

11/19/2013

# Data Quality Report

Renal Adult

2004 - 2012

Manitoba Centre  
for Health Policy



UNIVERSITY  
OF MANITOBA

Say Hong  
University of Manitoba



# Contents

- Overview of Database..... 3
- Accuracy: Completeness and Correctness (VIMO Table)..... 4
  - Incident ..... 4
  - Hospital 1 ..... 6
  - Death..... 8
  - Hospital 2 ..... 10
  - Hospital 3 ..... 12
- Linkability ..... 13
- Phin Types ..... 13
- Agreement ..... 13
- Internal Validity: Trend Analysis ..... 14

## Overview of Database

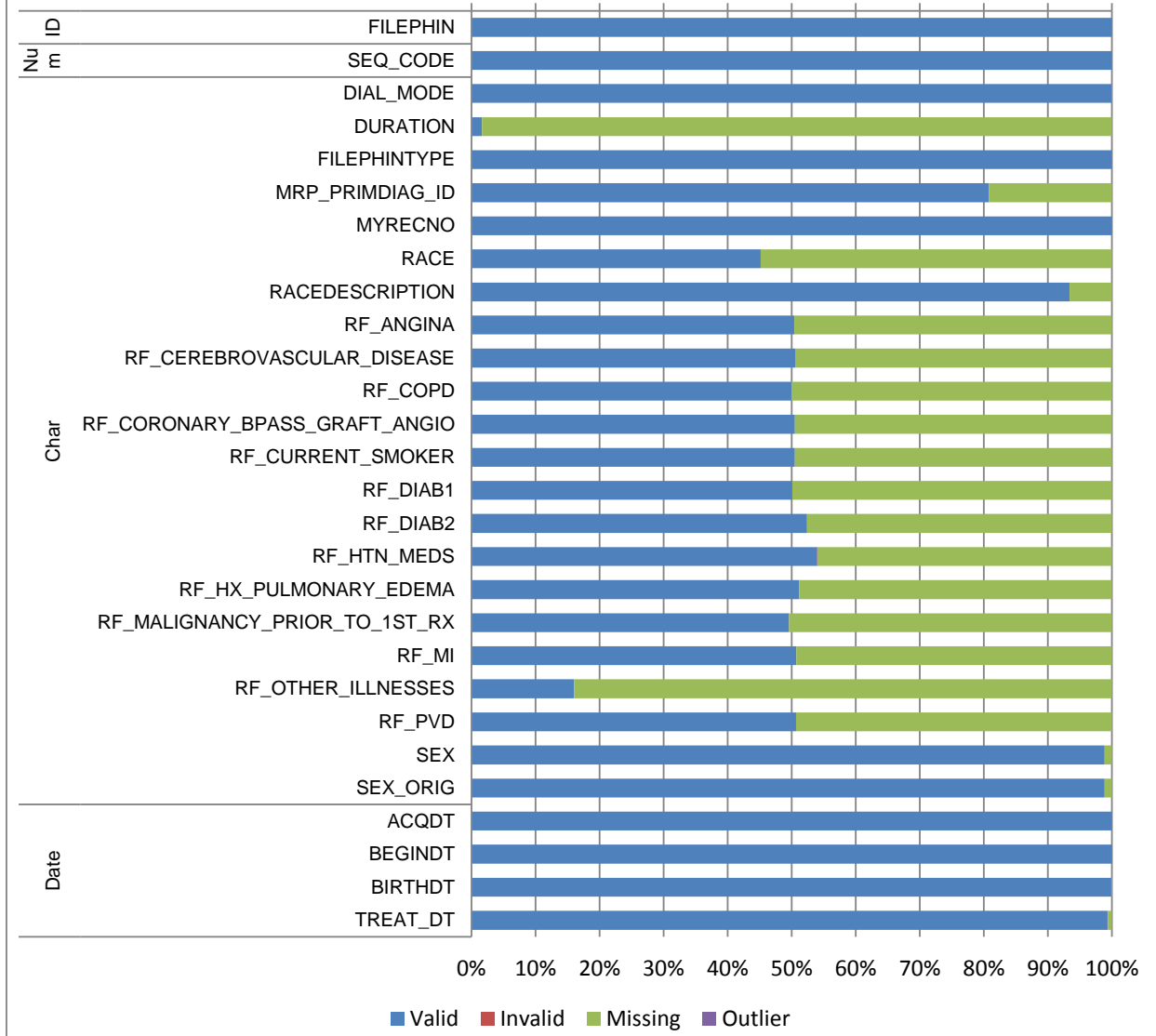
| Domain Name | SPDS Table                 | Dataset Label                | Number of Records | Number of Fields |
|-------------|----------------------------|------------------------------|-------------------|------------------|
| WORK        | CKD_2012_DEATHS_2004JAN    | Renal Adult - Deaths File    | 1147              | 6                |
| WORK        | CKD_2012_HOSPITAL1_2004JAN | Renal Adult - Hospital1 File | 8051              | 17               |
| WORK        | CKD_2012_INCIDENT_2004JAN  | Renal Adult - Incident File  | 2401              | 27               |
| WORK        | CKD_2012_HOSPITAL2_2004JAN | Renal Adult - Hospital2 File | 970               | 15               |
| WORK        | CKD_2012_HOSPITAL3_2004JAN | Renal Adult - Hospital3 File | 1750              | 7                |

