OMPRO
A Healthcare Quality Resource

Final Report

Oregon Chronic Disease Data Clearinghouse

August 19, 2005

Submitted to:

Oregon Health Care Quality Corporation

OMPRO
A Healthcare Quality Resource

Final Report

Oregon Chronic Disease Data Clearinghouse

August 19, 2005

Submitted to:

Oregon Health Care Quality Corporation

Oregon Chronic Disease Data Clearinghouse

Table of Contents

Oregon Chronic Disease Data Clearinghouse

Final Report

OMPRO had the privilege of acting as the clearinghouse for the chronic disease data clearinghouse pilot in Oregon. The purpose of the pilot was to test the legal and technical issues of getting electronic information into physician's hands to help them provide quality care to their patients with chronic disease. Specifically, the goals of the project as described in the contract were to

- use health plan data to identify patients who may be appropriate for asthma and/or diabetes care management from their physician or clinic
- provide recent inpatient, outpatient, laboratory, and pharmacy information about those individuals in a common integrated format to the appropriate physician or clinic
- design an electronic and paper format for this patient-level and aggregated information that is useful to the physician
- obtain physician feedback to the plans on the accuracy of the patient-level information
- identify legal and logistical issues that need resolution to continue the effort beyond the pilot stage

This report presents OMPRO's experience as the chronic disease data clearinghouse (CDDC), lessons learned, and recommendations for future development of a statewide clearinghouse.

Summary of Phase I (February 2003 – January 2004)

OMPRO was selected as the clearinghouse vendor by the Oregon Health Care Quality Corporation (OHCQC) on February 19, 2003 and signed a contract with OHCQC on September 10, 2003. OMPRO joined Oregon Department of Human Services' (DHS) Asthma and Diabetes Programs, and Riley Research to become the clearinghouse staff.

The nine-month delay between contract award and effective date were due to OHCQC's inability to secure the necessary funds for the pilot project. As a result, the pilot project was divided into two phases, with funding secured for Phase I of the pilot at the time of contract signing. In spite of the funding delay, the clearinghouse staff began meeting in April 2003 to begin planning the pilot project. The project plan was finalized by September 2003.

Data Specifications

Data submission specifications were finalized during Phase I as a result of close collaboration between the health plans and clearinghouse staff. The support and willingness of all the health plans to build the technical specifications was invaluable. In particular, Care Oregon, Providence, and Regence provided one-on-one consultation. The diabetes specifications were based on HEDIS criteria; the asthma specifications were based on the Oregon's Asthma Data Workgroup criteria.

Final data specifications were sent to interested health plans on November 7, 2003. The health plans were asked to submit two data sets: (1) the first data pull to build and test the clearinghouse processes and (2) the second data pull, at a later date, to generate the final reports that would be sent to the physicians. Receipt of Data Pull 1 was set for November 2003. This date was later moved to May 2004 due to delays for data processing at the health plans.

Proof-of-Concept Evaluation

By end of October 2003, the CDDC Steering Committee finalized the evaluation questions for the clearinghouse pilot. Of the six evaluation questions to be answered at the end of the pilot for proof of concept, OMPRO was responsible for questions 2, 3, and 4. The responses to these evaluation questions are in Appendix A. Evaluation of the clearinghouse pilot was later expanded beyond the proof-of-concept requirements. An independent evaluator was hired to more extensively review the clearinghouse pilot.

HIPAA Compliance

A model data sharing agreement was drafted by OMPRO with advice from legal counsel. The data sharing agreement clearly defined the health plans as the covered entities who owned the data, with OMPRO as their business associate. The agreement recognized the legal and operational agreements between the health plans and the participating physicians, and as such, authorized OMPRO to act on behalf of the health plans in sending patient-specific reports to the physicians who treat members of the health plans.

The agreement was sent to the health plans in mid-November 2003. Five health plans signed and returned the model agreement. Six health plans modified the agreement as members of the Oregon Healthcare Payers Forum and signed the modified data sharing agreement.

The two data sharing agreements differed in the section(s) applicable to data dissemination. The agreement modified by the members of the Oregon Healthcare Payers Forum developed a more restrictive dissemination clause than the model agreement drafted by OMPRO. Since the data were the property of the health plans, the plans had the authority to define the parameters of data dissemination by which the clearinghouse was required to abide.

HIPAA privacy regulations became effective in April 2003, just as the clearinghouse pilot was beginning. Most healthcare entities were still grappling with their implications and building operational and management protocols to address the privacy requirements. In spite of the many unknowns of HIPAA, the clearinghouse succeeded in addressing the necessary legal requirements effectively and efficiently, without the delays or complexities that were originally anticipated at the onset of the pilot.

By the end of phase I, OMPRO received data sharing agreements from two health plans only. The remaining data sharing agreements continued to arrive at OMPRO, with the final agreements received in May 2004. Six plans submitted the executed data sharing agreements before sending the data files; six health plans did not. From a legal and administrative standpoint, all health plans should have first finalized and executed the data sharing agreements prior to delivering data to the clearinghouse. Close collaboration between the health plans' QI and legal departments will be necessary to minimize the risk of noncompliance with HIPAA and other statutory laws governing data transmission, privacy, and security.

Data ownership and data dissemination will continue to be critical areas to address in future clearinghouse initiatives. Unfortunately, there is no easy solution for these issues. Much will depend on interpretation of HIPAA and state privacy regulations by legal experts, precedence set in case law, and agreed-upon terms by all stakeholders. These areas warrant further investigation, discussion, and legal opinions well before moving into any new phases of the clearinghouse initiative.

OMPRO's work for Phase I was completed by January 2004. By the end of phase I, OMPRO received data sets from 6 health plans. Although the clearinghouse staff originally targeted at least four health plans, it was surprised by the large number of willing health plans, and continued to enroll interested health plans.

Summary of Phase II (June 2004 – March 2005)

Funding for Phase II was again a challenge and resulted in a four-month unfunded period from January 2004 through May 2004. OMPRO and OHCQC signed a contract for a portion of the Phase II funding in June 2004. The remaining portion was provided by the asthma and diabetes programs at DHS. OMPRO signed an additional contract with DHS in June 2004.

The clearinghouse planning committee set a target of recruiting at least four health plans for data submission. By May 2004, 12 health plans submitted data. The large number of participating health plans validated the community's high interest in the project. For purposes of piloting the clearinghouse concept, however, the number of participating health plans created increased complexity in data management and the development and testing of the imputation and flagging algorithms. This had an impact on the projected scope, resources, and timelines. Data cleaning, merging, and matching took more time than originally estimated, which impacted subsequent pilot milestones. These areas will be discussed in more detail further into the report.

