Technical and Operational Issues in the Delivery and Integration of Retrospective Lab Result Data: Analysis and Recommendations

Final Report

Sujansky & Associates, LLC

April 23, 2007

Contents

1 Executive Summary

Laboratory test results are an important source of electronic data for quality-measurement and quality-improvement activities. Given the limited adoption of EHRs among clinicians today, many provider organizations and health plans acquire lab data through retrospective batch files provided by hospital and reference laboratories. These batch files ostensibly contain the full set of test results performed for a defined patient population (the covered population) over a defined period of time (the reporting period). Upon receipt, provider organizations and health plans associate the lab-result data with diagnosis, procedure, pharmacy, and other data that they posses to compute aggregate quality measures and to identify individual patients in need of clinical services. These analyses constitute important components of pay-for-performance programs and other clinical quality initiatives.

The effective use of retrospective lab data requires the correct execution of two separate processing steps: *Data Routing* and *Data Matching*. Data routing is the process by which laboratories determine which test results should appear in whose batch files (i.e., to which provider organization and/or health plan a result should be routed). Data matching is the process by which provider organizations and health plans, upon receipt of batch files, associate the test results with the individual patients in their covered populations (i.e., determine to which patient in a disease registry or data warehouse each result should be matched).

For various reasons, errors occur in both the data routing and data matching steps. The most common errors are *False Exclude Routing Errors* and *False Negative Matching Errors.* False Exclude Routing Errors cause laboratories to erroneously exclude test results from batch files in which they belong (i.e., the results are not sent to the appropriate recipients). False Negative Matching Errors cause the recipients of batch files to erroneously miss matches between the test results they've received and the patients in their data repositories (i.e., the results are discarded and their absence compromises the accuracy of quality measurement).

We studied the incidence and the causes of these errors through stakeholder interviews and through a set of experiments using actual lab result data and eligibility files. Although our findings indicate that *False Exclude Routing Errors* and *False Negative Matching Errors* occur in the minority of cases (5%-10% of batch files, in most cases), their incidence can be as high as 20%. Additionally, the rate of error varies significantly across provider organizations and health plans, and even among individual clinicians. Our findings also indicate that (1) provider organization with integrated processes and information systems experience fewer routing and matching problems, (2) probabilistic matching tools can significantly reduce false negative matching errors, and (3) health plan member IDs are currently not useful for patient matching, but they could be very useful if provider organizations and health plans changed the way they manage these identifiers.

We recommend a number of changes to improve the routing and matching of lab result data, divided into near-term, medium-term, and long-term actions. The near-term actions focus on the greater use of probabilistic matching tools. The medium-term actions focus on better management of patient demographic information by physician offices and health plans. The long-term actions focus on the development and uniform adoption of health plan ID cards with two-dimensional bar codes. Such cards would minimize discrepancies between the demographic data placed on lab orders and the demographic data included in eligibility files, thereby improving data routing and data matching. The automated reading of bar codes would also eliminate most data-entry errors. Over time, we believe these actions will improve the quality of lab-result data and increase the effectiveness of quality-improvement programs.

2 Introduction

The California HealthCare Foundation (CHCF) and the California Association of Physician Groups (CAPG) have funded this project to analyze data-quality issues in retrospective lab result data, particularly issues that affect the correct routing of results to provider and payer organization and the correct matching of results to patient registries. We have worked with health plans, provider organizations, and labs to understand their processes for routing and matching lab results. When these organizations are unable to route and match lab results correctly, it creates data quality issues and compromises Pay-for-Performance and HEDIS reporting. We have researched these problems and their various causes, with goal of understanding the causes and identifying viable solutions.

This report is divided into three sections. The first section provides general information about the problem we are addressing and the environment in which it occurs. The second section presents our research findings, based on interviews with stakeholders and on analysis of a large data set. The third section lists specific recommendations for improving lab-result routing and lab-result matching, based on our findings.

3 Background and Problem Formulation

The primary role of clinical laboratories is to provide test results to clinicians for the direct care of individual patients. Given the "mission critical" nature of test ordering and result reporting for both clinicians and laboratories, processes are in place to deliver the right result information to the right clinician with nearly 100% accuracy.

Many clinical laboratories also produce retrospective batch files of test results for their provider-organization and health-plan customers to support Pay-for-Performance and HEDIS reporting. These data files typically contain all of the test results for a defined population of patients (those covered by the receiving organization) for a defined period of time (usually one month). The files are delivered not to individual physicians, but to information technology groups within provider organizations[1] and health plans, which integrate and analyze the data en masse. As opposed to the very accurate process for reporting individual results to clinicians, significant errors occur in the process of sending batched result files to provider and payer organizations and in the process of associating those results with other patient data maintained by these organizations.

Data Routing

In providing batch files, the task of the laboratory is to correctly identify which historical test results should be sent to which receiving organizations ("*data routing*"). This task is very different than that of determining which clinician should receive a test result, because

(1) multiple clinicians may see patients covered by the same IPA, medical group, or health plan (necessitating the aggregation of results across clinicians when batch data is sent to an IPA/medical group/health plan),

[1] The provider organizations that receive these files are usually IPAs or large medical groups that contract with health plans for capitated lives and that contract with labs for testing services.

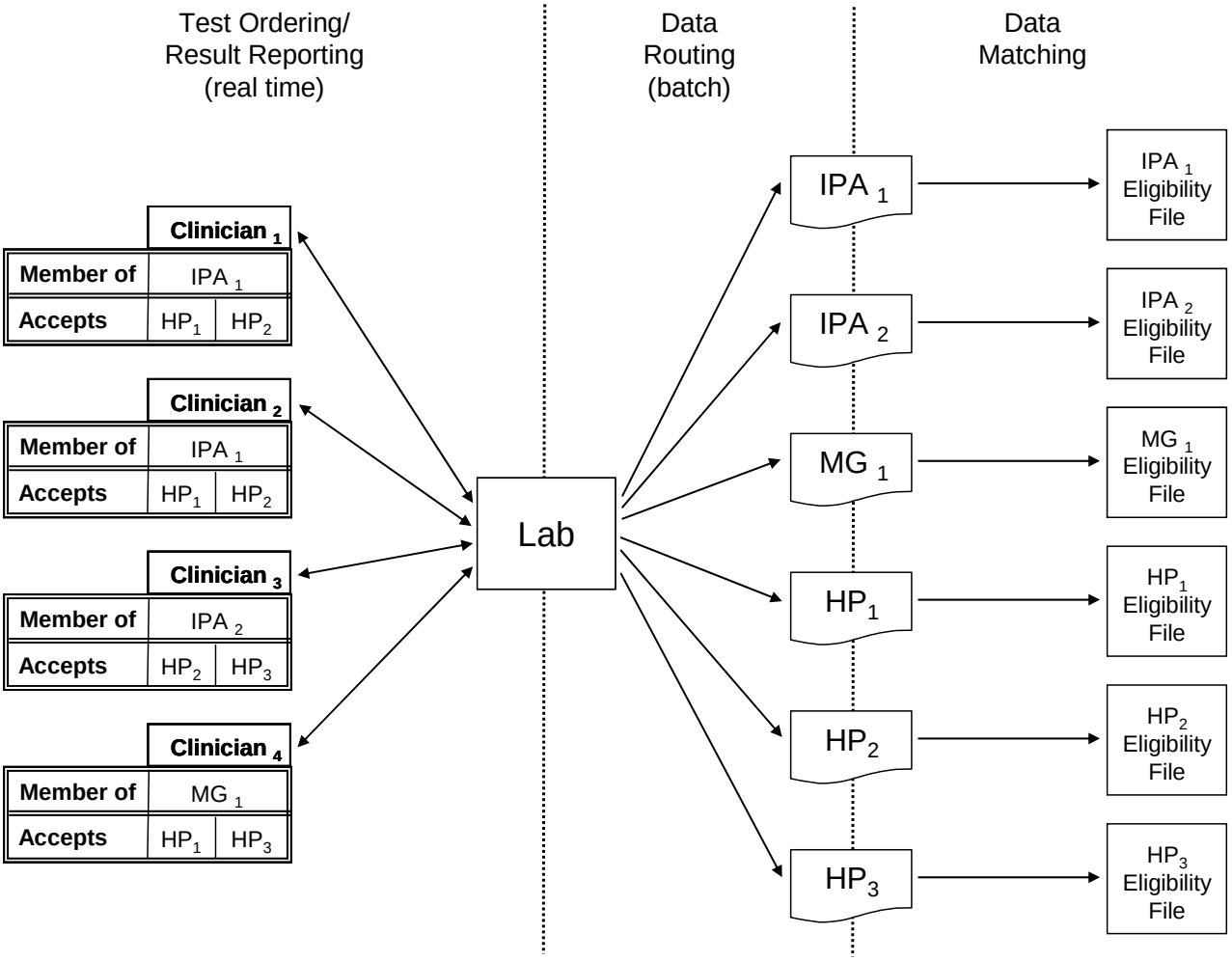
(2) not all of the results ordered by a clinician should be included in a retrospective batch file sent to that clinician's IPA or medical group (necessitating the filtering of test results based on the patient's insurance type), and

