

MANAGING LONGITUDINAL RESEARCH STUDIES:

OSHPD FORMAT LIBRARY

By

Linda L Remy, MSW PhD

Ted Clay, MS

UCSF Family Health Outcomes Project
Geraldine Oliva, MD MPH, Director
Jennifer Rienks, PhD, Associate Director
Linda L Remy, MSW PhD, Research Director

500 Parnassus Ave. Room MU-337
San Francisco, California 94143-0900
Phone: 415-476-5283
Fax: 415-476-6051
Web: <https://fhop.ucsf.edu/>

August 2020

TABLE OF CONTENTS

Overview	1
Structure of Formats Program	2
Clinical, Facility, and Geography History	3
OSHPD Facility Labels	4
Pay Plans	7
Centers for Medicare and Medicaid Services (CMS)	8
Diagnosis Related Group (DRG) and Major Diagnostic Category (MDC)	8
ICD-9-CM and ICD-10 CM Diagnoses and Procedures	10
Clinical Classification System	11
CCS ICD-9-CM Classification.....	12
CCS Outpatient Procedure Classification	15
CCS ICD-10 Classification	16
General Equivalence Mappings	19
Resources	20
Endnotes	21

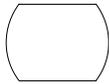
Table of Figures

Figure 1.	Summarize annual clinical, facility and geography history	3
Figure 2.	Check for new hospitals, closures, moves, current name	5
Figure 3.	Prepare Pay Plan formats	7
Figure 4.	MDC and DRG formats 1983-2017.....	9
Figure 5.	Summarize annual ICD sets	10
Figure 6.	Read ICD-9-CM CCS into SAS.	12
Figure 7.	Integrate ICD-9-CM patient DX and PX counts with CMS and CCS formats.....	14
Figure 8.	Integrate ICD-10-CM DX and PX codes with CCS codes	17
Figure 9.	Make ICD-10 DX and PX formats, and CCS crosswalk formats	18
Figure 10.	Read GEM into SAS.....	19
Figure 11.	Prepare GEM formats.....	20

Table of Tables

Table 1. Relationship between FHOP and CCS Classifications **Error! Bookmark not defined.**

Table of Legends

	From program creating incoming file or name of next program using file
	Excel file input or output
	SAS file input or output
	Flat or text file input or output
	SAS program with brief description of steps.

Suggested Citation

Remy L, Clay T. (2020) Managing Longitudinal Research Studies: OSHPD Format Library. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: <http://fhop.ucsf.edu/data-management-methods>.

ACRONYMS

AHRQ	Agency for Healthcare Research and Quality
ASC	Ambulatory Surgery Center
CCS	Clinical Classification Software
CPT	Current Procedural Terminology System
DX	ICD-9-CM Diagnosis
DXCH	CCS high-level DX group
DXCL	CCS single-level DX group (low-level)
ED	Emergency Department
GEM	General Equivalency Mappings
HCPC	Healthcare Common Procedure Coding System
HCUP	Healthcare Cost and Utilization Project
CMS	Centers for Medicare and Medicaid Services
FHOP	Family Health Outcomes Project
ICD-9	International Classification of Diseases, 9th Revision, Clinical Modification
ICD-10	International Classification of Diseases, 10th Revision, Clinical Modification
OSHPD	Office of Statewide Health Planning and Development
PD or PDD	Patient discharge (data)
PX	Procedure
PXCH	CCS high-level PX group
PXCL	CCS single-level PX group (low-level)
PXCLS	CCS Major or minor, therapeutic or surgical procedure classification
UCSF	University of California, San Francisco

OSHPD FORMAT LIBRARY

This document describes methods the UCSF Family Health Outcomes Project (FHOP) uses to maintain the format library for patient-level datasets distributed by California's Office of Statewide Health Planning and Development (OSHPD). It covers formats for patient discharges (PD), emergency department (ED) visits, and Ambulatory Surgery Center (ASC) visits. Format libraries for OSHPD's Annual Hospital Disclosure Report, for Vital Statistics files (births, deaths, fetal deaths), and geography-related libraries are described in separate documents.

We are making this basic methodology and its associated software public to help population health researchers understand the nature of data management for complex longitudinal research. This also should provide a background to users of our longitudinal DataBook and EpiHosp products.

We hope this will help people better understand the foundation behind longitudinal research. Data analysts working in local health jurisdictions and researchers interested in longitudinal research may find this helpful.

All work is in SAS, assisted by Microsoft Excel and Visio.

We assume that the user of this document has read other documents describing the foundation of our methodology:

Volume One: The Basic Operating Environment

Volume Two: Standardizing Variables Over Time

Volume Three: Preparing Master Files

These and related documents are available on our website:

<https://fhop.ucsf.edu/data-management-methods>

OVERVIEW

Formats give SAS information about how to read or write data [1]. SAS has many standard formats for reading and writing character, numeric, and date or time values. We take advantage of standard formats when we read master files into SAS.

Formats provide meaning to the strings of numbers and characters in a database. Consider a variable SEX with the values 1 and 2. The values have no meaning in and of themselves. The format tells the user that 1 means Male and 2 means Female.

Our focus here is user-defined formats. These convert a value to a different form for output. User-defined formats can have many sources. Dataset distributors such as OSHPD traditionally provide documents that describe the meaning of values for each variable. Data users such as FHOP create other formats consistent with their analytic needs. Formats from external sources expand our analytic capacity.

In this document, we describe how we make the format library for OSHPD datasets. These are complex datasets, and maintaining the format library is time consuming. FHOP's format library for OSHPD files is available at <https://fhop.ucsf.edu/data-management-methods>.

Contracts with various departments in the State of California require us to provide the funding agencies with an annual backup of all programs, logs, listings and output files. This creates an audit trail of our work. Since we do not know where or who will use the programs and/or resulting files, we try to write code that will run in any environment and provide as much internal documentation as possible. Because public funding supports development of these programs, they are in the public domain. This is why we are making them available.

