

# Introducing icd9: working with ICD-9 codes and comorbidities in R

Jack O. Wasey

July 7, 2014

## 1 Introduction

This package is designed to be used with a variety of input data, including multiple possible formats of ICD-9 codes, but some assumptions are made. There are many ways of misinterpreting ICD-9 codes, especially when dealing with ranges. The code in this package carefully considers a wide range of possibilities. **ICD-9 codes are not numeric.** Using numeric values for either decimal or non-decimal form will cause serious problems, hence the predominantly string-based processing here, and a robust set of unit tests.

When calculating which patients have which comorbidities, the input data is typically structured as follows:

```
patientData

##   visitId  icd9 poa
## 1    1000 27801   Y
## 2    1000  7208   N
## 3    1000 25001   Y
## 4    1001 34400   N
## 5    1001  4011   Y
## 6    1002  4011   N
```

Only the visitId column is propagated to the results. If 'present-on-arrival' is needed, it must be separated out first. The implicit default, therefore, is to ignore it, and give ICD-9 code regardless of POA status.

The comorbidities can be determined as follows:

```
icd9Comorbidities(icd9df = patientData)[, 1:5]

##   visitId   CHF VALVE PULMCIRC PERIVASC
## 1    1000 FALSE FALSE    FALSE    FALSE
## 2    1001 FALSE FALSE    FALSE    FALSE
## 3    1002 FALSE FALSE    FALSE    FALSE
```

The following shows the same code with default options written out:

```
icd9Comorbidities(icd9df = patientData,  
  visitId = "visitId",  
  icd9Field = "icd9",  
  icd9Mapping = ahrqComorbid,  
  validateMapping = FALSE,  
  shortMapping = TRUE)
```

## 2 Converting ICD-9 codes between types

These functions were designed with the problem of incorrectly or bizarrely formatted ICD-9 codes in mind. These functions make the reasonable assumption that short codes of three or fewer characters must be describing only the 'major' part.

`keepLoneDecimal`

allows retention of the decimal point even if there are no subsequent characters.

```
icd9DecimalToShort(c("10.20", "100", "123.45"))  
  
## [1] "01020" "100" "12345"  
  
icd9ShortToDecimal(icd9DecimalToShort(c("10.20", "100", "123.45")))  
  
## [1] "010.20" "100" "123.45"  
  
# only a subset of short codes can suffer dropping of leading zeroes.  
icd9DecimalToShort(c("1", "22", "22.44", "100"))  
  
## [1] "001" "022" "02244" "100"  
  
icd9ShortToDecimal(icd9DecimalToShort(c("1", "1.2", "123.45")))  
  
## [1] "001" "001.2" "123.45"  
  
icd9ShortToDecimal(icd9DecimalToShort(c("1", "2.2", "100")))  
  
## [1] "001" "002.2" "100"
```

## 3 Validation of ICD-9 codes

```

icd9ValidDecimal("V10.2")

## [1] TRUE

icd9ValidShort(c("099.17", "-1"))

## [1] FALSE FALSE

icd9ValidShort(c("1", "001", "100", "10023"))

## [1] TRUE TRUE TRUE TRUE

```

Validation forces the package user to provide character format ICD-9 codes. If great care is taken, passing some integers could be valid, but given the high chance of mistakes, and the simplicity of dealing entirely with character input, character is enforced:

```

icd9ValidShort(100) # throws an error

```

## 4 Ranges of ICD-9 codes

These functions generate syntactically valid ICD-9 codes, without including parent codes when the range limit would subset the parent. E.g. "100.99" %i9d% "101.01" does not include "100" or "100.0", both of which imply large subsets than requested by the range command.

```

"10099" %i9s% "10101"

## [1] "10099" "10100" "10101"

"V10" %i9d% "V10.02"

## [1] "V10"      "V10.0"    "V10.00"   "V10.01"   "V10.02"

# "E987" %i9d% "E988.9"

```

Another way of specifying ranges are to use function calls. These are exactly equivalent to the %i9s% and %i9d% range operators. This example shows the result when the user specifies a range which would include parents but not all their children:

```

icd9ExpandRangeShort("V100", "V1002")

## [1] "V100"   "V1000" "V1001" "V1002"

```

Although V100 would include ten children, the range only returns 4 values. In all other cases, parents are omitted to avoid the range returning overly broad classifications than intended. A planned feature is to optionally enable returning these parent codes, which would then follow a more numerical pattern (although still distinguishing trailing zeroes).