(3) each clinician may see patients covered by various health plans (necessitating the disaggregation of results that were reported to a single clinician when the data is sent to health plans).

Data Matching

Upon receipt of a batch result file, the task of the receiving organization is to correctly associate each test result with an individual in its eligibility file of covered patients or members ("*data matching*"), so that a patient-centric database may be created and patient-centric analyses performed. This task is made difficult by the absence of universal identifiers or other demographic data elements that are unique and consistent across test result files and eligibility files.

The following diagram illustrates the data flows described above:



IPA = Independent Practice Association
MG = Medical Group
HP = Health Plan

A few aspects of these data flows are noteworthy:

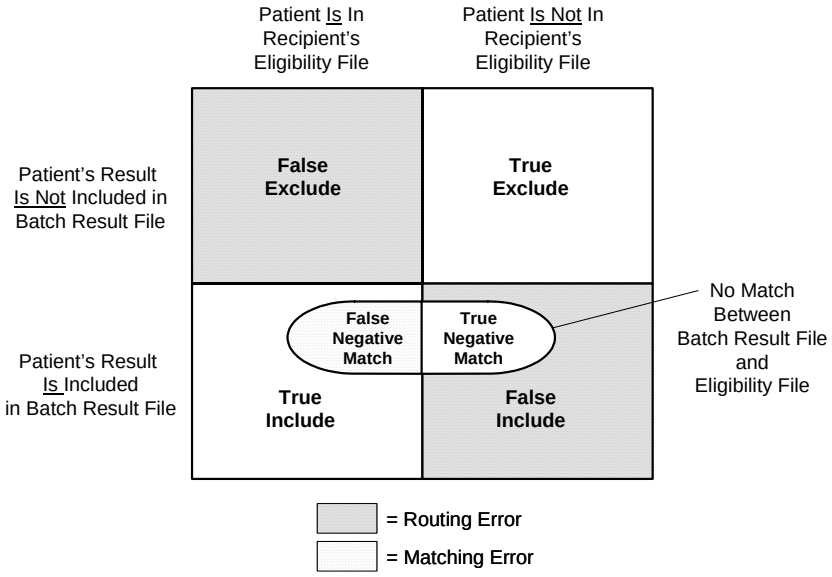
- Clinicians typically belong to one IPA or medical group. They see patients insured by various health plans.
- Certain of the patients that each clinician sees are *capitated*, i.e., the clinician's IPA or medical group is contracted by a health plan to provide most care for the patient, including lab services. Other patients are *fee-for-service*, i.e., whatever health services the patient receives from various providers (including labs) are reimbursed to those providers on a service-by-service basis by the health plan. In the former case, the lab usually bills and receives payment directly from the IPA or medical group for performing the tests; in the latter case, the lab usually bills and receives payment from the health plan for performing the test.
- IPAs typically are loose affiliations of independent physician practices, in which each practice is a separate business entity that uses separate information-processing workflows and tools (for example, for ordering lab tests). Medical groups are often single business entities that employ their staff clinicians and (sometimes) use integrated information-processing workflows and tools (for example, enabling them to standardize the lab-ordering process).
- Clinicians often order lab tests from the same lab for all of the patients they see, i.e., across different health plans and different coverage types (i.e., capitated versus fee-for-service). When generating retrospective batch files, the lab must determine which of the tests to send to the IPA/medical group of the ordering clinician and which to send to various health plans (and, sometimes, which to send to both). These decisions depend on the patient's insurance carrier and type of insurance, rather than the patient's physician and type of test.
- Labs send retrospective batch files of test results to provider organizations (IPAs and medical groups) as well as to health plans. Provider organizations and health plans request test results for their covered populations, i.e., their capitated patients (in the case of IPAs and medical groups) or their insured members (in the case of health plans). In certain cases, the same test result may be included in one file sent to the patient's health plan and in another file sent to the provider organization that provides care for the patient.
- When the recipient of a batch file is the same organization that pays for the test, labs typically use billing information to select which tests should be included in the file. When the recipient of a batch file is an organization other than that paying for the test (for example, the patient's health plan when the patient is capitated), labs typically match patient demographic information in the test to patient demographic information in available eligibility files to select which tests should be included in the file.

3.1 Types of Routing and Matching Errors

Because of the nature of information provided on laboratory test orders and limitations in the processes employed by laboratories and receiving organizations, both data routing and data matching are subject to significant error rates today. In certain cases, test results that should be

included in a receiving organization's batch file are omitted ("*false exclude"* routing errors). In other cases, test results that should not be included in an organization's batch file are included ("*false include*" routing errors). Additionally, receiving organizations may get the right test results, but be unable to match the results to individuals that appear in their eligibility files ("*false negative*" matching errors). These latter errors occur because the results contain insufficient, incorrect, or inconsistent patient demographic data. In certain cases, receiving organizations may erroneously match results to the wrong individuals in their eligibility files ("*false positive*" matching errors). Although all of these errors occur in the minority of cases[2], they are significant because they compromise the accuracy of quality reports, undermine the credibility of quality-improvement programs in the eyes of physicians, and may even constitute violations of patient-privacy regulations.

The following table illustrates the types of data-routing and data-matching errors that occur (note that the size of each cell is not proportional to the relative incidence of each error type):



Information gathered to date indicates that various factors account for these errors:

False Exclude Routing Errors

- Insurance information is missing from the order and unavailable from the patient, preventing the laboratory from ascertaining which provider organization or health plan should get the test result in a batch file
- Insurance information is incorrect on the order or incorrectly supplied by the patient
- Insurance information is incorrectly entered or interpreted at time of data entry by the lab
- Demographic information is missing from and/or incorrect on the order, preventing the laboratory from associating the result with a corresponding record in an eligibility file
- Demographic information is incorrectly entered at time of data entry by the lab

[2] Our research to date indicates that error rates range from 2% to 20%, although most organizations report error rates of 5% - 10%.