STRUCTURE OF FORMATS PROGRAM

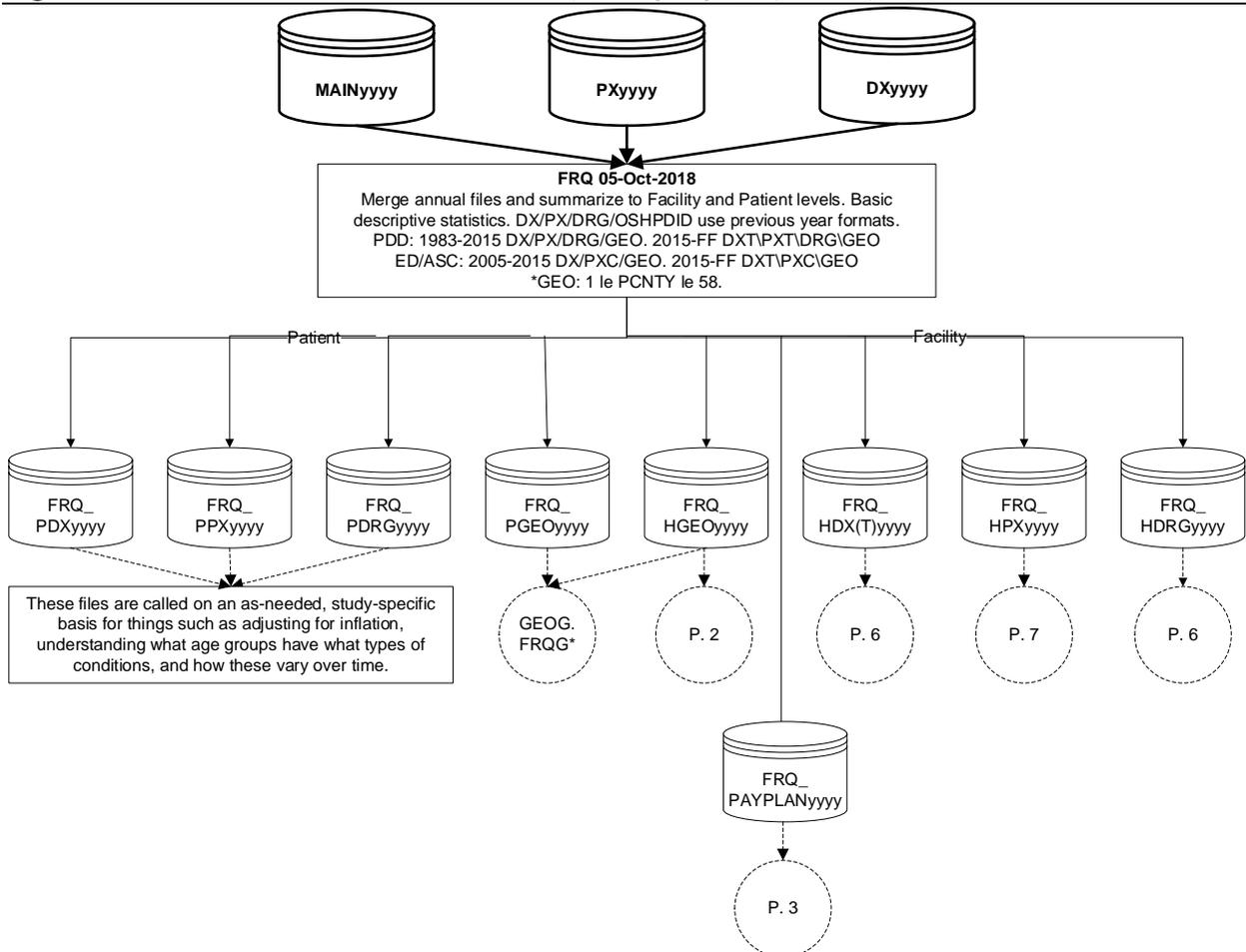
<pre>%LET PGMNAME=FORMATS; %STARTUP cards4; PURPOSE: Maintain the format library DESCRIPTION: The program can be re-run as many times as needed to be sure it contains all updated formats.</pre>	<p>FHOP uses the standard file name FORMATS.SAS to organize format definitions for all datasets and special studies.</p>
<pre>-- SEX SEX.FMT SEX SEXOM1F FEMALE *\$SEX %inc 'SEX.FMT';</pre>	<p>Related formats are grouped into text files with the .FMT extension. Formats in a given FMT file are identified. Formats preceded with an asterisk (*) are no longer active but code is still in the FMT file if we decide to reactivate it.</p>
<pre>-- Race/Ethnicity RACE.FMT RACE (1984-1994) RACEN HISPANIC (1995-FF) RACEDOF RACEF *\$RACE. \$RACEOP \$HISPANOP (ED/ASC) %inc 'RACE.FMT';</pre>	<p>Some formats (SEX, RACEN, HISPANIC, \$RACEOP, \$HISPANOP) are from OSHPD. Others reflect groupings made by FHOP.</p>
<pre>WARNING: <<< AVOID THE TEMPTATION OF CREATING FORMATS ON THE FLY >>></pre>	<p>We discourage making formats “on the fly”. Time has taught us that these can be dangerous to our mental health.</p>
<pre>INPUT: Various depending on formats OUTPUT: D:\PDD\SAS\FORMATS SOURCE: D:\PDD\FORMATS.OLD ;;; %STOREDOC</pre>	<p>The documentation section ends with our standard input, output, and source statements, followed by four semi-colons and calling the macro STOREDOC.SAS</p>

<pre>*EDITED: Linda Remy in Belvedere on MEHITABEL 27Jan2006; *%inc 'SEX.FMT';</pre>	<p>The analyst, location, computer, and date the format was last edited is added above the line calling the FMT file.</p>
<pre>* EDITED: Linda Remy in Belvedere on MEHITABEL 07Dec2007 Adding ED/ASC \$RACEOP \$HISPANOP Adding 0 to RACEDOF in Belvedere on MEHITABEL 01Jun2008; *%inc 'RACE.FMT';</pre>	<p>Specific changes to a format are identified. The asterisk (*) before the %inc statement comments out the line calling the FMT file. It signifies that we will not execute this portion of the format program at this time.</p>

CLINICAL, FACILITY, AND GEOGRAPHY HISTORY

We begin the process of making longitudinally consistent formats for the OSHPD files by running the program FRQ.SAS from the main directory for each source. This program calls a macro FRQOSH.SAS, stored in TOOLS\FHOP. The macro merges the annual master files then summarizes and outputs small annual files for each set of patient and facility variables. Figure 1 shows the annual incoming and outgoing datasets.

Figure 1. Summarize annual clinical, facility and geography history



In the box identifying the program, the first line shows the program name and date we ran the program. The output file names follow our standard. FRQ_ identifies the name of the program making the file, followed by P(atient) or H(ospital) then a 2 or 3 letters identifying the type of summary ((DX (diagnosis), PX (procedure), DRG (Diagnosis related group), GEO (geography)), then the file year. The dashed circle below each file figure shows the next section that will call these files. For example, we use FRQ_PGEOyyyy files to prepare geography formats, described in the geography chapter of our longitudinal methods documentation. Notation (P. n) identifies the next page in the Visio file (currently FORMATS_2019.VSDX) documenting steps to make formats. Figure 1 is from P. 1 in that file.

