

PCCF+ etc.

**Exploiting postal codes and census geography
in population-based research**

Russell Wilkins

**Health Analysis and Measurement Group
Statistics Canada**

**Public Health Agency of Canada,
Thursday 27 January 2005**

Seminar introductions

- **Name**
- **Institution/agency and department/section**
- **Specialty, occupation or profession**
- **Reason for interest in postal codes ?!**
- **Any previous work with PCCF or PCCF+?**

Plan

- **Objectives—a wide variety of possible uses**
- **Why PCCF+? Canadian postal codes can be tricky**
- **Features of PCCF+**
- **Getting started—a quick run through the process**
- **Results—looking at the outputs**
- **Strategies for use—after the initial coding is done**
- **Pitfalls of automated coding—examples**
- **Other useful tools—complementary to PCCF+**
- **Access to PCCF+, and Where to get help**

Possible follow-up presentation

- Technical facts about PCCF+
- Pre-editing the postal code field
- Using the problem file listing
- Understanding the diagnostic fields produced
- Explanation of the supplementary geographic codes
- Translating across vintages of census geography
- Evaluating coding accuracy

Objectives

- **A wide variety of possible uses – mostly related to widening the uses of existing data**

The population-health relevance of existing data files can be greatly enhanced through the addition of small-area geographic codes, neighbourhood socio-economic characteristics (contextual variables) and geographic point co-ordinates.

- **After coding, the enhanced data can be used to describe health conditions for neighbourhoods, local administrative districts or program service areas.**
- **Place of residence can be used as an indicator of neighbourhood socio-economic status (SES), community size group, metropolitan influence zone (MIZ), etc.**
- **This information can help in setting public health priorities, and in fixing attainable health objectives (based on differentials by region and socio-economic characteristics).**

- **The associated geographic co-ordinates can be used to determine distances to health service providers, and to examine possible effects of environmental hazards, based on location and distance relative to known sources of pollution or environmental contamination.**
- **“Easy-to-use” software and reference files are now available to automate much of the work required to make your data useable by geographic information systems (GIS), and linkable to small-area census data on a wide range of population characteristics.**

Why PCCF+?

Canadian postal codes can be tricky

- Population weights
- Diagnostics
- Imputations
- Supplemental codes
- Reproducible, documented processing

Features of PCCF+

- **Overview and history**
- **Major objectives**
- **Supplementary uses**
- **Other features**

***PCCF+* (Geocodes/PCCF)**

- **GEORES.SAS** Intended for residential coding
- **GEOINS.SAS** Intended for coding of medical offices and health facility locations

- **Version 1** 1986 census geography; equal weight to each duplicate record
- **Version 2** 1991 census geography; 2B (20% sample) household weights for most duplicate records
- **Version 3** 1996 census geography; 2A (100% sample) population weights for most duplicate records
- **Version 4** 2001 census geography; EA=>DA+BLK

Major problems which are dealt with by *PCCF+*

- Postal codes serving several DAs or blocks (especially in rural areas)
- Postal codes used by businesses or public institutions
- Postal codes which the regular PCCF only links to post office geography (rather than place of residence or business)

Supplementary uses of automated coding from postal codes

- To help identify residents of long-term care facilities (such as nursing homes, chronic care hospitals)
- To identify records where the postal code does not refer to a place of usual residence (commercial buildings; children's aid society offices; public trustees; social services headquarters, etc.)
- To help establish or confirm record linkages
- To permit studies of migration over time (for exposure or SES histories, or for better access to services, etc.) when longitudinal files are available

Other features of *PCCF+*

- Use FSA to impute or partially impute geographic coding where the postal code is not found or is only linked to PO geography
- Deal with retired postal codes, taking into account problems related to *previous* DMT
- Flag records for non-residential addresses, major health facilities, schools and prisons
- Provide information that may help in correcting erroneous or problematic postal codes, or for finding geographic codes by other means (if possible); try to furnish enough information so the user can decide whether to accept or reject the coding suggested (if correction of the underlying problem not possible or feasible)
- Provide for translation across different vintages of census geography*

Getting started

- a quick run through the process ...
in 4 simple steps!