- Eligibility data provided to labs by IPAs, medical groups, or health plans is incomplete, incorrect, and/or out of date, such that no match to the test result is possible
- Insurance and/or demographic data on the lab order or provided by the patient is inconsistent with that appearing in the eligibility files (especially, the health plan ID and patient name), such that no match to the eligibility file is possible.

False Include Routing Errors

- Insurance information is incorrect on the order or incorrectly supplied by the patient, such that the wrong organization is billed
- Insurance information is incorrectly entered or interpreted at time of data entry by lab, such that the wrong organization is billed
- Eligibility data provided by an IPA, medical group, or health plan is incorrect and/or out of date, such that a patient is matched to the wrong receiving organization

False Negative Matching Errors

- Demographic information is missing from and/or incorrect written on the order, preventing the recipient of the data from correctly associating the result with a corresponding record in an eligibility file
- Demographic information is incorrectly entered at time of data entry by the lab
- Insurance and/or demographic data on lab orders or provided by patient is inconsistent with that appearing in eligibility files (especially, the health plan ID and patient name), such that no match to the test result is possible.
- Matching algorithm used by IPA, medical group, or health plan is insufficiently powerful to resolve minor discrepancies in the data (e.g., name variations, character transpositions, absence of some data elements)

False Positive Matching Errors

- Demographic information contains common first and last names, and correspondence on health plan ID is not required to trigger a match. In these cases, if the date of birth is incorrectly entered or coincidentally the same for identifically named patients, a false positive match can result.
- In an effort to avoid missing matches, the matching algorithm is configured to be highly sensitive and less selective. For example, correspondence on only the month of birth and only the first three characters of the first name are required to trigger a match.

  Note: From the information we received, false positive matches are uncommon and rarely lead to the incorrect classification of patients as chronically ill. Typically, false positive matches compromise the accuracy of quality measures to a minor degree.

3.2 Problems out of Scope for this Report

This report focuses on the technical and workflow issues that impede the correct routing and matching of batched test results. Other types of problems prevent provider organizations and health plans from receiving batch result data that they would like to receive. These problems typically are related to contractual arrangements among labs, provider organizations, and health plans, as well as to certain privacy regulations. Although such issues are significant with respect to quality improvement and P4P measures, this report does not address them.

3.3 Variations in Workflow for Ordering, Processing, and Reporting of Lab Tests

A number of variations exist in various components of the process for ordering, recording, routing, and matching lab results. It's useful to be aware of these variations, insofar as they account for differences in data-routing and data-matching errors seen by different organizations and may suggest "best practices" for minimizing such errors. The following table lists the most important of these variations.

Variations in test order creation	• Paper order is generated, with demographic and insurance data entered differently by each physician practice belonging to an IPA • Paper order is generated, with demographic and insurance data entered consistently by physicians using a centralized information system in an integrated medical group • Electronic order is generated via a lab-ordering interface, with enforcement of required data elements during the order submission • Electronic order is generated via a lab-ordering interface, with real-time eligibility lookup during the order submission (planned)
Variations in place/time of specimen collection	• Specimen collected *at lab*, with lab afforded opportunity to inspect patient's health plan ID card and to interact with patient to verify demographic and insurance information • Specimen collected *in physician offic*e, with lab limited to data appearing on the test order and no opportunity to interact with the patient to verify demographic and insurance information
Variations in use of eligibility files during order processing by lab	• Patient's demographic and insurance information are entered by lab directly from test order, with no subsequent verification against eligibility data • Patient's demographic and insurance information are verified by lab retrospectively against available eligibility data, at the time that billing takes place • Patient's demographic and insurance information are verified by lab in real time against available eligibility data, at the time that the specimen is collected from the patient

Variations in availability of eligibility files	• IPA, medical group, or health plan promptly and regularly provide updated eligibility files to contracted laboratories • IPA, medical group, or health plan provide updated eligibility files to contracted laboratories at irregular intervals and/or with some time delay • IPA, medical group, or health plan do not provide eligibility files to contracted laboratories
Variations in organization receiving the data versus organization billed for the test	• Batch data sent to IPA/medical group for tests that were billed to same IPA/medical group (HMO patients, delegated model) – data routing based on billing information • Batch data sent to health plan for tests that were billed to an IPA/medical group (HMO patients, delegated model) – data routing based on match to eligibility file • Batch data sent to IPA/medical group for tests that were billed to health plan (PPO/POS/FFS patients) – data routing based on affiliation of ordering provider • Batch data sent to health plan for tests that were billed to health plan (PPO/POS/FFS patients) – data routing based on billing information
Variations in data-matching processes used by data recipients	• Different sets of data elements used for matching (especially variations in use of health plan ID) • Different matching techniques, such as full-name matching versus partial (“Alexander” vs. “Ale”), synonym substitution (“Robert” vs. “Bob”), transposition checking (“120274” vs. “122074” or “120247”) • Different matching algorithms (single probabilistic match versus successive deterministic matches) • Different degrees of manual effort expended on matching (automated-only matching versus manual review of “residual” unmatched data)
Variations in health plans’ ID management	• Consistent representation of health plan ID in eligibility files and on member cards versus inconsistent representation • Assignment of unique identifiers to each covered member versus assignment of unique identifiers to subscribers only (with dependents distinguished by name)

4 Specific Findings and Analysis

Our research consisted of (1) interviews with laboratories, provider organizations, and health plans and (2) analysis of detailed matching data provided to us. This research yielded findings in several areas and helped to inform the specific recommendations that we make later in this report.

4.1 Evidence of the Problem

As noted earlier, data routing problems occur when lab result records cannot be assigned to the correct payer (whether a health plan or contracted provider organization). Data matching problems occur when lab result records cannot be matched to a correct record in the member eligibility file (whether that of a health plan or a contracted provider organization). We learned of both types of problems.

Data Routing Problems

Inability to assign the correct payer causes the affected test results to be erroneously excluded from batch report files. Several organizations reported evidence of this problem.

- Data analysts at an IPA discovered that approximately 20% of the tests performed by a local hospital for the IPA were omitted from the batch data files the hospital provides. These tests were located using an online query tool that accesses the hospital database directly. Inquiries with the hospital suggested that these omissions are caused by assignment of the wrong payer code to outpatients at the time that they are registered.
- A vendor operating a disease registry allows physicians to manually add tests that do not appear in the registry database if the physician received a record of the result directly from the lab. Physicians that use the registry have supplemented approximately 5% of the tests within it because these tests were not provided in the batch result files that labs send to the vendor.
- An IPA found that a significant number of test performed by a reference lab on behalf of a subsidiary group were being routinely omitted from its batch result files. Upon inquiry, the lab discovered that, although the lab was correctly identifying the subsidiary group as the payer of the tests, the lab was incorrectly excluding the subsidiary's results from the batch file sent to the parent IPA. The error was easily corrected.
- Another IPA that uses a reference lab identified several thousand tests for which it was billed that did not appear in its batch report files. The cause of the discrepancy has yet to be identified, but it implies that the correct identification of the paying organization does not guarantee that the clinical result will be included in that organization's batch result file.
- On a number of tests ordered by a medical group, the name of an affiliated IPA was erroneously specified rather than the group's own name. As a result, the laboratory designated the IPA as the payer, rather than the medical group. This designation caused the tests to be erroneously omitted from the batch result file sent to the medical group.
- A reference lab reported that it can assign the correct paying organization to a lab test 95-97% of the time when the specimen is collected at the lab's facility. This in-person visit allows the lab to review and verify of the patient's insurance and demographic

information, which significantly improves the completeness and accuracy of the data collected. When the lab receives a specimen that was collected earlier and has no opportunity to review and verify the patient's insurance information, the rate at which the lab immediately assigns the correct paying organization goes down to 90-95%. This necessarily results in delays in identifying the correct paying organization and including the test results in that organization's batch result file (although the clinical results are always reported quickly to the appropriate ordering physician)

Inability to match lab result records to a *health plan's* member eligibility file causes test results to be erroneously excluded from batch report files sent to that plan, specifically when the paying organization is not the health plan (as with sub-capitated arrangements).

