Summary

Overview

An extensive literature documents a high prevalence of errors in clinical diagnosis discovered at autopsy. Multiple studies have suggested no significant decrease in these errors over time. Despite these findings, autopsies have dramatically decreased in frequency in the United States and many other countries. In 1994, the last year for which national U.S. data exist, the autopsy rate for all non-forensic deaths fell below 6%. The marked decline in autopsy rates from previous rates of 40-50% undoubtedly reflects various factors, including reimbursement issues, the attitudes of clinicians regarding the utility of autopsies in the setting of other diagnostic advances, and general unfamiliarity with the autopsy and techniques for requesting it, especially among physicians-in-training.

The autopsy is valuable for its role in undergraduate and graduate medical education, the identification and characterization of new diseases, and contributions to the understanding of disease pathogenesis. Although extensive, these benefits are difficult to quantify. This systematic review studied the more easily quantifiable benefits of the autopsy as a tool in performance measurement and improvement. Such benefits largely relate to the role of the autopsy in detecting errors in clinical diagnosis and unsuspected complications of treatment. It is hoped that characterizing the extent to which the autopsy provides data relevant to clinical performance measurement and improvement will help inform strategies for preserving the benefits of routinely obtained autopsies and for considering its wider use as an instrument for quality improvement.

This report does not attempt to address the roles of the autopsy in medical education; furthering medical research; quality control within pathology; verification, second-opinion consultations, and legal documentation of findings; the bereavement process for surviving family members; or other benefits that are described in many of the sources listed in the bibliography (Appendix F). In addition to being difficult to quantify, these benefits apply primarily to teaching hospitals. To address the role of the autopsy as an outcome measure and tool for quality improvement, the report focuses on benefits likely to apply to all hospitals, such as the detection of important diagnostic errors and related quality problems.

Reporting the Evidence

This report synthesizes the autopsy literature as it relates to the following four key questions:

1. To what extent does the autopsy reveal important diagnoses that were clinically unsuspected prior to death?
2. To what extent does the autopsy provide a useful performance measure or audit of clinical diagnosis in general?
3. What impact do autopsy findings have on clinical performance improvement?
4. To what extent are vital statistics compromised by low autopsy rates?

To address the above questions adequately, we also sought evidence pertaining to the properties of the autopsy as a diagnostic test. Specifically, we looked for any information describing autopsy quality, accuracy, and precision or reproducibility.

It is important to note that, though the phrase "diagnostic error" appears throughout this report, the discrepancies between clinical and autopsy diagnoses to which we refer do not necessarily represent errors in the sense of mistakes, "slips," or other such terms. Some of these discrepancies do undoubtedly result from failures to consider an appropriately broad differential diagnosis, misinterpretation of test results, and other quality problems, so that resulting discrepant diagnoses detected at autopsy do warrant the label "diagnostic errors." However, other such discrepancies clearly represent acceptable limits to clinical diagnosis, based on the performance of current technologies or the occurrence of atypical clinical presentations. (In fact, one of the areas of future research identified by this report involves characterizing the relative distribution of these two types of clinical-autopsy diagnostic discrepancies.) Despite these considerations, we use the term "diagnostic errors" because it appears so commonly in the autopsy literature.

Target Population

The patient population covered in this report includes all patients (e.g., adult and pediatric, male and female, and so on) in various settings, although predominantly consisting of hospitalized patients. We did not specifically exclude medical examiner cases, but few studies from the forensic literature addressed the specific questions posed in this report.

Search Strategy

We conducted an extensive search of the MEDLINE® database, supplemented by hand searches of article bibliographies and consultation with experts in the field. For articles published in languages other than English, we reviewed the abstract (if available) to determine whether or not the study reported methodologies or findings qualitatively different from those described in the English-language literature.

Study Inclusion Criteria

The autopsy literature consists entirely of observational studies, rendering problematic the development of appropriate inclusion and exclusion criteria, as the vast majority of systematic reviews involve at least some randomized controlled trials. In the absence of relevant and well-established quality scoring systems, we adopted fairly minimal inclusion and exclusion criteria. For studies reporting diagnostic error rates detected at autopsy, we required:

- Well-defined patient samples consisting of consecutive or randomly sampled autopsies meeting explicit criteria—convenience samples were excluded.

- Clinical diagnoses derived from autopsy request forms submitted by clinicians or chart review performed by the study investigators—clinical diagnoses derived solely from death certificates were excluded.

- Classification schemes for discrepancies between clinical and autopsy diagnoses conforming to one of three categories—potentially treatable causes of death ("Class I"), other major missed diagnoses, and discrepant disease categorizations based on standard international classification coding. These classifications (defined further in the report) encompass the majority of studies reported in the literature. Studies that reported clinical diagnoses simply as "correct/incorrect" or "confirmed/unconfirmed" were excluded.