## 5 Human-readable ICD-9

```
icd9ExplainDecimal("1.0")

##      ICD-9              Diagnosis              Description
## 1 001.0 Cholera due to vibrio cholerae Cholera d/t vib cholerae

icd9Explain("1.0", isShort = FALSE)

##      ICD-9              Diagnosis              Description
## 1 001.0 Cholera due to vibrio cholerae Cholera d/t vib cholerae

icd9ExplainDecimal("001.1")

##      ICD-9              Diagnosis              Description
## 1 001.1 Cholera due to vibrio cholerae el tor Cholera d/t vib el tor

icd9ExplainDecimal(icd9ShortToDecimal("0019"))

##      ICD-9              Diagnosis Description
## 1 001.9 Cholera, unspecified Cholera NOS

icd9ExplainShort("0019")

##      ICD-9              Diagnosis Description
## 1 0019 Cholera, unspecified Cholera NOS

# named list(s) of codes
icd9ExplainDecimal(list(cholera = c("001", "001.0", "001.1", "001.9")))

## $cholera
##      ICD-9              Diagnosis              Description
## 1 001.0 Cholera due to vibrio cholerae Cholera d/t vib cholerae
## 2 001.1 Cholera due to vibrio cholerae el tor Cholera d/t vib el tor
## 3 001.9 Cholera, unspecified Cholera NOS

# same using decimal codes without a list
icd9ExplainDecimal(c("001", "001.0", "001.1", "001.9"))
```

```
## ICD-9 Diagnosis Description
## 1 001.0 Cholera due to vibrio cholerae Cholera d/t vib cholerae
## 2 001.1 Cholera due to vibrio cholerae el tor Cholera d/t vib el tor
## 3 001.9 Cholera, unspecified Cholera NOS

# 001/cholera doesn't itself have an explanation: TODO walk down children to get next level
icd9ExplainDecimal(list(cholera = "001", rheumatic_heart = "390"))

## $cholera
## ICD-9 Diagnosis Description
## 1 001.0 Cholera due to vibrio cholerae Cholera d/t vib cholerae
## 2 001.1 Cholera due to vibrio cholerae el tor Cholera d/t vib el tor
## 3 001.9 Cholera, unspecified Cholera NOS
##
## $rheumatic_heart
## ICD-9 Diagnosis Description
## 1 390 Rheumatic fever without mention of heart involvement
##
## 1 Rheum fev w/o hrt involv
```

Now try to explain on a non-existent (but 'valid') ICD-9 code:

```
icd9ExplainDecimal("001.5")

## [1] ICD-9 Diagnosis Description
## <0 rows> (or 0-length row.names)

icd9ExplainDecimal(list(cholera=c("001.0", "001.1", "001.9"))) @
```

## 6 Chaining commands

With the

`magrittr`

package installed, commands can be chained together in a convenient and readable manner:

```
c("001.1", "391") %>% icd9DecimalToShort %>% icd9ExplainShort

## ICD-9 Diagnosis Description
## 1 0011 Cholera due to vibrio cholerae el tor
## 2 3910 Acute rheumatic pericarditis
## 3 3911 Acute rheumatic endocarditis
## 4 3912 Acute rheumatic myocarditis
## 5 3918 Other acute rheumatic heart disease
```

```
## 6 3919 Acute rheumatic heart disease, unspecified
##           Description
## 1 Cholera d/t vib el tor
## 2 Acute rheumatic pericard
## 3 Acute rheumatic endocard
## 4 Ac rheumatic myocarditis
## 5 Ac rheumat hrt dis NEC
## 6 Ac rheumat hrt dis NOS
```

