

KDB抽出ファイル活用のためのマニュアル作成

岡本 悦司 国立保健医療科学院 国際保健支援研究分野 統括研究官

研究要旨

全国の市町村に配備された国保データベース（KDB）は、国保、後期高齢者医療制度の医療レセプトにくわえて介護保険レセプトならびに特定健康診査・保健指導データも含む、地域包括ケア推進の上できわめて有用な情報源として期待される。KDBは必要なデータをcsv形式で抽出できるようになっており、市町村が抽出データを分析・加工してケアカンファレンス等ですることが求められる。しかしクロス表形式のcsvファイルはそのままでは活用しにくく、それをキューブ化してExcelのピボットテーブルを用いて柔軟に分析するためのマニュアルを開発した。また、地域保健・健康増進報告や医療給付実態調査等の公開データを用いて使用法を実演した。

A．研究目的

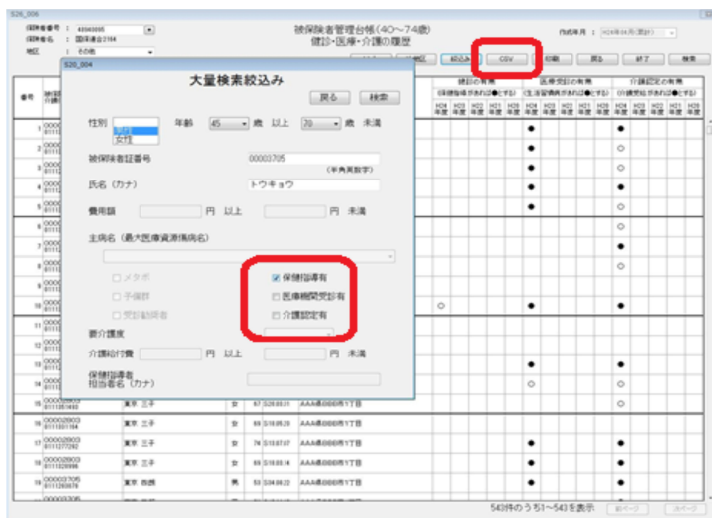
KDBを地域包括ケアやケアカンファレンス等で有効に活用するために、KDBより抽出されたcsvファイルExcelのピボットテーブル上で活用するためのマニュアルを開発する。

B．研究方法

csvファイルをキューブ形式（Excelピボットテーブルで活用できる形式）に変換する方法を開発しe-STAT上で公開されているcsvファイル（地域保健・健康増進事業報告ならびに医療給付実態調査）を実例として使用法を示す。

KDBのcsvファイル抽出機能

KDBの有用な機能として、必要なデータをcsvとして取り出す機能がある。抽出されたcsvファイルを効果的に活用するためのマニュアルを作成した。

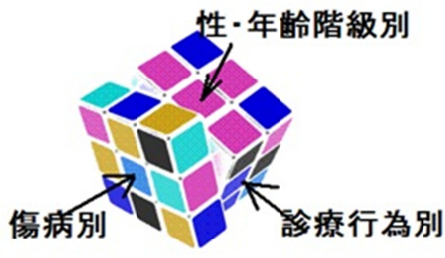


キューブ形式とは

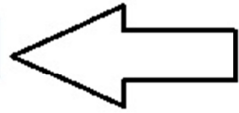
下記のように、行見出しが県、列見出しが性と調査年のような統計表を、県、性、調査年そしてデータとしてタテ列に並べた形式をキューブ形式と呼ぶ。キューブ形式ならExcelのピボットテーブル機能を使って、県、性そして調査年という属性を自由に入れ換えて表示できる。ちょうどルービックキューブのように、あらゆる次元に自由に回転できるのでキューブ形式と呼ばれている。たとえば医療費デ

ータなら、性・年齢階級、傷病そして診療行為といった次元で自由に作表できる。

| | 男 | | | | 女 | | | | 県名 | 年 | 性 | データ |
|----|------|------|------|------|------|------|------|------|----|------|---|-----|
| | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 | | | | |
| A県 | 55 | 88 | 99 | 10 | 59 | 46 | 91 | 11 | A県 | 2008 | 男 | 55 |
| B県 | 24 | 29 | 12 | 92 | 13 | 70 | 40 | 94 | B県 | 2008 | 男 | 24 |
| C県 | 85 | 8 | 95 | 29 | 34 | 80 | 18 | 53 | C県 | 2008 | 男 | 85 |
| D県 | 35 | 1 | 0 | 49 | 10 | 86 | 28 | 85 | D県 | 2008 | 男 | 35 |
| E県 | 14 | 82 | 1 | 69 | 34 | 17 | 91 | 16 | E県 | 2008 | 男 | 14 |
| | | | | | A県 | 2009 | 男 | 88 | | | | |
| | | | | | B県 | 2009 | 男 | 29 | | | | |



| 県名 | C県 | | |
|---------|-----|-----|-----|
| 合計 / 列ラ | 女 | 男 | 総計 |
| 2008 | 34 | 85 | 119 |
| 2009 | 80 | 8 | 88 |
| 2010 | 18 | 95 | 113 |
| 2011 | 53 | 29 | 82 |
| 総計 | 185 | 217 | 402 |



レポートに追加するフィールドを選択してください:

- 県名
- 年
- 性
- データ

次のボックス間でフィールドをドラッグしてください:

レポートフィールド: 県名, 年, 性, データ

列ラベル: 性

行ラベル: 年

Σ 値: 合計 / データ

C. 研究結果

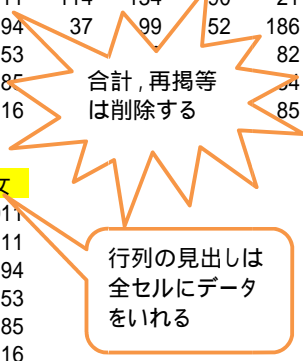
キューブ化の方法

統計表をキューブに加工する方法を説明する。

- 1) 前準備として、行列の全ての項目が MECE(mutually exclusive collectively exhaustive, 重なりがなく合計すると総数となること)になるように「合計」「総計」「再掲」等は削除する。たとえば男、女、計とあったら計を削除する。また行列の項目が全て埋まるように

| | 男 | | | | 女 | | | | 計 | | | |
|-----|------|------|------|------|------|------|------|------|-------------|------|------|------|
| | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 |
| 病院A | 55 | 88 | 99 | 10 | 59 | 46 | 91 | 11 | 114 | 134 | 90 | 21 |
| 病院B | 24 | 29 | 12 | 92 | 13 | 70 | 40 | 94 | 37 | 99 | 52 | 186 |
| 病院C | 85 | 8 | 95 | 29 | 34 | 80 | 18 | 53 | 合計、再掲等は削除する | | | |
| 病院D | 35 | 1 | 0 | 49 | 10 | 86 | 28 | 85 | | | | |
| 病院E | 14 | 82 | 1 | 69 | 34 | 17 | 91 | 16 | | | | |

| | 男 | 男 | 男 | 男 | 女 | 女 | 女 | 女 |
|-----|------|------|------|------|------|------|------|------|
| | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 |
| 病院A | 55 | 88 | 99 | 10 | 59 | 46 | 91 | 11 |
| 病院B | 24 | 29 | 12 | 92 | 13 | 70 | 40 | 94 |
| 病院C | 85 | 8 | 95 | 29 | 34 | 80 | 18 | 53 |
| 病院D | 35 | 1 | 0 | 49 | 10 | 86 | 28 | 85 |
| 病院E | 14 | 82 | 1 | 69 | 34 | 17 | 91 | 16 |



- 2) 表の最上行と最左列に通番を挿入する。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|------|------|------|------|------|------|------|------|
| | 男 | 男 | 男 | 男 | 女 | 女 | 女 | 女 |
| | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 |
| 1 病院A | 55 | 88 | 99 | 10 | 59 | 46 | 91 | 11 |
| 2 病院B | 24 | 29 | 12 | 92 | 13 | 70 | 40 | 94 |
| 3 病院C | 85 | 8 | 95 | 29 | 34 | 80 | 18 | 53 |
| 4 病院D | 35 | 1 | 0 | 49 | 10 | 86 | 28 | 85 |
| 5 病院E | 14 | 82 | 1 | 69 | 34 | 17 | 91 | 16 |

- 3) 下図の黄, 緑, 青の部分にそれぞれ「列見出し」「行見出し」そしてデータを抽出したい「本体」と名前をつける。たとえば黄色い部分をハイライトし, 「数式」 「名前の定義」をクリックし下のよ
うに名前に「列見出し」といれてOKする。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|------|------|------|------|------|------|------|------|
| | 男 | 男 | 男 | 男 | 女 | 女 | 女 | 女 |
| | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 |
| 1 病院A | 55 | 88 | 99 | 10 | 59 | 46 | 91 | 11 |
| 2 病院B | 24 | 29 | 12 | 92 | 13 | 70 | 40 | 94 |
| 3 病院C | 85 | 8 | 95 | 29 | 34 | 80 | 18 | 53 |
| 4 病院D | 35 | 1 | 0 | 49 | 10 | 86 | 28 | 85 |
| 5 病院E | 14 | 82 | 1 | 69 | 34 | 17 | 91 | 16 |

新しい名前

名前(N): 列見出し

範囲(S): ブック

コメント(Q):

参照範囲(R): =Sheet1!\$C\$17:\$J\$19

OK キャンセル

- 3) Shift+F11 で新しいワークシートを開き, 上の例では「本体」部分は 5 行×8 列だから A,B 列に本
体の行と列番号を挿入する。すなわち A 列に 1,2,3,4,5 を 8 回, B 列に 1,1,1,1,1,2,2,2,2,2,3,3,3...と
それぞれ 40 の数値を挿入する。このような単純作業を Excel は「データ分析」というアドインプログラ
ムの「乱数発生」機能で行うことができる。「データ」 「データ分析」 「乱数発生」で分布=「パ
ターン」指定し A 列では「1 から 5 までステップ 1 ずつの数値を 8 回繰り返す」B 列では「1 から 8 ま
でステップ 1 ずつ数値を 5 回繰り返す」として生成する。(正しく生成されたら 5X8=40 行できている
はず。要確認)。

乱数発生

変数の数(V): 1

乱数の数(E): 1

分布(D): パターン

パラメータ
1 から(E) 5 まで(I) ステップ(P) 1

数値の繰り返し回数(E) 1

数列の繰り返し回数(Q) 8

ランダム シード(R):

出力オプション
 出力先(O): \$A\$1

乱数発生

変数の数(V): 1

乱数の数(E): 1

分布(D): パターン

パラメータ
1 から(E) 8 まで(I) ステップ(P) 1

数値の繰り返し回数(E) 5

数列の繰り返し回数(Q) 1

ランダム シード(R):