Data Collection and Analysis

Articles identified from the literature search were stored in a reference database and categorized according to the study questions addressed. Structured abstraction forms were then used to collect demographic data (pertaining to patients and institutions), salient methodologic features and results. Each article was abstracted by at least two of the four reviewers, including three physicians and one non-physician research assistant. One of the physicians reviewed all of the articles.

Findings

To address the first key question pertaining to the extent to which autopsies reveal clinically unsuspected important diagnoses, we reviewed studies assessing the performance of the autopsy as a diagnostic test. Given the generally accepted role of the autopsy as the ultimate diagnostic standard for many aspects of clinical care, the test characteristics of the autopsy have received surprisingly little attention.

- The quality of the autopsy has received little systematic study, with the only evidence pertaining to perinatal autopsies, where two studies show that deficiencies relative to reporting standards (i.e., a proxy measure for potentially inadequate quality) appear to be common.

- The potential for error or disagreement in autopsy interpretations has been assessed in only one small study. In relation to the determination of principal diagnoses relating to the cause of death in technically adequate autopsy, diagnostic uncertainty persists in 1-5% of cases, although rates of up to 40% have been reported, depending on the type of autopsy cases, e.g., perinatal. Importantly, errors in classification of autopsy diagnoses involving even a few percent of cases substantially distort estimates of the performance of clinical diagnosis when autopsy is used as the gold standard.

- The reproducibility of judgments about errors in clinical diagnosis as indicated by autopsy findings has only been mentioned in passing in the autopsy literature. Studies from the health care quality and medical error literature suggest that reproducibility of similar types of judgments is likely fair to moderate at best.

There is insufficient literature to address: a) the quality of the autopsy, b) the technical adequacy in interpreting autopsy findings, and c) the reliability of judgments made regarding autopsy detected discrepancies. There is also no literature that addresses the quality of training in autopsy pathology or the ability of physicians to utilize autopsy findings.

In terms of the four main study questions:

1. To what extent does the autopsy reveal important diagnoses that were clinically unsuspected prior to death?

- The chance that autopsy will reveal a misdiagnosis that may have affected outcome (i.e., a Class I error) was 10.2% (95% CI: 6.7-15.3%) using data from all studies and the base values of time (1980), autopsy rate (overall mean rate of 44.3%), country (U.S.) and case mix (general autopsies). Restricting the analysis to data from U.S. institutions only yielded a slightly higher point estimate but almost entirely overlapping confidence interval, 11.2% (95% CI: 6.9-17.5%). Adjusting for changes in autopsy rates, and the effects of case mix and the country, the probability of a Class I error showed a relative decrease of 26.2% per decade (p=0.10).

- The base probability of the autopsy detecting a major error in a given case was 25.6% (95% CI: 20.8-31.2%) when data from all institutions were included. Using data from U.S. institutions only, the probability of the autopsy detecting a major error in a given case was slightly lower at 24.0%, but with an almost entirely overlapping 95% CI of 17.6-31.5%. Major error rates also showed a similar decrease over time, but, in contrast to the results for Class I errors, this relationship was statistically significant. Relative to the base rate in 1980, the prevalence of major errors exhibited a relative decrease of 28.0% (95% CI: 9.8-42.6%) per decade.

- The regression analysis supported the expected inverse correlation between error rate and autopsy rate (i.e., that lower autopsy rates produce higher error rates due to selection of diagnostically challenging cases), but this effect is relatively modest. Specifically, every 10% increase in the autopsy rate is associated with a relative decrease in Class I errors of 7.8% (p=0.18). For major errors, this relationship was more substantial and statistically significant, with every 10% increase in autopsies associated with a relative decrease in major errors of 12% (p=0.0003).

- Using the regression model to compute rates of autopsy-detected diagnostic errors over a range of autopsy rates and as a function of time, contemporary (year 2000) autopsies detect Class I errors in 3.8-7.9% of cases and major errors in 8.0-22.8%, of cases. These ranges reflect variations in autopsy rates from 5-100%.

- The weak relationship between autopsy rates and error rates in the general analysis was supplemented by review of studies specifically addressing the issue of clinical selection of diagnostically challenging or uncertain cases. These studies indicated that clinicians cannot reliably predict which autopsies will be of high diagnostic yield, reinforcing the conclusion that the relatively unchanged diagnostic error rates do not simply reflect competing effects of medical progress (leading to fewer errors) and fewer autopsies (leading to selection for cases likely to have errors).