In June 2004, under Phase II, OMPRO began working with the Data Pull 1 to begin building and testing clearinghouse processes for (1) the request and receipt of data; (2) data management; (3) disease flag calculations; (4) PCP imputation; and (5) report generation. Once these processes were built, tested, and refined, the health plans would submit a second data set with more current data for the clearinghouse to use to generate patient-level data for the participating physicians.

By August 2004, OMPRO processed Data Pull 1 and sent patient lists to participating physicians for verification. Obtaining the lists back from the physicians took longer than expected, resulting in reports being sent to physicians in November 2004. Feedback from the physicians on errors in patient matching helped OMPRO discover an error in the disease flagging algorithm. This was corrected and OMPRO began processing Data Pull 2 in order to provide more timely patient information in the next set of reports.

By February 2005, the second list of patients was sent to participating physicians for verification. The physicians verified the accuracy of the lists, thereby validating the corrected algorithms. Reports were sent to the physicians in March 2005. March 2005 marked the end of funding for Phase II of the pilot, however, further reconciliation of the data in the physician reports and additional evaluation activities conducted by the evaluation consultant moved the end of the project until August 2005.

The pilot, originally scheduled for one year, took 2.5 years to complete. Of those 2.5 years, 18 months were unfunded.

Data Management

Data processing and transmission

As the health plans began processing the data request, OMPRO worked directly with their analysts. Each health plan's timeline and approach for data processing was unique and required one-on-one advice from OMPRO. Production of Data Pull 1 took considerable time at the health plan level. This was due in part to each plan's internal processes and resources, and in part to the new legal requirements under HIPAA. Once these issues were resolved and a system was in place, the health plans were able to produce Data Pull 2 significantly faster and with minimal assistance from OMPRO.

The clearinghouse offered several options for data transmission to the participating health plans. The preferred method was the use of encryption software for secure transmission via email. The software was provided at no cost to a health plan. Upon request from a health plan, OMPRO sent this easy-to-use software, a user's guide, and instruction on how to transmit the data. Of the 12 participating health plans, 7 requested a copy of the software, and five used it to transmit data to the clearinghouse. The remaining plans either used their own software, or simply sent the data on a compact disc.

Data storage

For each data pull, the clearinghouse received a minimum of four data sets from each health plan. The data were loaded onto OMPRO's MS-SQL server. Any data received on a disk were stored in a locked room after upload for security and privacy compliance. In anticipation of the data, OMPRO developed a program to upload the data to its MS-SQL server. The health plan data, however, came in many different formats, requiring OMPRO to modify the program to accommodate the uniqueness of each plan's data and to facilitate the upload of each data set depending on file format, field names, and field size. The data formats received from the health plans included the following file formats: .cvs, SAS, .mdb, .txt, .xls, .dbf.

Originally, the MS-SQL server had 20 gigabytes of data volume reserved for clearinghouse data based on the assumption that only a few health plans would participate in the pilot. After receipt of the second data pull, however, the server volume was quickly depleted. As a result, OMPRO upgraded the server's capacity to 90 gigabytes, 55 of which were allocated to the clearinghouse data. This alleviated the data storage constraints and allowed for more efficient storage and retrieval processes. Until the upgrade, however, data processing was a slow and cumbersome process. In the end, the clearinghouse data required approximately 35 gigabytes of server volume.

Data verification

The clearinghouse staff underestimated the number of hours required to clean, format, and standardize the data both due to the more-than-expected number of participating health plans and the lack of data standards across the health plans. Each plan (n = 12) submitted at least four data sets (patient demographics, provider demographics, claims, and pharmacy), with 15 to 30 variables, and up to 1.5 million records.

Data Pull 1

Upon receipt of the Data Pull 1, OMPRO set up queries and programs to verify the data. The health plan analysts were contacted if OMPRO had questions, concerns, or required clarification during the verification process.

The data tables varied across all health plans, which required OMPRO to revise the query for each set of files. This lack of data standardization required a manual verification process, which created a substantial disadvantage for any automated clearinghouse processes and prevented the use of any automated data verification process where data could be divided into sets of met or unmet data. An analyst was required to physically be present to run the queries and validate each result. To illustrate:

> The lack of standardization necessitated OMPRO to write 12 different codes to import the health plan data. In some instances, field names were either different from the clearinghouse specifications or coded differently. In other instances, health plan data included additional fields that the clearinghouse did not request or did not add blank columns when the data fields were not available. In other instances yet, data fields were merged where the specifications required two or three separate fields (i.e., patient name was merged into one field in contrast to a field for first name and a field for last name.) As a result, OMPRO reviewed each data table and modified the program to deal with the uniqueness of each data table.

Although each plan received the data specifications from the clearinghouse, few health plans submitted the variables in the format requested. Each health plan had developed its own definitions and formats for capturing the same data. In an effort to reduce the level of effort for the health plans, the clearinghouse attempted to ascertain data definitions of each health plan and match them to the required specifications. For example, four data sets provided claims data that could not match patient and/or provider demographics. Each data set had its own unique set of circumstances and within each set, no pattern for data definition could be identified. Other examples of data issues were date fields, and procedure and diagnoses codes. For a more complete review of data submitted by health plans see Appendix B.

The first Data Pull contained data that could not be matched using patient and/or provider demographics. Four plans had data where claims data could not be matched to demographic data using ID numbers. One case involved patient IDs. Claims data from four different claims files all contained fields named “UniquePatientID”. Patient demographics file contained a field named “UniquePatientID”. However, the “UniquePatientID” from claims actually matched a field “PatientID_Alt”, which contained “UniquePatientID” embedded in the variable (e.g., claim has “UniquePatientID”= 123ABC, “PatientID_Alt”= 123ABC-017, with no apparent pattern to the extension), something that wasn’t discovered until further investigation. One case involved provider IDs. Claims data from medical claims files contained field for “ProviderID”. Provider demographics file contained a field named “ProviderID”. However, the “ProviderID” from claims contained a truncated version of “ProviderID” from the provider demographics file (e.g., claim has “ProviderID”= 123ABC, provider demographics file has “ProviderID”= 123ABC-017, with no apparent pattern to the extension). One case involved most provider IDs not being in provider demographics files. Another case involved a plan where provider IDs, patient IDs, and other fields did not seem to match in any particular way.