出力オプション
 出力先(O): \$B\$1

- 4) 新しいワークシートには AB 列に行列番号が挿入されている。下のように VLOOKUP, HLOOKUP,
INDEX 関数を挿入する。これを 40 行コピーすると一気に行列の見出しとデータが作成される(この
例では列見出しは性と年と二つある点に注意)。

| | A | B | C | D | E | F |
|---|---|---|-----------------------|-----------------------|-----------------------|------------------|
| 1 | 1 | 1 | =VLOOKUP(A1,列見出し,2,0) | =HLOOKUP(B1,列見出し,2,0) | =HLOOKUP(B1,列見出し,3,0) | =INDEX(本体,A1,B1) |
| 2 | 2 | 1 | | | | |

- 5) 必ず値複写(コピー 形式を選択して複写 値)して一番上に行を挿入し,見出しをつける。ピボット
テーブルでは一番上の行に見出しが必要である。

| 行番号 | 列番号 | 病院名 | 性 | 年 | データ |
|-----|-----|-----|---|------|-----|
| 1 | 1 | 病院A | 男 | 2008 | 55 |
| 2 | 1 | 病院B | 男 | 2008 | 24 |
| 3 | 1 | 病院C | 男 | 2008 | 85 |
| 4 | 1 | 病院D | 男 | 2008 | 35 |
| 5 | 1 | 病院E | 男 | 2008 | 14 |

6) 「挿入」 「ピボットテーブル」とすればできあがり。

The screenshot shows an Excel spreadsheet with a PivotTable. The PivotTable is structured as follows:

| 行ラベル | 女 集計 | | | | 男 集計 | | | | 男 集計 | 総計 | |
|------|------|------|------|------|------|------|------|------|------|-----|------|
| | 2008 | 2009 | 2010 | 2011 | 2008 | 2009 | 2010 | 2011 | | | |
| 病院A | 59 | 46 | 91 | 11 | 207 | 55 | 88 | 99 | 10 | 252 | 459 |
| 病院B | 13 | 70 | 40 | 94 | 217 | 24 | 29 | 12 | 92 | 157 | 374 |
| 病院C | 34 | 80 | 18 | 53 | 185 | 85 | 8 | 95 | 29 | 217 | 402 |
| 病院D | 10 | 86 | 28 | 85 | 209 | 35 | 1 | 0 | 49 | 85 | 294 |
| 病院E | 34 | 17 | 91 | 16 | 158 | 14 | 82 | 1 | 69 | 166 | 324 |
| 総計 | 150 | 299 | 268 | 259 | 976 | 213 | 208 | 207 | 249 | 877 | 1853 |

The PivotTable Field List on the right shows the following configuration:

- レポートに追加するフィールドを選択してください:
- 行番号
- 列番号
- 病院名
- 性
- 年
- データ

Field Settings for '性' (Gender):

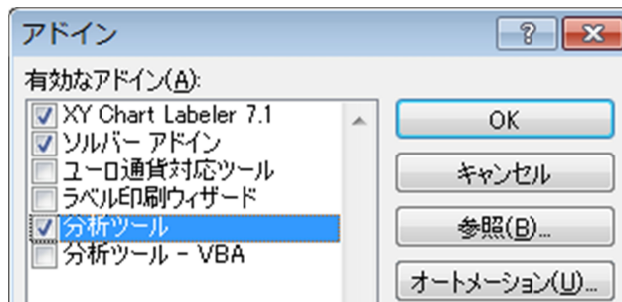
- レポート フィールド: 性
- 列ラベル: 年

Field Settings for '年' (Year):

- 行ラベル: 年
- Σ 値: 合計 / データ

Excel のアドインプログラム

Excel はインストールしただけでは使えない隠し機能がある。アドインと呼ばれる上級者向けプログラムがそれだ。使うためには「ファイル」「オプション」「アドイン」「設定」でチェックボックスにチェックを入れる。とりわけ「データ分析」アドインは、多変量解析等豊富な分析機能を含んでいるのでインストール直後に追加しておくことをおすすめする。



アドインが追加されたら「データ」タブの右端に「データ分析」というアイコンが表示される(下図では他にソルバーというアドインも追加されている)。



実例

キューブ化による効果を平成 25 年度地域保健・健康増進事業報告の表 16-2.csv を例に説明する。
ダウンロードした csv ファイルをそのまま Excel で開くと下記のように表示される。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|----|---|---------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|--------|
| 1 | 平成25年度 | 地域保健・健康増進事業報告(健康増進編)市区町村表 | | | | | | | | | | | | |
| 2 | 閲覧 (健康増進編 第16-2表 胃がん検診受診者数, 市区町村, 検診回数・検診方式・年齢階級別(男)) | | | | | | | | | | | | | |
| 3 | 注: 検診回数の初回・非初回については、計数不詳の市区町村があるため、総数と一致しない場合がある。 | | | | | | | | | | | | | |
| 4 | | 総数 | | | | | | | | | | | | |
| 5 | | 総数 | | | | | | | | | | 集団検診 | | |
| 6 | | 総数 | 40~44歳 | 45~49歳 | 50~54歳 | 55~59歳 | 60~64歳 | 65~69歳 | 70~74歳 | 75~79歳 | 80歳以上 | 総数 | 40~44歳 | 45~49歳 |
| 7 | 00 全国 | 1572882 | 86882 | 71908 | 78094 | 96191 | 225234 | 340093 | 333772 | 213781 | 126927 | 1135151 | 61517 | 51570 |
| 8 | 01 北海道 | 68574 | 3718 | 3371 | 4028 | 4921 | 10406 | 13823 | 13860 | 9196 | 5251 | 56411 | 2872 | 2700 |
| 9 | 01100札幌市 | 13978 | 608 | 550 | 537 | 658 | 1893 | 3253 | 3243 | 2047 | 1189 | 8836 | 305 | 299 |
| 10 | 01202函館市 | 1487 | 91 | 82 | 107 | 114 | 280 | 310 | 260 | 189 | 54 | 1084 | 67 | 56 |
| 11 | 01203小樽市 | 724 | 40 | 34 | 31 | 30 | 102 | 150 | 163 | 112 | 62 | 724 | 40 | 34 |
| 12 | 01204旭川市 | 3388 | 122 | 134 | 102 | 176 | 579 | 809 | 719 | 470 | 277 | 3388 | 122 | 134 |
| 13 | 01205室蘭市 | 138 | 3 | 3 | 3 | 2 | 17 | 20 | 41 | 30 | 19 | - | - | - |
| 14 | 01206釧路市 | 2380 | 189 | 135 | 133 | 169 | 337 | 423 | 479 | 324 | 191 | 1357 | 67 | 73 |
| 15 | 01207帯広市 | 2153 | 98 | 75 | 105 | 123 | 270 | 433 | 465 | 357 | 227 | 2153 | 98 | 75 |
| 16 | 01208北見市 | 1570 | 74 | 59 | 89 | 81 | 286 | 310 | 289 | 231 | 151 | 1570 | 74 | 59 |