- A health plan reported a study it performed of several hundred patients who should have received certain lab tests according to P4P criteria. The health plan received batch lab results from these patients from a reference lab, but also had access to an online tool to directly search the lab's database. For patients apparently lacking the relevant tests, the health plan manually searched the lab's system using the online tool. In a significant number of instances, the results were located via the tool, although they had not appeared in the batch file (no statistics were provided).

Data Matching Problems

Inability to match lab results to a *provider organization's* member eligibility file prevents test results that are included in batch result files from being used in the disease registry, because the results cannot be associated with any appropriate patient. This problem was reported by several organizations, as well as directly seen in the lab data we analyzed.

- An IPA reported that it matches 95% of the result records it receives to its member eligibility file, using automated comparison methods based on demographic information. These methods are deterministic (i.e., they do not apply probabilistic algorithms[3]). The IPA routinely performs an exhaustive manual comparison for all result records that remain unmatched. This manual process increases the match rate to 99%, suggesting that (1) very few "false includes" are sent to the IPA and (2) improved matching techniques that apply probabilistic methods may achieve more complete matching without resource-intensive manual reviews.
- In processing a set of 136,000 lab-result records for three provider organizations, we achieved an overall match rate of 79% using deterministic methods. This rate increased to 93% with the application of probabilistic methods. There were, however, significant variations in the match rates across the provider organizations:

Provider Organization	Number of Tests	Match Rate (Deterministic)	Match Rate (Probabilistic)	Contribution of Probabilistic Methods
#1	12,194	96%	98%	2%
#2	98,252	79%	95%	16%
#3	25,560	72%	82%	10%
Total	136,000	79%	93%	14%

[3] The deterministic and probabilistic techniques that we used are described later in the report.

Further analyses of this data set as well as the possible reasons for these variations are discussed in the following sections.

4.2 Background: Patient Matching Techniques

Before reviewing the specific findings of our data analysis, it is useful to review the techniques we used to match lab-result records to member eligibility files. A variety of techniques are used by various provider organizations, labs, and health plans for such matching. Most of the techniques used are *deterministic*, i.e., they generate a binary "yes/no" determination of whether two records correspond to the same individual. Most of these techniques involve matching on some subset of the following demographic data elements:

Last Name
First Name
Date of Birth
Gender
Health Plan Member ID

A common algorithm involves the sequential application of increasingly relaxed matching criteria until the maximum number of lab-result records are matched. We used the following deterministic algorithm for the initial matching of our test data to the eligibility files:

1. *Find all records that match exactly on Last Name, First Name, Date of Birth, Gender, and Health Plan Member ID.*
2. *Of the remaining unmatched records, find all that match exactly on Last Name, First Name, Date of Birth, and Gender only.*
3. *Of the remaining unmatched records, find all that match exactly on Last Name, First Name, and Date of Birth.*
4. *Of the remaining unmatched records, find all that match exactly on Last Name, the first 8 characters of the First Name, and Date of Birth.*

These algorithms, which are simple to program, have the advantage of selecting the most specific matches first, and then relaxing the match criteria for those records that remain unmatched. The greater the relaxation that occurs, the more records ultimately match, but the greater the risk of incorrect matches (false positive matches). For example, certain organizations relax the criteria for matching the First Name further, comparing the first three letters only. This approach successfully reconciles "Ann" with "Anne", but may also erroneously match "Robert" to "Roberta".

Because deterministic techniques require exact or near exact matches, they have the disadvantage of missing matches when the data include data entry errors, common spelling variations, nicknames, and/or missing data elements. Probabilistic matching techniques are more resilient in these situations, because they allow certain data elements to differ if enough other elements match exactly. We leveraged a commercial identity-matching application to match the lab results that remained unmatched following application of the deterministic algorithm. This application identified a significant number of additional matches, which were "missed" by the deterministic method. For example, the probabilistic algorithm recognized the following pairs of records as matches (the data have been somewhat modified for privacy reasons):

	LName	Fname	DOB	Gender	Health Plan ID
Lab Result	Wong Jr.	Samuel	3/6/1954	M	NCF599A33958
Eligibility File	Wong	Samuel	3/6/1954	M	599A33958-10

	LName	Fname	DOB	Gender	Health Plan ID
Lab Result	Velasquez	Alba	7/17/1934	F	125621301
Eligibility File	Velasquez	Alba	7/7/1934	F	1256213-01

	LName	Fname	DOB	Gender	Health Plan ID
Lab Result	Reynes	Tony	2/25/1945	M	654638201
Eligibility File	Reynes	Antonio	2/25/1945	M	6546382-01

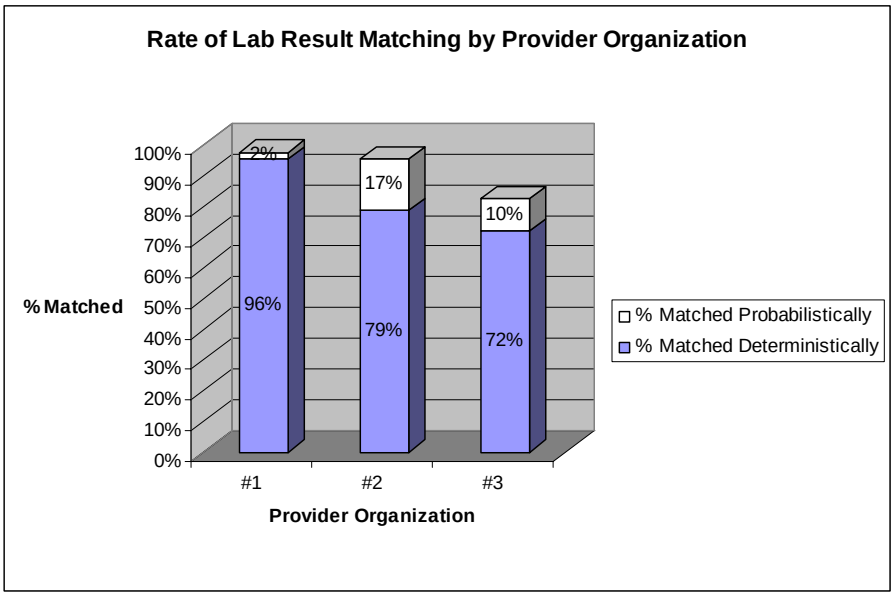
	LName	Fname	DOB	Gender	Health Plan ID
Lab Result	Jopherlin	Majorie	5/6/1971	M	
Eligibility File	Jopherlin	Marjorie	5/6/1971	M	6558382-01

A number of commercial tools are available that perform probabilistic matching. Certain of these tools are moderately priced and designed to run on desktop computers[4].

