Chapter 2. Methods

Writing as an epidemiologist with an interest in the autopsy, Saracci[58, 68, 69] suggested several changes in the approach to the autopsy in order to foster its valid use as a "monitor of the quality of clinical diagnosis." First, he suggested shifting away from the common use of agreement or confirmation rates, and instead using autopsy findings to calculate the sensitivity and specificity of clinical diagnoses. Second, he stressed the importance of well-defined protocols for autopsy performance and a systematic approach to assessing discrepancies between autopsy findings and clinical diagnoses. Third, he outlined steps that ought to be taken to account for the distorting effects of non-random selection of cases for autopsy. Such steps included tracking the severity of illness and spectrum of diseases seen at the hospital in question and the percentage of cases with different diagnoses that undergo autopsy.

The issues highlighted by Saracci[58] inform much of the methods used to evaluate the literature summarized in this report. For instance, as described below, autopsy studies that merely dichotomize clinical diagnoses as confirmed/not confirmed (or correct/incorrect, etc.) were not included in the section on diagnostic errors detected at autopsy. Second, we review the literature addressing the properties of the autopsy as a diagnostic test. Thus, we consider questions relating to the quality of autopsy performance, variability among pathologists in interpreting autopsy results, and the frequency with which diagnoses remain unclear despite adequate postmortem examination. Third, in reviewing the literature on diagnostic errors detected autopsy, we discuss in detail issues related to the non-random selection of cases for autopsy

In establishing the specific questions addressed in this report, we started with an initial set of objectives proposed by the College of American Pathologists to the Agency for Healthcare Research and Quality (AHRQ). Preliminary reviews of the literature, discussions with a Technical Expert Advisory Group (Appendix F) for the project, and input from AHRQ resulted in refinement of the original questions. For example, determination of the cost-effectiveness of autopsy was proposed in the initial questions, but not included in the final set of study objectives. To accomplish a cost-effectiveness analysis for the autopsy requires quantifying the benefits associated with autopsy or interventions using information derived from autopsy findings. As stated previously, many of the benefits of the autopsy are difficult to quantify at all. The benefit with the most potential for quantification to the extent required of a cost-effectiveness analysis is the impact of autopsy on reducing diagnostic errors or complications of care. However, no study has ever addressed this question, so the effectiveness of the autopsy is unknown in this regard. In addition, preliminary reviews of the literature retrieved inadequate data on the true costs of the autopsy. It is possible that data on autopsy-detected errors in clinical diagnosis collected for this report might provide part of the foundation for some modeling of the possible cost effectiveness of the autopsy, but generating such estimates was beyond the scope of the current project which is primarily a systematic review of the literature.

Search Strategy

We conducted an extensive electronic literature of the MEDLINE® database, supplemented by hand searches of article bibliographies and information from experts in the field. The

MEDLINE® search strategy involved the Medical Subject Heading (MeSH) terms *autopsy* and *postmortem changes*, as well as the following title words: *autopsy*, *post-mortem*, *postmortem*, *necropsy*, and *posthumous*. We searched the more than 35,000 citations retrieved with this initial broad search using terms that captured aspects of study design (e.g., *epidemiologic studies, clinical trials, comparative study)* or topics relating to diagnosis (e.g., *diagnostic errors, diagnostic techniques and procedures, diagnosis, differential)* or error (e.g., *medical error, iatrogenic disease, sentinel surveillance, safety*).

The Cochrane Library was searched using similar key terms and title words. Preliminary searches of the CINAHL® database revealed extensive overlap with articles already retrieved from MEDLINE®, so this database was not searched further.

Reference lists from all relevant articles were reviewed to identify additional studies. The extensive search was completed in September 2000, but periodic electronic searches were conducted throughout the project until November 1, 2001 as relevant studies continued to appear during this time. During the final preparation of the report, four additional studies were identified after a repeat search was conducted in April 2002.[319-322]

Unlike systematic reviews of treatment efficacy, reviews related to health services research must take into account differences in practice patterns between countries, as well as different economic structures to health care systems. We were concerned that U.S. clinicians might regard patients in other countries as less likely to undergo extensive sophisticated imaging procedures (e.g., magnetic resonance imaging) and thus question the relevance of these studies to U.S. practice. On the other hand, we did not want to miss the opportunity to review as much of the world literature as possible, especially as this view likely represents a misconception. As a result, we did not rule out the possibility of including non-English studies, especially when English language abstracts were available and indicated relevant data addressing a research question not answered by data from English-language articles. However, we did not search other databases known to capture more of the non-English published literature (e.g., EMBASE).