列見出しの挿入

キューブ化にあたっては全ての列に見出しが入っていないなければならない。そこで下記のように空白に左の見出しを挿入する。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|----|---|---------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|--------|--------|
| 1 | 平成25年度 | 地域保健・健康増進事業報告(健康増進編)市区町村表 | | | | | | | | | | | | | |
| 2 | 閲覧 (健康増進編 第16-2表 胃がん検診受診者数, 市区町村, 検診回数・検診方式・年齢階級別(男)) | | | | | | | | | | | | | | |
| 3 | 注: 検診回数の初回・非初回については、計数不詳の市区町村があるため、総数と一致しない場合がある。 | | | | | | | | | | | | | | |
| 4 | | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 |
| 5 | | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 総数 | 集団検診 | 集団検診 | 集団検診 | 集団検診 |
| 6 | | 総数 | 40~44歳 | 45~49歳 | 50~54歳 | 55~59歳 | 60~64歳 | 65~69歳 | 70~74歳 | 75~79歳 | 80歳以上 | 総数 | 40~44歳 | 45~49歳 | 50~54歳 |
| 7 | 00 全国 | 1572882 | 86882 | 71908 | 78094 | 96191 | 225234 | 340093 | 333772 | 213781 | 126927 | 1135151 | 61517 | 51570 | 57327 |
| 8 | 01 北海道 | 68574 | 3718 | 3371 | 4028 | 4921 | 10406 | 13823 | 13860 | 9196 | 5251 | 56411 | 2872 | 2700 | 3219 |
| 9 | 01100札幌市 | 13978 | 608 | 550 | 537 | 658 | 1893 | 3253 | 3243 | 2047 | 1189 | 8836 | 305 | 299 | 276 |
| 10 | 01202函館市 | 1487 | 91 | 82 | 107 | 114 | 280 | 310 | 260 | 189 | 54 | 1084 | 67 | 56 | 78 |
| 11 | 01203小樽市 | 724 | 40 | 34 | 31 | 30 | 102 | 150 | 163 | 112 | 62 | 724 | 40 | 34 | 31 |
| 12 | 01204旭川市 | 3388 | 122 | 134 | 102 | 176 | 579 | 809 | 719 | 470 | 277 | 3388 | 122 | 134 | 102 |
| 13 | 01205室蘭市 | 138 | 3 | 3 | 3 | 2 | 17 | 20 | 41 | 30 | 19 | - | - | - | - |

MECE 化

キューブ化にあたっては全ての合計が一致しなければならない。そこで「総数」は削除し、「再掲」は「総計」から引き算して Mutually exclusive collectively exhaustive (MECE) 化を行う。この例の場合、検診は集団検診か個別検診か、受診回数は初回か非初回か、年齢階級は 10 歳きざみなので「総数」とある例は全て削除する。

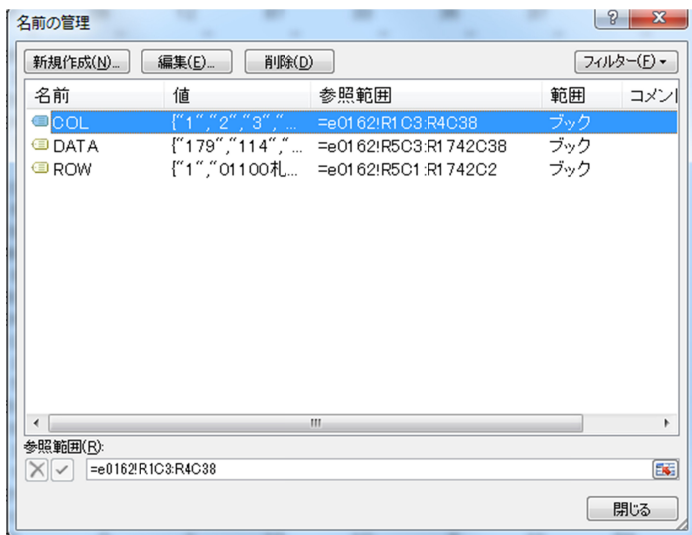
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|---|--------|--------|--------|--------|--------|--------|--------|--------|-------|--------|--------|--------|
| 1 | 平成25年度 | | | | | | | | | | | | |
| 2 | 閲覧 (健康増進編) 市区町村表 | | | | | | | | | | | | |
| 3 | 注: 検診回数の初回・非初回については、計数不詳の市区町村があるため、総数と一致しない場合がある。 | | | | | | | | | | | | |
| 4 | | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 |
| 5 | | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 個別検診 | 個別検診 |
| 6 | | 40~44歳 | 45~49歳 | 50~54歳 | 55~59歳 | 60~64歳 | 65~69歳 | 70~74歳 | 75~79歳 | 80歳以上 | 40~44歳 | 45~49歳 | 50~54歳 |
| 7 | 00 全国 | 29137 | 14930 | 14747 | 16814 | 43924 | 47331 | 28129 | 14489 | 7865 | 12376 | 6657 | 62 |
| 8 | 01 北海道 | 1130 | 668 | 707 | 864 | 2233 | 2357 | 1663 | 878 | 482 | 261 | 117 | 11 |
| 9 | 01100札幌市 | 179 | 114 | 112 | 127 | 426 | 618 | 426 | 182 | 104 | - | - | - |
| 10 | 01202函館市 | 26 | 12 | 19 | 19 | 70 | 64 | 53 | 34 | 4 | 13 | 10 | - |
| 11 | 01203小樽市 | 15 | 3 | 10 | 1 | 18 | 32 | 16 | 12 | 5 | - | - | - |
| 12 | 01204旭川市 | 51 | 43 | 25 | 51 | 246 | 249 | 104 | 52 | 31 | - | - | - |

