Optimal Policies for Clinical Laboratory Quality Control

AHCPR Grant # R0S HS09329-01

James C. Benneyan, Ph.D.
334 Snell Engineering Center
Industrial Engineering and Operations Research
Northeastern University
Boston, MA  02115
e-mail: benneyan@coe.neu.edu

EXECUTIVE SUMMARY

Summary of Research

This research is concerned with the overall sensitivity, specificity, and cost of clinical laboratory screening processes, with particular focus on the policy currently required by the Clinical Laboratory Improvements Amendments Act (CLIA) for screening Pap smears for early indications of cervical cancer. The objectives are to develop and apply statistical and economic methods, similar to those used in industry, in order to help analyze and improve laboratory screening quality, to maximize overall sensitivity and specificity, and to minimize total associated costs. Motivated by recent concerns in the media, several costly litigation settlements, and the fact that treatment is much less successful if not detected early, several improvements to CLIA have been proposed in the literature, including the use of higher rescreening rates, multiple reviews of each smear, automated rescreening technology, alternate smear preparation products, and others. The relative merits of these approaches, however, have remained largely uncertain, with little theoretic investigation into the overall sensitivity, overall specificity, and overall costs of each possible policy.

Several mathematical and economic models therefore have been developed to help examine inherent tradeoffs between sensitivity, specificity, and all associated costs of each approach. Results show that CLIA never is optimal by any criteria and always increases total costs, that the overall sensitivity of CLIA never can be improved beyond certain mathematical bounds, that no amount of partial rescreening ever is optimal, and that in some cases multiple evaluations of each smear is optimal. Similar results show that in certain scenarios the proposed use of automated rescreening technology recently approved by the FDA can dramatically increase overall costs without significantly improving sensitivity, despite widespread marketing by manufacturers to the contrary. Preliminary analyses indicate that the improvements possible by switching to the optimal non-automated policy range from 90,000 to 165,000 fewer false-negatives and $250 to $750 million savings per year nationwide. While the current project focuses primarily on cervical cancer, results are expected to be equally beneficial if extended to mammography and other screening policies at a later date.

Problem Background, Motivation, and Research Questions

Introduced by Dr. Papanicolaou more than 50 years ago, the cervical cytologic (or Pap) smear has been widely used in the U.S. since the 1950s as the standard means of screening females for early detection of cervical cancer, with approximately 52 million now annually performed in the U.S. and over
132 million per year worldwide. In this screening process, a sample of tissue cells is obtained periodically (e.g., annually for women in certain age brackets) from the endocervix and ectocervix, prepared on a slide, and examined under a microscope by one or more individuals for any indications of cervical cancer, its early precursors, or other abnormal cells. Although cure rates generally are very high when detected early, treatment is much less successful if allowed to progress undetected. As a consequence of media coverage and public concern regarding false-negative Pap smear results, congressional hearings were held in 1976 and again in 1988 on medical laboratory practices. In both instances, these hearings resulted in legislative action directed toward improving quality policies of laboratories involved in cytologic screening for cervical cancer. For example, the Clinical Laboratory Improvement Amendments Act of 1988 [1] requires that a random sample of at least ten percent of all smears that a laboratory technician initially judges to be negative then be rescreened by either a pathologist or a senior technician. Additionally, all slides judged as positive after a single examination should be reviewed a second time for verification. Patients confirmed by this process to have positive indications typically undergo a subsequent more costly verification test, such as a colposcopy, biopsy, or spectroscopy, and then medical treatment if deemed necessary.

Although clinical laboratories must demonstrate adherence to CLIA in order to be eligible for Medicare reimbursement, this policy has received little support in the professional clinical community, with numerous editorials and authors criticizing the mandatory 10% rescreening requirement in particular. Recent studies, however, have estimated that although 71% of all laboratories (including 93% of private laboratories) continue to apply the 10% rescreening rule as a primary method of quality assurance, 1.5% of all pathologists are involved in malpractice lawsuits regarding false-negative Pap smears. Of particular interest in the present research, therefore, is examining whether and by how much CLIA can be improved upon. The approach taken herein is to develop analytic models that help a decision-maker or analyst determine and compare the overall sensitivity, specificity, and cost of CLIA to several alternatives proposed in the literature. These alternatives include the use of higher rescreening rates, various amounts of multiple reviews of presumed-negatives, alternate slide preparation products, industrial quality control methods, and automated technology to rescreen presumed-negatives. For example, some facilities randomly rescreen negative slides at rates up to 4 times higher than the minimum required 10%, read all smears a minimum of two times, or pay for additional external rescreening of some partial amount of smears [2,3]. As an alternative to increased human rescreening, automated technology also recently has been proposed, with any slides initially determined to be negative now being subjected to non-human rescreening in order to redirect any suspicious smears back for a second human evaluation. Two such automated rescreening systems recently were approved by the Hematology and Pathology Devices Panel of the Medical Devices Advisory Committee to the FDA [4]. Other companies have taken a different approach by developing new methods to improve the preparation of slides for interpretation, essentially based on the feeling that overall accuracy might be improved more significantly and at lower costs by increasing the accuracy of the primary human interpretation.