Find all ICD-9 codes matching 'heart' or 'cardiac' in the short or long descriptions:

```
cardiac <- unique(c(
  icd9CmDesc[
    grepl(
      pattern="(heart)|(cardiac)",
      x = icd9CmDesc[["descLong"]],
      ignore.case = TRUE
    ),
    "icd9"],
  icd9CmDesc[
    grepl(
      pattern="(heart)|(cardiac)",
      x = icd9CmDesc[["descShort"]],
      ignore.case = TRUE
    ),
    "icd9"]
  ))
```

then explain the list, just showing the first ten:

```
cardiac %>% icd9ExplainShort %>% extract(2) %>% head(10)

##           Diagnosis
## 1 Chagas' disease with heart involvement
## 2 gonococcal heart disease
## 3 Malignant neoplasm of heart
## 4 Benign neoplasm of heart
## 5 Rheumatic fever without mention of heart involvement
## 6 Other acute rheumatic heart disease
## 7 Acute rheumatic heart disease, unspecified
## 8 Rheumatic chorea with heart involvement
## 9 Rheumatic chorea without mention of heart involvement
## 10 heart disease, unspecified
```

More examples, this time also demonstrating automatic summarization of a long list of ICD-9 codes into the minimum set of explanatory parent codes:

```

quanDeyoComorbid[["Dementia"]] %>%
  icd9ExplainShort() %>%
  extract(c("ICD-9", "Description"))

##      ICD-9      Description
## 1    2900  Senile dementia uncomp
## 2    29010      dementia
## 3    29011      delirium
## 4    29012      delusion
## 5    29013      depression
## 6    29020      delusion
## 7    29021      depressive
## 8    2903    Senile delirium
## 9    29040      dementia, uncomp
## 10   29041      dementia w delirium
## 11   29042      dementia w delusion
## 12   29043      dementia w depressn
## 13   2908    Senile psychosis NEC
## 14   2909  Senile psychot cond NOS
## 15   29410      w/o behav dist
## 16   29411      w behavior dist
## 17   3312    Senile degenerat brain

# use a range with more than two hundred ICD-9 codes

length("390" %i9d% "392.1")

## [1] 244

"390" %i9d% "392.1" %>% icd9DecimalToShort() %>% icd9ExplainShort()

##      ICD-9      Diagnosis
## 1    390  Rheumatic fever without mention of heart involvement
## 2    3910      Acute rheumatic pericarditis
## 3    3911      Acute rheumatic endocarditis
## 4    3912      Acute rheumatic myocarditis
## 5    3918      Other acute rheumatic heart disease
## 6    3919      Acute rheumatic heart disease, unspecified
## 7    3920      Rheumatic chorea with heart involvement
##      Description
## 1  Rheum fev w/o hrt involv
## 2  Acute rheumatic pericard
## 3  Acute rheumatic endocard
## 4  Ac rheumatic myocarditis
## 5    Ac rheumat hrt dis NEC
## 6    Ac rheumat hrt dis NOS
## 7  Rheum chorea w hrt invol

```

```
"390" %i9d% "392.1" %>% icd9ExplainDecimal()

##      ICD-9                                     Diagnosis
## 1    390 Rheumatic fever without mention of heart involvement
## 2 391.0                                     Acute rheumatic pericarditis
## 3 391.1                                     Acute rheumatic endocarditis
## 4 391.2                                     Acute rheumatic myocarditis
## 5 391.8                                     Other acute rheumatic heart disease
## 6 391.9                                     Acute rheumatic heart disease, unspecified
## 7 392.0                                     Rheumatic chorea with heart involvement
##
##      Description
## 1 Rheum fev w/o hrt involv
## 2 Acute rheumatic pericard
## 3 Acute rheumatic endocard
## 4 Ac rheumatic myocarditis
## 5 Ac rheumat hrt dis NEC
## 6 Ac rheumat hrt dis NOS
## 7 Rheum chorea w hrt invol
```

## 7 AHRQ comorbidity classification

The AHRQ keeps an updated version of the Elixhauser classification of ICD-9-CM codes into comorbidities, useful for research. They provide the data in the form of SAS code. This package provides just enough code to parse the SAS source code provided by the AHRQ (but probably not much other SAS code), and generate a list of ICD-9 codes for each comorbidity.

```
ahrqComorbid <- parseAhrqSas()
```

```
head(summary(ahrqComorbid))
```

```
##      Length Class      Mode
## CHF      " 120" "-none-" "character"
## VALVE    " 545" "-none-" "character"
## PULMCIRC " 130" "-none-" "character"
## PERIVASC " 585" "-none-" "character"
## HTN      "  27" "-none-" "character"
## HTNCX    "  72" "-none-" "character"
```