Step 1: Getting set up

- SAS must have already been installed
- Copy all of the files into your own directory
 - for example, C:\PCCF4D\
- Edit the program :
 - GEORES4D.SAS (for residential coding)
 - or
 - GEOINS4D.SAS (for offices and institutions)

Step 2: Tell SAS where to find your input file (with postal codes to be assigned geography)

```
filename HLTHDAT 'c:\pccf4d\sampldat.can';/*your input file*/

DATA HLTHDAT0; INFILE HLTHDAT MISSOVER;
INPUT
  @ 5 ID $CHAR8. /* UNIQUE IDENTIFIER OR REGISTRAT NUMBER */
                /* -- UP TO 12 CHARACTERS IN LENGTH */
  @ 88 FSA $CHAR3. /* FSA (ANA)--FIRST 3 CHARACTERS OF PCODE */
  @ 92 LDU $CHAR3.; /* LDU (NAN)--LAST 3 CHARACTERS OF PCODE */
PCODE=FSA||LDU; /* POSTAL CODE (ANANAN) */
```

Step 3: Tell SAS where to put the two output files produced

```
filename HLTHOUT `c:\pccf4d\sampldat.geo';  
                /* the main output file */
```

```
filename GEOPROB `c:\pccf4d\sampldat.prb';  
                /* the problem file      */
```

Step 4: Submit the program

- Page setup:
 - Orientation: Landscape
- Print setup:
 - Font: SAS Monospace 8 pt
- Submit program
- Review results: log, list, then files

Results

- **Looking at the outputs**
 - **Printouts (SAS listings)**
 - **Files generated**
 - **HLTHOUT (.GE0)**
 - **GEOPROB (.PRB)**

Coded output files (HLTHOUT+GEOPROB)

GEOG CODING

- ID (<=12), PCODE
- PR, CD, CSD
- CMA, CT; HR, SUB
- DA, BLK; EA96UID
- LAT, LONG
- QAIPPE, CSIZE,
- SACTYPE, NSREL
- EA96UID, FED1996, FED2003
- ER, AR, CCS, BLKURB, DPL

DIAGNOSTICS

- DMT, DMTDIFF
- LINK (PROB)
- SOURCE
- NCSD, NCD
- RPF, SERV, PREC
- BLDG NAME+ADR*
- CSDNAME+TYPE*
- CPCCODE
- RESFLG, INSTFLG

APPENDIX D:
SAMPLE OUTPUTS
FROM THE PCCF+ PACKAGE

SUMMARY OF AUTOMATED CODING RESULTS USING PCCF+ VERSION 4

| RECORDS | PERCENT | PROB | MESSAGE | ACTION |
|---------|---------|------|--|--------|
| 3996 | 100.00 | | TOTAL RECORDS INPUT FROM HLTHDAT (ID + PCODE) | |
| 131 | 3.28 | 0 | ERROR: NO MATCH TO PCCF---CHECK PCODE/ADDRESS &OR CODE MANUALLY | |
| 5 | 0.13 | 1 | ERROR: LINKED TO PO GEOG--CODE MANUALLY IF RESID ADD AVAILABLE | |
| 3 | 0.08 | 2 | WARNING: NON-RESIDENTIAL--CHECK PCODE/ADDRESS (LEGITIMATE RES?) | |
| 3 | 0.08 | 3 | WARNING: BUSINESS BLDG----CHECK PCODE/ADDRESS (LEGITIMATE RES?) | |
| 241 | 6.03 | 4 | WARNING: COMMERC/INSTITU--CHECK PCODE/ADDRESS (LEGITIMATE RES?) | |
| 65 | 1.63 | 5 | WARNING: RETIRED PCODE----CHECK PCODE/ADDRESS IF OLD DMT UNKNOWN | |
| 1 | 0.03 | 6 | NOTE: MULT MATCH CSD-PCCF-DISTRIBUTED AMONG APPLIC DA/BLK/BLKF | |
| 535 | 13.39 | 7 | NOTE: MULT MATCH CSD-WCF--DISTRIBUTED BY POP WEIGHTS OBSERVED | |
| 3012 | 75.38 | 9 | NO PROB (ERR,WARN,NOTE)---NO ACTION REQUIRED | |
| 8 | 0.20 | | NOT CODED AT ALL | |
| 39 | 0.98 | | PARTIALLY CODED TO PR ONLY | |
| 2 | 0.05 | | PARTIALLY CODED TO PR + (CD OR CMA)--& APPROX LAT LONG | |
| 12 | 0.30 | | PARTIALLY CODED TO PR+CD+CMA--AND APPROX LAT LONG | |
| 8 | 0.20 | | PARTIALLY CODED TO PR+CD+CMA+CSD--AND APPROX LAT LONG | |
| 3927 | 98.27 | | FULLY CODED TO PR+CD+CMA+CSD+CT+BLK--AND DA/BLK/BLKFACE LAT LONG | |