同様に左端列の全国や都道府県も削除し、最上行と最左列に通番を挿入する。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|----|-----------|--------|--------|--------|--------|--------|--------|--------|--------|-------|--------|--------|
| 1 | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| 2 | | | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 | 初回 |
| 3 | | | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 集団検診 | 個別検診 | 個別検診 |
| 4 | | | 40~44歳 | 45~49歳 | 50~54歳 | 55~59歳 | 60~64歳 | 65~69歳 | 70~74歳 | 75~79歳 | 80歳以上 | 40~44歳 | 45~49歳 |
| 5 | 1 | 01100札幌市 | 179 | 114 | 112 | 127 | 426 | 618 | 426 | 182 | 104 | - | - |
| 6 | 2 | 01202函館市 | 26 | 12 | 19 | 19 | 70 | 64 | 53 | 34 | 4 | 13 | |
| 7 | 3 | 01203小樽市 | 15 | 3 | 10 | 1 | 18 | 32 | 16 | 12 | 5 | - | - |
| 8 | 4 | 01204旭川市 | 51 | 43 | 25 | 51 | 246 | 249 | 104 | 52 | 31 | - | - |
| 9 | 5 | 01205室蘭市 | - | - | - | - | - | - | - | - | - | - | 2 |
| 10 | 6 | 01206釧路市 | 20 | 15 | 5 | 7 | 29 | 31 | 31 | 29 | 12 | 87 | : |
| 11 | 7 | 01207帯広市 | - | - | - | - | - | - | - | - | - | - | - |
| 12 | 8 | 01208北見市 | 41 | 17 | 25 | 20 | 109 | 47 | 35 | 18 | 9 | - | - |
| 13 | 9 | 01209夕張市 | 1 | 2 | 4 | 5 | 4 | 3 | 3 | 5 | 2 | - | - |
| 14 | 10 | 01210岩見沢市 | 34 | 16 | 10 | 20 | 54 | 42 | 29 | 18 | 7 | - | - |
| 15 | 11 | 01211網走市 | 14 | 8 | 7 | 8 | 20 | 25 | 24 | 9 | 6 | - | - |

COL, ROW そして DATA 範囲の命名

見出しとデータ本体にそれぞれ ROW, COL, DATA の名前をつける。数式 名前管理で下記のようにになっていることを確認する。



3) 乱数発生により, 1738 行 × 36 列の乱数を発生させる。

4) 下記のように vlookup, hlookup そして index 関数を挿入

| | A | B | C | D | E | F |
|---|---|---|--------------------|--------------------|------------------|--------------------|
| 1 | 1 | 1 | =+VLOOKUP(\$A1,ROW | =+VLOOKUP(\$A1,ROW | =+HLOOKUP(B1,COL | =+INDEX(DATA,A1,B1 |
| 2 | 2 | 1 | | | | |

5) 値複製した上で最上行を挿入し, 見出しをつける。

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----|----|---|-----------|-------|---------|--------|------|
| 1 | R | C | CITY | KAISU | KENSHIN | AG | DATA |
| 2 | 1 | 1 | 01100札幌市 | 初回 | 集団検診 | 40~44歳 | 179 |
| 3 | 2 | 1 | 01202函館市 | 初回 | 集団検診 | 40~44歳 | 26 |
| 4 | 3 | 1 | 01203小樽市 | 初回 | 集団検診 | 40~44歳 | 15 |
| 5 | 4 | 1 | 01204旭川市 | 初回 | 集団検診 | 40~44歳 | 51 |
| 6 | 5 | 1 | 01205室蘭市 | 初回 | 集団検診 | 40~44歳 | - |
| 7 | 6 | 1 | 01206釧路市 | 初回 | 集団検診 | 40~44歳 | 20 |
| 8 | 7 | 1 | 01207帯広市 | 初回 | 集団検診 | 40~44歳 | - |
| 9 | 8 | 1 | 01208北見市 | 初回 | 集団検診 | 40~44歳 | 41 |
| 10 | 9 | 1 | 01209夕張市 | 初回 | 集団検診 | 40~44歳 | 1 |
| 11 | 10 | 1 | 01210岩見沢市 | 初回 | 集団検診 | 40~44歳 | 34 |
| 12 | 11 | 1 | 01211網走市 | 初回 | 集団検診 | 40~44歳 | 14 |