4.3 Variations in Patient Matching by Provider Organization

Our matching experiments showed that significant variations in the rate of successful matching exist among provider organizations, even when the same matching algorithms are applied. There are some relevant attributes of the data itself, therefore, that vary across provider organizations.

The variations in match rates among provider organizations are shown in the following graph.



[4] For example, see *Patient Data Matching Software: A Buyer's Guide for the Budget Conscious* (California HealthCare Foundation). http://www.chcf.org/documents/ihealth/PatientDataMatchingBuyersGuide.pdf.

Several of the organizations we interviewed suggested that part of this variation is due to the way that lab orders are generated by different provider organizations. "Integrated" medical groups, like Provider Organization #1, have a single information system that all of their physicians use to print laboratory orders. This system contains exactly the same eligibility data that the organization uses later to match lab-result records. This concordance is possible because the medical group both holds the contract with the health plan and operates the information systems for the practice.

In contrast, physicians practicing in IPAs, such as Provider Organizations #2 and #3, are more loosely affiliated and use separate information systems. These information systems all have their own databases of patient information, usually collected directly from patients at the time they register with the practice. These data, which are later included on lab orders, may or may not correspond to the member eligibility files that the IPA uses to match lab results.

These variations in test-ordering processes are clearly visible in the data. The following table shows the rates, by provider organization, at which various demographic data elements correspond between matching lab results and member eligibility records:

Provider Organization	Overall Match Rate	When Matched, Rate of Agreement On…					
		Last Name	First Name (8)	DOB	Gender	Last + First(8) + DOB + Gender	Member ID
#1	98%	98%	100%	100%	100%	98%	16%
#2	95%	95%	88%	97%	96%	79%	8%
#3	82%	95%	94%	97%	99%	86%	32%

Note that the rate of agreement on each of the demographic data elements (except Member ID) is consistently higher for Provider Organization #1 than for the other two organizations. More strikingly, 98% of the matched records agreed exactly on all four of the primary demographic data elements (Last Name, first 8 characters of First Name, Date of Birth, and Gender).

At the same time, Provider Organization #2 is able to achieve an overall match rate close to that of #1, although agreement on certain data elements (especially the first 8 characters of the first name) is markedly lower. This ability is explained by the effectiveness of the probabilistic matching algorithms, which increase the match rate for organization #2 from 79% (with deterministic methods only) to 95% (supplemented by probabilistic methods). Note from the bar chart above, however, that organization #1 achieved an even higher match rate (96%) even before the application of probabilistic methods, owing to the high rate of concordance on the demographic data elements that the deterministic algorithm compares.

Also interesting is the fact that the Health Plan Member IDs rarely correspond in the matched records of Provider Organization #1 (16%). In fact, further analysis showed that a member ID is entirely omitted from 75% of the lab orders originating from organization #1 (versus omitted from only 2% and 1% of the lab orders from Provider Organizations #2 and #3, respectively). Apparently, if other demographic information is consistently represented on lab orders, a high match rate can be achieved without the Member ID.

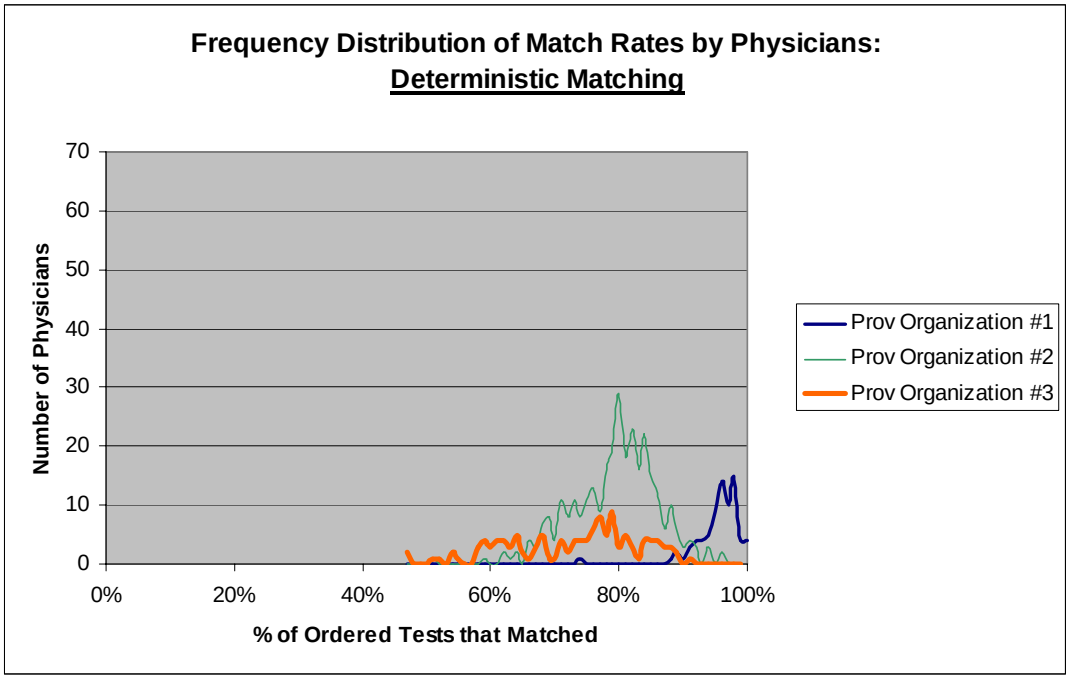
4.4 Variations in Patient Matching by Ordering Physician

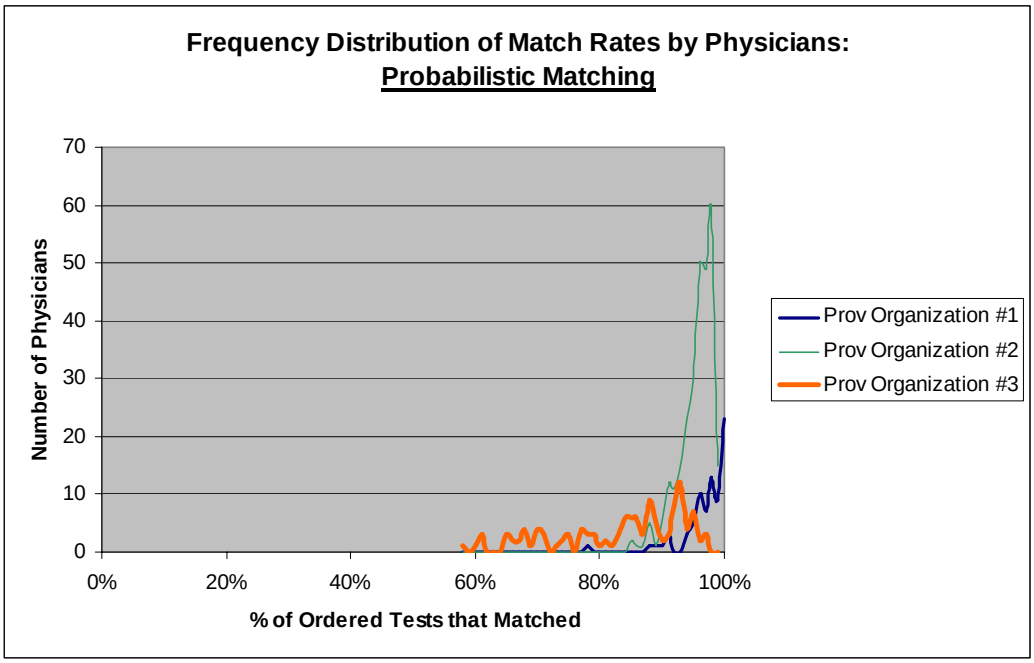
Within each provider organization, there is also significant variability *across physicians* in the rate at which the lab tests they ordered match to eligibility records. Further, this variability, itself, differs by the provider organization to which the physician belongs. The following table

shows the distribution of match rates (both deterministic and probabilistic) among the ordering physicians in the three provider organizations. Note the differences in average match rates and in standard deviations.