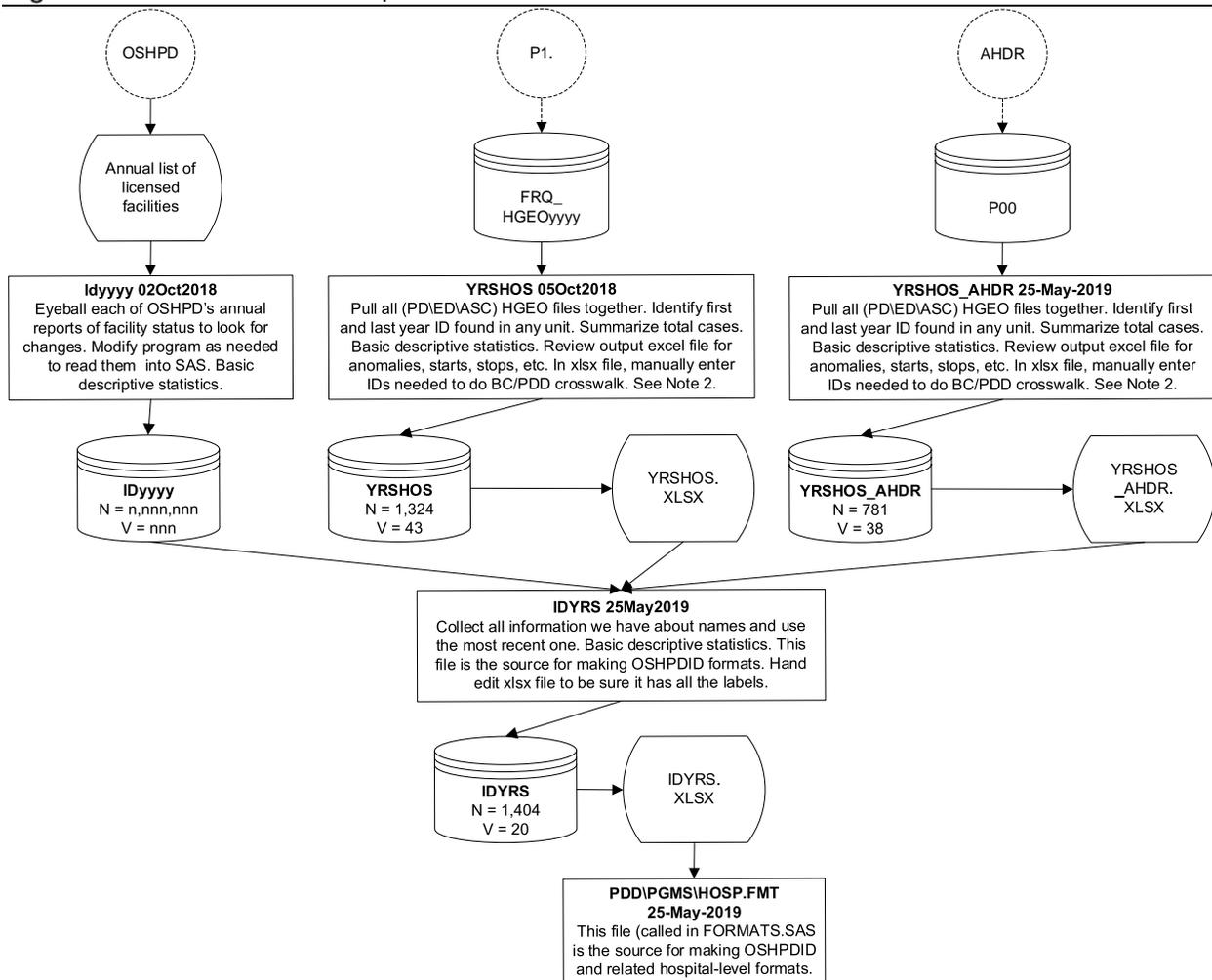
OSHPD FACILITY LABELS

Most people are unaware that hospitals change their names over time. When we report hospital names, we report the last known name for the last time we had any information for that facility. This is consistent with decision rules made in California's first reports on risk-adjusted patient outcomes [2,3]. This section describes steps to maintain a current list of OSHPD facility labels.

Our starting facility format OSHPDID contains the last label for every facility discharging patients from 1983 to the last time we updated them. OSHPD distributes annual lists of facilities (PD, ED, ASC) discharging patients with the current name and OSHPD's identifier. Facilities that closed or consolidated reporting are not on the list. OSHPD also provided FHOP with several variations on a theme of historical lists of hospitals, many of which closed since 1983. These identifier-related files have different structures and contents. We read each OSHPD file into SAS (IDyyyy.SAS), where yyyy is the year we received the file.

The program IDYRS.SAS merges the IDyyyy files by year, retains the chronologically last label, and merges with YRSHOS (from Volume 3: Preparing Master Files [4]) and with YRSHOS_AHDR (from Volume 8: Annual Hospital Disclosure Reports). We convert the old labels from YRSHOS into upper case. If an OSHPDID is not in the IDyyyy file, we use the older label. OSHPD licensed a number of facilities since we last updated this. Figure 2 summarizes the process to update these formats.

Figure 2. Check for new hospitals, closures, moves, current name



The resulting file IDYRS is output to Excel for manual review. Here, we make minor edits such as spelling. In HOSP.FMT, we import the edited IDYRS.XLS back into SAS, and use it as the basis for the new OSHPDID formats. The relevant section has the following code:

<pre> proc import datafile = "D:\PDD\XLS\IDYRS.XLSX" out = user.oshpid replace; sheet = "IDYRS"; getnames = yes; run; </pre>	<p>Programs need to specify the output file (out =) as user.(fmtname) to save the imported temporary file to PERMWORK [5]. The replace statement over-writes any previous version. We distinguish permanent files and variables with upper case and use lower case for temporary files and variables. If you need to resubmit the rest of the program, comment out this step because user.oshpid is in PERMWORK.</p>
--	--

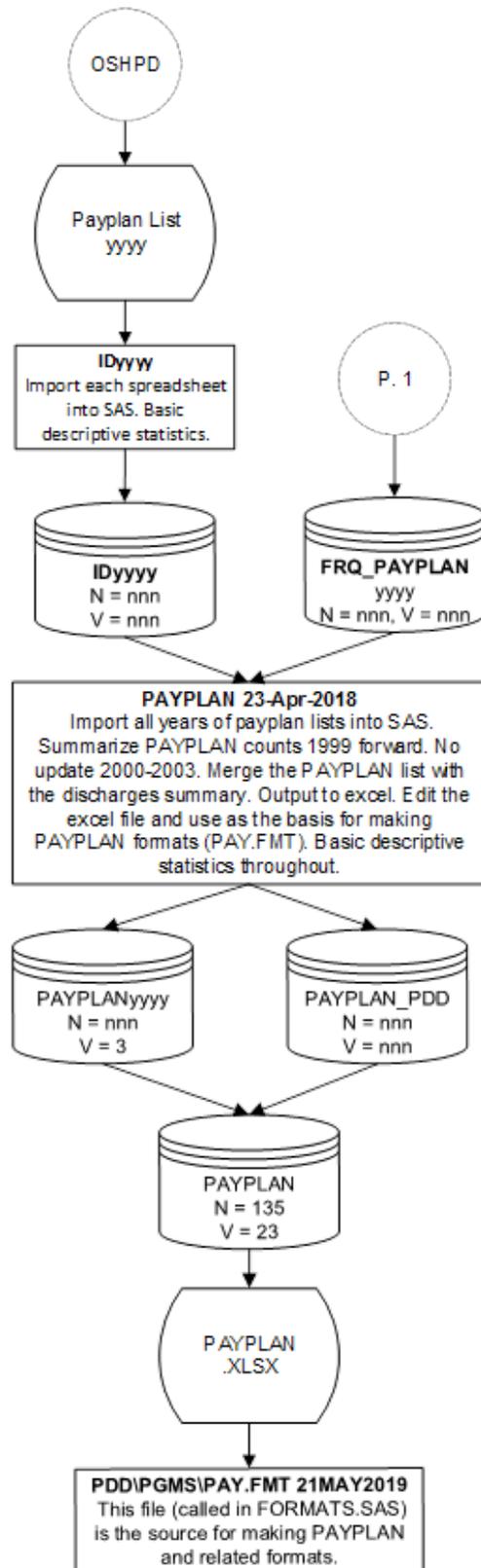
<pre>%makefmt(data = oshpid, values = OSHPDID, labels = OSHPDIDL, fmtname = OSHPDID, library = studylib);</pre>	<p>The MAKEFMT [6] macro does the necessary work to write information to the format library. A given value can have many records, but MAKEFMT stops if it finds more than one label per value.</p>
---	--

Another way to think of the values parameter is that it is the variable on the left contains values to format, while the labels parameter is the variable on the right with values to be assigned. Again notice our naming conventions: The variable to use to make the format label is the name of the variable with the suffix “L”, and the format name is the same as the variable it is formatting.

One more technical note. A number of files we download have minor structural errors that cause problems when we read them into SAS. We manually repair these and resubmit the program. At the end, we do a lot of manual work to insert labels when we do not have them, or to fix typographic errors. Many processes summarized in this document involve over-writing previous values. When updating, we take great care not to over-write older information until we verify this is valid to do. Old labels not replaced are always in upper case, and new ones are always in mixed case.

PAY PLANS

Figure 3. Prepare Pay Plan formats



OSHPD introduced a variable identifying insurance plans paying for inpatient care in 1999. As part of its standard documentation, it includes a list with the provider numbers and names. In 2004, OSHPD began to update the list. Over time, carrier names changed and new carriers started paying for care.