PCCF+ VERSION 4 -- SAMPLE OUTPUT FROM THE HLTHOUT DATASET (.GEO FILE)

| ID | PCODE | PRCDCSD | CMA | CT | DABLK | LAT | LONG | DPL | DIAG | VER | COMM | HRSUB | C | Q | S | N | U | FED | ER | AR | CCS | EA9GUID |
|------------|--------|---------|-----|--------|--------|-------------------|------|-----------|------|------|-------|-------|---|---|---|---|-----|-----|----|-----|----------|---------|
| 1304183010 | H1A5H8 | 2466025 | 462 | 580.03 | 000601 | 45689925073486893 | 000 | A9D111172 | R4A | 3276 | 06 | 1 | 3 | 1 | S | 1 | 044 | 40 | 06 | 025 | 24045417 | |
| 1304183033 | H1A5G4 | 2466025 | 462 | 582.01 | 292702 | 45653189073503887 | 000 | A9D111176 | R4A | 3276 | 06 | 1 | 3 | 1 | S | 1 | 044 | 40 | 06 | 025 | 24045358 | |
| 1304183332 | G1H2C1 | 2423030 | 421 | 273.01 | 082102 | 46856140071245151 | 000 | A9D111117 | R4A | 2587 | 03 | 2 | 2 | 1 | S | 1 | 015 | 20 | 03 | 030 | 24016455 | |
| 1304183333 | G1H7B3 | 2423030 | 421 | 273.01 | 081902 | 46850294071240870 | 000 | A9F111191 | R4A | 2587 | 03 | 2 | 2 | 1 | S | 1 | 015 | 20 | 03 | 030 | 24016452 | |
| 1304183632 | G8T8L9 | 2437055 | 442 | 200.00 | 015910 | 46367087072500828 | 000 | B9D111171 | R4A | 2561 | 04 | 3 | 1 | 1 | S | 1 | 014 | 70 | 04 | 050 | 24014354 | |
| 1304184533 | J8V2P3 | 2481015 | 505 | 841.03 | 037904 | 45515264075736270 | 000 | A9D111176 | R4A | 2752 | 07 | 2 | 3 | 1 | S | 0 | 023 | 60 | 08 | 015 | 24015556 | |
| 1304185031 | G1P1H6 | 2423025 | 421 | 039.02 | 065901 | 46822089071329615 | 000 | A9D111117 | R4A | 3313 | 03 | 2 | 1 | 1 | S | 1 | 052 | 20 | 03 | 025 | 24054103 | |
| 1304185033 | G2E5Y7 | 2423055 | 421 | 140.03 | 047503 | 46806119071370503 | 000 | A9D111173 | R4A | 2859 | 03 | 2 | 4 | 1 | S | 1 | 052 | 20 | 03 | 060 | 24054063 | |
| 1601001210 | L1G3Y1 | 3518013 | 532 | 015.00 | 008602 | 43937498078876105 | 000 | A9D111117 | R4A | 5227 | 0330 | 3 | 2 | 1 | S | 1 | 016 | 30 | 03 | 013 | 35016270 | |
| 1601002733 | L8V3V5 | 3525005 | 537 | 005.01 | 059702 | 43217763079851251 | 000 | A9F111191 | R4A | 4809 | 0837 | 2 | 1 | 1 | S | 1 | 030 | 50 | 01 | 005 | 35030108 | |
| 1601005410 | R2G0E6 | 4611040 | 602 | 141.02 | 071402 | 49937939097087637 | 000 | A9D111117 | R4A | 6221 | 10 | 2 | 2 | 1 | S | 1 | 013 | 50 | 09 | 040 | 46008417 | |
| 1601007832 | P7A5G4 | 3558004 | 595 | 015.00 | 014505 | 48438993089226888 | 000 | A9F111191 | R4A | 5549 | 1662 | 3 | 1 | 1 | S | 1 | 087 | 95 | 05 | 004 | 35084320 | |