Data Pull 2

Many data issues that were present with Data Pull 1 did not occur in Data Pull 2. This was attributable to the one-on-one interactions with the plan analysts. Certain variables were recoded or reformatted to better suit the needs of the clearinghouse. Other issues surrounding data, however, remained from Data Pull 1. Data were not standardized, and files and variables remained in different formats. Significant data cleaning was still required. Internally, the clearinghouse developed the concept of "good" and "useable" data. Useable data were those that, with some attention, reformatting, and cleaning, could be used in analysis (e.g., ICD-9-CM codes that needed to have decimals removed or dates in text format that needed to be changed to date/time format). Good data were those that were ready to run through the processing algorithms. There was very little "bad" (un-useable) data. Of the 12 health plans that submitted data, only one health plan's data could not be used in the pilot. Due to time constraints and deadlines, the clearinghouse could no longer delay data merging in order to permit the health plan to resolve its data file issues.

Lessons learned

It is important to have a better estimate of the number of participating health plans in order to allocate both labor and financial resources. The technical infrastructure developed by the clearinghouse, although scalable, assumed that no more than four health plans would participate in the pilot. Upgrading the infrastructure to support two data pulls from 12 health plans took time and additional resources on the part of the clearinghouse.

Understanding the variation across health plan data consumed more resources than were originally budgeted. Without data standardization across the health plans, it was critical for clearinghouse staff to understand the uniqueness of each health plan's data and how it correlated with the clearinghouse data specifications. This process required in-depth analyses and interactions with the data analysts from the health plans.

The process for accepting data that do not meet the clearinghouse specifications should be tightened, giving the clearinghouse authority to reject data that do not meet requirements. This approach would create additional resource outlays for the health plans, but would allow the clearinghouse to automate the data verification process and avoid rewriting the program each time, for each file, for each health plan. It will also allow the clearinghouse to run automated programs overnight and in batches to minimize production slow downs. An automated program also allows clearinghouse staff to produce a log that identifies which verification checks failed or were missed, and allows for expedient follow up with the health plans to correct these issues.

Finally, the clearinghouse should establish one method for data submission, using a secure and encrypted Internet tunnel. This allows the clearinghouse to receive data according to HIPAA security standards and to maintain an automated log of what was received, what was uploaded, and who was notified. OMPRO has since built a secure Internet environment through a virtual private network, but this capability was not available during the pilot's data transmission phase.

Data merging and matching

Data were merged and matched using either SAS or MS Access, depending on data volume and task requirements. The clearinghouse staff approached this activity with the goal of automating data processing beyond the pilot and into a statewide initiative, even with disparate data file formats from the health plans.

Patient demographics file

The ability to accurately merge the patient files and eventually match the patients to providers relied on a unique identifier for each patient. When assessing the patient demographic file, the clearinghouse faced a challenge—in many files, patient names were not separated by first, last, and middle names; and in many files, patient names were contained in one field.

In the case of full names contained in one field, some were separated by a comma, one or two spaces, or another method. In some cases, a middle initial was included; in some it was not. The order of names was inconsistent—in many cases the order was first name, last name, and middle initial, in other cases the order was last name, first name, and middle initial, or many different combinations. This variation and inconsistency prevented any automated data processing approach, since the clearinghouse had to modify or rewrite its code to handle all of the data variations.

Although the name of the patient was critical to the patient-physician matching process, it could not be relied on as a unique patient identifier. The initial thought was to use a plan's unique patient identifier and match this with the plan's provider identifier. This, however, would not have produced an accurate match for several reasons. First, patient crossover was not only significant between managed care plans and the Oregon Health Plan, but also among managed care plans. Second, it appeared that many health plans used duplicate patient identifiers. Third, it also appeared that when patients changed health plans, they received a new patient identifier at the new plan.

For these reasons, the clearinghouse could not use the health plans' patient identifiers and explored other methods of identifying unique patients, regardless of misspellings or various formats of the same name. Other approaches the clearinghouse considered were (1) assigning a random number to each patient, but this would prevent the capture of patients as they moved among health plans; and (2) using the social security number, but many health plans did not submit this variable, submitted portions of the number, or variations thereof.

Consequently, the clearinghouse developed the following schema that took several demographic variables associated with a patient and created a clearinghouse patient identifier:

first 5 characters of the last name + first 4 characters of the first name + gender + date of birth

This method provided an accurate patient identifier. Duplicate identifiers were possible based on this schema, but the chances of this occurring were very small. Also any patient name changes would void this schema.

Provider file

The provider file presented both similar and unique challenges. Perhaps the biggest challenge inherent to the provider file was the multiple definitions of providers among the health plans. These definitions included either clinics, systems, facilities, or physicians based on the health plan's contracting relationships. Additionally, there was no apparent pattern to how a provider would be defined. Any or all of the above definitions could have appeared in the same provider name field in one health plan's file.

When a physician's name was used as the provider name, that physician was not always the primary care physician. In one case, the physician identified was a medical director that had left a clinic four years ago, although his name was still being used for billing purposes. Because the intent of the clearinghouse pilot was to deliver patient-specific reports to an imputed primary care physician, the multiple definitions and incorrect name entries prevented the clearinghouse from accurately identifying a physician as the primary care provider.

The clearinghouse approached this challenge similarly to its approach of the patient file. It designated a unique provider identification regardless of how the health plans coded or defined a provider. The clearinghouse took into account physician names, clinic names, healthcare center names, hospital names, laboratory names, and many other variations occurring in the health plan files. Consequently, the clearinghouse developed the following schema for a clearinghouse provider identifier:

first 5 characters of the last name + last 3 characters of the last name + first 4 characters of the first name

In cases where the provider name did not appear to be an individual physician, the entire name was used as the *last name*. Before the clearinghouse could apply the formula, all unknown, blank or non-provider entries were manually removed. This schema appeared to accurately identify individual physicians within the multiple variations of provider definitions. Other providers, such as clinics, were not always accurately identified. Overall, the process was manual and cumbersome.

Associating a physician to a facility, clinic, or system was another challenge. It is common for physicians to practice at multiple locations and to have different provider identifiers, even within the same health plan. Clinics with resident physicians bill under the same physician identifier.

Claims and pharmacy files

These files required standardization as well. For example, before merging all claims files, the clearinghouse standardized date fields, and diagnosis, revenue, and CPT codes across all files. For the pharmacy files, the clearinghouse calculated dispensing types and age before merging.