## Accuracy: Completeness and Correctness (VIMO Table)

### Incident

| Dataset Label: Renal Adult - Incident File           |                               | Records: 2401                      |        | Legend (Data Quality Problems) : |                   |                      |   |            |      |        |     |  |
|--|-------------------------------|------------------------------------|--------|----------------------------------|-------------------|----------------------|---|------------|------|--------|-----|--|
| Dataset Name: Ckd_2012_incident_2004jan              |                               | Period: 2004-1012                  |        | None or Minimal<br>< 5%          | Moderate<br>5-30% | Significant<br>> 30% | Unknown<br>or N/A   |            |      |        |     |  |
| SUPPRESSED = Variables being suppressed in data file |                               |                                    |        |                                  |                   |                      |   |            |      |        |     |  |
| Type   | Variable Name                 | Variable Label                     | Valid  | Invalid                          | Missing           | Outlier              | Min   | Max        | Mean | Median | STD | Comment                                |
| ID   | FILEPHIN                      | MH SCrumbled PHIN                  | 100.00 |                                  | .00               |                      |   |            |      |        |     |  |
| Num  | SEQ_CODE                      | SEQ_CODE                           | 100.00 |                                  | .00               | .00                  | 1.00  | 1.00       | 1.00 | 1.00   | .00 |  |
| <b>Observed Values</b>                               |                               |                                    |        |                                  |                   |                      |   |            |      |        |     |  |
| Char   | DIAL_MODE                     | Modality                           | 99.96  |                                  | .04               |                      | SBDU, BDU, CDU, CAPD7, CAPD, SOAKS, KENORA, CCPD, CCPD7, PAE                    |            |      |        |     |  |
|  | DURATION                      | Length of total treatment to death | 1.62   |                                  | 98.38             |                      | 238, 200, 21, 34, 40, 32, 145, 207, 240, 152, 0, 218, 250, 31, 27, 50, 134, 117 |            |      |        |     |  |
|  | FILEPHINTYPE                  |                                    | 100.00 |                                  | .00               |                      | 4, 0  |            |      |        |     |  |
|  | MRP_PRIMDIAG_ID               | MRP diagnosis                      | 80.80  |                                  | 19.20             |                      | DN-2 Clinical, PCKD Adult, TIN HTN nephrosclerosis no Bx, GN DN-2 clinical,     |            |      |        |     |  |
|  | MYRECNO                       | Database patient identifier        | 100.00 |                                  | .00               |                      | 04883, 05056, 05069, 05154, 06271, 06283, 06309, 06405, 06455, 06532, 086       |            |      |        |     |  |
|  | RACE                          | Old Race description               | 45.15  |                                  | 54.85             |                      | <b>SUPPRESSED</b>   |            |      |        |     |  |
|  | RACEDESCRIPTION               | MRP race description               | 93.38  |                                  | 6.62              |                      | <b>SUPPRESSED</b>   |            |      |        |     |  |
|  | RF_ANGINA                     | CORR co morbid condition yes or no | 50.40  |                                  | 49.60             |                      | No, Yes, Unknown, no  |            |      |        |     |  |
|  | RF_CEREBROVASCULAR_DISEASE    | RF-Cerebrovascular Disease         | 50.52  |                                  | 49.48             |                      | No, Yes, Unknown, y   |            |      |        |     |  |
|  | RF_COPD                       | CORR co morbid condition yes or no | 49.94  |                                  | 50.06             |                      | Yes, No, Unknown  |            |      |        |     |  |
|  | RF_CORONARY_BPASS_GRAFT_ANGIO | CORR co morbid condition yes or no | 50.48  |                                  | 49.52             |                      | No, Yes, Unknown  |            |      |        |     |  |
|  | RF_CURRENT_SMOKER             | CORR co morbid condition yes or no | 50.48  |                                  | 49.52             |                      | No, Yes, Unknown  |            |      |        |     |  |
|  | RF_DIAB1                      | CORR co morbid condition yes or no | 50.02  | .04                              | 49.94             |                      | No, Yes, Unknown, NY  |            |      |        |     |  |
|  | RF_DIAB2                      | CORR co morbid condition yes or no | 52.39  |                                  | 47.61             |                      | No, Yes, Unknown  |            |      |        |     |  |
|  | RF_HTN_MEDS                   | CORR co morbid condition yes or no | 53.98  | .04                              | 45.98             |                      | Yes, No, ny, Unknown, Y   |            |      |        |     |  |
|  | RF_HX_PULMONARY_EDEMA         | CORR co morbid condition yes or no | 51.19  |                                  | 48.81             |                      | No, Yes, nn, Unknown, NN, y   |            |      |        |     |  |
|  | RF_MALIGNANCY_PRIOR_TO_1ST_RX | CORR co morbid condition yes or no | 49.60  |                                  | 50.40             |                      | No, Yes, Unknown  |            |      |        |     |  |
|  | RF_MI                         | CORR co morbid condition yes or no | 50.73  |                                  | 49.27             |                      | No, Yes, Unknown, no  |            |      |        |     |  |
|  | RF_OTHER_ILLNESSES            | CORR co morbid condition yes or no | 15.99  |                                  | 84.01             |                      | emphysema, glaucoma, sarcoidosis, obesity;fatty liver disease, Pancreatitis, i  |            |      |        |     |  |
|  | RF_PVD                        | CORR co morbid condition yes or no | 50.69  |                                  | 49.31             |                      | No, Yes, Unknown  |            |      |        |     |  |
| SEX  | Gender                        | 98.88                              |        | 1.12                             |                   | 1, 2                 |   |            |      |        |     |  |
| SEX_ORIG   | Original sex values           | 98.88                              |        | 1.12                             |                   | M, F                 |   |            |      |        |     |  |
| Date   | ACQDT                         | Date record was acquired at MCHP   | 100.00 |                                  | .00               |                      | 2012-11-01  | 2012-11-01 |      |        |     |  |
|  | BEGINDT                       | Date associated with modality      | 99.96  | .04                              | .00               |                      | 2004-01-02  | 2012-12-23 |      |        |     | 2012-12-23 ( 1 Invalid Obs. in total ) |
|  | BIRTHDT                       | Date of birth                      | 99.92  |                                  | .08               |                      | 1911-01-17  | 2012-06-30 |      |        |     |  |
|  | TREAT_DT                      | First treatment date               | 99.33  | .04                              | .62               |                      | 2004-01-02  | 2012-12-09 |      |        |     | 2012-12-09 ( 1 Invalid Obs. in total ) |

