt K, Pautot V, *et al*. Declining autopsy rate in a French hospital: physicians' attitudes to the autopsy and use of autopsy material in research publications. *Arch Pathol Lab Med* 2000;**124**:739–45.
- 23 **Burton EC**, Nemetz PN. Medical error and outcomes measures: where have all the autopsies gone? *Medscape General Medicine*. 2000;2000:E8, Available at <http://www.medscape.com/medscape/GeneralMedicine/journal/2000/v2002.n2002/mgm0428.burt/mgm0428.burt.html>.
- 24 **Burton EC**, Nemetz PN, Fisher WPJ, *et al*. Institutional characteristics associated with hospital autopsy performance: a survey of Louisiana hospitals. *Am J Clin Pathol* 2001;**116**:590.
- 25 **Bedell SE**, Fulton EJ. Unexpected findings and complications at autopsy after cardiopulmonary resuscitation (CPR). *Arch Intern Med* 1986;**146**:1725–8.
- 26 **Papadakis MA**, Mangione CM, Lee KK, *et al*. Treatable abdominal pathologic conditions and unsuspected malignant neoplasms at autopsy in veterans who received mechanical ventilation. *JAMA* 1991;**265**:885–7.
- 27 **Tai DY**, El-Bilbeisi H, Tewari S, *et al*. A study of consecutive autopsies in a medical ICU: a comparison of clinical cause of death and autopsy diagnosis. *Chest* 2001;**119**:530–6.
- 28 **Twigg SJ**, McCrerrick A, Sanderson PM. A comparison of post mortem findings with post hoc estimated clinical diagnoses of patients who die in a United Kingdom intensive care unit. *Intensive Care Med* 2001;**27**:706–10.
- 29 **Ryu JH**, Olson EJ, Pellikka PA. Clinical recognition of pulmonary embolism: problem of unrecognized and asymptomatic cases. *Mayo Clin Proc* 1998;**73**:873–9.
- 30 **Burke MC**, Aghababian RV, Blackburne B. Use of autopsy results in the emergency department quality assurance plan. *Ann Emerg Med* 1990;**19**:363–6.
- 31 **Bernicker EH**, Atmar RL, Schaffner DL, *et al*. Unanticipated diagnoses found at autopsy in an urban public teaching hospital. *Am J Med Sci* 1996;**311**:215–20.
- 32 **Mitchell ML**. Interdepartmental quality assurance using coded autopsy results. *Mod Pathol* 1993;**6**:48–52.
- 33 **Tse GM**, Lee JC. A 12-month review of autopsies performed at a university-affiliated teaching hospital in Hong Kong. *Hong Kong Med J* 2000;**6**:190–4.
- 34 **Bobrowitz ID**. Active tuberculosis undiagnosed until autopsy. *Am J Med* 1982;**72**:650–8.
- 35 **Lee JK**, Ng TH. Undiagnosed tuberculosis in hospitalized patients: an autopsy survey. *J R Soc Health* 1990;**110**:141–3.
- 36 **Pineda LA**, Hathwar VS, Grant BJ. Clinical suspicion of fatal pulmonary embolism. *Chest* 2001;**120**:791–5.
- 37 **Nichols L**, Aronica P, Babe C. Are autopsies obsolete? *Am J Clin Pathol* 1998;**110**:210–8.
- 38 **Saracci R**. Is necropsy a valid monitor of clinical diagnosis performance? *BMJ* 1991;**303**:898–900.
- 39 **Shojania KG**, Burton EC, McDonald KM, *et al*. *The autopsy as an outcome and performance measure*. Evidence Report/Technology Assessment No 58. Prepared by the UCSF-Stanford Evidence-based Practice Center under AHRQ contract no.290-97-0013. Publication No.02-E020. Rockville, MD: Agency for Healthcare Research and Quality, 2002. Available at <http://www.dhcrp.gov/clinic/evrpfifiles.htm#autopsy>.
- 40 **Sullivan PR**, Wolfson AB, Leckey RD, *et al*. Diagnosis of acute thoracic aortic dissection in the emergency department. *Am J Emerg Med* 2000;**18**:46–50.
- 41 **Rosman HS**, Patel S, Borzok S, *et al*. Quality of history taking in patients with aortic dissection. *Chest* 1998;**114**:793–5.
- 42 **Goldhaber SZ**, Visani L, De Rosa M. Acute pulmonary embolism: clinical outcomes in the International Cooperative Pulmonary Embolism Registry (ICOPER). *Lancet* 1999;**353**:1386–9.
- 43 **DeRiemer K**, Rudoy I, Schecter GF, *et al*. The epidemiology of tuberculosis diagnosed after death in San Francisco, 1986–1995. *Int J Tuberc Lung Dis* 1999;**3**:488–93.
- 44 **Baker PB**, Zarbo RJ, Howanitz PJ. Quality assurance of autopsy face sheet reporting, final autopsy report turnaround time, and autopsy rates: a College of American Pathologists Q-Probes study of 10003 autopsies from 418 institutions. *Arch Pathol Lab Med* 1996;**120**:1003–8.