Lessons learned

Data standardization is essential to develop efficient, cost-effective, and automated processes for the clearinghouse. Again, this will require an additional resource outlay at the health plan level, but will create efficiencies for the clearinghouse. Definitions for variables, such as *provider,* will need to be developed and followed in order to determine a primary care provider or clinic accurately. The goal of the clearinghouse was to assign a physician as the primary care provider and provide patient-level reports to that physician. Health plan data, as it is currently collected, does not allow for accurate assignment at the physician level.

Physician identification at the health plan level and at the clinic/practice level appears to be highly unreliable. Based on the data submitted by the health plans, there was no clear solution for reconciling a provider identifier, especially with the goal of identifying a physician as the primary care provider.

Another consideration would be to assign the medical practice as the primary care provider rather than an individual physician. This approach would change the focus from delivering reports directly to physicians who care for patients to medical practices who care for patients. In a paper-based environment, the practices would be responsible for distributing the patient-level reports to the physicians. In an electronic (Internet) environment, the practice would match the physician to the patient before generating any patient-specific reports.

Finally, the clearinghouse may need to begin compiling a validated provider list that will grow as more participants join the initiative. This list should be shared with the health plans as a means of information validation. Validating the list will be resource intensive, but may eventually provide the most accurate source of information.

Disease flagging algorithms

The clearinghouse staff selected HEDIS 2004 criteria as the method of identifying patients with diabetes. Many methods were explored, but in the end HEDIS was selected because (1) plans were familiar with HEDIS methods; and (2) HEDIS diabetes measures were widely used and accepted as a standard for diabetes care. It is important to note that not all HEDIS criteria were used. Criteria that were not necessary for clearinghouse purposes, such as age, eligibility, and exclusion criteria, were not included. For asthma, the clearinghouse staff used Oregon's Asthma Data Workgroup criteria.

The clearinghouse requested the health plans to submit data on patients they believed had diabetes or asthma, using the data specifications for submission guidelines. For asthma, each plan was to identify all patients with any asthma-related ICD-9-CM diagnosis code (493.xx). For diabetes, plans were to identify all patients for any of the diabetes-related ICD-9-CM codes (250, 357.2, 362.0, 366.41, 648.0). For pharmacy data, plans were to submit any claims that contained codes for drugs related to asthma or diabetes as defined by HEDIS criteria. After identifying these patients, the clearinghouse requested that the health plans submit all claims associated with those patients for two consecutive years. For Data Pull 1, the timeframe was between July 1, 2001 and July 31, 2003. For Data Pull 2, the timeframe was between April 1, 2002 through March 31, 2004.

The first data pull was used to program and test the disease flagging algorithms. The first test of the diabetes algorithm discovered a flagging error. One filtering criterion was inadvertently eliminated, thereby counting outpatient visits for all diagnoses, not just diabetes, in one of the queries of the flagging algorithm. This caused patients to be flagged with a diabetes visit when they actually had an outpatient visit for some other diagnosis. This error was discovered when the patient reports were sent to the test-cycle physicians who reported back on the inaccuracies of the over-counting of patients with diabetes. The error in the algorithm was found and corrected, and for Data Pull 2, the disease flagging algorithms were highly accurate—90% for diabetes and 80% for asthma. COPD was the most commonly mis-flagged condition for patients with asthma.

Lessons learned

Future clearinghouse specifications may need to incorporate HEDIS exclusion criteria. This may help in reducing the number of false positive for patients with diabetes or asthma, since HEDIS excludes certain conditions like COPD.

Imputation of primary care provider

The clearinghouse began developing its imputation algorithm during Phase 1 of the pilot, after finalizing the data specifications. This process included collecting and evaluating existing imputation algorithms used at OMPRO and elsewhere. Clearinghouse staff interviewed analysts and IT staff from health plans and learned that while some used algorithms to impute the PCP, others did not. Again, definitions for PCP varied across the health plans. For some, it was the clinic, for others it was the individual physician, and yet for others it was any healthcare provider.

A review of all known methods revealed that every imputation algorithm had its limitations. For example, one health plan used a three-tiered imputation logic by associating (1) a PCP with the patient's last visit based on encounter data, or (2) a PCP closest to the patient's home ZIP code, or (3) a PCP closest to patient's work ZIP code.

Another health plan imputed to a clinic only, not to an individual physician provider as necessary for the clearinghouse. A third health plan applied a three-tiered algorithm that was condition specific, applied a tight list of specialty codes and specialty services. This algorithm also weighted the encounter data in order to assign the PCP.

This third algorithm closely met the needs of the clearinghouse, however, the clearinghouse could not apply it fully for several reasons. First, filtering for condition-specific codes required a unique algorithm for each condition, was resource intensive, and prohibitive to the clearinghouse due to limited resources and time. Second, although the clearinghouse originally considered applying a tight list of specialty codes, it found that not all health plans provided these data. Also, those that did, often times provided conflicting data in the fields that captured specialty type and specialty name. For example, one entry captured *hospital* as specialty type and *John Smith* as specialty name. By excluding this entry, the clearinghouse would have inadvertently excluded a possible physician (*John Smith*) as a patient's primary care provider, due only to incorrect coding or data entry.

As a result of this review and in an attempt to develop an automated algorithm that could be applied to any clinical condition or clearinghouse indicator without resource-intensive revisions, the clearinghouse staff developed an algorithm that combined the different approaches into the following PCP imputation algorithm:

1. Identify unique patient and first day of service.
2. Identify the provider who provided the majority of the care for that patient.
3. Assign provider as the PCP. Use plan definition of PCP (e.g., physician, clinic, etc.).
4. If the visits are evenly distributed between multiple PCPs, then assign the PCP with the most recent service provided.

The algorithm also accounted for any PCP assignments determined by a health plan. In these instances, if patients belonged to a single health plan and that plan assigned a PCP, the clearinghouse would apply that assignment.

Analysis of the fields associated with health plan assignment, however, revealed several inconsistencies and inaccuracies that prevented the clearinghouse from using the health plan assignment. For example, for those plans that assigned a contractual PCP, start and end dates for

assignment were anywhere between 1994 and 2004, with some assignments beginning and ending on the same day. In other instances, a provider's name was not provided even though an assignment was entered. In yet other instances, the plan-imputed provider was a facility and not an individual physician. Accuracy of current assignment, therefore, could not be validated without new data or information from the health plans.

As a result, plan assignment of PCP was not used as originally intended or anticipated. Instead, the clearinghouse applied its algorithm to match all patients to a PCP. This provided an opportunity to test the agreed-upon clearinghouse algorithm and determine if it were a feasible approach in a broader, statewide clearinghouse initiative for any condition and in cases where plan assignments appeared questionable.