Provider Organization	# Ordering Providers (>= 50 tests)	Match Rate Deterministic: Avg (St. Dev.)	Match Rate Probabilistic: Avg (St. Dev.)
#1	77	96% (4%)	98% (3%)
#2	312	80% (7%)	96% (3%)
#3	125	73% (10%)	84% (10%)

The table shows not only that the match rates for tests ordered by physicians at Provider Organizations #2 and #3 (the IPAs) are lower than for those ordered by physicians at #1, but that greater variation exists in the match rates as well, as evidenced in the higher standard deviations. The frequency distributions below further illustrates these variations. Note that the spread of values is significantly greater for Provider Organizations #2 and #3. The differences in variation are not unexpected, given the absence of centralized information systems and uniform lab-ordering processes across the physician offices of Provider Organizations #2 and #3.





Because the same lab reported all of the test data we analyzed, the significant variability in match rates across physicians and across provider organizations suggests that impediments to accurate matching are due, at least in part, to differences in information management at the physician offices and provider organizations, rather than solely to the result of data-entry errors or other processing problems at the lab. Systematic laboratory errors that prevent patient matching would probably occur with equal frequency across ordering physicians, a phenomenon not seen in the data.

4.5 Role of Health Plan Member IDs

In theory, health plan member IDs could be very useful for patient matching, because they are more specific to individual patients than gender, first name, last name, or even date of birth. In practice, however, member IDs are not useful for matching because the IDs that are reported with lab results are often entirely incorrect or at variance with their representations in the patients' eligibility record.

As discussed earlier, lab orders from Provider Organization #1 lack member IDs entirely 75% of the time. This presents an obvious problem for matching based on member IDs, so lab orders from that provider organization are excluded from the analysis below.

For Provider Organizations #2 and #3, the following statistics are noteworthy regarding the degree of agreement between health plan member IDs in matched records.

Provider Organization	Member ID NULL in Lab Result	Member ID No Match	Member ID Exact Match	Member ID "Near" Match
#2	0.6%	61.7%	7.3%	30.9%
#3	1.1%	29.7%	26.6%	43.7%
Combined	0.7%	54.0%	12.0%	34.0%

As the data show, the problem with matching on member ID is not that this data element is unavailable. Indeed, a member ID value is omitted from fewer than 1% of the lab results reported to these two provider organizations. Rather, the value of the data is frequently incorrect or inconsistent. These mismatches occur for a number of reasons.

In 54% of the results analyzed, the member ID in the lab result bears no resemblance to that in the patient's eligibility record. Based on discussions with provider organizations and laboratories, these types of errors occur for at least two reasons:

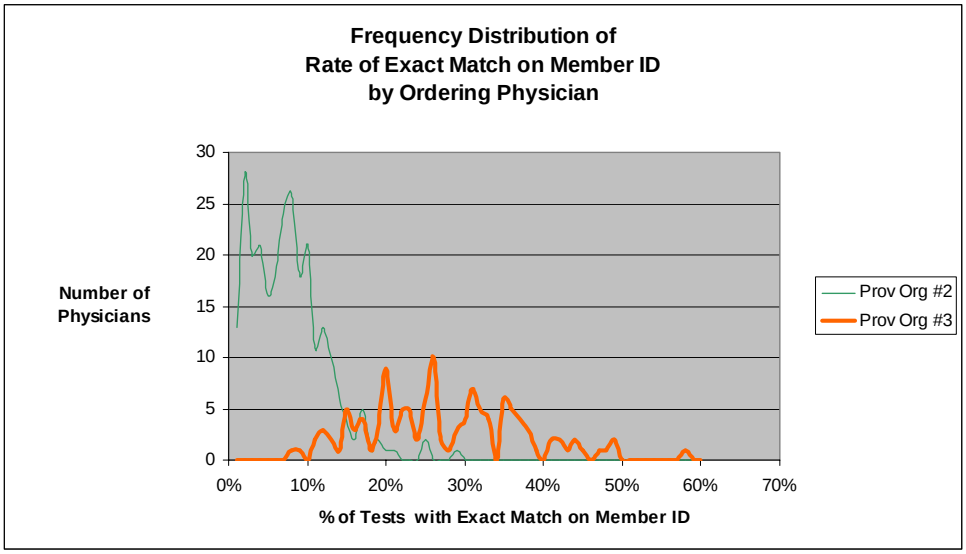
1. The patient's ID has changed because of a change in health plan coverage or dependent status. However, the physician office has not updated its records or the patient continues to use an outdated health plan ID card, resulting in placement of the outdated member ID on lab orders. Because most of these patients are capitated, payment for services does not depend on tracking of correct health plan member IDs, so these errors can go unnoticed and uncorrected.
2. The physician office places the patient's medical record number or social security number on lab orders, rather than the patient's health plan ID. For capitated patients, physician offices often maintain these "local" identifiers in their information systems, rather than the patients' health plan IDs (even storing them in the "Health Plan ID" field!). These numbers are then used on lab orders instead of the correct Health Plan IDs.

Only 12% of the lab results contain a health plan member ID that matches exactly to that in the patient's eligibility record. However, this rate varies significantly between the two provider organizations. As the table above shows, only 7% of the lab results from one of the provider organization match on member ID, versus 27% of the results from the other organization.

Within the provider organizations, the inter-physician rate of exact matching also varies significantly. The following table shows the average rate at which member IDs match exactly across ordering physicians with at least 50 ordered tests (as well as the standard deviation).

Provider Organization	# Ordering Physicians	Exact Match Rate Average	Exact Match Rate Std Dev
#2	265	7%	5%
#3	124	27%	10%
Combined	389	13%	11%

This variance is also illustrated in the following chart, which shows the frequency distribution of exact ID matches across these physicians. Note the spread of values, especially among the physicians in Provider Organization #3.



Finally, for 34% of results, the member IDs in the test result and the eligibility file do not match exactly, but clearly are derived from the same ID (i.e., "near" matches). The cause of discrepancies in near matches include omission of leading characters, omission of trailing characters, and/or addition of extraneous hyphens. Several examples of such near matches are shown below.