Here are a couple of the shorter ICD-9 groups listed in ahrqComorbid:

SAS source code has a strong whiff of the 1970s about it. A fragment of a recent AHRQ SAS comorbidity mapping SAS FORMAT is as follows. Note the mix of character and numeric-only ranges, isolated values, all in 'short' ICD-9 code form.



```

PROC FORMAT LIB=library fmtlib;
VALUE $RCOMFMT
"2780 ",
"27800",
"27801",
"27803",
"64910"-"64914",
"V8530"-"V8539",
"V8541"-"V8545",
"V8554",
"79391"          = "OBESE"      /* Obesity      */

"3004 ",
"30112",
"3090 ",
"3091 ",
"311 "          = "DEPRESS"

```

This is parsed using:

```
parseAhrqSas()
```

resulting in a named list. here is an extract.

```

ahrqComorbid[c("OBESE", "DEPRESS")]

## $OBESE
## [1] "2780" "27800" "27801" "27802" "27803" "27804"
## [7] "27805" "27806" "27807" "27808" "27809" "V8554"
## [13] "79391" "64910" "64911" "64912" "64913" "64914"
## [19] "V8530" "V8531" "V8532" "V8533" "V8534" "V8535"
## [25] "V8536" "V8537" "V8538" "V8539" "V8541" "V8542"
## [31] "V8543" "V8544" "V8545"
##
## $DEPRESS
## [1] "3004" "30040" "30041" "30042" "30043" "30044"
## [7] "30045" "30046" "30047" "30048" "30049" "30112"
## [13] "3090" "30900" "30901" "30902" "30903" "30904"
## [19] "30905" "30906" "30907" "30908" "30909" "3091"
## [25] "30910" "30911" "30912" "30913" "30914" "30915"
## [31] "30916" "30917" "30918" "30919" "311" "3110"
## [37] "3111" "3112" "3113" "3114" "3115" "3116"
## [43] "3117" "3118" "3119" "31100" "31110" "31120"
## [49] "31130" "31140" "31150" "31160" "31170" "31180"
## [55] "31190" "31101" "31111" "31121" "31131" "31141"
## [61] "31151" "31161" "31171" "31181" "31191" "31102"

```

```

## [67] "31112" "31122" "31132" "31142" "31152" "31162"
## [73] "31172" "31182" "31192" "31103" "31113" "31123"
## [79] "31133" "31143" "31153" "31163" "31173" "31183"
## [85] "31193" "31104" "31114" "31124" "31134" "31144"
## [91] "31154" "31164" "31174" "31184" "31194" "31105"
## [97] "31115" "31125" "31135" "31145" "31155" "31165"
## [103] "31175" "31185" "31195" "31106" "31116" "31126"
## [109] "31136" "31146" "31156" "31166" "31176" "31186"
## [115] "31196" "31107" "31117" "31127" "31137" "31147"
## [121] "31157" "31167" "31177" "31187" "31197" "31108"
## [127] "31118" "31128" "31138" "31148" "31158" "31168"
## [133] "31178" "31188" "31198" "31109" "31119" "31129"
## [139] "31139" "31149" "31159" "31169" "31179" "31189"
## [145] "31199"

lapply(ahrqComorbid[c("OBESE", "DEPRESS")], icd9ChildrenShort)

## $OBESE
## [1] "2780" "27800" "27801" "27802" "27803" "27804"
## [7] "27805" "27806" "27807" "27808" "27809" "V8554"
## [13] "79391" "64910" "64911" "64912" "64913" "64914"
## [19] "V8530" "V8531" "V8532" "V8533" "V8534" "V8535"
## [25] "V8536" "V8537" "V8538" "V8539" "V8541" "V8542"
## [31] "V8543" "V8544" "V8545"
##
## $DEPRESS
## [1] "3004" "30040" "30041" "30042" "30043" "30044"
## [7] "30045" "30046" "30047" "30048" "30049" "30112"
## [13] "3090" "30900" "30901" "30902" "30903" "30904"
## [19] "30905" "30906" "30907" "30908" "30909" "3091"
## [25] "30910" "30911" "30912" "30913" "30914" "30915"
## [31] "30916" "30917" "30918" "30919" "311" "3110"
## [37] "3111" "3112" "3113" "3114" "3115" "3116"
## [43] "3117" "3118" "3119" "31100" "31110" "31120"
## [49] "31130" "31140" "31150" "31160" "31170" "31180"
## [55] "31190" "31101" "31111" "31121" "31131" "31141"
## [61] "31151" "31161" "31171" "31181" "31191" "31102"
## [67] "31112" "31122" "31132" "31142" "31152" "31162"
## [73] "31172" "31182" "31192" "31103" "31113" "31123"
## [79] "31133" "31143" "31153" "31163" "31173" "31183"
## [85] "31193" "31104" "31114" "31124" "31134" "31144"
## [91] "31154" "31164" "31174" "31184" "31194" "31105"
## [97] "31115" "31125" "31135" "31145" "31155" "31165"
## [103] "31175" "31185" "31195" "31106" "31116" "31126"
## [109] "31136" "31146" "31156" "31166" "31176" "31186"
## [115] "31196" "31107" "31117" "31127" "31137" "31147"