A larger sample of clinics will be necessary to conclusively validate the accuracy of the clearinghouse's imputation algorithm. But preliminary findings show that the clearinghouse PCP imputation was successful at linking patients to providers. This linkage, however, did not always establish the correct patient-physician match due mostly to billing inaccuracies. The findings also revealed that when imputing, it may have been more effective to use the clinic as unit of analysis rather than an individual physician, since patients who were inaccurately linked to one physician, were actually patients of a colleague at the same clinic.

Lessons learned

The clearinghouse pilot afforded an opportunity to review many imputation algorithms and to test the agreed-upon algorithm developed by the clearinghouse. It is clear that more exploration is necessary to continue to improve the ability to impute a patient to a provider, whether the provider is a physician, clinic, or other entity. The following examples are only a few issues that could be considered in future clearinghouse initiatives.

Lesson 1. The broadly defined HEDIS criteria allow patients to be identified by pharmacy flags alone, therefore, a patient without any other claim information, would not be imputed to a provider.

Possible solution: Modify the HEDIS criteria to exclude pharmacy information, or use a different identification method.

Implications: HEDIS is well established criteria. Eliminating pharmacy information will not provide a clear picture of asthma and diabetes in Oregon.

Lesson 2. Many non-specialty providers who are ***not*** providing diabetes or asthma treatment may be imputed as the PCP, because the volume of treatment they are providing is higher than the volume of diabetes or asthma care provided by another provider.

Possible solution: Isolate only outpatient claims (assuming patient reports are for physicians or medical practices and not for hospitals or ER units) using billed HEDIS codes for outpatient visits that are asthma or diabetes related. It may be necessary to eliminate some codes from the HEDIS list since some are for office follow-up visits that may not be related to asthma or diabetes. Also a combination of procedures, and primary or secondary asthma and diabetes diagnoses could be considered.

Implications: Patients who have claims for emergency or hospital visits only, or who are identified as having asthma through pharmacy data and have no follow-up visits with a physician, will not be assigned to a PCP. In these instances, the plan-assigned PCP would be useful, if the data issues around plan assignment are resolved.

Lesson 3. Oregon patients who receive care elsewhere than Oregon, especially those with fee-for-service coverage, will not be captured for reporting to Oregon providers.

Possible solution: Remove all services provided outside of Oregon before applying the algorithm based on provider city and state.

Implications: This approach may not accurately capture the correct primary care provider, especially for FFS patients that live along Oregon borders with California, Washington, and Idaho. If the clearinghouse does account for "border" care, then parameters must be set for distance into the bordering state to be included into the clearinghouse data. On the one hand, this approach adds complexity to data analysis and reporting. On the other hand, access to the clearinghouse data could be opened to any provider who treats Oregon patients.

Lesson 4. The clearinghouse requires a clear definition of primary care providers. Certain specialty and provider types should be eliminated, such as obstetrics/gynecology, otherwise these providers may be imputed as a primary care provider.

Possible solution: There is currently no easy way to eliminate certain specialty or provider types.

Implications: This issue will require close discussion and collaboration with the health plans and other stakeholders.

Lesson 5. The current clearinghouse data specifications make it difficult to group data at the clinic or group practice level.

Possible solution: Future data specifications must include another layer of information and data table structure, where each individual physician is separated from a clinic in one data table and clinic information from each plan is provided in another. These two tables are then matched to link each physician with a clinic.

Implications: This issue will require close discussion and collaboration with the health plans and other stakeholders.

Lesson 6. There is no clear definition of how each health plan contracts or assigns a PCP. Assignment start and end dates in the current health plan data seem to be unreliable data sources.

Possible solution: Clearinghouse staff need a better understanding of why providers have multiple identifiers in health plan data. It may be that health plans have additional fields that clarify this occurrence. The clearinghouse must

ensure that it receives complete data specifications from all participating health plans, in advance, in order understand what data are unique to plans, to determine what data are available across all health plans, and what should be included in the clearinghouse data specifications.

Implications: This issue will require close discussion and collaboration with the health plans and other stakeholders. This may also be an ongoing activity as data collection approaches and information needs may change over time.

Lesson 7. Claims information and plans' provider files do not provide accurate address and mailing information. Addresses are formatted differently, many physicians appear to have multiple addresses, and it is unclear which address is correct.

Possible solution: The clearinghouse may need to keep an updated master file with current address and mailing information for all Oregon providers, or run an address validation through a third-party vendor to detect incorrect addresses.

Implications: This is an added cost and requires additional financial and human resources. Even if using a third-party vendor, matching or linking plans' data to vendor data may be cumbersome and will rely on an identifier that is common to both systems, such a UPIN, tax ID, or SSN, and common across all plans.

Lesson 8. More than one provider can be assigned as a primary care provider for many reasons. How does the clearinghouse reconcile this assignment, verify appropriate assignment, and remain HIPAA compliant?

Possible solution: Strive toward a simplified, uniform, and transparent approach by developing consensus among all stakeholders (healthcare providers and health plans) on use and disclosure of protected health information within the clearinghouse infrastructure. Consider establishing a Privacy Board to create a framework and oversee compliance across the clearinghouse activities.

Implications: HIPAA compliance, especially with respect to patient privacy, has big implications for future clearinghouse initiatives. As participation in a clearinghouse expands to include others, data ownership will become an even more delicate balance, and will require education, public trust, and transparency in a many-to-many environment.

Data reports

The data reports were the final outcome of the pilot clearinghouse. Indeed the primary focus of the pilot was to deliver patient-specific reports to physicians. The original scope of work called for paper-based and electronic patient reports. The clearinghouse staff agreed to produce paper-based reports only due to (1) insufficient funding; (2) uncertainty around the new HIPAA regulations; and (3) the limited time (12 months) originally planned for the pilot. Subsequent scopes of work would expand upon the lessons learned of this pilot and build the next level of reporting to incorporate electronic reporting.

Reports were generated for both data pulls. The reports generated with Data Pull 1 revealed the disease flagging algorithm discussed in earlier sections of this report. Because of the delays caused by merging and matching Data Pull 1, the error in the diabetes disease flag, and the time to verify the patient list, the data were out of date. Instead of continuing to reconcile these reports with data from clinics and the clearinghouse, the staff decided to retire Data Pull 1 due to the aged data and begin working with Data Pull 2. Once the data were cleaned, merged, matched, and the test-cycle physicians verified the patient lists, the clearinghouse produced paper-based reports for three clinics and 10 physicians.