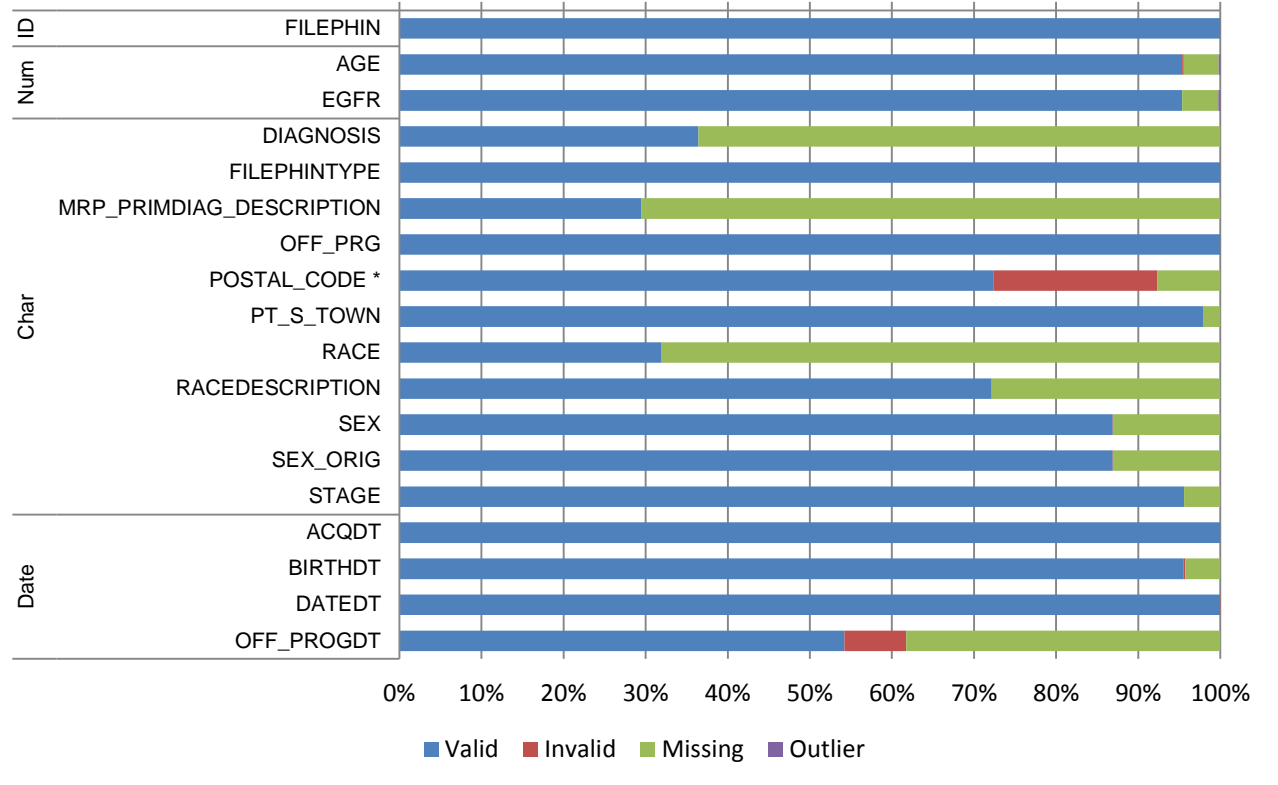
## Ckd\_2012\_incident\_2004jan



## Hospital 1

| Dataset Label: Renal Adult - Hospital1 File   |                          | Records: 8051                    |        | Legend (Data Quality Problems) : |                   |                      |   |            |       |        |       |  |
|---|--------------------------|----------------------------------|--------|----------------------------------|-------------------|----------------------|---|------------|-------|--------|-------|--|
| Dataset Name: Ckd_2012_hospital1_2004jan  |                          | Period: 2004-1012                |        | None or Minimal<br>< 5%          | Moderate<br>5-30% | Significant<br>> 30% | Unknown<br>or N/A   |            |       |        |       |  |
| <b>SUPPRESSED = Variables being suppressed in data file</b><br>* = All postal codes listed here have frequency count > 20 |                          |                                  |        |                                  |                   |                      |   |            |       |        |       |  |
| Type  | Variable Name            | Variable Label                   | Valid  | Invalid                          | Missing           | Outlier              | Min   | Max        | Mean  | Median | STD   | Comment  |
| FILEPHIN  | FILEPHIN                 | MH SCrambled PHIN                | 100.00 |                                  | .00               |                      |   |            |       |        |       |  |
| Num ID  | AGE                      | Age at last eGFR                 | 95.37  | .20                              | 4.25              | .19                  | -18.79  | 935.33     | 62.23 | 64.10  | 22.42 | 16 invalid obs. out of [0, 110] range                          |
|   | EGFR                     | estimated glom filt rate         | 95.34  |                                  | 4.42              | .24                  | 2.02  | 377.62     | 44.61 | 36.88  | 34.19 |  |
| <b>Observed Values</b>  |                          |                                  |        |                                  |                   |                      |   |            |       |        |       |  |
| Char  | DIAGNOSIS                | Primary diagnosis-old            | 36.42  |                                  | 63.58             |                      | RVD, Sarcoid, Membranous GN, Diabetic/HTN, Unknown, Hypertens   |            |       |        |       |  |
|   | FILEPHINTYPE             |                                  | 100.00 |                                  | .00               |                      | 4, 0  |            |       |        |       |  |
|   | MRP_PRIMDIAG_DESCRIPTION | Mrp Primary diagnosis-in use     | 29.46  |                                  | 70.54             |                      | SLE Bx, RVD HTN (Biopsy proven), no renal disease, Vas (P-ANCA)   |            |       |        |       |  |