- Because of the recent interest in medical error and patient safety, we specifically looked for studies that reported the proportion of autopsies that detected clinically unsuspected complications of care. These data were usually mentioned in passing in these studies, with no study specifically focusing on this issue. Thus, the extent to which these complications contributed to death (and even the extent to which they were truly unsuspected) was often unclear. For this reason, and because of the heterogeneity of the case mix in the relatively small sample of studies reporting the relevant data, we did not pool estimates for rates of autopsy-detection of unsuspected complications of care. Nonetheless, the 11 studies that did provide data on this point indicated that approximately 1-5% of autopsies disclose unsuspected complications of care.

2. To what extent does the autopsy provide a useful performance measure or audit of clinical diagnosis in general?

- Autopsy studies commonly report diagnostic "error rates," but these error rates involve autopsied cases only. It is commonly assumed that the true denominator of interest is all deaths; hence the interest in increased autopsy rates. However, the denominator of interest for clinical performance measurement is, in fact, all patients receiving care during the autopsy observation period. Only one autopsy study provides any data on clinical diagnoses for patients discharged alive from the hospital during the same observation period as for the autopsy series. Because of the importance of this question, we searched extensively for studies outside the autopsy literature *per se* for potentially relevant studies.

- Specifically, we looked for studies reporting clinical diagnoses and other follow-up data on cohorts of patients (e.g., all patients admitted to a given hospital during a defined observation period), not just the diagnoses obtained for patients who died and went to autopsy. Supplementing autopsy findings with the results of ante mortem diagnostic testing and/or clinical follow-up for patients who did not die permits determination of the numerator and denominator required to assess the sensitivity of clinical diagnosis. Despite an extensive search, we found appropriate studies for only five target conditions: pulmonary embolism (PE), acute myocardial infarction (MI), acute appendicitis, aortic dissection, and active tuberculosis.

- Among these five conditions, the performance of clinical diagnosis exhibited substantial variation, with excellent performance only for acute MI and to a lesser extent PE. Even for these two conditions, the high sensitivities obtained likely overstate clinical performance, as focusing on the dichotomous outcome of correct or incorrect identification of one target condition (PE or MI) obscures the extent to which other important conditions are missed once these target diagnoses are ruled out. A patient who is correctly identified as not having an MI counts as a success, regardless of whether or not the underlying cause of the patient's presenting complaint is ever diagnosed.

3. What impact do autopsy findings have on clinical performance improvement?

- No intervention study has directly addressed the impact of autopsy findings on clinical practice or performance improvement. Consequently, the study objectives in this regard were

not met, including not being able to perform a cost effectiveness analysis, as the effectiveness of the autopsy in reducing errors and other quality problems remains unknown. This does not invalidate the potential role of the autopsy in relation to clinical practice or performance improvement, but does reveal an important gap in the literature.

4. *To what extent are vital statistics compromised by low autopsy rates?*

- Major error rates detected by autopsy indicate substantial inaccuracies in death certificates and hospital discharge data, both of which play important roles in epidemiologic research and health care policy decisions. Previous studies have suggested that these errors roughly cancel each other out (i.e., for a given condition, false positive and false negative diagnoses are roughly equal). However, this finding has not been consistent across studies. Even when present, this balancing effect applies only when considering the most general of diagnostic categories (i.e., cardiovascular, neoplastic, infectious, metabolic, and so on). Thus, the current evidence is adequate to suggest that the epidemiologic data for important diseases such as myocardial infarction, breast cancer, pneumonia, stroke, and so on, all contain substantial inaccuracies—in the 20-30% range reported for major errors.

Conclusions

The findings of this review have different implications depending on the level of analysis—individual clinicians, hospitals, or the health care system as a whole. From the point of view of the individual clinician, the chance that autopsy will reveal important unsuspected diagnoses in a given case remains significant. Moreover, clinicians do not seem able to predict reliably cases in which such findings are more likely to occur. Thus, clinicians have compelling reasons to request autopsies far more often than currently occurs.

At the institutional level, the role of the autopsy is less clear. The prevalence of missed diagnoses among autopsied patients (or even all deaths) provides a numerator, but not a denominator with which to assess the rate at which patients with a given condition remain undiagnosed until death. Using autopsy results to track hospital quality requires not only explicitly defined error rates, but also data on the number of patients discharged alive with diagnoses that appear among the list of conditions first detected at autopsy. Clearly, though, the unexpected findings at autopsy in specific cases are of interest to institutions as a whole and not just the individual treating clinicians. However, no study has ever examined the impact of performing autopsies (and communicating autopsy findings back to clinicians) on institutional performance improvement. This represents a major area for future research, but should not detract from the finding that many institutions perform too few autopsies to allow any meaningful assessment of local diagnostic performance and other quality problems, no matter how communication and feedback to clinicians occurs.