| 1601007833 | P7B3H1 | 3558004 | 595 | 011.01 | 031611 | 48421824089235996 | 000 | A9F111191 | R4A | 5549 | 1662 | 3 | 1 | 1 | S | 1 | 087 | 95 | 05 | 004 | 35084410 | |
| 1601009010 | M6S4Y8 | 3520005 | 535 | 050.01 | 147401 | 43637293079471415 | 000 | B9F111191 | R4A | 5562 | 0495B | 1 | 4 | 1 | S | 1 | 064 | 30 | 03 | 005 | 35063258 | |
| 1601009033 | M6P2H9 | 3520005 | 535 | 100.00 | 140201 | 43664058079462540 | 000 | A9F111191 | R4A | 5562 | 0495E | 1 | 3 | 1 | S | 1 | 064 | 30 | 03 | 005 | 35098002 | |
| 1601010231 | K7M7B4 | 3510010 | 521 | 014.00 | 013602 | 44250712076533691 | 000 | B9D111171 | R4A | 4951 | 0241 | 3 | 1 | 1 | S | 1 | 036 | 15 | 04 | 010 | 35037506 | |
| 1601011533 | L5C3S8 | 3521005 | 535 | 527.08 | 069101 | 43577841079654532 | 000 | A9D111172 | R4A | 5106 | 0653 | 1 | 3 | 1 | S | 1 | 046 | 30 | 02 | 005 | 35049404 | |
| 1601011910 | S0E1E0 | 4714076 | 000 | 000.00 | 002410 | 53349268104019508 | 000 | W7C934459 | R4A | 6735 | 08 | 5 | 1 | 0 | R | 1 | 006 | 50 | 8A | 072 | 47002573 | |
| 1601013832 | L7R4M7 | 3524002 | 537 | 207.01 | 053802 | 43334767079821521 | 000 | B9F111191 | R4A | 4458 | 0636 | 2 | 3 | 1 | S | 1 | 010 | 50 | 02 | 002 | 35008115 | |
| 1601014733 | L2G3E7 | 3526043 | 539 | 203.01 | 006904 | 43070976079095668 | 000 | A9F111191 | R4A | 5177 | 0946 | 3 | 2 | 1 | S | 1 | 052 | 50 | 01 | 043 | 35051016 | |
| 1601015931 | L4W1L1 | 3521005 | 535 | 527.05 | 032501 | 43624059079608402 | 000 | A9F111191 | R4A | 5106 | 0653 | 1 | 1 | 1 | S | 1 | 047 | 30 | 02 | 005 | 35047351 | |
| 1601016133 | L2S2M9 | 3526053 | 539 | 003.01 | 037804 | 43145861079253296 | 000 | A9F111191 | R4A | 5473 | 0946 | 3 | 1 | 1 | S | 1 | 051 | 50 | 01 | 053 | 35090216 | |
| 1601017132 | L4N2V4 | 3543042 | 568 | 005.00 | 038106 | 44367352079679190 | 000 | A9F111191 | R4A | 4358 | 0560 | 3 | 5 | 2 | S | 1 | 002 | 40 | 02 | 042 | 35079159 | |
| 1601017421 | N7S5L7 | 3538030 | 562 | 102.02 | 015804 | 42973744082365802 | 000 | A9F111191 | R4A | 5391 | 1242 | 4 | 3 | 2 | S | 1 | 071 | 70 | 01 | 030 | 35072209 | |
| 1601017633 | M4K1C1 | 3520005 | 535 | 069.00 | 383001 | 43669948079342406 | 000 | A9F111191 | R4A | 5562 | 0495I | 1 | 2 | 1 | S | 1 | 008 | 30 | 03 | 005 | 35006061 | |
| 1601017910 | N4B2W4 | 3528052 | 547 | 000.00 | 008009 | 42756837080558774 | 000 | H9C114259 | R4A | 4613 | 1034 | 4 | 4 | 3 | S | 0 | 027 | 50 | 01 | 052 | 35018012 | |