|   | OFF_PRG                  | Off renal program yes or no      | 100.00 |                                  | .00               |                      | 1, 0  |            |       |        |       |  |
|   | POSTAL_CODE *            | Postal code                      | 72.36  | 19.99                            | 7.65              |                      | R0C, R0B0J0, R0B1B0, R0B, R2W, R0B1J0, R3B, R2G, R0E1M0, R2V, R2P, R0B, P0X, ... (1609 invalid obs. in total) |            |       |        |       |  |
|   | PT_S_TOWN                | Pt's Town                        | 97.90  |                                  | 2.10              |                      | Brandon, Elkhorn, Sioux Narrows, Winnipeg, Sandy Lake, Sandy Lak  |            |       |        |       |  |
|   | RACE                     | Historic race description        | 31.95  |                                  | 68.05             |                      | <b>SUPPRESSED</b>   |            |       |        |       |  |
|   | RACEDESCRIPTION          | MRP Race description             | 72.14  |                                  | 27.86             |                      | <b>SUPPRESSED</b>   |            |       |        |       |  |
|   | SEX                      | sex                              | 86.91  | .01                              | 13.08             |                      | 1, 2, 0   |            |       |        |       |  |
|   | SEX_ORIG                 | Original sex values              | 86.91  | .01                              | 13.08             |                      | M, F, Male, m, male, female, FEMALE, Female, malr, f, 0   |            |       |        |       |  |
| STAGE   | ckd stage                | 95.58                            |        | 4.42                             |                   | 1, 3, 2, 4, 5        |   |            |       |        |       |  |
| Date  | ACQDT                    | Date record was acquired at MCHP | 100.00 |                                  | .00               |                      | 2012-11-01  | 2012-11-01 |       |        |       |  |
|   | BIRTHDT                  | Date of Birth                    | 95.55  | .17                              | 4.27              |                      | 1902-06-21  | 2029-12-21 |       |        |       | 14 invalid obs. out of [1900-01-01, 2012-11-01] range          |
|   | DATEDT                   | Date of last eGFR                | 99.96  | .04                              | .00               |                      | 1999-12-01  | 2022-05-16 |       |        |       | 2012-11-09, 2013-08-12, 2022-05-16 ( 3 Invalid Obs. in total ) |
|   | OFF_PROGDT               | Off Program date                 | 54.22  | 7.50                             | 38.28             |                      | 2001-09-04  | 2020-04-20 |       |        |       | 604 invalid obs. out of [2004-01-01, 2012-11-01] range         |