From the many discussions in the literature and at the recent FDA hearings, however, it is clear that the relative benefits of each possible approach and the overall effectiveness of any rescreening policy (either human or automated) remain largely uncertain. In particular, no previous study has analytically investigated the impacts that increasing the required minimum 10% human rescreen rate, using multiple reviews per smear, or using automated rescreening will have simultaneously on the overall process sensitivity, specificity, and cost. For example, a recent article in the Business and Medical Report of In Vivo described the history of automated cervical smear screening and stated:
But these companies have yet to prove to potential users that their instruments save money in the long term and greatly improve the quality of a test that many laboratory directors believe already has a relatively low error rate.

Moreover, while some have argued that multiple rescreenings will increase cost and workload, others have suggested that additional primary screening ultimately potentially could reduce total consequences of false-negatives and false-positives and thus total costs. With respect to the possibility of repeated re-examinations of mammograms, however, leading researchers recently argued that it is essential for future research to analytically address the inherent tradeoffs between sensitivity, specificity, and workload costs, stating [5].

An issue that must be addressed is whether the increase in sensitivity is worth the associated increase in false-positive findings. Determination of whether this type of outcome is worthy of the additional costs required by double reading should be required components of future research.

**Research Approach and Methods**

The above quotes and background largely motivate the approach to the present research, namely to develop quantitative tools to help better understand the overall performance of screening policies and ultimately to identify the optimal policy that minimizes total associated medical and operational costs. Specifically, statistical and economic models have been developed in order to study, in a structured and methodical manner, the overall costs, false-negative rate, false-positive rate, workload, and throughput of each possible laboratory process. These models are very useful for comparing the total economics and accuracies of different cancer screening policies, for identifying the optimal amount of partial rescreening, and for determining when multiple readings of each smear becomes cost effective. These results also have been used to develop an algorithm that guarantees the optimal policy in any given scenario. The intention is that these tools be used by risk managers, clinicians, and regulatory bodies to help develop a better understanding of the current screening policy and, ultimately, to inform the redesign of more effective and minimum cost screening policies and regulations.

In order to develop these tools, a thorough understanding of the entire process logic first was developed and documented, working in conjunction with several laboratory and pathology personnel familiar with various aspects of the screening process, CLIA, and the proposed automated and new slide preparation approaches. Analytic and computer models then were developed following standard statistical and economic modeling approaches taken in similar scenarios in many other industries. These results include analytic models, for each possible screening policy, of the:

- overall sensitivity (defined as the probability of correctly detecting a true-positive),
- overall specificity (defined as the probability of correctly classifying a true-negative),
- expected number of correct-detects per 100 (or per n) truly positive smears,
- expected number of correct-nondetects per 100 (or per n) truly negative smears,
- expected number of false-negatives per 100 (or per n) smears of unknown status,
- expected number of false-positives per 100 (or per n) smears of unknown status,
- expected laboratory workload demands and throughput to process 100 (or n) smears in order to assist in staff planning to meet policy revision needs, and
- standard deviations and probability distribution functions for each of the above measures.
All results were rigorously validated for accuracy using computer simulation, with several also being rederived independently for further verification. Simulation models also were developed to obtain a few results which proved to be analytically intractable, such as the probability distribution and standard deviation of total costs. Throughout this process, feedback actively was sought from several researchers in this area through periodic review meetings, working papers, and presentations in order to ensure the relevancy and accuracy of process, individual accuracy, and cost input assumptions. In the interest of space, further detail on these models, their mathematical derivation, example applications, and assumptions can be found elsewhere [6-8].