- 45 **Stein PD**, Henry JW. Prevalence of acute pulmonary embolism among patients in a general hospital and at autopsy. *Chest* 1995;**108**:978–81.
- 46 **Rieder HL**, Kelly GD, Bloch AB, *et al*. Tuberculosis diagnosed at death in the United States. *Chest* 1991;**100**:678–81.
- 47 **Yokoe DS**, Subramanyan GS, Nardell E, *et al*. Supplementing tuberculosis surveillance with automated data from health maintenance organizations. *Emerg Infect Dis* 1999;**5**:779–87.
- 48 **PIOPED Investigators**. Value of the ventilation/perfusion scan in acute pulmonary embolism. Results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). *JAMA* 1990;**263**:2753–9.
- 49 **Wells PS**, Ginsberg JS, Anderson DR, *et al*. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med* 1998;**129**:997–1005.
- 50 **Quiroz R**, Kucher N, Zou KH, *et al*. Clinical validity of a negative computed tomography scan in patients with suspected pulmonary embolism: a systematic review. *JAMA* 2005;**293**:2012–7.
- 51 **Gordin FM**, Slutkin G, Schecter G, *et al*. Presumptive diagnosis and treatment of pulmonary tuberculosis based on radiographic findings. *Am Rev Respir Dis* 1989;**139**:1090–3.
- 52 **Greenaway C**, Menzies D, Fanning A, *et al*. Delay in diagnosis among hospitalized patients with active tuberculosis: predictors and outcomes. *Am J Respir Crit Care Med* 2002;**165**:927–33.
- 53 **Sherman LF**, Fujiwara PI, Cook SV, *et al*. Patient and health care system delays in the diagnosis and treatment of tuberculosis. *Int J Tuberc Lung Dis* 1999;**3**:1088–95.
- 54 **Mathur P**, Sacks L, Auten G, *et al*. Delayed diagnosis of pulmonary tuberculosis in city hospitals. *Arch Intern Med* 1994;**154**:306–10.
- 55 **Butler J**, Ormerod OJ, Giannopoulos N, *et al*. Diagnostic delay and outcome in surgery for type A aortic dissection. *Q J Med* 1991;**79**:391–6.
- 56 **Spittell PC**, Spittell JA Jr, Joyce JW, *et al*. Clinical features and differential diagnosis of aortic dissection: experience with 236 cases (1980 through 1990). *Mayo Clin Proc* 1993;**68**:642–51.
- 57 **Trowbridge RL**, Araoz PA, Gotway MB, *et al*. The effect of helical computed tomography on diagnostic and treatment strategies in patients with suspected pulmonary embolism. *Am J Med* 2004;**116**:84–90.
- 58 **Bui C**. PIOPED II: Is spiral CT the best and only test for suspected pulmonary embolism? Highlights of the Radiological Society of North America (RSNA) 90th Scientific Assembly and Annual Meeting. *Medscape General Medicine*, 2004. Available at <http://www.medscape.com/viewarticle/499489>.
- 59 **Veress B**, Gadaleanu V, Nennesmo I, *et al*. The reliability of autopsy diagnostics: inter-observer variation between pathologists, a preliminary report. *Qual Assur Health Care* 1993;**5**:333–7.
- 60 **Morrell MT**, Dunnill MS. The post-mortem incidence of pulmonary embolism in a hospital population. *Br J Surg* 1968;**55**:347–52.
- 61 **Matsubara O**, Sato T, Kitagawa M, *et al*. Pulmonary thromboembolism: frequency, site, and vascular change in 103 autopsy cases. *Am J Cardiovasc Pathol* 1989;**2**:321–8.
- 62 **Wagenvoort CA**. Pathology of pulmonary thromboembolism. *Chest* 1995;**107**:10–75.
- 63 **Ro A**, Kageyama N, Tanifuji T, *et al*. Histopathological study of pulmonary arteries in 14 autopsy cases with massive pulmonary thromboembolism. *Leg Med (Tokyo)* 2003;**5**(Suppl 1):S315–7.
- 64 **Mehta RH**, O'Gara PT, Bossone E, *et al*. Acute type A aortic dissection in the elderly: clinical characteristics, management, and outcomes in the current era. *J Am Coll Cardiol* 2002;**40**:685–92.
- 65 **Gould DA**, Birkmeyer JD. Efficacy versus effectiveness of carotid endarterectomy. *Effect Clin Pract* 1999;**2**:30–6.
- 66 **Soumerai SB**, McLaughlin TJ, Ross-Degnan D, *et al*. Effectiveness of thrombolytic therapy for acute myocardial infarction in the elderly: cause for concern in the old-old. *Arch Intern Med* 2002;**162**:561–8.
- 67 **Connors AF**, Speroff T, Dawson NV, *et al*. The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *JAMA* 1996;**276**:889–97.
- 68 **Rao PM**, Rhea JT, Novelline RA, *et al*. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *N Engl J Med* 1998;**338**:141–6.
- 69 **Flum DR**, Morris A, Koepsell T, *et al*. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. *JAMA* 2001;**286**:1748–53.
- 70 **Mulherin SA**, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;**137**:598–602.
- 71 **Ransohoff DF**, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;**299**:926–30.