### Ckd\_2012\_hospital1\_2004jan

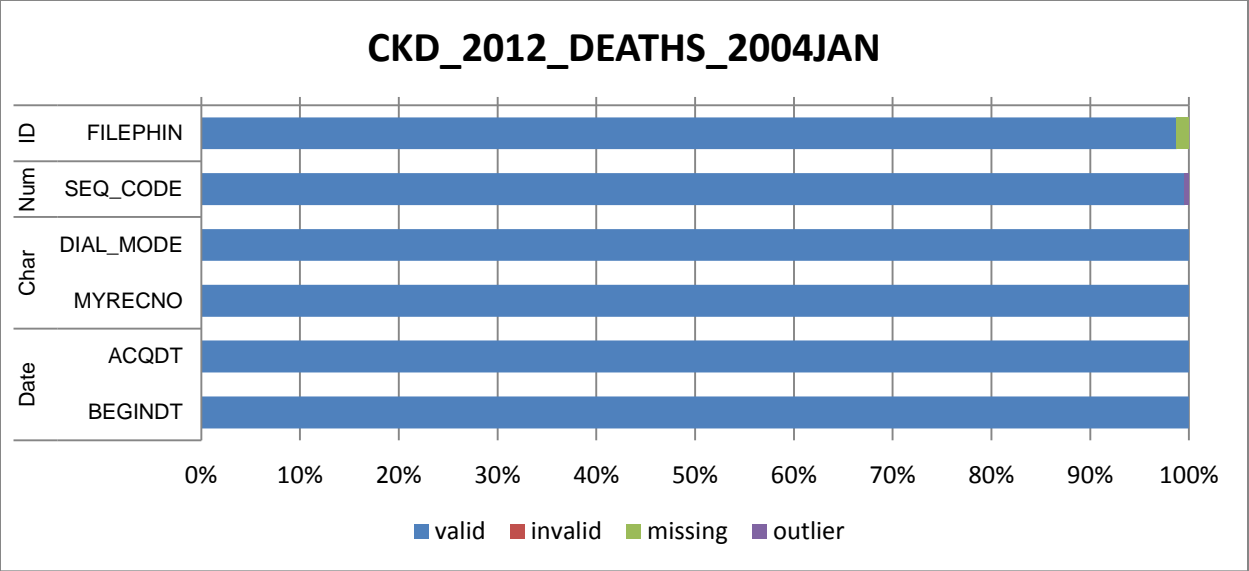


## Death

| Dataset Label: Renal Adult - Deaths File |           | Records: 1147                    |        | Legend (Data Quality Problems) : |                   |                      |   |            |      |        |      |         |
|--|-----------|----------------------------------|--------|----------------------------------|-------------------|----------------------|---|------------|------|--------|------|---------|
| Dataset Name: CKD_2012_DEATHS_2004JAN    |           | Period: 2002-2012                |        | None or Minimal <<br>5% Missing  | Moderate<br>5-30% | Significant<br>> 30% | Unknown<br>or N/A   |            |      |        |      |         |
| type                                     | varname   | varlabel                         | valid  | invalid                          | missing           | outlier              | min   | max        | mean | median | std  | Comment |
| ID                                       | FILEPHIN  | MH SCrambled PHIN                | 98.78  |                                  | 1.22              |                      |   |            |      |        |      |         |
| Num                                      | SEQ_CODE  | Sequence                         | 99.56  |                                  | .00               | .44                  | 2.00  | 16.00      | 3.07 | 3.00   | 1.40 |         |
| <b>Observed Values</b>                   |           |                                  |        |                                  |                   |                      |   |            |      |        |      |         |
| Char                                     | DIAL_MODE | Discont or Death                 | 100.00 |                                  | .00               |                      | DIED, DISCONT   |            |      |        |      |         |
|  | MYRECNO   | Database patient identifier      | 100.00 |                                  | .00               |                      | 02027, 02362, 02519, 03931, 04006, 04122, 04204, 04474, 04480 |            |      |        |      |         |
| Date                                     | ACQDT     | Date record was acquired at MCHP | 100.00 |                                  | .00               |                      | 2012-11-01  | 2012-11-01 |      |        |      |         |
|  | BEGINDT   | Date of change                   | 100.00 |                                  | .00               |                      | 2004-01-13  | 2012-09-10 |      |        |      |         |



