Author: Melisa Litchfield

Reviewer: Andrea Schaffer

Dataset: PBS 10% Data

Background
The PBS 10% sample dataset reflects 100% of the patient PBS claiming activity for approximately 10% of the Australian PBS eligible population. The 10% sample includes PBS item codes, supply date, number of repeats, quantity of medicine dispensed, and the type of prescriber. The dataset also includes patient sex, patient year of birth, patient year of death, and pharmacy location by state. The PBS 10% sample cannot be linked to any other dataset. From July 2012, the sample also contains under co-payment medicines.

The PBS 10% sample dataset is based on a 'one in ten' random sample of patients within the PBS whose patient ID ends in a predetermined digit, for example 7 or any other digit from zero to nine (the same digit is always used). The dataset is longitudinal so every patient selected in the sample will remain in the 10% sample throughout the entire period. If the person is deceased, they will remain in the sample, however no claiming activity should be seen against these patients (there is a year of death field). The cohort is a dynamic cohort so more patients will be added if they have the predetermined digit in their patient ID.

Updates are received quarterly:

Update name	Month update received	Claims processing months
Summer	April	Dec/Jan/Feb
Autumn	July	March/April/May
Winter	October	June/July/August
Spring	January	Sept/Oct/Nov

The PBS 10% sample consists of three main datasets.
1. Claims data, the dispensing details of the medicine (Table 2 data dictionary)
2. Patient identifiers, gender and year of birth and death (Table 3 data dictionary)
3. Authority data, the medicines requiring authorisation prior to supply (Table 4 data dictionary)

1. **Claims data**
   Due to the size of the data, the analysis files are stored by year based on the date of supply.

2. **Patient Identifiers**
   The file 'patient_ids' contains the patient id number (scrambled), the patient sex, the patient year of birth, and patient year of death. This file will be updated with each quarterly update as new people are added to the dataset, and new dates of death become available. If there are differences in demographics, the newest data is retained. Previous versions of this file are stored in the folder 'Patient identification' in the main directory. The previous files are dated with their creation date.

3. **Authority data**
   The file 'Authority' contains the medicines that require authorisation from Department of Human Services prior to supply. The medicines listed in the authority data will appear in the supply data only if dispensed. The data link on Patient ID, prescriber ID, item code, and Authority approval date links to the prescription date in the supply data. The authority

Approved: Sallie Pearson   Signed: Date:

Author: Melisa Litchfield

Reviewer: Andrea Schaffer

dataset contains the variable restriction number to help identify the reason the authority medicine was prescribed.

Data custodian
Data is supplied by the Services Australia

Author: Melisa Litchfield

Reviewer: Andrea Schaffer

Table 2 – Claims datasets						
Variable list	**Type**	**Len**	**Format**	**Label**	**Description/Notes**	**Data Values**
PAT_ID	Num	8		Patient ID	Unique identifier for each patient (encrypted)	10 digit ID
ITEM_CODE	Char	6		Item code	PBS Item Code of drug supplied	
AUTH_FORM_ID	Num	8			Prescription number generated for authority items only	0 or . if not authority item or up to 8 digit number
PHARMACY_STATE	Char	3			State of Pharmacy where script was supplied	eg NSW, SA, VIC etc
SUPP_DATE	Num	8	ddmmyy10.	Supply date	Date the Pharmacy supplied the script +/- 14 days	
PRESC_DATE	Num	8	ddmmyy10.	Prescribe date	Date the Doctor prescribed the script +/- 14 days (Data Y2012M09 onwards only)	
PRESC_ID	Num	8		Prescriber ID	Unique identifier for each prescriber (encrypted)	
DERIVED_SPECIALTY	Char	3	$Special		Prescribers Derived Specialty	See Metadata for codes
FORM_CAT	Char	2	$Form_cat		Type of Script supplied – original, repeat, etc	eg AR, OR, OS – See table 5
CONCESSIONAL_STATUS	Char	1	$Con_stat		Concessional Status of patient	Eg A-M – See table 6
QTY	Num	8			Quantity Dispensed	Range 3approx.. 1 – 30,000
REPEATS	Num	8			Number of repeat scripts Some files contained *, this was recoded to 99	Range 3approx.. 0-60 99 = greater or equal 10
PREV_SUPP_NUM	Num	8			Number of previous scripts supplied (Data Y2012M09 onwards only)	Range 3approx.. 0-30
SCRIPTS	Num	8			Number of scripts supplied Some files contained *, this was recoded to 99	Range 3approx.. 1-20 99 = greater or equal 10
COPAY_IND	Char	5	$Copay		Classifies each script as either CoPayment or Under-CoPayment (Data processing date June 2012 onward)	COPAY = CoPayment UNDER=Under_CoPayment
STREAM_AUTH_CODE	Num	8			Identifies streamline authority scripts, with code (Data processing date June 2012 onward)	Mostly 4 digit eg 4172 Range 0 – 48756. Missing = . '0' is also in data
REG25_IND	Char	1			Indication of use of Regulation 25 Indicator (Data Y2012M09 onwards only)	Y = Yes N = No