Summary of Results-to-Date

While the primary focus of this project was to develop the above tools and demonstrate their application, preliminary use and analyses already have shown that very significant improvements in accuracy and cost are possible over the current CLIA policy. All results fall into one of two categories, those which are entirely independent of cost and accuracy assumptions and those that are dependent on these estimates (which may differ by geographic location, patient demographics, or laboratory). Because many inputs are not known precisely (and in particular as consensus about false-negative costs is unlikely), several sensitivity analyses also have been conducted. Complete discussion of all results-to-date can be found in the above references, with the most significant general conclusions being that:

- CLIA never is optimal by any criteria and in particular never minimizes the overall false-negative rate nor total costs;
- The overall sensitivity of CLIA never can be improved beyond certain bounds due to the pathologist verification of presumed positives;
- Neither the required CLIA 10% nor any other amount of partial rescreening ever is optimal, with either 0% or 100% rescreening always being optimal in all cases;
- Multiple 100% re-examinations of each smear can be optimal in some cases;
- Neither CLIA nor automated rescreening mathematically always must result in higher sensitivity or lower costs than the other, with the best policy being situation-dependent;
- In some realistic cases, however, the use of automated rescreening can dramatically increase overall costs without significantly improving sensitivity;
- The optimal non-automated policy, moreover, almost always is preferable to automated rescreening;
- The optimal policy and amount of improvement can vary significantly dependent upon assumptions about individual accuracies, costs, and incidence rate; and
- Further applied research using these tools is in order to investigate general conditions when each policy is preferable and to make specific policy revision recommendations.

To illustrate a few of the above uses and conclusions based on the developed models, Figure 1 shows the impact of alternate rescreen rates and multiple examinations of each smear for three different lab technician sensitivities. As shown, in each case the 10% requirement results in negligible improvement in overall policy sensitivity, especially as compared to the much greater benefits possible by higher rescreen rates or a second 100% examination of all smears. As this figure also demonstrates, it can be shown mathematically that the overall policy sensitivity never can be greater than that of the verifying pathologist who confirms all positive smears. This result suggests that greater improvements
are possible either by redirecting current resources and effort focused on technician accuracy towards improving pathologist accuracy or perhaps by removing this verification step altogether. Similar results also have been obtained for the overall specificity and total costs of each policy in order to examine inherent tradeoffs between false-positives and false-negatives. These models have been used to prove mathematically that, dependent upon the specific scenario, either the CLIA or automated screening policies can be preferable over the other. That is, it is not possible to claim that either policy or

Figure 1. Examples of Effect of Rescreening Rate and Multiple Screenings on Overall Policy Sensitivity (Upper Bound on Overall Policy Sensitivity = Pathologist Sensitivity = 0.95)
approach always must have lower costs or greater accuracy than the other, despite widespread marketing by certain manufacturers to the contrary. Given cost and accuracy estimates in the literature, in fact, non-human rescreening sometimes can result in significantly higher overall costs and lower overall sensitivity, as shown by the examples in Figure 2 [6]. Several mathematical conditions therefore

![Graph of Overall Policy Sensitivity vs. Sensitivity of Redirected Rescreening Device]

![Graph of Total Cost per Smear vs. Redirected Cost per Screening]

**Figure 2.** Example Comparisons of the Overall Sensitivity and Total Costs of the CLIA, Automated, and Optimal Cervical Cancer Laboratory Screening Policies
have been derived under which each policy will be best in terms of overall sensitivity, overall specificity, and ultimately total cost. These results can be used to help determine at what point proposed policy changes and introduction of technology are beneficial, as illustrated above. Also note that in most situations examined to-date, the optimal human screening policy was found to be more cost effective than both CLIA and non-human screening (for example, see Figure 2).

Additionally, optimization methods have been developed to help identify the overall optimal screening policy and partial rescreening rate in any given situation. Interestingly, these results have led to a mathematical proof that in no case is any amount of partial rescreening ever the optimal minimum cost policy, regardless of any and all input assumptions, despite the CLIA 10% requirement and widespread practice to the contrary. Although several cost and other model inputs are difficult to estimate precisely, exploratory analyses indicate that quite dramatic improvements are possible by switching to the derived optimal policy. For example, to illustrate the possible savings by revising the CLIA requirement and the effect of different assumptions, Table 1 summarizes four hypothetical examples using inputs based on the literature and discussions with pathology practitioners. The scenario in column 3 represents fairly accurate technicians and pathologists with high sensitivities and specificities, resulting in a single evaluation and 0% rescreening being the minimum total cost policy. Note that a savings of approximately $400 per 100 smears would result from changing to this optimal policy, equating to roughly $208 million per year if extrapolated to the 52 million smears examined annually in the U.S. Under CLIA, however, this policy is not permitted. (Also note that the CLIA 10% policy has a higher total cost than two independent evaluations with 0% rescreening.) Conversely, the scenario in column 4, with higher technician and pathologist false-negative rates, results in two independent evaluations and a 0% rescreening rate now being optimal, with a savings of over $832 per 100 smears, or roughly $418 million and 93,600 fewer false-negatives per year.