We read each annual file into SAS, then save the most recent records. We merge these with the PAYPLAN frequencies, export to an excel file, and edit if needed.

As before, we call this file in PAY.FMT, and update the list of numbers and names.

CENTERS FOR MEDICARE AND MEDICAID SERVICES (CMS)

The remainder of this document focuses on externally defined formats for OSHPD datasets. These start with products obtained from the Centers for Medicare and Medicaid Services (CMS), the Clinical Classification System (CCS) from the Agency for Healthcare Research and Quality (AHRQ), and FHOP formats that build on those to focus on longitudinal public health surveillance and monitoring.

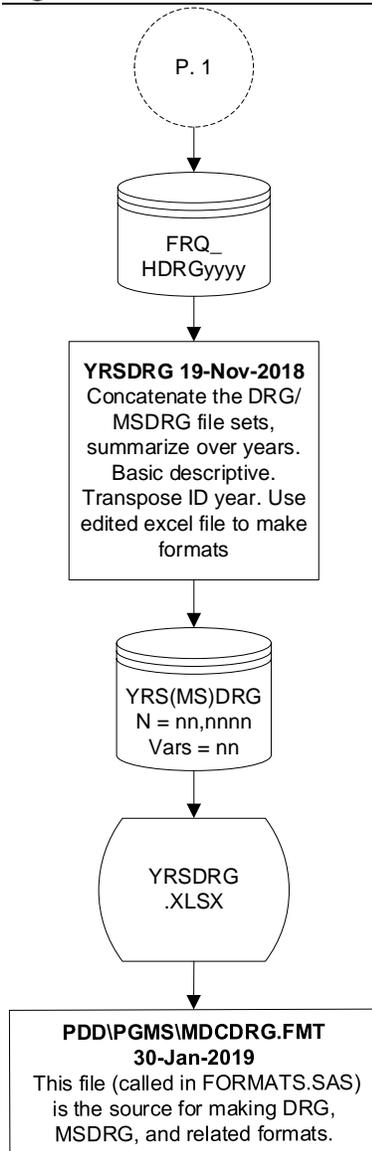
Since 1983, patient diagnoses (DX) in OSHPD files and procedure (PX) codes in PD files have been coded using the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9). The PD also includes ICD-9 derived codes for Major Diagnostic Category (MDC) and Diagnosis Related Group, used to bill CMS for inpatient care. CMS began to issue annual updates in 1986. Version 32, issued for use beginning 01-Oct-2014, was the end of ICD-9 [7]. On 01-Oct-2015, CMS and OSHPD converted to the ICD-10-CM [8].

Diagnosis Related Group (DRG) and Major Diagnostic Category (MDC)

The Diagnosis Related Group (DRG) is the basis for the inpatient Prospective Payment System for hospitals treating Medicare and Medicaid patients. The Medicare Severity Diagnosis Related Group (MS_DRG) replaced the DRG in 2008. Each DRG derives from the principal DX and PX, augmented by other information summarizing the event to evaluate severity. DRGs have a severity index, with payment based on severity plus hospital-specific factors. DRGs are further stratified by whether they are primarily medically or surgically based, or other. DRGs group to a Major Diagnostic Category (MDC) based on the primarily affected body system. Some complex DRGs are not assigned to an MDC.

AHRQ relies on DRGs for the Inpatient Quality Indicators (IQI), Patient Safety Indicators (PSI), Prevention Quality Indicators (PQI), and Pediatric Quality Indicators (PDI) software [9]. FHOP uses these for special studies, but we do not discuss those formats in this document.

Figure 4. MDC and DRG formats 1983-2017



From the longitudinal DRG summaries we concatenated the annual files and used the resulting file with DRGs crosswalked to their associated MDC, classified as to whether the DRG is medical, surgical, or other (MEDSURG). The Excel file required hand editing. MDCDRG.FMT calls it to make formats.

In early years of the PDD, OSHPD provided the DRG but not the MDC. If we have DRG but not MDC, the following code will create it:

```
if MDC = . then MDC = 1 * put(DRG,$drgmdc.);
```

Again, note the format naming convention. In the MAKEFMT macro [6] when we use data to create data, we typically assign format names based on two parts: the name of the variable we have (leftvar), followed by the name of the variable we are assigning (rightvar). This makes format names easy to recall. Format names have a limit of 8 characters, so variable names underlying them often must be shortened.

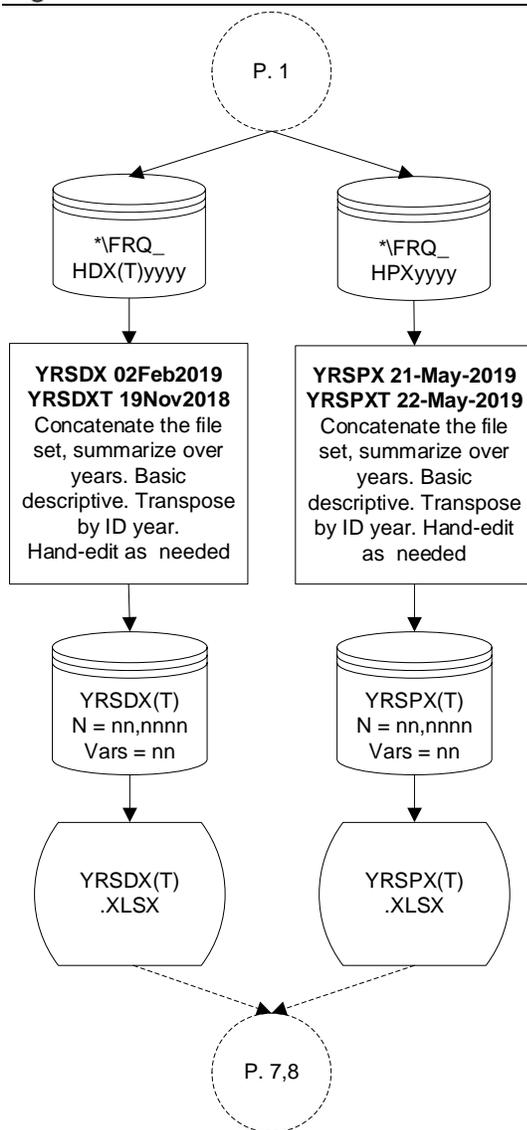
From 1983 through 2007, CMS updated the DRG annually, discontinuing some codes, adding others, and adjusting the severity index. A few years ago, CMS and others conducted a number of studies. These investigations concurred that the index sensitivity had declined, and hospital payments no longer aligned appropriately. Effective 2008, hospitals use the MS-DRG.

An important point is that we lost longitudinal continuity and will have to develop methods to address this challenge. CMS/AHRQ reportedly is developing a crosswalk between the old and new DRGs, which AHRQ needs to maintain longitudinal consistency for their various software products. When this becomes available, we will find it helpful moving forward.

ICD-9-CM and ICD-10 CM Diagnoses and Procedures

The World Health Organization (WHO) Ninth Revision, International Classification of Diseases (ICD-9) was the basis for the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). The United States (US) used the ICD-9 to code and classify mortality data from death certificates until 1998, then shifted to ICD-10 coding from 1999 forward. The US did not implement the ICD-10-CM until 01-Oct-2015, when it became the official system for assigning codes associated with hospitalization. The National Center for Health Statistics (NCHS) and Centers for Medicare and Medicaid Services are the US agencies responsible for overseeing all changes and modifications to the ICD.