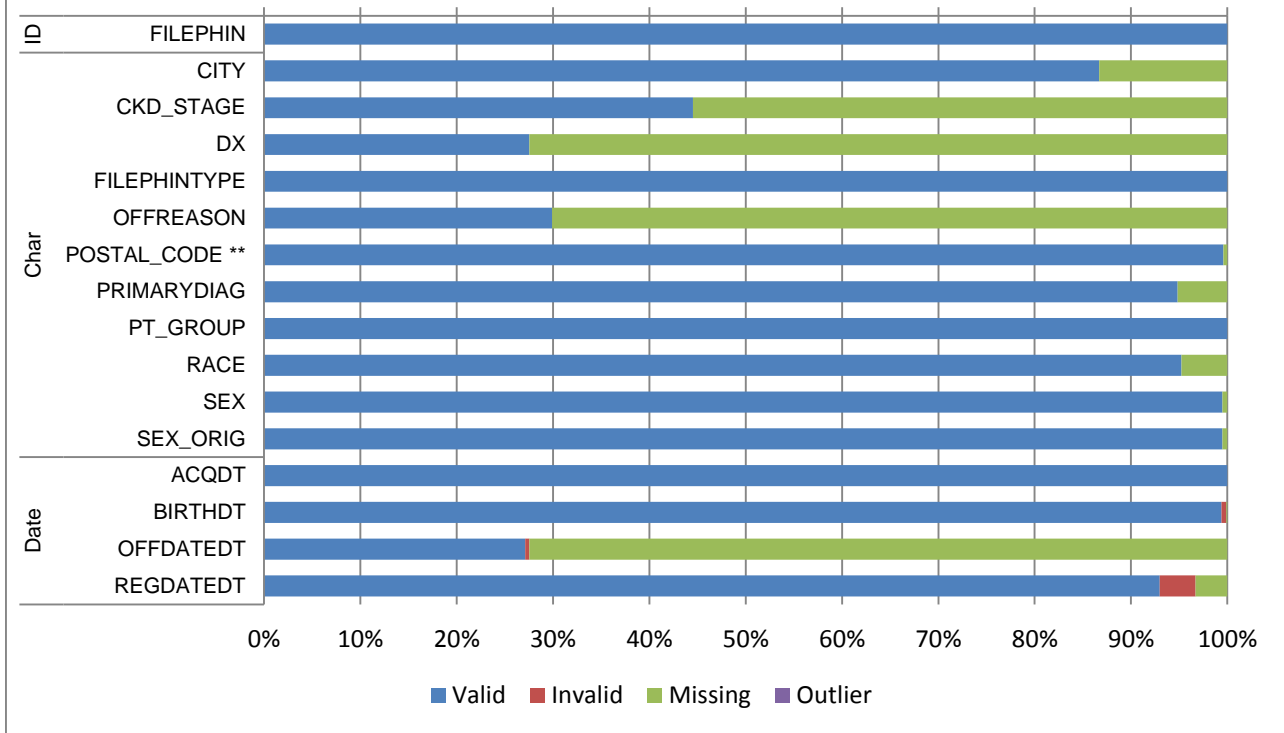
### CKD\_2012\_DEATHS\_2004JAN



## Hospital 2

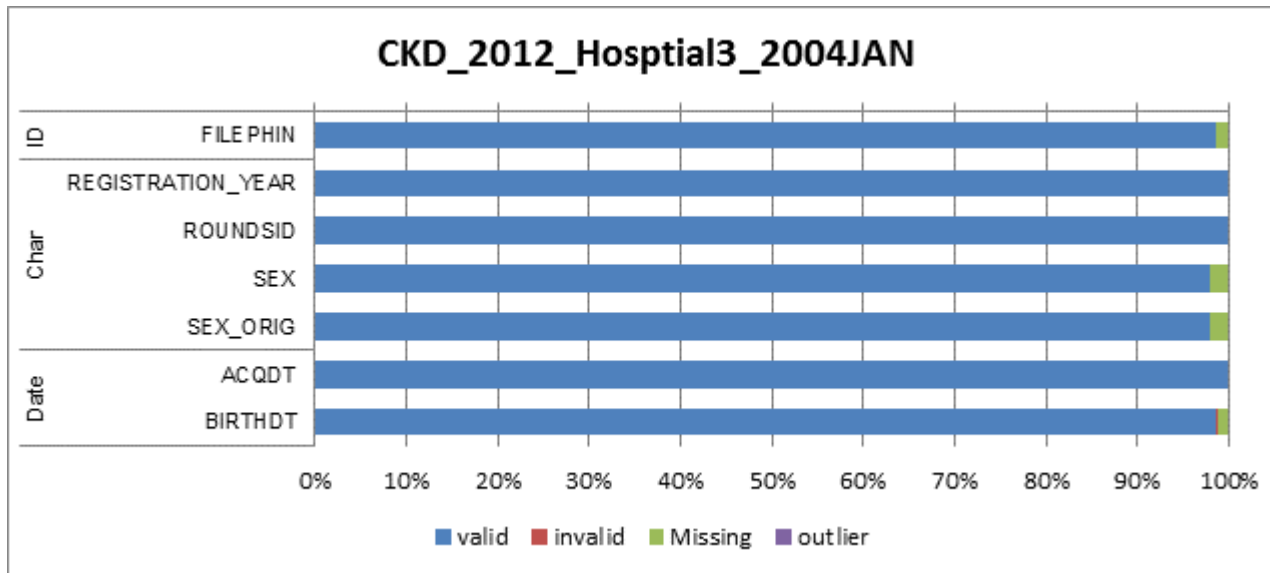
| Dataset Label: Renal Adult - Hospital2 File               |                        | Records: 970                     | Legend (Data Quality Problems) : |                   |                      |                   |   |            |      |        |     |  |
|---|------------------------|----------------------------------|----------------------------------|-------------------|----------------------|-------------------|---|------------|------|--------|-----|--|
| Dataset Name: Ckd_2012_hospital2_2004jan                  |                        | Period: 2004-1012                | None or Minimal<br>< 5%          | Moderate<br>5-30% | Significant<br>> 30% | Unknown<br>or N/A |   |            |      |        |     |  |
| SUPPRESSED = Variables being suppressed in data file      |                        |                                  |                                  |                   |                      |                   |   |            |      |        |     |  |
| ** = Postal codes suppressed due to small frequency count |                        |                                  |                                  |                   |                      |                   |   |            |      |        |     |  |
| Type  | Variable Name          | Variable Label                   | Valid                            | Invalid           | Missing              | Outlier           | Min   | Max        | Mean | Median | STD | Comment  |
| ID  | FILEPHIN               | MH SCrambled PHIN                | 100.00                           |                   | .00                  |                   |   |            |      |        |     |  |
| Char  | <b>Observed Values</b> |                                  |                                  |                   |                      |                   |   |            |      |        |     |  |
|   | CITY                   | City                             | 86.70                            |                   | 13.30                |                   | RAINY RIVER, ARVIAT, NU, BALMERTOWN, VERMILLION,                |            |      |        |     |  |
|   | CKD_STAGE              | ckd stage                        | 44.54                            |                   | 55.46                |                   | 4, 1, 2, 3, 5   |            |      |        |     |  |
|   | DX                     | Dx                               | 27.53                            |                   | 72.47                |                   | 1, 0, 4, 7  |            |      |        |     |  |