Report generation revealed several lessons learned and steps to consider for future clearinghouse initiatives. The first was report production. The reports were designed using Excel software, which was a suboptimal tool for report production, especially when importing data from Access or SAS software. Future clearinghouse initiatives will need to allocate adequate resources to obtain a better report-generating tool, especially for paper-based reporting. The process of importing data to the reports required manual manipulation before they could be printed. For example, if a patient list was longer than the designated page length, it would spill into a second page with only the table footer showing on page 2. These issues appear relatively insignificant, but added considerable time to report production. This process will be unsustainable in a statewide initiative.

A second, more serious issue with the reports was the potential for gaps in the data. The data were based on HEDIS indicators, but represented information derived from claims data only. As a result, the reports did not reflect a full HEDIS data set (chart information, lab results, etc.). Additional gaps in the data were also likely because of gaps in the data received from the health plans. If physicians who received the reports were seeking a comprehensive and detailed report on their patients with asthma or diabetes, they may have been unimpressed. In preliminary discussions with a subset of test-cycle physicians, gaps were indeed identified. In the case of diabetes care, LDL testing seemed to be missed frequently. One test clinic explained that LDL tests may have been missed because the tests were not coded as diabetes related, but as a lipid disorder. Initial investigation of the data revealed that most nephropathy tests (>70%) and HbA1c tests (>80%) were being coded as diabetes related, while only about 40% of LDL tests were coded as diabetes related. This appears to be more a characteristic of billing rather than treatment practices.

A third issue that arose was missing patients or patients who were not matched to the appropriate physician. The clearinghouse staff worked closely with one of the test-cycle clinics, Legacy Clinic–Good Samaritan, to investigate this finding by reviewing a list of 322 diabetes patients generated from the clinic's registry against data in the clearinghouse. The clearinghouse staff first attempted to locate these patients in the clearinghouse patient files received from the health plans. Of the 322 patients, only 152 (47%) were found in the clearinghouse patient files. Of those, 146 (96%) were flagged with the appropriate disease flag. Of the 146 patients, 144 (99%) were imputed to a primary car provider. There may be several reasons why such a large number of patients did not get reported back to the physicians in the clearinghouse reports. First, the health plans may not have submitted the data. Second, the patients may have had no claims submitted during the time frame selected for Data Pull 2, even though they are patients of the clinic. Third, the sample from the clinic's registry may include patients who are inactive. Fourth, patients in the registry may be members of nonparticipating health plans, therefore, the data would not be in the clearinghouse files.

Next, the clearinghouse staff attempted to identify all physicians who practiced at the clinic and then match those patients found in the clearinghouse files with physicians from the clinic. The investigation resulted in the following findings:

- One health plan had the correct address, but the address was incorrectly linked to many other facilities associated with the system to which the clinic belonged.
- One health plan had the address misspelled.
- One health plan had multiple addresses for the clinic.
- One health plan used a post office box address for all the clinics associated with the system, including the clinic.
- One health plan provided no UPIN or facility information, so neither physicians nor a facility could be identified.
- One health plan used multiple provider identifiers for each physician and did not provide any facility information.
- Two health plans had no match.
- One health plan had one successful match, but the field was blank for the provider name.

These findings underscored the inaccuracies of the provider files discussed in earlier sections of this report and the challenges of not only identifying physicians, but clinics as well.

Project management—Lessons learned

In hindsight, the 12-month timeframe originally set aside for the clearinghouse pilot was not only ambitious, but also unrealistic. The pilot took 2.5 years to complete. Funding for the pilot was also insufficient. With funding at $70,000 for the entire pilot, the scope should have been better contained and managed to ensure optimal operations and maximum results for the price. It is important to note, that OMPRO contributed approximately $40,000 of in-kind support to help ensure pilot completion.

Although funding was limited, the biggest challenge was the lack of continuity in funding. This created tremendous operational challenges at OMPRO. The clearinghouse project consistently fell into “drift” mode when funding was unavailable or unclear. Staff and other resources could not be allocated consistently to the pilot project when other funded priorities surfaced. OMPRO staff needed additional training in more advanced programming and analytic tools, but this was difficult to justify in a project that was under funded and over budget.

OMPRO faced other internal operational issues that impacted its ability to meet deliverables independent of the clearinghouse pilot. Some of these were internal process and quality control, and appropriate management protocols and oversight. These issues have been addressed and appropriate steps were taken to ensure better operational and management oversight. Nonetheless, OMPRO was challenged to meet agreed-upon pilot deliverables in a timely manner and in a way the best represented its ability to complete the work.

The clearinghouse pilot included staff from DHS, OMPRO, and Riley Research Associates. While each partner was responsible for portions of the pilot, it was unclear who was accountable for overall project management and oversight. Unclear roles, responsibilities, and incomplete communication often led to unclear expectations, rework, scope creep, and delays. Staff turnover at all organizations (OMPRO, DHS, Quality Corporation) exacerbated this problem. Although OMPRO can only speak to the challenges it faced internally as a result of these circumstances, it is likely that all three organizations faced similar challenges.

Limiting the number of health plans that participated in the pilot would have provided a more manageable scope, allowed for more effective test cycles, enabled more rapid course correction, and would have allowed more in-depth analysis and discussion among the participating health plans and the clearinghouse staff. Perhaps, more thought should have been given to selecting a small, representative group of health plans to conduct the pilot, with opportunities to share the lessons learned with all interested health plans. These interactions could have also laid the framework for discussing data standardization across all health plans. The same is true for the selection of the participating test-cycle clinics. These clinics were selected late in the pilot and mostly based on their willingness to participate. Interaction with these clinics was limited due to the lateness of pilot deliverables and the need to complete the pilot. As a result, only a few clinics received the reports, and the clearinghouse staff worked with only one clinic to better understand the results of those reports.

The health plan data were complex; yet understanding these data is foundational for future clearinghouse initiatives. The clearinghouse staff needed more time to understand the data and interact with the health plans, but this was not feasible due to time limitations, funding, operational inefficiencies, and pilot focus. The primary goal of the clearinghouse pilot was to generate reports for physicians and to obtain feedback on the usefulness of the information contained in these reports. Secondary to this were the operational, technical, and legal requirements. In hindsight, perhaps a more appropriate first goal would have been to deal with the data challenges and reconciling the inaccuracies that occurred across the health plan data.