Figure 5. Summarize annual ICD sets



In addition to the DRG and MDC, OSHPD files have several sets of clinical variables: principal and up to 24 DX, principal and up to 4 external E-Codes, and principal and up to 20 PX. From 1983 through 30-Sep-2015, OSHPD used ICD-9-CM to classify classified DX and PX in the PD files. As of 01-Oct-2015, the PD uses the ICD-10-CM to code DX and the ICD-10-PCS (Procedure Coding System) to code PX. As of Oct-2015, the ED and AS files use the ICD-10-CM to code DX and continue to use a shortened version of Current Procedural Terminology (CPT-4) to code PX. These changes required major revisions of many programs, formats and macros. From Figure 5, note that the process is similar for ICD-9 and ICD-10 (T).

Regardless of version, these systems for coding diagnoses and procedures consist of lists of disease code numbers in tabular form; an alphabetical index to the disease entries; and a classification system for surgical, diagnostic, and therapeutic procedures (alphabetic index and tabular list).

Clinical codes can change annually. Each code has long and short labels, current for the given year. As medical knowledge advances, some codes discontinue while others begin. The annual set does not include discontinued codes.

We download annual DX and PX files from CMS, saving them in the non-confidential drive (See Volume 1: The Basic Operating Environment [5]). We store these files under AHRQ because the CMS products are the base for the AHRQ Clinical Classification System (CCS). We read these into SAS on our work drive under the directory AHRQ. We return to these after we prepare the AHRQ files.

CLINICAL CLASSIFICATION SYSTEM

AHRQ is the health services research arm of the U.S. Department of Health and Human Services. Its research centers specialize in major areas of health care research such as quality improvement and patient safety, outcomes and effectiveness of care, clinical practice and technology assessment, and health care organization and delivery systems. AHRQ's Healthcare Cost and Utilization Project (HCUP) is a family of health care databases and related tools for research and decision making [10], developed in coordination with WHO.

For current purposes, the most relevant are the free tools developed to use with administrative databases such as OSHPD files. These tools include Clinical Classification Software (CCS) DX and PX modules for ICD-9-CM and now ICD-10-CM [11,12], and the CPT and Healthcare Common Procedure Coding (HCPC) system [13].

The CCS clusters patient DX and PX into a manageable number of clinically meaningful categories. CCS offers the ability to group conditions without having to sort through thousands of codes. This "clinical grouper" makes it easier to understand illness and treatment patterns so that local health jurisdictions, health plans, policy makers, and researchers can analyze costs, utilization, and outcomes associated with particular illnesses and treatments.

The CCS groups codes and classifies them consistently from 1980 forward. This permits longitudinal analyses to examine for trends. Without the CCS, the large number of codes makes statistical analysis and reporting difficult and time consuming. The CCS is very useful for ranking DX and PX and directly integrates into other risk adjustment, patient access, and patient quality software that AHRQ distributes. CCS enables jurisdictions to compare, directly and uniformly, patient access and outcomes to other regions, the state, or nation.

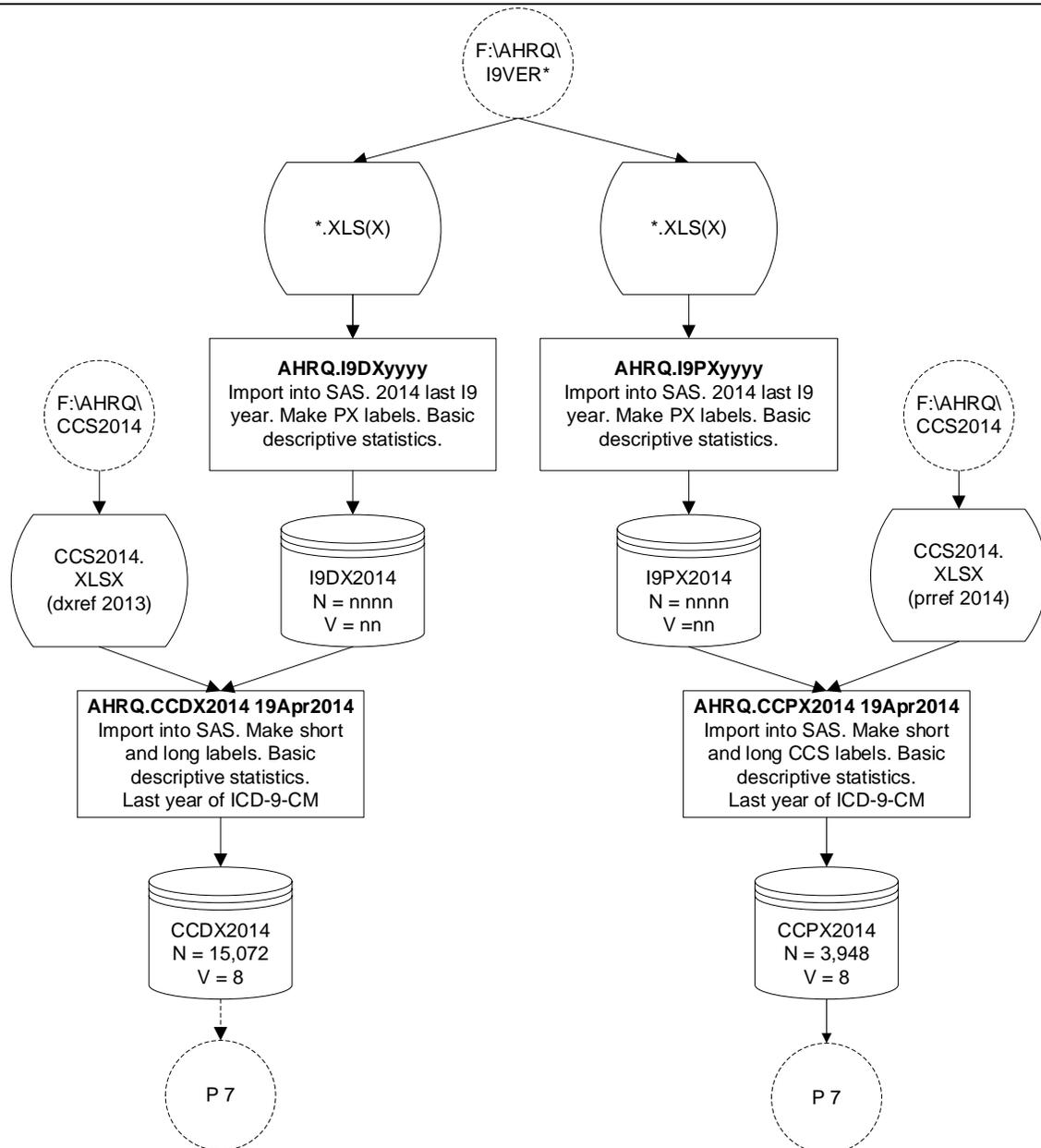
The CCS's most important contribution is reducing the enormous number of codes in the various classification systems to a small number of clinically meaningful groups, divorced from the DRG with its focus on payment, and consistent across systems and time. Its developers engaged very high-level thinking. They brought clarity to near chaos.

CCS consists of two related classification systems, single-level and multi-level, designed to meet different needs. Single-level CCS is most useful for ranking DX and PX and for direct integration into risk adjustment and other software. Multi-level CCS is ideal for evaluating larger aggregations or exploring groupings in more detail. [14]. The ICD-9 CCS is valid for the period January 1980 through September 2015. The ICD-10 CCS began Oct-2015.

CCS ICD-9-CM Classification

Figure 6 shows preliminary steps to integrate the CCS for ICD-9-CM diagnoses and procedures.

Figure 6. Read ICD-9-CM CCS into SAS.



CCS identified 15,072 ICD-9 DX codes ever used from January 1980 through September 2015. The single-level CCS aggregates illnesses and conditions into 285 mutually exclusive DX categories. We refer to these as DXCL, or low-level. To work longitudinally, CCS needs to know when ICD-9 codes started and ended. We primarily use the CCS short labels, which show this.