At the level of the entire health care system, existing literature provides two compelling reasons to pursue autopsies. First, results for the five conditions examined in this report suggest that clinical diagnosis in routine practice may not perform as well as is generally believed by clinicians or as suggested by the literature assessing specific aspects of clinical diagnosis (e.g., new tests) in research settings. Better characterizing the performance of clinical diagnosis for common conditions would clearly benefit the entire health system and identify important targets for quality improvement that could be pursued in a concerted manner.

The second benefit to the health care system as a whole relates to vital statistics and other epidemiologic data. Vital statistics impact important decisions about allocation of funding for research and other aspects of health care policy. The existing literature demonstrates that clinical diagnoses, whether obtained from death certificates or hospital discharge data, contain major inaccuracies compared with diagnoses generated from postmortem findings. The use of autopsy data to correct inaccuracies in epidemiologic data would likely confer multiple benefits on the health care system as a whole.

Future Research

1. Various aspects of the performance of the autopsy as a diagnostic test (e.g., the reproducibility of findings between pathologists) remain undefined and represent areas for further research. More specifically relevant to the present review is the inter-rater reliability for error classifications in specific cases, i.e., establishing the extent to which pathologists, clinicians or other peer reviewers agree that a particular case does or does not involve a clinically important diagnostic error.

2. The causes of important diagnostic discrepancies remain uncharacterized. This represents a very important area of investigation. Discrepancies between efficacy and effectiveness (i.e., differences between the performance of a diagnostic or therapeutic procedure in routine practice compared to the result in the research literature) have diverse causes. Broadly speaking, though, discrepancies are caused by a) quality problems related to *underuse*, *overuse* and *misuse* of diagnostic or therapeutic procedures, and b) patient factors, including atypical presentations and complex interactions between comorbid conditions and patient demographic factors. Neither of these categories are captured in the "efficacy literature" (i.e., clinical trials), as the nature of research settings make underuse, overuse or misuse unlikely, and stringent patient selection reduces the complexities of comorbid conditions and multiple competing diagnostic considerations.

   Autopsy data provide a window into discrepancies between efficacy and effectiveness both for therapeutics (by detecting clinically unsuspected complications of care) and diagnostics (by detecting the diagnostic discrepancies discussed in this report). In both cases, but perhaps especially the latter, the autopsy can play a pivotal role in spearheading investigations into the causes of these discrepancies. Where discrepancies prove to present quality problems, the institution benefits and, where they reflect differences between the types of patients receiving care in routine practice and clinical trials, the whole health system may benefit from awareness of these findings.

3. Future research should establish strategies for optimizing the utility of the autopsy at the institutional level. No study has ever directly assessed the impact of detecting errors in clinical diagnosis on subsequent clinical performance. Thus, future research should establish optimal methods of involving clinicians in the autopsy process (or communicating its results to them) and effective ways of stimulating change based on autopsy findings. Until such research is performed it is not clear to what extent autopsy rates need to be increased as opposed to achieving improvements in communication and utilization of information generated from autopsies performed at current rates.

4. Future research should establish the optimal means of using autopsy data to provide more accurate vital statistics and other important epidemiologic data. The first step might be to validate the findings suggested in this review, namely that current vital statistics contain substantial inaccuracies. Such an undertaking might involve funding a small number of demographically diverse institutions to achieve high institutional autopsy rates, with prospectively determined protocols for autopsy performance and error classification. Even one year's worth of data from such a project would likely document substantial inaccuracies in vital statistics. Continuing such a project could also provide ongoing epidemiologic data, as well as more meaningful error rates that could be used to fuel quality improvement efforts throughout the health system. Such a program would *not* replace autopsies as routinely performed elsewhere, that is, this suggested research program would not be equivalent to a system of regional autopsy centers performing autopsies on behalf of other institutions. Rather, these centers would act as surveillance centers for basic causes of death and detection of quality problems and present numerous opportunities for basic research into the pathogenesis of acute and chronic illnesses.

Availability of the Full Report

The full evidence report from which this summary was taken was prepared for the Agency for Healthcare Research and Quality (AHRQ) by the University of California at San Francisco-Stanford Evidence-based Practice Center (EPC), Stanford, CA, under Contract No. 290-97-0013. It is expected to be available in the fall 2002. At that time, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requesters should ask for Evidence Report/Technology Assessment No. 58, *The Autopsy as an Outcome and Performance Measure*. In addition, Internet users will be able to access the report and this summary online through AHRQ's Web site at www.ahrq.gov.