7) ピボットテーブルを挿入する

| CITY | | 22206三島市 | | | | | | | |
|-----------|------------|-------------|--|-------------|-----------|------------|---------|------------|-------------|
| 合計 / DATA | 列ラベル | 個別検診 | | 個別検診 集計 | 集団検診 | | 集団検診 集計 | 総計 | |
| 行ラベル | 初回 | 非初回 | | | 初回 | 非初回 | | | |
| 40～44歳 | 90 | 69 | | 159 | 11 | 4 | | 15 | 174 |
| 45～49歳 | 61 | 113 | | 174 | 5 | 4 | | 9 | 183 |
| 50～54歳 | 61 | 138 | | 199 | 2 | 7 | | 9 | 208 |
| 55～59歳 | 58 | 151 | | 209 | 6 | 6 | | 12 | 221 |
| 60～64歳 | 133 | 279 | | 412 | 13 | 17 | | 30 | 442 |
| 65～69歳 | 134 | 597 | | 731 | 25 | 42 | | 67 | 798 |
| 70～74歳 | 99 | 790 | | 889 | 4 | 59 | | 63 | 952 |
| 75～79歳 | 79 | 652 | | 731 | 5 | 38 | | 43 | 774 |
| 80歳以上 | 36 | 556 | | 592 | 3 | 20 | | 23 | 615 |
| 総計 | 751 | 3345 | | 4096 | 74 | 197 | | 271 | 4367 |

傷病別データのキューブ化

KDB では傷病分類は ICD10 の中分類が提供されている。この他，KDB 独自の細小分類も提供されているが，これは ICD10 ではなく KDB 独自の分類なので ICD10 との整合性は無い(たとえば中分類ではウイルス肝炎とあるのが，A,B,C 型肝炎に分類されている)。ここでは中分類のみを扱う。中分類と大分類とは以下のように中分類が大分類に包含される関係になっているので，ドリルダウン，ドリルアップといった OLAP 操作が可能になる。

| BUNRUIMEI | CHUBUNRUI | DAIBUNRUI |
|--------------------|-----------------------|-----------|
| 腸管感染症 | 001腸管感染症 | 感染症及び寄生虫症 |
| 結核 | 002結核 | 感染症及び寄生虫症 |
| 主として性的伝播様式をとる感染症 | 003主として性的伝播様式をとる感染症 | 感染症及び寄生虫症 |
| 皮膚及び粘膜の病変を伴うウイルス疾患 | 004皮膚及び粘膜の病変を伴うウイルス疾患 | 感染症及び寄生虫症 |
| ウイルス肝炎 | 005ウイルス肝炎 | 感染症及び寄生虫症 |
| その他のウイルス疾患 | 006その他のウイルス疾患 | 感染症及び寄生虫症 |
| 真菌症 | 007真菌症 | 感染症及び寄生虫症 |
| 感染症及び寄生虫症の続発・後遺 | 008感染症及び寄生虫症の続発・後遺症 | 感染症及び寄生虫症 |
| その他の感染症及び寄生虫症 | 009その他の感染症及び寄生虫症 | 感染症及び寄生虫症 |
| 胃の悪性新生物 | 010胃の悪性新生物 | 新生物 |
| 結腸の悪性新生物 | 011結腸の悪性新生物 | 新生物 |
| 直腸S状結腸移行部及び直腸の悪 | 012直腸S状結腸移行部及び直腸の悪性新 | 新生物 |
| 肝及び肝内胆管の悪性新生物 | 013肝及び肝内胆管の悪性新生物 | 新生物 |
| 気管、気管支及び肺の悪性新生物 | 014気管、気管支及び肺の悪性新生物 | 新生物 |
| 乳房の悪性新生物 | 015乳房の悪性新生物 | 新生物 |
| 子宮の悪性新生物 | 016子宮の悪性新生物 | 新生物 |
| 悪性リンパ腫 | 017悪性リンパ腫 | 新生物 |
| 白血病 | 018白血病 | 新生物 |
| その他の悪性新生物 | 019その他の悪性新生物 | 新生物 |
| 良性新生物及びその他の新生物 | 020良性新生物及びその他の新生物 | 新生物 |

傷病別のドリルダウン，ドリルアップの実際を平成 24 年度医療給付実態調査のデータを用いて説明する。点数や件数といったデータ(メジャー)を大中分類別に示すためには，下記のように行ラベルに大分類を上，中分類を下に配置することによってドリルダウン，ドリルアップが可能となる。

ピボットテーブルの表示は以下の通りである(中分類は各大分類の下に折り畳まれているため「+」のマークが表示されている)。なお以下は，2012 年度医療給付実態調査の国保分の外来件数である。