The multi-level CCS expands the hierarchical structure. It splits single level CCS categories to provide more detail about particular groupings of codes. We use multi-level codes only to get the broadest categories (e.g., Infectious Diseases, Mental Disorders, Injury, etc), loosely related to the MDC. In our nomenclature, we refer to these as DXCH, or high-level.

Before we import CCS files into SAS, we edit column headings to be consistent with our needs. The final file (CCDX2014) has the list of ICD-9 DX codes, DXCH and DXCL codes associated with each DX, with short DX labels and short and long labels for DXCH and DXCL.

CCS identified 3,948 ICD-9 PX codes ever used from January 1980 through September 2014. The single-level CCS aggregates to 231 mutually exclusive PX groups. We refer to these as PXCL, or low-level. To work longitudinally, CCS again needs to know when ICD-9 codes started and ended. We primarily use short labels (PXCL), which show this. As with the CCS DX classification, a higher-level PX grouper somewhat related to body system has 16 categories (PXCH).

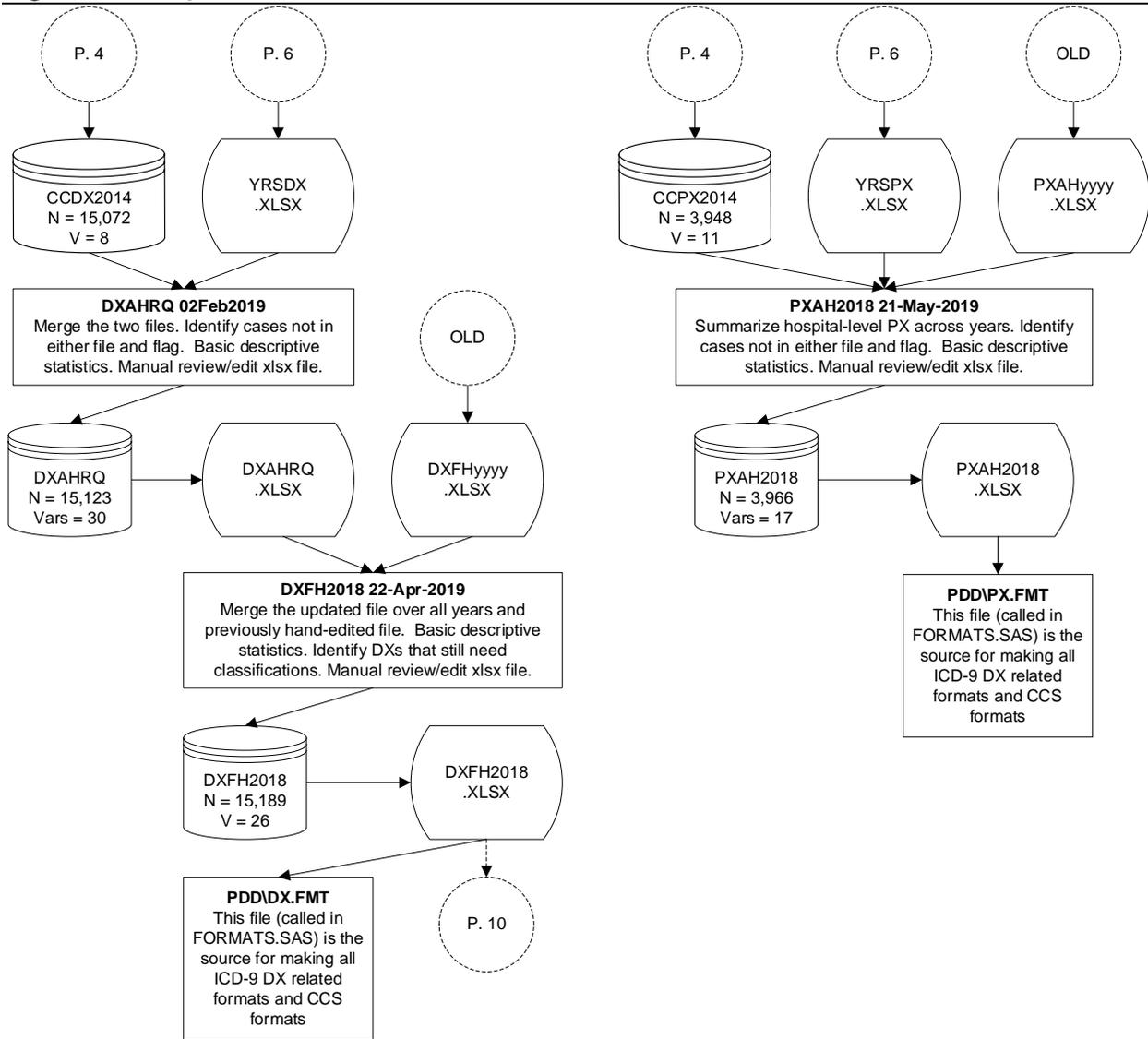
PX also can be grouped by whether they are used for diagnosis or treatment; and if they are operating room (OR) or non-operating room PX [15]. We refer to these as PXCLS. They group as follows:

1. Minor Diagnostic - Diagnostic non-OR PX (87.03 CT scan of head)
2. Minor Therapeutic - Therapeutic non-OR PX (02.41 Irrigate ventricular shunt)
3. Major Diagnostic - OR PX performed for diagnostic reasons (01.14 Open brain biopsy)
4. Major Therapeutic - OR PX performed for therapeutic reasons (39.24 Aorta-renal bypass).

As before, we edit column headings to be consistent with our needs. The final file (CCPX2014) has the list of ICD-9 PX codes, PXCH, PXCL, and PXCLS codes associated with each PX, with short PX labels and short and long labels for PXCH and PXCL.

At this point, Figure 7 shows we have summarized DX and PX patient data (YRSDX, YRSPX) from 1983 through Sep-2015, prepared the new CCS lists (CCDX2014, CCPX2014), and our previously edited lists (DXFH2014, PXAH2014). Each file has a different number of codes reflecting their status at these different periods.

Figure 7. Integrate ICD-9-CM patient DX and PX counts with CMS and CCS formats



In the end, we output excel files (DXFH2018, PXA2018), then manually edit the main Excel tab, to make corrections as needed. Note that old labels not in the newer files, and unlabeled codes that we entered manually, have upper case labels for emphasis.

Below is an example showing how to use formats to convert a character-based variable (here principal PX) with thousands of possible values to the parallel numeric-based PXCLS with 4 values. If we were doing this over the full range of PX, the statement would be expressed as two parallel sets of arrays. Notice the use of our format naming conventions. We are making a numeric variable PXCLS0 by applying the character format \$pxclsn (with 4 values) to the character PX0 variable, then formatting the numeric result with the numeric format pxcls. To avoid messages in the log, we are only doing this where a procedure code is present, because many patients have no procedures.

```

length PXCLS0 3;
*Convert ICD-9 principal PX with thousands of values to numeric PXCLS with 4 values;
if PX0 gt ` ` then PXCLS0 = 1 * put(PX0,$pxcls.);
*Many patients have no PX;
else PXCLS0 = 0;
format PXCLS0 pxcls.;

```

CCS Outpatient Procedure Classification

Outpatient PX are grouped using the CPT/HCPC system [16]. CCS assigns CPT/HCPC ranges of categories to groups, for example, the range of codes 61000 to 61055 are on a single record, assigned to PXCL 1, Incision and excision of CNS. In the CCS ICD-9 system, PXCL 1 includes individual PX 0101 to 0159, where each code is on a separate record. Using ranges, CCS converted 9,310 CPT/HCPC code ranges to 244 groups. Despite different classification systems in different settings, the CCS makes it possible to summarize PX uniformly across settings.