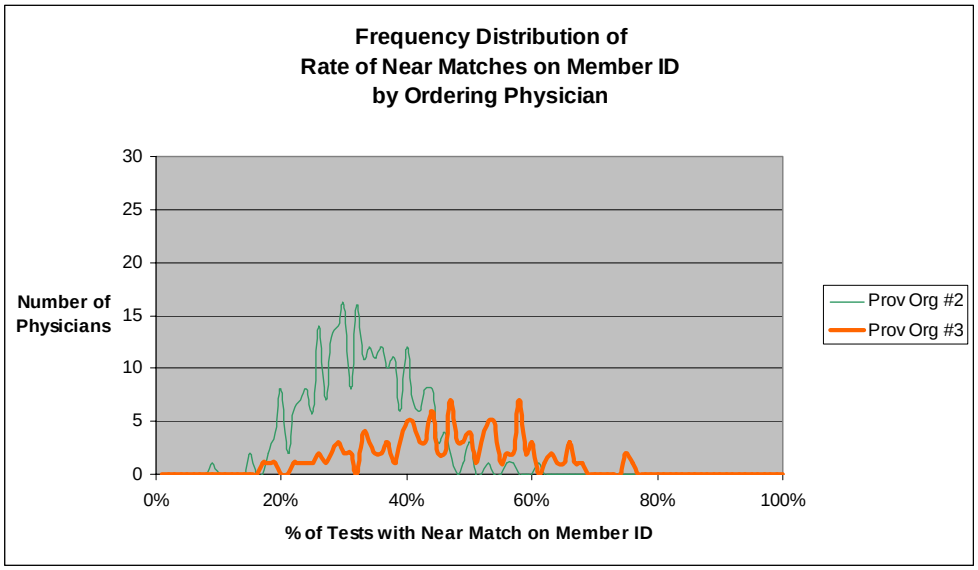
	ID in Lab Result	ID in Eligibility File
Example 1	XEEG9202726301	G9202726301
Example 2	R04779543	R04779543-FM1
Example 3	XDS826A55769	826A55769-20
Example 4	730488502	7304885-02

Near matches are significant for two reasons. First, they can contribute useful information to the matching process if the matching algorithm can reconcile discrepancies between nearly matching IDs (we did this in our analysis via simple string comparisons). Second, they suggest that much more exact matching on member IDs could be achieved if the reasons for the "near match" discrepancies could be corrected. These reasons include:

- Inconsistencies between the member ID that appears in eligibility files and the ID printed on member cards (see examples 1 and 3 above)
- Omission of trailing "family member" suffixes on member cards, although such suffixes are included in eligibility files (see examples 2 and 3 above).
- Inconsistent transcription of member IDs into the practice management systems of physician offices and/or onto lab orders (see example 4 above)

As with exact matching, significant variations in the rate of near matches also exist across provider organizations and across ordering physicians. These variations suggests that the specific practices of physician offices with respect to test ordering and the specific practices of provider organizations with respect to eligibility-file management influence the degree of

agreement between member IDs on test results and in eligibility files. The following chart shows the frequency distribution of such "near matches" in the data we analyzed.

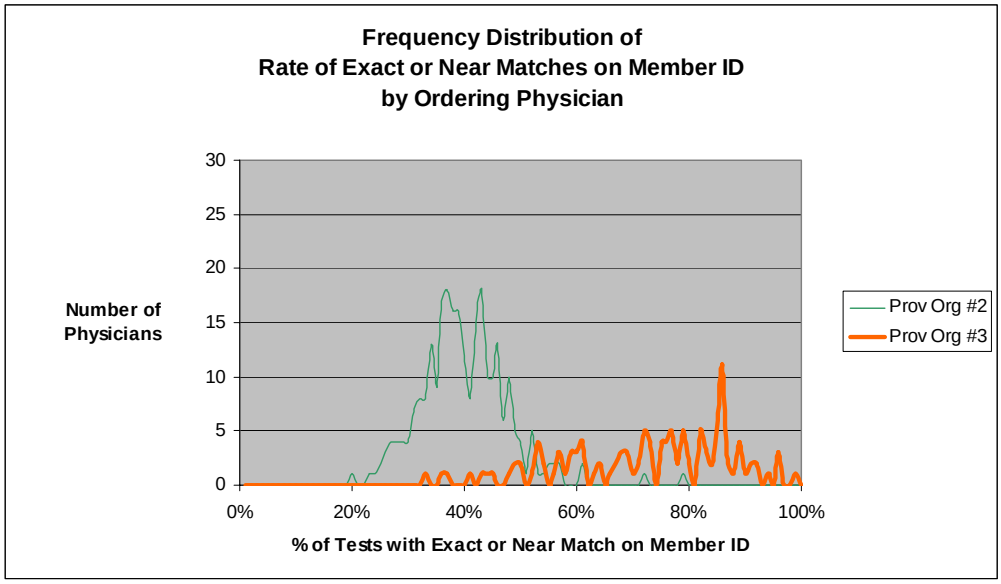


As a result of the frequent discrepancies in Health Plan Member ID described above, most organizations do not rely on this data element when matching lab-result data. The IDs rarely match exactly and most organizations do not apply string manipulations to detect near matches. However, the data show significant variability in the rate of matches and near matches across provider organizations and across physicians within provider organizations, and this variability suggests that *potential* exists for health plan member IDs to match with much greater frequency. Specifically, a minority of ordering physicians are able to achieve very high rates of ID matching or near matching for the tests they order. The following chart shows the range of values for an "exact-or-near-match rate"[5] across all of the ordering physicians. The chart shows, for example, that one physician achieved an exact-or-near-match rate of 99%. Nine other physicians (2.3%) achieved exact-or-near-match rates greater than 90%.

	Lowest Rate	Hightest Rate	Average Rate
Exact Match	0%	58%	13%
Near Match	8%	76%	36%
Exact or Near Match	20%	99%	50%

The following chart illustrates the distribution of the match-or-near-match across ordering providers with > 50 test orders.

[5] This rate represents the proportion of tests in which the member ID on the lab result either matches exactly or matches approximately to the member ID in the corresponding eligibility record.



These data suggest that, if the causes of the variations between the highest and lowest matching rates for member IDs can be addressed and the highest rates can be achieved by many more physicians, member IDs in lab test data could match those in eligibility files with a much higher average frequency, perhaps exceeding 90%. Such a result would greatly increase the value of member IDs for patient matching and could significantly improve the overall match rates for lab result records, especially for organizations that do not use probabilistic matching methods and cannot rely on centralized information systems to generate consistent lab orders.

4.6 Interpretation of Findings

Several tentative conclusions can be drawn from the information we gathered and the data we analyzed.

- The problem is real. A relatively small but significant number of lab results get omitted from the batch result files sent by labs to provider organizations and to health plans. Additionally, provider organizations and health plans cannot correctly match the results they receive to their patient populations in a small but significant proportion of cases. Both problems cause relevant lab results to be excluded from databases used for quality measurement and quality improvement.
- Lab result records are incorrectly omitted from the batch files sent to provider organizations for a variety of reasons. Sometimes incorrect information regarding the insurance status of the patient is placed on the lab order. Sometimes no information regarding insurance status is placed on the lab order and the lab is unable to ascertain the correct payer via matching to various eligibility files. Sometimes the information about insurance status is correctly placed on a lab order but incorrectly entered by the lab. Sometimes labs do not correctly associate the patient's insurance status with the right provider organization (although the insurance status was accurately recorded on the lab order and correctly captured by the lab). All of these causes ultimate relate to data-management practices that are susceptible to human error.