Finally, the additional evaluation component of the pilot conducted by an independent evaluator, although invaluable, was not originally part of the scope of work signed by OMPRO. The evaluation began later in the pilot, which added additional, unfunded requirements to be completed by the clearinghouse staff.

Conclusion

This report provided an objective and critical look at the clearinghouse pilot. For all its challenges, the pilot was successful in testing the first building blocks of a chronic disease clearinghouse. The pilot offers tremendous preliminary insight into health plan data, warehousing operations, data reporting, and technical and legal requirements. These lessons should not be forgotten, but built upon in future clearinghouse initiatives.

Sufficient and ongoing funding will be critical to build a successful clearinghouse. No other issue challenged the clearinghouse operations more than the lack of adequate funding.

Building a clearinghouse is a difficult and costly endeavor. There are many statewide initiatives underway across the country that are attempting to connect and exchange healthcare data. All of these, without exception, face similar challenges as the clearinghouse pilot. But it is those communities that succeed in building dialog and trust among all stakeholders will move beyond the challenges and will succeed in building a vision of promoting the use of data to improve the quality of care for patients with chronic conditions.

Appendix A. Proof of Concept Evaluation
Questions 2 – 4 Answered by OMPRO

2. What is the magnitude of the data merging?	How many health plans participated? **12 plans submitted data to the Clearinghouse. The plans ranged in size and geographic location.**
	How many patients with diabetes and asthma were identified? For each health plan, what percent of their patient population did this represent? **There were 62,634 patients identified with asthma and 88,248 patients identified with diabetes. The clearinghouse could not calculate a percentage of patient population, because the plans did not submit data on all patients.**
	How many patients had data from multiple physicians? **About 40% had claims from multiple providers. As noted below, the clearinghouse could not always ascertain if providers were physicians or facilities.**
	How many patients had data from multiple plans? **Most patients belonged to two health plans — 16% of patients with asthma and 14% of patients with diabetes were enrolled in two or more health plans.**
	How many physicians had data from multiple plans? **This was difficult to define because providers were hard to identify. While in some cases it was clear which providers were physicians and which were facilities, in some cases it was impossible to determine without going through record by record. However, concerning the providers that received patient lists, all of those providers had data from multiple plans.**
	What were the characteristics of the patients, providers, and plans included in the pilot data? **The plans, providers, and patients came from across the state. While the majority were in the Portland area, that was not surprising given the population of the state. As well, two plans serving OHP members were included so that this population was represented in the data as well.**
	What were the challenges to merging patient, provider, and claims data? **The challenges are described in great detail in the report. Among the problems: lack of standardized data, data fields did not match technical specifications, missing information.**

3. What is the quality of the individual-level data?	
a. Matching to physician	What was the degree of agreement between the Clearinghouse/plan designation of primary care provider and whether the provider regarded him/herself as that patient's primary care provider? What were the challenges in designating the primary care physician? **The rate of identification was above 90% for the two sets of patient lists provided to physicians. The main challenge was trying to identify physicians only for imputation purposes. However, in many cases, no physician providers were identified for a patient, only facilities. The plan-assigned PCP, when available, seemed unreliable due to coding errors.**
b. Diagnosis	What was the degree of agreement between the Clearinghouse/plan identification of patients with diabetes and asthma and the physicians' diagnosis? What were the challenges in determining a correct diagnosis? **In the first set of patient lists, there was a programming error that caused patients with diabetes to be incorrectly identified. This problem was corrected for the second set of lists. The percentage correctly identified for the second lists was more than 90% for diabetes and more than 80% for asthma. The most common mis-flagging was for patients with COPD to be identified as having asthma.**
c. Completeness	How many health plans included inpatient, outpatient, and emergency department visit data? How many patients did this represent? **All plans provided data for inpatient, outpatient, and ED visits. Overall, there were 581,834 patients and 259,572 providers in the data received from the plans.**
	How many plans included laboratory test and test results data? How many patients did this represent? **All plans provided data for laboratory tests (such as LDL, HbA1c, etc.). Plans did not explicitly send "lab" data. Lab data were included in claims for inpatient/outpatient/ED visits. Most plans identified claims as lab claims in their data. No plans provided lab test results.**
	How many plans included pharmacy data? How many patients did this represent? **All plans provided data for pharmacy claims.**
d. Currency	How current were the inpatient, outpatient, emergency department, laboratory and pharmacy data the plans were able to provide? What were the challenges to obtaining current data? **The data submitted was to be old enough for the majority of claims to have been filed (typically 90 days old). The data due for Data Pull 2 on August 31, 2004 were up through March 31, 2004. This gave claims enough time to be submitted and then data were taken and manipulated for submission. The challenge was to get the data in a timely manner. Since there is lag time between claim happening and being submitted/processed, "real time" time data were impossible to obtain.**

4. What are the practice summary/ benchmarking statistics?	What percent of the patients met HEDIS and other eligibility criteria for inclusion in practice summary and benchmarking statistics? **Using statistics from Data Pull 2, roughly 25% of patients that were included in Clearinghouse data submissions were identified as having diabetes or asthma. These patients were included in summary reports.**
	What summary statistics and benchmarking data were provided to physicians about their practice? **A comparison between provider, all identified providers at their clinic, and all Clearinghouse patients was provided for asthma and diabetes indicators.**
	How much variation was there by physician group (what was the range, blinded by group)? **The clearinghouse reviewed only three clinic groups. There was some variation, though, given variation in plan submission of test information. Further analysis would need to be done to provide meaningful comparisons.**
	How much variation was there by plan (what was the range, blinded by plan)? **While statistics were not calculated for plans, there did appear to be variation by plan as far as appearance of diabetic tests was concerned. Patients from some plans had more test information than other plans. Further investigation of plan-specific data may answer questions about this.**
	Is the number of cases large enough at the individual provider and group-level to be meaningful? **In some cases, the number of patients was too small for a provider to provide meaningful feedback. In addition to the fact that some patients were identified as being patients of another provider, the numbers available were not always appropriate for making comparisons. Perhaps at this stage, a comparison of clinic versus Clearinghouse may have been more appropriate.**
	What unanticipated barriers were identified? **Overall, diabetes test rates were very low. It is unclear what the source of this could be. Since the clearinghouse only processed claims data, it may be that there were missing data from medical charts. It could be that some lab data were not submitted as a medical claim as inpatient or outpatient data are. Perhaps internal coding practices led some tests being unidentified.**