The example in column 5 may be the most likely in terms of technician and pathologist sensitivities, again resulting in the optimal policy being to use 2 independent screenings and a 0% rescreening rate. In this case, changing from CLIA to the optimal policy would result in $1,092 savings per 100 smears and an increase in overall process sensitivity by over 20%, or roughly $568 million and 120,400 fewer false-negatives per year. For completeness, column 6 illustrates a more extreme false-negative cost, resulting in 4 independent evaluations and 0% rescreening being optimal by several magnitudes, with a savings over CLIA of almost $15,000 per 100 smears and an increase in overall process sensitivity by over 28%, or roughly $7.4 billion per year and 164,000 fewer false-negatives per year. This last case illustrates that scenarios theoretically can exist where something more than double rescreening is optimal. Of course, implementation feasibility still needs to be considered, as laboratory managers have suggested that administering and coordinating anything more than two screenings may be logistically infeasible and as the nationwide pool of trained technicians and pathologists is limited. The expected number, variances, and probability distributions of technician and pathologist workloads (i.e., the number of readings per \( n \) smears) therefore also have been derived in order to help anticipate changes when switching to the optimal policy.

The above results are similar to industrial experiences in which no amount of partial sampling ever is optimal, multiple inspections significantly improve overall outgoing quality, and one to four multiple inspections of presumed nonconforming items often minimizes total expected costs. As the above examples illustrate, of course, the exact results may differ significantly in any particular setting and specific conclusions therefore should not be extrapolated loosely to all other scenarios. For example, in the third example above (column 5), an increase in the cervical cancer incidence rate from .005 to .025
Table 1. Comparison of Total Expected Cost per 100 Smears Using Human Screening
(Four Examples to Illustrate Impact of Assumptions on the Optimal Policy)

<table>
<thead>
<tr>
<th>Number of Technicians</th>
<th>Rescreen Rate (%)</th>
<th>Average Cost per False-Negative = $5,000</th>
<th>$50,000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Technician Sens = .95 Pathologist Sens = .95</td>
<td>Technician Sens = .60 Pathologist Sens = .95</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$1,273 *</td>
<td>$5,337</td>
</tr>
<tr>
<td></td>
<td>10 (&quot;CLIA&quot;)</td>
<td>$1,673</td>
<td>$5,593</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$1,434</td>
<td>$4,760 *</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>$1,845</td>
<td>$5,102</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>$2,257</td>
<td>$5,444</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>$2,669</td>
<td>$5,785</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>$3,080</td>
<td>$6,127</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$3,492</td>
<td>$6,468</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>$3,903</td>
<td>$6,810</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>$4,315</td>
<td>$7,152</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>$4,727</td>
<td>$7,493</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>$5,138</td>
<td>$7,835</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$5,550</td>
<td>$8,176</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>$1,890</td>
<td>$4,804</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>$2,283</td>
<td>$5,168</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>$2,676</td>
<td>$5,532</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>$3,068</td>
<td>$5,896</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>$3,461</td>
<td>$6,260</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$3,853</td>
<td>$6,624</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>$4,246</td>
<td>$6,988</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>$4,639</td>
<td>$7,352</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>$5,031</td>
<td>$7,716</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>$5,424</td>
<td>$8,080</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$5,816</td>
<td>$8,444</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>$2,339</td>
<td>$5,082</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$6,070</td>
<td>$8,697</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>$2,766</td>
<td>$5,440</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$6,310</td>
<td>$8,938</td>
</tr>
</tbody>
</table>

Cost per single technician evaluation = $3  
Cost per single pathologist evaluation = $9  
Cost per false-positive = $750

** = Optimal policy

Technician Specificity = 95%  
Pathologist Specificity = 95%  
Incidence Rate of Cervical Cancer = 1.5%
changes the optimal number of evaluations from 1 to 3, resulting in approximately 290,000 additional detections and a $490 million reduction in total costs (nationwide per year) over the current CLIA practice. Note that use of automated rescreening in this case increases total annual nationwide costs by $195 million over CLIA and $685 million over the most cost effective policy and reduces overall sensitivity by roughly 9.5%, equating to 115,800 additional false-negatives per year in comparison to the optimal policy.