|   | FILEPHINTYPE           |                                  | 100.00                           |                   | .00                  |                   | 4, 0  |            |      |        |     |  |
|   | OFFREASON              | OffReason                        | 29.90                            |                   | 70.10                |                   | deceased, no follow - up, no follow-up, transferred to TB, PD S |            |      |        |     |  |
|   | POSTAL_CODE **         | Postal Code                      | 99.59                            |                   | .41                  |                   | ...   |            |      |        |     |  |
|   | PRIMARYDIAG            | PrimaryDiag                      | 94.85                            |                   | 5.15                 |                   | Obstructive Nephropathy, Other ESRD, Hypertension, Congen       |            |      |        |     |  |
|   | PT_GROUP               | Pt Group                         | 100.00                           |                   | .00                  |                   | GENSO, OFF, RHSO  |            |      |        |     |  |
| RACE  | Race                   | 95.26                            |                                  | 4.74              |                      | <b>SUPPRESSED</b> |   |            |      |        |     |  |
| SEX   | Gender                 | 99.48                            |                                  | .52               |                      | 1, 2              |   |            |      |        |     |  |
| SEX_ORIG  | Original sex values    | 99.48                            |                                  | .52               |                      | M, F              |   |            |      |        |     |  |
| Date  | ACQDT                  | Date record was acquired at MCHP | 100.00                           |                   | .00                  |                   | 2012-11-01  | 2012-11-01 |      |        |     |  |
|   | BIRTHDT                | DateofBirth                      | 99.38                            | .52               | .10                  |                   | 1910-01-03  | 2023-08-01 |      |        |     | 2014-10-09, 2017-11-10, 2019-04-06, 2023-07-09, 2023-08-01 ( 5 Invalid Obs. in total ) |
|   | OFFDATEDT              | OffDate                          | 27.11                            | .41               | 72.47                |                   | 1998-12-09  | 2029-09-10 |      |        |     | 2017-11-09, 2020-01-19, 2021-10-09, 2029-09-10 ( 4 Invalid Obs. in total )             |
|   | REGDATEDT              | ReqDate                          | 92.99                            | 3.71              | 3.30                 |                   | 1930-01-08  | 2029-11-10 |      |        |     | 36 invalid obs. out of [1900-01-01, 2012-11-01] range                                  |