Appendix B. Preliminary Data Verification and Inspection

Plan	Summary of Data Inspection
Plan A	Data tests: all patient IDs appearing in claims are in the patient demographics table, 32 provider IDs appeared in claims but weren't in the provider demographics table. Fields for concern: Provider ID in provider demographics table. Data cleaning needs: Medical claims—ICD-9 codes need decimal removed. Provider demographics—need to remove duplicate and non-physician providers. There are many facilities and "unknown" in the provider name field. A provider may have more than one ID. Often, a nine-character ID number will have a different two- or three-character extension. Also, the provider name comes in one field instead of separated into First, Last, Middle. Other issues: N/A
Plan B	Data tests: all patient IDs appearing in claims are in the patient demographics table, only about 25% of claims had a provider ID that appeared in the provider demographics table (this may be due to the fact that non-physician providers appear in claims but not provider demographics; this will be explored more). Fields for concern: Provider ID in medical claims (see above). Data cleaning needs: Medical claims—ICD-9 codes need decimal removed, each Dx code appears on a different line, with a flag, so they need to be put into one record. Provider demographics—need to remove duplicate providers. Patient demographics—need to remove duplicate patients. Pharmacy claims—claims came from 3 different files and had to be merged. All files—date fields need to be changed from text to date/time field. Other issues: N/A

Plan C	Data tests: 1 provider ID appeared in claims but wasn't in the provider demographics table. 20 patient IDs in claims files don't match patient demographic file. Fields for concern: N/A Data cleaning needs: Medical claims—ICD-9 codes need decimal removed. Other issues: N/A
Plan D	Data tests: 2413 patient IDs appeared in claims but weren't in the patient demographics table, 153 provider IDs appeared in claims but weren't in the provider demographics table. Fields for concern: Provider name (see below) Data cleaning needs: Provider demographics—need to remove non-physician providers, need to separate first name, last name, middle initial in provider name field. Pharmacy claims—leading zeros missing from NDC code field. Other issues: Finding unique physicians is a problem. Non-physicians often need to be removed manually; provider type does not always help to eliminate non-physicians. Provider name is all in one field (last name, first name). Non-physicians are also in this field. Table structure is different from last data pull.
Plan E	Data tests: all patient IDs appearing in claims are in the patient demographics table, all provider IDs that appeared in claims appeared in the provider demographics table. Fields for concern: N/A Data cleaning needs: Medical claims—inpatient and outpatient claims were separate and had to be merged, ICD-9 codes need decimal removed, revenue codes have an extension of "0" in front of them which needs to be removed. Provider demographics—some physicians have more than one ID number, Ids relate to clinic sites, so duplicate provider records need to be removed. All files—date fields need to be changed from text to date/time field. Other issues: N/A

Plan F	Data tests: all patient IDs appearing in claims are in the patient demographics table, 426 provider IDs appeared in claims but weren't in the provider demographics table. Fields for concern: N/A Data cleaning needs: Pharmacy claims—date fields have to be changed from text to date/time. Provider demographics—need to remove duplicate and non-physician providers. There are many facilities and blank names in the provider name field. The blank names still have a ProviderID with them. Also, a provider may have more than one ID. Often, a seven-character ID number will have a different two-character extension. In the provider last name field there are many "half names". For example, FName="Good Facili, LName="a Clinic". It will take time to remove non-physician providers of this type. Patient demographics— date fields have to be changed from text to date/time. Other issues: N/A
Plan G	Data tests: all patient IDs appearing in claims are in the patient demographics table, about 75% of claims had a provider ID that appeared in the provider demographics table. I will look into this more. There are UPINs as well as provider IDs in the Provider field in the claims, as well as provider names. Fields for concern: Provider ID in medical claims file (see above). Data cleaning needs: Medical claims—ICD-9 codes need decimal removed. Provider demographics—Non-physicians often need to be removed manually, provider type does not always help to eliminate non-physicians. Other issues: Large number of medical claims per member.
Plan H	Data tests: all claims had a patient ID that was in patient demographics table, <1% of claims had a provider ID that wasn't in the provider demographics table. Fields for concern: Provider name fields (see below). Data cleaning needs: Provider demographics—provider name is a problem (see below), many out-of-state providers are included, non-physicians often need to be removed manually, provider type does not always help to eliminate non-physicians. Other issues: There is an issue with provider names: there is no consistency in data entry. A name could appear as FName=John, LName=Doe or FName=Doe, LName=John or FName=Doe, John, etc. As we test with a small number of providers, this shouldn't be an issue, but as we do larger data reporting, this is a big concern.

Plan I	Data tests: 28 patient IDs appeared in claims but weren't in the patient demographics table, all claims had a provider ID that appeared in the provider demographics table. Fields for concern: N/A Data cleaning needs: Medical claims—ICD-9 codes need decimal removed. Provider demographics—need to remove duplicate providers (see below). Patient demographics—need to remove duplicate patients. All files—date fields need to be changed from text to date/time field. Other issues: Finding unique physicians is a problem. Non-physicians often need to be removed manually, provider type does not always help to eliminate non-physicians.
Plan J	Data tests: 99 patient IDs appeared in claims but weren't in the patient demographics table, no provider IDs from the provider demographics table appeared in claims files. Fields for concern: Provider ID in claims files (see above). Data cleaning needs: Medical claims—ICD-9 codes need decimal removed. All files—date fields need to be changed from text to date/time field. Other issues: There are two medical claims files. One includes fields for procedure codes, revenue codes, and DRGs; the other only has one procedure code field. It appears only asthma and diabetes claims are present. The data files could not be matched and this health plan's files were not merged into the clearinghouse files.
Plan K	Data tests: 96 patient IDs appeared in claims but weren't in the patient demographics table, 16 provider IDs appeared in claims but weren't in the provider demographics table. Fields for concern: N/A Data cleaning needs: Medical claims—Revenue codes have an extension of "R" in front of them which needs to be removed. Provider demographics—need to remove non-physician providers. Other issues: Finding unique physicians is a problem. Non-physicians often need to be removed manually, provider type does not always help to eliminate non-physicians.

Plan L	Data tests: 77 patient IDs appeared in claims but weren't in the patient demographics table, 7 provider IDs appeared in claims but weren't in the provider demographics table. Fields for concern: N/A Data cleaning needs: Medical claims—procedure code field contains DRGs, revenue codes, and CPT codes. Using prefixes to identify each, these data elements will need to be separated into unique fields. ICD-9 codes need decimal removed. Provider demographics—some physicians have two ID numbers, one has an extension of "01" which needs removing. Other issues: It appears only asthma and diabetes claims are present.