Study Selection

The autopsy literature consists entirely of observational studies, rendering problematic the development of appropriate inclusion and exclusion criteria as most systematic reviews usually involve at least some randomized controlled trials. In the absence of quality scoring systems or other well-established criteria, we adopted fairly minimal inclusion and exclusion criteria.

Inclusion Criteria

For studies of diagnostic errors detected at autopsy, we required the following:

1. Well-defined patient samples consisting of consecutive or randomly sampled autopsies meeting explicit criteria (e.g., all adults dying after hospital arrival during a specified time period of time);

2. Clinical diagnoses derived from autopsy request forms submitted by clinicians or chart review performed by the study investigators;

3. Classification schemes for discrepancies between clinical and autopsy diagnoses conforming to one of the following three categories:

    - *Class I equivalent*—missed diagnoses that "would," "could," possibly" or "might" have affected patient "prognosis" or "outcome" had they been detected during life. (At a minimum, such impact involved discharge from the hospital alive). This category also includes any scheme explicitly identified as equivalent to the classification proposed by Goldman et al [51] or the subsequent modification by Battle et al.[70] Error schemes that make no distinction between changes in management and changes in outcome were classified under "major" below.

    - *Major errors*—missed diagnoses that, while important, likely had no effect on patient outcome; changes in management could be assessed as possible or even expected, but impacts on outcome for errors in this category were judged to have been "equivocal," "doubtful," "unlikely," or otherwise unexpected. (Errors for which expected impacts on patient outcomes were explicitly restricted to symptom palliation were included in this category.) This category also includes the sum of Class I and Class II errors from articles that explicitly reference the classification proposed by Goldman et al [51] or the subsequent modification by Battle et al.[70] The reason for including Class I and Class II together in this category is that this seemed more comparable to the many articles that did not distinguish between changes in management and changes in outcome – i.e., errors that, had they been diagnosed would likely have led to changes in treatment or improved survival.

    - *Discrepant ICD major disease categories*—errors in this category include cases in which major clinical and autopsy diagnoses fall in different disease categories in the International Classification of Diseases (ICD). These categories consist of groupings of three-digit ICD codes into general headings such as "infective," "neoplastic," "cardiovascular," and so on. Although this classification of clinical-autopsy discrepancies makes no mention of changes in outcome or treatment, such changes would be likely in many, if not most cases. Moreover, the classification scheme is well defined and likely involves less subjectivity than the other two error categories above.

Exclusion Criteria

For studies of diagnostic errors detected at autopsy, we excluded those with any of the following features:

- An autopsy series equivalent to a "convenience sample" (consecutive series missing more than 20% of eligible cases were also excluded)
- Assessments of clinical diagnoses based primarily on death certificates. (This criterion did not apply to death certificates created especially for a study, for instance, by having an investigator code new death certificates for all included cases using clinical information.[14])

- Dichotomous error classification schemes for clinical diagnoses such as "correct/incorrect" and confirmed/unconfirmed."

Studies excluded from consideration with regard to the performance of clinical diagnosis could still be considered in addressing other study questions, such as the performance of the autopsy as a diagnostic test, trends in autopsy rates and attitudes towards the autopsy.

Data Collection and Analysis

Articles identified from the literature search were stored in a reference database and categorized according to the study questions addressed (Appendix C). Structured abstraction forms (Appendix D) 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 four reviewers, including three physicians and one non-physician research assistant. One of the physicians reviewed all of the articles.

Mean error rates were calculated using logistic regression methods. Specifically, the probability of autopsy detected errors served as the dependent variable in a regression model that included as predictors: study period, autopsy rate, case mix and country (US or non-US). Appendix B provides further details regarding the statistical methods.

Peer Review

A preliminary draft of the report was submitted for peer review and commentary by experts chosen by the Evidence-based practice center (EPC) team as well as by the College of American Pathologists. These reviewers provided many helpful comments and suggestions, which were incorporated into the final draft. These reviewers, as well as the many advisors to this project, are listed in the Acknowledgments (Appendix G).