- The rate at which provider organizations match lab results to their eligibility files varies significantly by provider organization, ranging from a low of 82% to a high of 98% in the data we analyzed.
- In most cases, probabilistic matching techniques can significantly increase the proportion of lab result records that match to eligibility files. We observed that the additional proportion of records matched after the application of probabilistic techniques was as high as 16%.
- "Integrated" provider organizations that use centralized systems and consistent processes to generate lab test orders can significantly increase the proportion of result records that they match to eligibility files, even in the absence of applying probabilistic matching techniques.
- At provider organizations that are not integrated (e.g., IPAs), the rate at which lab results can be matched to eligibility files varies significantly based on the specific physician who ordered the test. These variations presumably reflect the way that different practices manage patient information and generate lab orders.
- The health plan member IDs that currently get reported with lab results are not useful for patient matching because they are inconsistent about 90% of the time with the member IDs that appear in eligibility files. In about 40% of these cases, the inconsistencies are minor formatting differences ("near matches"). In about 60% of these cases, the inconsistencies are grossly incorrect member IDs.
- The rate at which member IDs in test results correspond to member IDs in eligibility files varies significantly based on the physician who orders the test. Typically, the rate is below 50%. However, for tests ordered by a handful of physicians, the member IDs match or nearly match to eligibility files at a very high rate (> 90%). The management of patient information and the test-ordering processes in physician offices, therefore, seem to significantly affect the chances that the member IDs reported with test results will match those in eligibility files.

5 Recommendations

Based on our findings, we believe the following specific measures would significantly improve the accuracy of lab-result matching and lab-result routing. Realizing that stakeholders cannot pursue all of these proposals immediately, we categorize our recommendations by the time frame in which they may be possible.

5.1 Lab-Result Matching

Near Term

1. Provider organizations and health plans should avail themselves of probabilistic matching software. There exist a handful of cost-effective products, at least one of which has already demonstrated value in this specific domain.

Medium Term

1. Health plans should ensure that the same patient information appears on the ID cards that they provide to members and in the eligibility files they distribute to provider organizations and laboratories. This is particularly true for health plan member IDs. Additionally, plans should provide distinct cards to each family member covered, with the appropriate dependent code included in the ID on this card (i.e., each covered individual should have a distinct member ID). When members' demographic information changes (including member IDs), the health plans should distribute updated cards and eligibility files in a prompt and coordinated fashion. If possible, health plans should seek to avoid changing a member's ID during the member's entire period of continuous enrollment (regardless of changes in subscriber/dependent status, coverage type, etc.).
2. Health plans in California should standardize the file format and delivery mechanisms they use to distribute eligibility information. Standardization will allow provider organizations and labs that depend on this data to streamline their eligibility-update processes. Also, health plans should establish a frequent and regular update cycle for eligibility data, perhaps as frequently as weekly or bi-weekly. If the file format and delivery mechanism are standardized across health plans, provider organizations and labs will probably be able to accommodate more frequent eligibility updates. Such improvements will reduce the incidence of mismatches between the demographic information that appears on lab results and in the eligibility information used for matching.
3. Physician practices should strive to maintain within their practice management systems an accurate, complete and up-to-date record of their patients' demographic information *as provided by the patient's health plan*. The source of this information may be either an electronic eligibility file (as in the case of integrated medical groups) or a health plan ID card (as in the case of IPA affiliates). If ID cards are used, the information should be recorded exactly as it appears on the card, including the complete health plan ID of the individual patient.
4. Physician practices should include on all lab orders the relevant demographic information recorded in their practice management systems: Last Name, First Name, Date of Birth, Gender, and Health Plan Member ID. Practices should establish and follow specific processes to correctly and consistently record these data on lab orders. Such processes

may include pre-printed patient-identification stickers or patient-specific lab forms that include the relevant information.

5. IPAs should encourage improvements in the information-management and test-ordering practices of their member physicians by emphasizing the importance of accurate lab-result matching for pay-for-performance scores and quality-improvement efforts. Obviously, not all practices will be able to make the recommended changes, but IPAs should assist those that are willing by providing advice, eligibility data, and relevant analytical services.
6. Further study should be conducted of the provider organizations and the individual physicians whose test results already demonstrate high rates of matching to eligibility files, based on the data we have analyzed. The specific processes used by these organizations that improve patient matching of test results should be shared with other provider organizations and physician practices.

Long Term

1. All health plans should standardize on a member ID card that includes a two-dimensional bar code representing key patient-identification information. The Workgroup for Electronic Data Interchange (WEDI) is currently considering such a standard for health insurance cards (see healthdatamanagement.com/HDMSearchResultsDetails.cfm?articleId=14624). Bar-coding the information would avoid common errors and inconsistencies in data entry by the physician office or the lab. For printed lab requisitions, a photocopy or printed image of the two-dimensional bar code should be included with the order, allowing labs to directly capture the patient demographic information without manual entry. This bar-coding approach allows labs to leverage the information on the ID card even if the physician office lacks a bar code reader. Assuming consistency between the information encoded on the ID card and that appearing in the eligibility file, this approach would eliminate many of the sources of human error that currently impede effective data matching.

5.2 Lab-Result Routing

Near Term

1. Laboratories should also avail themselves of probabilistic matching software when matching test results to the eligibility files provided by health plans. This matching process, which is required to correctly route test results to health plans when patients are capitated, currently suffers the same inaccuracies as the process used to match results after they are received by provider organizations. The outcome is that relevant lab results are omitted from the files sent to health plans, a problem that could be mitigated by the use of more powerful matching techniques.

Medium Term

1. All of the medium-term measures to improve lab-result matching will also improve lab-result routing to health plans, because correct routing actually involves the matching of test results to the health plans' eligibility files. The greater consistency in patient information between lab test orders and eligibility files, the more accurate this matching will be.

2. All of the medium-term measures to improve lab-result matching noted above will also improve the routing of lab results to provider organizations. Certain labs attempt to match data in lab orders to data in the eligibility files of provider organizations to determine or verify the correct organization to bill for lab tests. Most such matching is done in an automated batch process rather than interactively, so greater consistency between the information on test orders and in eligibility files will improve the accuracy of this matching. That, in turn, will improve the completeness of the result files sent to provider organizations.

Long Term

1. For capitated patients, the standard bar coding of health plan ID cards should include the identity of the contracted IPA. If the bar-code image itself is available to the laboratory, this information will allow the laboratory to unambiguously determine the provider organization that should receive the result. Similarly, if the bar code also includes an appropriate identifier for the payer, plan and group, labs will be able to use the information to correctly route results to the appropriate health plan without needing to perform the error-prone lab-result matching.

6 Future work

The aim of additional work on this topic would be to better understand and better effect the changes needed to improve the current system, as recommended above. Specific areas of future work might include:

- Coordination with the WEDI group that is currently developing standards for health insurance cards, to ensure that the needs of stakeholders engaged in retrospective lab reporting are adequately represented in that activity.
- Study of the processes at the physician offices that have achieved the highest rates of matching to eligibility files, as well as the greatest consistency between health plan member IDs on their lab orders and in eligibility files.
- Analysis of match rates for test results performed by other laboratories. The results that we analyzed all originated from a single reference laboratory.
- Development and promotion of a single file format for eligibility data and a uniform update frequency for such data (this would be similar to the CALINX work previously undertaken for pharmacy utilization data, which also originates from the health plans).

7 Sources of Information and Data

Name	Type	Source
Blue Shield of California	Health Plan	Michael Higgins
Brown & Toland Medical Group	Provider Organization	Ann Hardesty Ray Birmingham
California Pacific Medical Center	Hospital Lab	Kathy Blankenship Ken Fountain
Greater Newport Physicians	Provider Organization	Wayne Sass
Intelligent Healthcare	Data Integrator for 20 Provider Organizations	Paul Katz
Laboratory Corporation of America	Reference Lab	Derek Walter
PacifiCare	Health Plan	Kathy Mota Gina Gasilan
Quest Diagnostics	Reference Lab	Andy Shaw Patrick Boyle
Sharp HealthCare	Provider Organization	Betsy Ellis