| データ型 | 件数 | | |
|---|---|-----------------------------|------------------|
| 入外 | 外来 | | |
| 年 | 2012 | | |
| 合計 / データ | | | |
| | 列ラベル | | 総計 |
| | <input type="checkbox"/> 国保 | <input type="checkbox"/> 組合 | |
| | <input checked="" type="checkbox"/> グループ1 | <input type="checkbox"/> 組合 | |
| 行ラベル | | 組合 | |
| <input checked="" type="checkbox"/> 01 感染症及び寄生虫症 | 8584535 | 801177 | 9385712 |
| <input checked="" type="checkbox"/> 02 新生物 | 10649205 | 644785 | 11293990 |
| <input checked="" type="checkbox"/> 03 血液及び造血器の疾患免疫機構の障害 | 1052490 | 89581 | 1142071 |
| <input checked="" type="checkbox"/> 04 内分泌、栄養及び代謝疾患 | 32439844 | 1538447 | 33978291 |
| <input checked="" type="checkbox"/> 05 精神及び行動の障害 | 14256316 | 594962 | 14851278 |
| <input checked="" type="checkbox"/> 06 神経系の疾患 | 7949601 | 405227 | 8354828 |
| <input checked="" type="checkbox"/> 07 眼及び付属器の疾患 | 23650662 | 1615662 | 25266324 |
| <input checked="" type="checkbox"/> 08 耳及び乳様突起の疾患 | 4137875 | 321374 | 4459249 |
| <input checked="" type="checkbox"/> 09 循環器系の疾患 | 58317219 | 2547446 | 60864665 |
| <input checked="" type="checkbox"/> 10 呼吸器系の疾患 | 29720758 | 3512145 | 33232903 |
| <input checked="" type="checkbox"/> 11 消化器系の疾患 | 16336890 | 1016516 | 17353406 |
| <input checked="" type="checkbox"/> 12 皮膚及び皮下組織の疾患 | 14166683 | 1389899 | 15556582 |
| <input checked="" type="checkbox"/> 13 筋骨格系及び結合組織の疾患 | 27777661 | 1420815 | 29198476 |
| <input checked="" type="checkbox"/> 14 腎尿路生殖器系の疾患 | 9056729 | 619594 | 9676323 |
| <input checked="" type="checkbox"/> 15 妊娠、分娩及び産じょく<褥> | 363410 | 59632 | 423042 |
| <input checked="" type="checkbox"/> 16 周産期に発生した病態 | 115594 | 18970 | 134564 |
| <input checked="" type="checkbox"/> 17 先天奇形、変形及び染色体異常 | 529263 | 56005 | 585268 |
| <input checked="" type="checkbox"/> 18 症状、徴候及び異常臨床所見・異常検査所見 | 4383012 | 314119 | 4697131 |
| <input checked="" type="checkbox"/> 19 損傷、中毒及びその他の外因の影響 | 6678020 | 586724 | 7264744 |
| #N/A | 45 | | 45 |
| 総計 | 270165812 | 17553080 | 287718892 |

詳細な中分類を表示させるためには上の「+」をクリックすればよい。たとえば 新生物をドリルダウンすると以下のように表示される。

| | | | |
|---|---|-----------------------------|------------------|
| 合計 / データ | | | |
| | 列ラベル | | 総計 |
| | <input type="checkbox"/> 国保 | <input type="checkbox"/> 組合 | |
| | <input checked="" type="checkbox"/> グループ1 | <input type="checkbox"/> 組合 | |
| 行ラベル | | 組合 | |
| <input checked="" type="checkbox"/> 01 感染症及び寄生虫症 | 8584535 | 801177 | 9385712 |
| <input checked="" type="checkbox"/> 02 新生物 | 10649205 | 644785 | 11293990 |
| 0201:胃の悪性新生物 | 1049846 | 48932 | 1098778 |
| 0202:結腸の悪性新生物 | 926545 | 43762 | 970307 |
| 0203:直腸S-状結腸移行部及び直腸の悪性新生物 | 410547 | 19341 | 429888 |
| 0204:肝及び肝内胆管の悪性新生物 | 195522 | 8513 | 204035 |
| 0205:気管、気管支及び肺の悪性新生物 | 684712 | 32016 | 716728 |
| 0206:乳房の悪性新生物 | 1208832 | 68719 | 1277551 |
| 0207:子宮の悪性新生物 | 343319 | 23734 | 367053 |
| 0208:悪性リンパ腫 | 262307 | 13055 | 275362 |
| 0209:白血病 | 117268 | 7241 | 124509 |
| 0210:その他の悪性新生物 | 2554379 | 118582 | 2672961 |
| 0211:良性新生物及びその他の新生物 | 2895928 | 260890 | 3156818 |
| <input checked="" type="checkbox"/> 03 血液及び造血器の疾患免疫機構の障害 | 1052490 | 89581 | 1142071 |
| <input checked="" type="checkbox"/> 04 内分泌、栄養及び代謝疾患 | 32439844 | 1538447 | 33978291 |
| <input checked="" type="checkbox"/> 05 精神及び行動の障害 | 14256316 | 594962 | 14851278 |
| <input checked="" type="checkbox"/> 06 神経系の疾患 | 7949601 | 405227 | 8354828 |
| <input checked="" type="checkbox"/> 07 眼及び付属器の疾患 | 23650662 | 1615662 | 25266324 |
| <input checked="" type="checkbox"/> 08 耳及び乳様突起の疾患 | 4137875 | 321374 | 4459249 |
| <input checked="" type="checkbox"/> 09 循環器系の疾患 | 58317219 | 2547446 | 60864665 |
| <input checked="" type="checkbox"/> 10 呼吸器系の疾患 | 29720758 | 3512145 | 33232903 |
| <input checked="" type="checkbox"/> 11 消化器系の疾患 | 16336890 | 1016516 | 17353406 |
| <input checked="" type="checkbox"/> 12 皮膚及び皮下組織の疾患 | 14166683 | 1389899 | 15556582 |
| <input checked="" type="checkbox"/> 13 筋骨格系及び結合組織の疾患 | 27777661 | 1420815 | 29198476 |
| <input checked="" type="checkbox"/> 14 腎尿路生殖器系の疾患 | 9056729 | 619594 | 9676323 |
| <input checked="" type="checkbox"/> 15 妊娠、分娩及び産じょく<褥> | 363410 | 59632 | 423042 |
| <input checked="" type="checkbox"/> 16 周産期に発生した病態 | 115594 | 18970 | 134564 |
| <input checked="" type="checkbox"/> 17 先天奇形、変形及び染色体異常 | 529263 | 56005 | 585268 |
| <input checked="" type="checkbox"/> 18 症状、徴候及び異常臨床所見・異常検査所見 | 4383012 | 314119 | 4697131 |
| <input checked="" type="checkbox"/> 19 損傷、中毒及びその他の外因の影響 | 6678020 | 586724 | 7264744 |
| #N/A | 45 | | 45 |
| 総計 | 270165812 | 17553080 | 287718892 |