As another example, using similar assumptions as in column 5 above, Figure 3 illustrates the effect of different average false-negative costs on the optimal number of evaluations per smear. Although this cost may be difficult to estimate precisely, note that in this example a single evaluation is optimal only if this cost is below $650, two independent readings is optimal if this cost is between $650 and $2,400, and three or more readings are optimal if this cost can be assumed to be greater than roughly $2,400. While additional analysis examining the effect of technician accuracy, population incidence, and other assumptions are explored further elsewhere, the above examples illustrate the need for further research to develop better cost and other estimates and to determine what factors most affect the optimal policy. A more general implication is that the optimal policy can differ between specific laboratories and between particular demographic and geographic populations. It even may be optimal for a given laboratory to follow different policies for smears obtained under different conditions.

![Optimal Screening Policy versus False-Negative Cost](image)

**Figure 3. Example of Effect of the Cost of a False-Negative on the Optimal Screening Policy**

**Discussion and Further Research**

While the above examples illustrate particular uses of the developed models, it is important to emphasize again that many other assumptions are possible and that all scenarios have not been explored sufficiently yet to report general recommendations. Specific results and the best screening policy to use may differ in other settings depending on the accuracy of the screening personnel, the prevalence of cancer in the demographic group being screened, the frequency with which women are screened, and other considerations. It is hoped that further research may produce some general conclusions that can
inform improved laboratory management and policy revision. Results-to-date, however, clearly illustrate the importance of accurate cost and other estimates, and further analysis thus is in order before suggesting general guidelines concerning the best overall policy in common scenarios. Ongoing research therefore currently is examining many input and process assumptions, including incidence rates and per screening costs, technician and pathologist accuracy, sensitivities and specificities for each type of abnormality, and dependence of error rates on the observed incidence rate. The current models also could be applied separately to each individual type of abnormality in the Papanicolaou and Bethesda classification systems or could be expanded to include different costs and incidences for each abnormality class. For example, higher false-negative costs usually result from higher grade abnormalities, whereas sensitivity typically is greater for detecting these more advanced disease states. These more detailed approaches also might be useful for more explicitly examining any relative benefits of alternate screening systems on particular abnormality classifications, alternate slide preparation products, or alternate classification scales. The current models also could be expanded to account for several related concerns, including errors in obtaining tissue samples, errors in preparing slides, and the rate of patient compliance to annual pelvic examinations.

Finally, one possible advantage of rescreening could be to monitor the process performance over time and to identify individuals in need of retraining, as improving the process by determining the root causes of the problem ultimately may lead to the greatest savings. As an alternate approach for this purpose, statistical quality control charts, which are the foundation of modern quality systems in many other industries, might be applied effectively to control laboratory quality, compare the accuracy of individual personnel, and identify retraining needs. While not the primary focus of the present research, new special-purpose statistical quality control charts have been proposed in order to detect changes in inter-evaluator and intra-evaluator agreement in such situations [9], although more work in this area clearly is needed. Finally, although this research focused primarily on cervical cancer, significant motivation exists to explore similar research in breast, prostate, and colorectal cancer screening. Given their greater incidences and subsequent health and mortality consequences, results here could be at least as significant as those obtained to-date for cervical cancer.

References

Publications and Presentations Related to this Research


A Probability Model of Cross-Contamination for Determining the Distribution of the Number of Abnormal Cells, working paper with D. Zanhiser, F. Kaminsky.

Automated and Redirected Rescreening in Laboratory Screening Policies, Chapter IV in *Some Approaches to Quality in the Presence of Inspection Error, With Application to Optimal Laboratory Cancer Screening Policies*, PhD dissertation, University of Massachusetts, Amherst MA, 1997.

Multiple Inspection and Optimal Laboratory Cancer Screening Policies, Chapter V in *Some Approaches to Quality in the Presence of Inspection Error, With Application to Optimal Laboratory Cancer Screening Policies*, PhD dissertation, University of Massachusetts, Amherst MA, 1997.