```

```
## [121] "31157" "31167" "31177" "31187" "31197" "31108"
## [127] "31118" "31128" "31138" "31148" "31158" "31168"
## [133] "31178" "31188" "31198" "31109" "31119" "31129"
## [139] "31139" "31149" "31159" "31169" "31179" "31189"
## [145] "31199"
```

## 8 Elixhauser co-morbidities

Elixhauser originally developed this set of co-morbidities to predict long term mortality based on hospital ICD-9-CM coding records. The AHRQ comorbidities are an updated version of this, however the original Elixhauser have been used in many publications. The ICD-9-CM codes have changed slightly over the years.

```
names(elixhauserComorbid)

## [1] "chf"          "arrhythmia"   "valve"
## [4] "pulm.circ"    "pvd"          "htn"
## [7] "htncx"        "paralysis"    "neuro.other"
## [10] "chronic.pulm" "dm.uncomp"    "dm.comp"
## [13] "hypothyroid"  "renal"         "liver"
## [16] "pud"          "hiv"          "lymphoma"
## [19] "mets"         "solid.tumor"  "rheum"
## [22] "coag"         "obesity"      "wt.loss"
## [25] "lytes"        "anemia.loss"  "anemia.def"
## [28] "etoh"         "drugs"        "psychoses"
## [31] "depression"
```

## 9 Quan

Quan's paper looked at indices using both ICD-10 and ICD-9-CM. Quan generated updated ICD-9-CM codes for all 30 of Elixhauser and all 17 of Charlson/Deyo's co-morbidities. Thus there are two 'Quan' comorbidity mappings.

```
names(quanDeyoComorbid)

## [1] "Myocardial Infarction"
## [2] "Congestive Heart Failure"
## [3] "Periphral Vascular Disease"
## [4] "Cerebrovascular Disease"
## [5] "Dementia"
```

```
## [6] "Chronic Pulmonary Disease"
## [7] "Connective Tissue Disease-Rheumatic Disease"
## [8] "Peptic Ulcer Disease"
## [9] "Mild Liver Disease"
## [10] "Diabetes without complications"
## [11] "Diabetes with complications"
## [12] "Paraplegia and Hemiplegia"
## [13] "Renal Disease"
## [14] "Cancer"
## [15] "Moderate or Severe Liver Disease"
## [16] "Metastatic Carcinoma"
## [17] "AIDS/HIV"
```

```
names(quanElixhauserComorbid)
```

```
## [1] "chf"           "arrhythmia"   "valve"
## [4] "pulm.circ"     "pvd"          "htn"
## [7] "htncx"         "paralysis"    "neuro.other"
## [10] "chronic.pulm"  "dm.uncomp"    "dm.comp"
## [13] "hypothyroid"   "renal"         "liver"
## [16] "pud"           "hiv"          "lymphoma"
## [19] "mets"          "solid.tumor"  "rheum"
## [22] "coag"          "obesity"      "wt.loss"
## [25] "lytes"         "anemia.loss"  "anemia.def"
## [28] "etoh"          "drugs"         "psychoses"
## [31] "depression"
```