データ型は件数と点数の二つがあり、また KDB は 2012 年度以降のデータを収載しているので、2012、13 年度を、内分泌疾患及び腎泌尿器系疾患で比較したものが以下の通りである。

| | | | |
|--------------------------|--------------------|--------------------|--------------------|
| 年 | 2012 | | |
| MDC | (複数のアイテム) | | |
| 制度 | 国保 | | |
| 合計 / データ 行ラベル | 列ラベル | | |
| | 入院 | 外来 | 総計 |
| 医療費 | 2.53152E+11 | 9.71467E+11 | 1.22462E+12 |
| 0401:甲状腺障害 | 4161603990 | 29765156500 | 33926760490 |
| 0402:糖尿病 | 82796146240 | 3.0608E+11 | 3.88876E+11 |
| 0403:その他の内分泌, 栄養及び代謝疾患 | 29955219840 | 1.76035E+11 | 2.05991E+11 |
| 1401:糸球体疾患及び腎尿管間質性疾患 | 18201239170 | 12460256440 | 30661495610 |
| 1402:腎不全 | 73594374440 | 3.58373E+11 | 4.31967E+11 |
| 1403:尿路結石症 | 10423471260 | 8356548420 | 18780019680 |
| 1404:その他の腎尿路系の疾患 | 8831675630 | 17881542750 | 26713218380 |
| 1405:前立腺肥大(症) | 8954114760 | 27155033720 | 36109148480 |
| 1406:その他の男性生殖器の疾患 | 2181853630 | 3015454480 | 5197308110 |
| 1407:月経障害及び閉経周辺期障害 | 253973850 | 7255345760 | 7509319610 |
| 1408:乳房及びその他の女性生殖器の疾患 | 13797973530 | 25089516610 | 38887490140 |
| 件数 | 589733 | 43654614 | 44244347 |
| 0401:甲状腺障害 | 10687 | 2187008 | 2197695 |
| 0402:糖尿病 | 200304 | 16102206 | 16302510 |
| 0403:その他の内分泌, 栄養及び代謝疾患 | 76566 | 15689077 | 15765643 |
| 1401:糸球体疾患及び腎尿管間質性疾患 | 46458 | 520194 | 566652 |
| 1402:腎不全 | 119742 | 1237748 | 1357490 |
| 1403:尿路結石症 | 32866 | 470302 | 503168 |
| 1404:その他の腎尿路系の疾患 | 23979 | 1662604 | 1686583 |
| 1405:前立腺肥大(症) | 26707 | 2331074 | 2357781 |
| 1406:その他の男性生殖器の疾患 | 8574 | 321447 | 330021 |
| 1407:月経障害及び閉経周辺期障害 | 1043 | 953682 | 954725 |
| 1408:乳房及びその他の女性生殖器の疾患 | 42807 | 2179272 | 2222079 |
| 総計 | 2.53152E+11 | 9.71511E+11 | 1.22466E+12 |

D. 結論

KDB はきわめて豊富なデータを含んでおり、市町村はそこから csv で抽出してケアカンファレンス等の地域包括ケアの推進のために活用することが期待される。しかしながら、クロス表形式の csv ファイルはそのままでは活用しにくい。

データを、性・年齢階級、傷病別、データ型別といった様々な次元で集計することを多次元 OLAP と呼び、ビッグデータ活用方法として注目されている。そのためにはクロス表形式ではよくなく、CUBE 形式する必要がある。たとえば全国健康保険協会は、レセプトや被保険者データを CUBE 形式で公開している。CUBE 形式では、そのまま Excel で開いても理解できないから、全国健康保険協会が最初から CUBE 形式でデータを公表していることは、少なくとも Excel のピボットテーブルや ACCESS を使いこなす上級者を念頭においていることがうかがえる。

残念ながら KDB では csv ファイルは全てクロス表形式であって CUBE 形式でとりだすことはできない。本研究で開発したマニュアルにしたがって処理すれば Excel 上で CUBE 形式に変換することが可能となり、データ活用が広がる。

今回の研究を通じて、KDB のデータ抽出機能にも制約があることが明らかとなった。

第一は、KDB は電子化された医科・調剤・介護レセプトを収集しているので、単なる傷病別や入院外来別といったおおまかな分析だけでなく、診療行為(たとえば腹腔鏡手術とか処方箋料の一般名加算等)、薬効分類(たとえばベンゾジアゼピン系向精神薬とか糖尿病薬)分析も可能なはずであるが、こうした詳細なデータ抽出機能は現在の KDB にはまだ備わっていない。こうした詳細なレセプトデータをキューブ化するためには、より綿密な前処理が必要となるためと考えられるが、今後の KDB の発展方向として検討すべき課題といえる。

本マニュアルが示した方法によって KDB の有効活用が可能となり、地域包括ケア推進に資することが期待される。