The CCS classification for CPT/HCPC has more categories than the ICD-9 CCS because outpatient care has more services (home care, durable medical equipment, or transportation) with no inpatient care parallel. Per email confirmation from AHRQ, two ICD-9 PXCL related to delivery do not have a CPT equivalent: “133 Episiotomy” and “136 Artificial rupture of membranes to assist delivery”.

The CPT/HCPC describes codes with a length 5, in different combinations of alphanumeric characters. Examples include a range ('00840-00848') or a single value displayed as a range ('J0970-J0970'). After we import the CPT/HCPC excel file into SAS, the following is sample code to make the variables START and END that the macro MAKEFMT needs to make formats based on ranges.

```

data cpt2;
  set cpt;
  *Convert RANGE to two character variables of length 5;
  length START END $5;
  START = substr(RANGE,2);
  END = substr(RANGE,8);
run;

```

We do not make a permanent SAS file for this. Rather, we make the format after preparing the range in a temporary data step as shown above. MAKEFMT uses different syntax to make formats based on ranges.

```

%makefmt(data = cpt2 (where = (START gt ' ')),
  values = START,
  values_hi = END,
  labels = PXCL,
  fmtname = $CPTCCS,
  library = library);

```

Even though inpatient and outpatient files use different PX codes, converting them to CCS makes it possible to study differences across settings.

CCS ICD-10 Classification

For patients discharged on or after 01-Oct-2015, OSHPD implemented ICD-10-CM DX and PX coding which incorporates much greater clinical detail and specificity than ICD-9-CM.

Terminology and disease classification were updated to be consistent with current clinical practice. The ICD-10-CM will provide better data for research and epidemiology studies, and to track public health. This makes it a very important new resource for groups like FHOP.

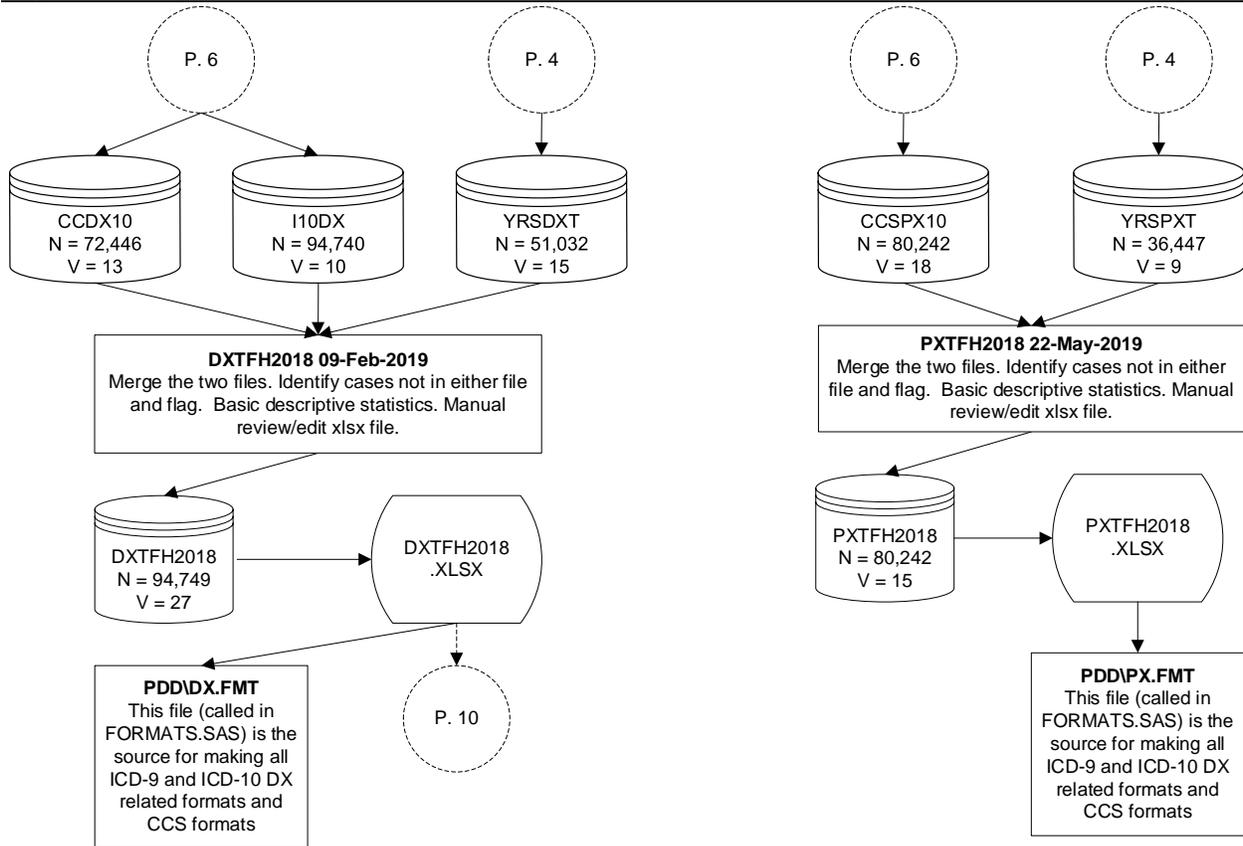
Figure 8 shows the steps taken to integrate the ICD-10-CM and the CCS codes and labels. Like the ICD-9-CM process, this work will be updated annually to reflect changes as codes and classifications are re-labeled, introduced, and discontinued. The work through 11-Nov-2018 included ICD-10-CM for the period 2013 through 2017, the most recent year of data we had received at that point.

Figure 8. Integrate ICD-10-CM DX and PX codes with CCS codes



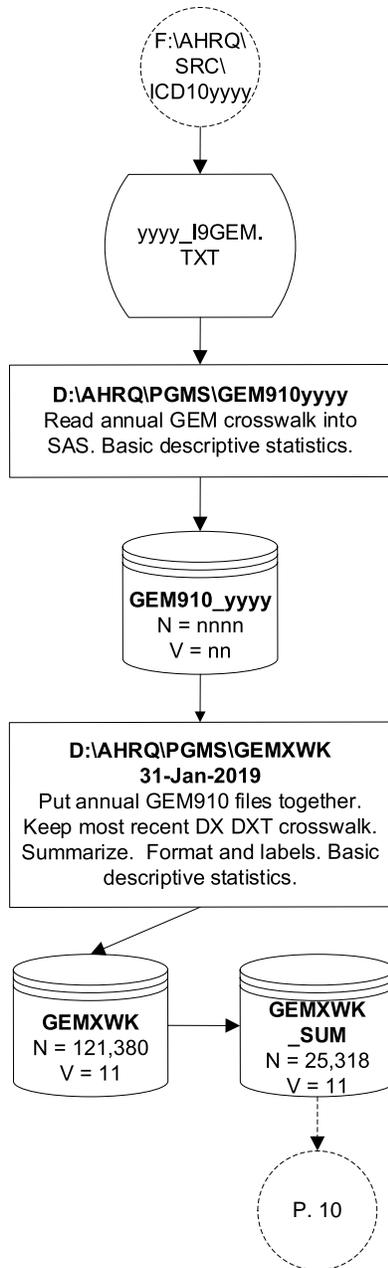
Figure 9 shows the last steps to make the ICD-10 formats and CCS crosswalk formats for them.

