

Conference Call 6/6/01

Unduplication

Karen's questionnaire that progresses from most to least restrictive matching parameters were sent out with responses requested by Friday 6/7. Although several have been returned, Karen is now waiting on four people. She is inputting the responses as she receives them and is assembling respondents' qualitative comments.

User Pops/Workload

A meeting was convened yesterday to determine if user pop numbers for FY99 are within acceptable limits of accuracy and how the results of that examination will impact any re-export process. We looked at user pop data for all Areas from FY96-FY99 with percentage changes in counts from FY91-FY99. Data for FY98 forward reflected three 'fixes' made by NPIRS personnel and EXACT unduplication by Last Name, First Name, SSN, and DOB. The fixes included some reloading of mainframe data, correction of a known registration problem, and changing unknown community codes to values pulled from the last visit. Based on this analysis, the following decisions were made:

- Karen will examine a couple of side issues raised, then we will publish her spreadsheet analysis to the DQA Team website.
- We will publish FY99 workloads and user pops (these using the EXACT undup and 3 fixes) now.
- We will publish workload and user pops frequently to the NPIRS/DQAT page where we can iterate them as we work out any problems with them. Then, aiming toward a predictable frequency, we will publish them to the NPIRS website when we feel confident with the numbers.
- When the remaining responses to Karen's unduplication questionnaire arrive, she will create the algorithm "gold standard" and then use it to recompute FY98 and FY99 and republish them to the NPIRS/DQAT website page. The team will revisit them at that time to decide about publishing them to the NPIRS website.
- Those Areas that came out with irregular numbers in Karen's analysis will be worked. We are not going to go back and fix FY97 data. We'll do a focused response to problems made clear in Karen's spreadsheet and pro-actively work the problems.
- Patricia will complete her review of FY2000 and FY2001 workloads and make her recommendation about publishing them.

- Patricia will submit request to publish user pops (650A and 650B) on a monthly basis rather than a FY basis. She will ask for input from Stat Officers and the Data Center about what "lag" time is needed in this report to account for data processing time.

Data Warehouse/Data Mart

Strategy: A meeting was held on 5/30 to determine strategy for pilot warehouse and the resulting coordination of IBM and SAS efforts. The pilot will be limited as follows:

- 3 Areas only (considering Phoenix, Nashville, and Albuquerque Areas)
- May pick out some CHS, but now only pcc (apcrun) and Dental (denrun) for a sufficient sampling of data to establish and test implementation of the Extract, Transformation, and Load processes (proviso: include one tribe – Norton Sound – to make sure we handle it right)
- FY97 to present data to satisfy user pop data for FY99 forward.

For the data we use in the pilot and in the full warehouse subject area, we're going to re-export all applicable data using improved/enhanced version of PCC extract/export process that will be prepared over the next couple of months (e.g., Patch 6 beta version). This should minimize the number of problems with data quality that result from the extract logic and export process. These export modifications are separate and apart from the regular export program and being prepared only for the pilot export. We will then marry the two export programs after the re-export, when we're ready to go to the full warehouse. We will separate the re-exported data and pilot warehouse from the existing NPIRS. We will do a second smaller export to include modified visits to see that we can handle them. Stephanie will contact the ISCs of the pilot Areas to set this up in a way that impacts them the least.

We'll have to plan continuing exports from the three pilot Areas. In fact, we may have to make changes based on learnings in the pilot that precipitate incremental re-exports from these areas or we may find we need to do a whole re-export of these Areas to support the full warehouse. We will make some of these decisions as we go and when we can see more of the circumstances.

Registration data: Recognizing that the Stat officer re-export plan may take 6 months to create and implement, we will create a pilot re-export of the selected Areas within a shorter time period to enable a standup warehouse in 4-5 months. Hopefully by the end of Sept or early Oct, we will examine whether we need to make another round of changes to Pt. Reg to support a re-export for the full warehouse. Note: Direct exporters will have to write their code for the re-export. A workgroup chaired by Patricia and Linda Querec is gathering information on direct, tribal, non-rpms re-exports.

Hard Drive Space: Additional storage space is being obtained, but we should have enough now for the pilot. We can also drop the test database (testdb) if necessary. The more normalized data structure that will be designed for the warehouse will use space more efficiently.

Technology Comparison: IBM and SAS will perform parallel processing on the pilot exports to enable a comparison of their technologies and facilitate a decision of which to use in what phase of processing and reporting. Both IBM and SAS consultants are preparing their proposals. Joan will work with the consultants to produce a coordinated project plan for this parallel processing effort.

NPIRS Access

Access and Use Criteria: Comments received from almost everyone are now being incorporated into the document with expected distribution to the DQA Team and Stat Officers this Friday, 6/8. An issue was raised in the feedback about discomfort with people having access to Area Identifiers (e.g., Navajo, when there hasn't been tribal approval). We'll need to determine processes that need to be implemented (e.g., facilitating tribal approval) and identify who should create and implement them.

Implementation via View Structures: Stan has distributed excerpts from the draft Access and Use Criteria the kinds of ways we might restrict access (by data, by fields, etc) so that we can begin to anticipate the kind of views that might be required to support the criteria. Once the criteria is finalized (at this stage), Stephanie and selected NPIRS folks will translate these 'business rules' into technical logic/rules for implementation. This implementation will be on the NPIRS production database and will be mirrored in the pilot warehouse.

CHS Information: We need to examine Stephanie's comparison of the CHS export data we get from the three sources and determine what decisions need to be made regarding what sources to use, store, and process. We also need to identify additional fields we want from the FI. Stephanie, Stan, and Joan will work this and bring recommendations back to group.

Documentation

NPIRS Documentation: User Population Process and Reporting document will be placed on the website this week. Joan is waiting for team review of the Workload Processing document and hopes to have it by end of this week. Startup on the second workload document dealing with the direct exports and other less frequently used update programs has been delayed until next week.

PCC Export Logic: PCC Statistical Export document from Stan describes the report logic, edits done, error checks edits, visit check edits (for exclusion from extract), etc. Stan would like to see all visits exported and those that did not meet the edit criteria flagged. We need to examine the export of visits that are excluded because they don't meet workload or user pop criteria.

Coordination

IHLC Presentation: When Stan makes his presentation at this meeting, he would like to request that members help in a couple specific ways. Team members are asked to think

about what we most need from members of this group. For example, one idea is to ask the Area Directors to ask their respective ISCs to send NPIRS a report of what's in an export when they send it to NPIRS, even if it's just a list of what's sent. Another idea is that they ask their ISCs to examine the report Melvin returns after receiving an export and give NPIRS feedback. We want to encourage two-way communication. For next meeting, team members are bring their ideas for IHLC requests.

There is an MPI product fair scheduled for July 31 - Aug 1 in Albuquerque. The MPI group is making selections and getting the purchase order ready to obtain a product by end of the year. We need to coordinate with this workgroup so we can be ready to marry our warehouse pilot with the MPI pilot. Karen is member of the MPI workgroup and will serve as point person on this effort. KC will ensure that there is a stat officer on this group.

Stat Officers are meeting in Portland on Aug. 6. Patricia and Debra will attend and represent the DQA Team.