Author: Melisa Litchfield

Reviewer: Andrea Schaffer

						S = Counts towards safety net (so Y also)
FILE_NAME	Char	8		Original file name	File name that data arrived in. It is also the processing date.	Eg Y2010M01
ATC_CODE	Char	7			ATC code of the medicine merged by the MPRU from the 'PBS Item drug map'. http://www.pbs.gov.au/info/statistics/dos-and-dop/dos-and-dop This is the Australian ATC code which differs slightly to the WHO ATC codes. http://www.whocc.no/atc_ddd_index/	Eg A10BA02

Table 3 - Patient IDs (PIN) dataset						
Variable list	Type	Len	Format	Label	Description/Notes	Data Values
PAT_ID	Num	8		Patient ID	Unique identifier for each patient (encrypted)	10 digit ID
PATIENT_SEX	Char	1		SEX	Sex of patient	M = Male F = Female
YEAR_BIRTH	Num	8		Year of birth	YYYY based on a randomly modified date of birth +/- 180 days	Range 1862 – 2019
YEAR_DEATH	Num	8		Year of death	YYYY based on randomly modified date of death +/- 180 days	Range 1915 – 2019
FLAG_100	Num	8		Flag 100	Year of birth more than 100 years ago DHS have removed their year of birth from the file	100 . = missing (under 100)
DATE_100	Num	Ddmmyy.			Date year_birth '100' appeared in the patient identifier file	Eg 08/05/2019

Table 4 - Authority dataset						
Variable list	Type	Len	Format	Label	Description/Notes	Data Values
PAT_ID	Num	8		Patient ID	Unique identifier for each patient (encrypted)	10 digit ID
AUTH_ITEM_CODE	Char	6		Item code	PBS Item Code of drug supplied or	

					Stoma Appliance Scheme code if CONCESSIONAL_STATUS = 'M'	
AUTH_APPR_DTE	Num	8	ddmmyy10.	Authority approval date	Date the Authority Approval for the script was supplied. Modified by a random offset (+/- 14 days) that is a constant for each PAT_ID	Min March 2008
PRESC_ID	Num	8		Prescriber ID	Unique identifier for each prescriber (encrypted)	
AUTH_RPT_NUM	Char	5		Authority repeat number	Number of repeat scripts	Single digit with * for 10 or more repeats
RSTR_NUM	Num	8	RESTRIC.		Restriction number of the script. Helps identify the reason for prescribing *note, metadata codes very incomplete	4 digit number See restriction number metadata for codes
ORIG_SUPP_IND	Char	2			Indication that an original script was supplied for an authority approval (Data Y2012M09 onwards only)	Y or missing Y indicates medicine dispensed