Figure 9. Make ICD-10 DX and PX formats, and CCS crosswalk formats



GENERAL EQUIVALENCE MAPPINGS

Figure 10. Read GEM into SAS



The CMS and the Centers for Disease Control and Prevention (CDC) developed the General Equivalence Mappings (GEM). The GEM are reference systems to facilitate both forward and backward mappings between ICD-9-CM and ICD-10-CM/PCS codes in the United States, with a particular eye to the needs of longitudinal research [17]. CMS/CDC updated the GEM annually through 30-Sep-2018 as a transition aid or bridge between the two systems [18].

Figure 10 summarizes our work preparing the GEM files. We first read the five years of files into SAS and then concatenated them (GEMXWK), summarizing the result to get unique codes (GEMXWK_SUM N = 25,318). Because of CMS changes from year to year in when they introduced codes and where they assigned codes, the file GEMXWK_SUM has 25,318 records. Of these, only 11,065 were unique combinations, that is, the DX9 walked to one and only one DX10. The remaining DX9 walked to 2 to 6 DX10 codes.

AHRQ also annually updated clinical grouping tools such as the CCS to use ICD-10-CM/PCS codes. These groupings can be useful to bridge the ICD-9-CM to ICD-10-CM/PCS transition, by relying on classes or categories of codes to create a bridge between code sets, instead of relying on individual codes.

Figure 11. Prepare GEM formats

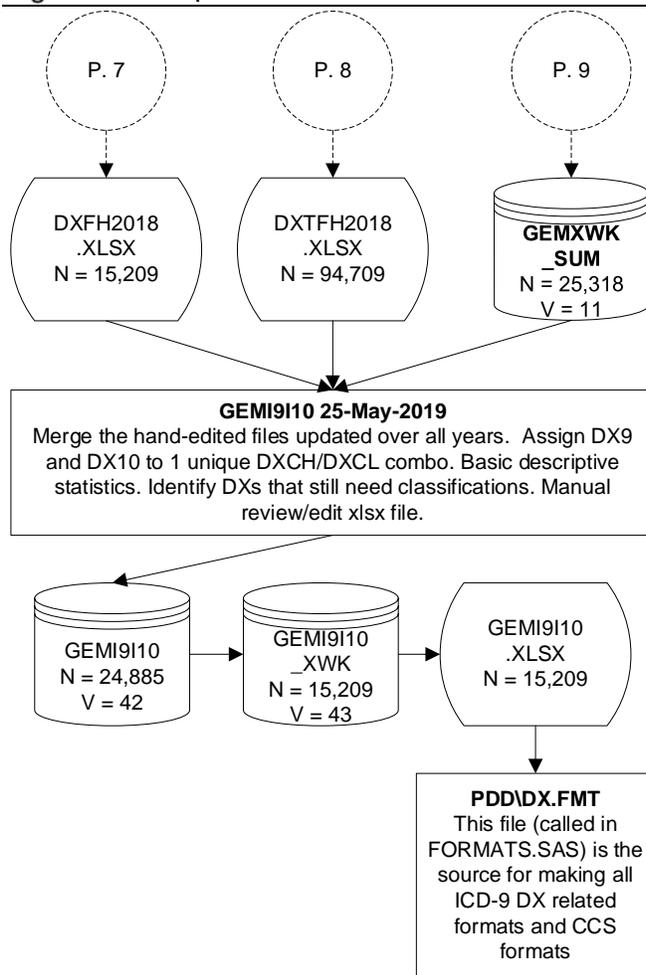


Figure 11 summarizes the steps to prepare the GEM crosswalk formats.

Based on limited experiences from other countries, researchers should expect a variety of effects on longitudinal research studies. Effects will range from insubstantial changes in which direct code mappings from ICD-9-CM to ICD-10-CM/PCS are available and simple methods such as concordance tables and comparability ratios may be employed to bridge the transition. However, in some instances coding discontinuities could create substantial issues for longitudinal research studies.

A separate document, available on the FHOP website summarizes changes for diagnosis groups of particular interest to our longitudinal research activities: DXCH11 (pregnancy, birth, and the puerperium), and DXCH06 (Mental Health) [19].

RESOURCES

We have focused on making formats from sources external to OSHPD files, particularly those from CMS, AHRQ, and developed by FHOP. All Excel files to make formats and the format libraries they produce are available on our [website](#). We strongly recommend that programmers join FHOP's SAS User Group. FHOP has only two people who can provide a limited amount of handholding to learn how to use these resources. Users will have to contract for more than one hour of support.

ENDNOTES

- 1 SAS Institute Inc. (1990) Chapter 18, The FORMAT Procedure. SAS Procedures Guide, Version 6, Third Edition, Cary NC: SAS Institute Inc., 1990, 705 pp.
- 2 Romano P, Luft HS, Remy L. (Dec 1993) Annual Report of the California Hospital Outcomes Project. Volume Two: Technical Appendix. California Health and Welfare Agency, Office of Statewide Health Planning and Development.
- 3 Romano P, Zache A, Luft HS, Rainwater, J, Remy L, Campa D. (Dec. 1995) The California Hospital Outcomes Project: Using administrative data to compare hospital performance. Joint Commission Journal on Quality Improvement, 21(12), 668-682.
- 4 Remy L, Clay T. (2017) Managing Longitudinal Research Studies: Preparing Master Files. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: <http://fhop.ucsf.edu/data-management-methods>
- 5 Remy L, Clay T. (2016) Managing Longitudinal Research Studies: The Basic Computing Environment. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: <http://fhop.ucsf.edu/data-management-methods>
- 6 Included in TOOLS.ZIP at:
http://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/custom_download/TOOLS.zip
- 7 ICD-9-CM Diagnosis and Procedure Codes: Abbreviated and Full Code Titles. (2014) Last access 25-May-2019 at:
<https://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/Downloads/ICD-9-CM-v32-master-descriptions.zip>.
- 8 International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). National Center for Health Statistics, Centers for Disease Control and Prevention. Last access 25-May-2019 at: <http://www.cdc.gov/nchs/icd/icd10cm.htm>
- 9 AHRQ Quality Indicators. Last access 25-May-2019 at:
<http://www.qualityindicators.ahrq.gov/software/>.
- 10 Databases and Related Tools from the Healthcare Cost and Utilization Project (HCUP) (2013). Last access 25-May-2019 at: <https://www.ahrq.gov/data/hcup/index.html>.
- 11 Last access 25-May-2019 at: <http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>.
- 12 Last access 25-May-2019 at: <http://www.hcup-us.ahrq.gov/toolssoftware/procedure/pc2013.csv> .
- 13 Last access 25-May-2019 at: http://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccsproc_downloading.jsp.
- 14 Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS), 2015. U.S. Agency for Healthcare Research and Quality. Last access 25-May-2019 at:
<https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp#download>
- 15 Source: <http://www.hcup-us.ahrq.gov/toolssoftware/procedure/pc2013.csv> Last access 25-May-2019.
- 16 Clinical Classifications Software for Services and Procedures (CCS-Services and Procedures) (2012). For a detailed description, see: <http://www.hcup->

us.ahrq.gov/toolssoftware/ccs_svcsproc/ccssvcproc.jsp. Last access 25-May-2019.
Downloaded File: fy2012_ccs_services_procedures.zip.

- 17 Gibson T, Casto A, Young J, Karnell L, Coenen N. Impact of ICD-10-CM/PCS on Research Using Administrative Databases. HCUP Methods Series Report # 2016-02 ONLINE. July 25, 2016. U.S. Agency for Healthcare Research and Quality. Last access 16-Jun-2019 at: <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp>.
- 18 2018 ICD-10 CM and GEMs. Last access 16-Jun-2019 at: <https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html>
- 19 Remy L, Clay T. (2020) Managing Longitudinal Research Studies: The crosswalk between ICD-9 AND ICD-10. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: <http://fhop.ucsf.edu/data-management-methods>.