## Ckd\_2012\_hospital2\_2004jan



## Hospital 3

| Dataset Label: Renal Adult - Hospital3 File |                   |                                  | Records: 1750     |         |         | Legend (Data Quality Problems) : |   |                      |                   |        |     |  |
|---|-------------------|----------------------------------|-------------------|---------|---------|----------------------------------|---|----------------------|-------------------|--------|-----|--|
| Dataset Name: CKD_2012_Hospital3_2004JAN    |                   |                                  | Period: 2004-2012 |         |         | None or Minimal<br>< 5% Missing  | Moderate<br>5-30%   | Significant<br>> 30% | Unknown<br>or N/A |        |     |  |
| type  | varname           | varlabel                         | valid             | invalid | Missing | outlier                          | min   | max                  | mean              | median | std | Comment  |
| ID  | FILEPHIN          | MH SCrambled PHIN                | 98.51             |         | 1.49    |                                  |   |                      |                   |        |     |  |
|   |                   |                                  |                   |         |         |                                  | Observed Values   |                      |                   |        |     |  |
| Char  | REGISTRATION_YEAR | Registration Year                | 100.00            |         | .00     |                                  | 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012  |                      |                   |        |     |  |
|   | ROUNDSID          | auto id in list                  | 100.00            |         | .00     |                                  | 0001, 0002, 0003, 0004, 0005, 0006, 0007, 0008, 0009, 0010, |                      |                   |        |     |  |
|   | SEX               | Gender                           | 97.89             |         | 2.11    |                                  | 1, 2  |                      |                   |        |     |  |
|   | SEX_ORIG          | Original sex values              | 97.89             |         | 2.11    |                                  | F, M  |                      |                   |        |     |  |
| Date  | ACQDT             | Date record was acquired at MCHP | 100.00            |         | .00     |                                  | 2012-11-01  | 2012-11-01           |                   |        |     |  |
|   | BIRTHDT           | Date of Birth                    | 98.51             | .34     | 1.14    |                                  | 1903-01-21  | 2029-09-11           |                   |        |     | 2017-04-07, 2018-10-15, 2020-12-08, 2026-06-19, 2029-04-29, 2029-09-11 (6 invalid obs. in total) |



## Linkability

| Dataset                    | Dataset Label                | Total Number of Records | Number of Linkable Records | % Linkable Records | Number of Linkable Individuals |
|----------------------------|------------------------------|-------------------------|----------------------------|--------------------|--------------------------------|
| CKD_2012_INCIDENT_2004JAN  | Renal Adult - Incident File  | 2401                    | 2363                       | 98.42              | 2362                           |
| CKD_2012_Hospital1_2004JAN | Renal Adult - Hospital1 File | 8051                    | 7670                       | 95.27              | 7654                           |
| CKD_2012_DEATHS_2004JAN    | Renal Adult - Deaths File    | 1147                    | 1133                       | 98.78              | 946                            |
| CKD_2012_Hospital2_2004JAN | Renal Adult - Hospital2 File | 970                     | 954                        | 98.35              | 953                            |
| CKD_2012_Hospital3_2004JAN | Renal Adult - Hospital3 File | 1750                    | 1724                       | 98.51              | 1724                           |

## Phin Types

| FILEPHINTYPE                              | CKD_2012_INCIDENT_2004JAN | CKD_2012_Hospital1_2004JAN | CKD_2012_DEATHS_2004JAN | CKD_2012_Hospital2_2004JAN | CKD_2012_Hospital3_2004JAN |
|---|---------------------------|----------------------------|-------------------------|----------------------------|----------------------------|
| 0 MH verified against concurrent registry | 98.42                     | 95.27                      | 98.78                   | 98.35                      | 98.51                      |
| 4 MCHP db specific ScrPHIN - No MH found  | 1.58                      | 4.73                       | 1.22                    | 1.65                       | 1.49                       |

## Agreement

| Dataset Name              | Degree of Agreement with Registry - SEX (Kappa Statistic) | Degree of Agreement with Registry - Date of Birth (Kappa Statistic) |
|---------------------------|---|---|
| CKD_2012_INCIDENT_2004JAN | 0.9803  | 0.9753  |

# Internal Validity: Trend Analysis

Trend Analysis for Dataset: projectCkd\_2012\_incident\_2004jan

Date Variable: treat\_dt, By Variable: \_ALL\_

By Variable=All Records