| 1601018131 | N6G2E5 | 3539036 | 555 | 044.04 | 035003 | 43006922081306309 | 000 | A9D111117 | R4A | 5013 | 1144 | 3 | 3 | 1 | S | 1 | 044 | 60 | 01 | 036 | 35045463 | |
| 1601019332 | L5G1J8 | 3521005 | 535 | 540.01 | 037901 | 43553413079585884 | 000 | B9F111191 | R4A | 5106 | 0653 | 1 | 1 | 1 | S | 1 | 048 | 30 | 02 | 005 | 35048068 | |
| 1601019721 | R2K0V9 | 4611040 | 602 | 133.00 | 070502 | 49927590097100976 | 000 | A9F111191 | R4A | 6221 | 10 | 2 | 2 | 1 | S | 1 | 014 | 50 | 09 | 040 | 46014203 | |
| 1601020010 | M4E3M6 | 3520005 | 535 | 022.00 | 379901 | 43677506079285931 | 000 | A9D111117 | R4A | 5562 | 0495K | 1 | 5 | 1 | S | 1 | 003 | 30 | 03 | 005 | 35002068 | |
| 1601020131 | T7P1A3 | 4813031 | 000 | 000.00 | 004620 | 54164822113845804 | 000 | A9F112181 | R4A | 7709 | 26 | 5 | 4 | 0 | R | 1 | 001 | 70 | 06 | 028 | 48001057 | |
| 1601020432 | N4G4T7 | 3532004 | 546 | 000.00 | 007010 | 42876846080729595 | 000 | B9F112181 | R4A | 5555 | 1152 | 4 | 4 | 3 | S | 1 | 063 | 60 | 01 | 012 | 35062064 | |
| 1601020610 | M1C1K9 | 3520005 | 535 | 362.02 | 374802 | 43788038079163502 | 000 | A9D111117 | R4A | 5400 | 0495M | 1 | 5 | 1 | S | 1 | 075 | 30 | 03 | 005 | 35077052 | |
| 1601025533 | T5H2X1 | 4811061 | 835 | 046.00 | 020303 | 53550678113501115 | 000 | A9F111191 | R4A | 7229 | 25 | 2 | 1 | 1 | R | 1 | 015 | 60 | 05 | 061 | 48012253 | |
| 1601026631 | K1V9K4 | 3506008 | 505 | 002.05 | 087501 | 45347074075665245 | 000 | B9F111191 | R4A | 5230 | 0151 | 2 | 3 | 1 | S | 1 | 060 | 10 | 04 | 008 | 35059014 | |
| 1601027832 | S4V0G7 | 4706027 | 705 | 008.02 | 019701 | 50432251104564832 | 000 | A9D111117 | R4A | 6814 | 04 | 3 | 5 | 1 | S | 1 | 013 | 10 | 2B | 027 | 47007161 | |
| 1601028831 | N7S4X8 | 3538030 | 562 | 102.02 | 015903 | 42970869082365165 | 000 | A9F111191 | R4A | 5391 | 1242 | 4 | 2 | 2 | S | 1 | 071 | 70 | 01 | 030 | 35072208 | |
| 1601028832 | N7T6J8 | 3538030 | 562 | 008.00 | 019504 | 42982172082396827 | 000 | A9F111191 | R4A | 5391 | 1242 | 4 | 2 | 2 | S | 1 | 071 | 70 | 01 | 030 | 35072164 | |
| 1601029531 | T1K4A4 | 4802012 | 810 | 019.00 | 016101 | 49678240112881944 | 000 | A9D111117 | R4A | 7414 | 20 | 4 | 2 | 2 | S | 1 | 018 | 10 | 02 | 011 | 48017419 | |
| 1601030710 | L5C3L4 | 3521005 | 535 | 527.08 | 069502 | 43576525079661365 | 000 | A9F111191 | R4A | 5106 | 0653 | 1 | 4 | 1 | S | 1 | 046 | 30 | 02 | 005 | 35049405 | |
| 1601030733 | L5A3T1 | 3521005 | 535 | 521.06 | 085901 | 43597525079626646 | 000 | B9F111191 | R4A | 5106 | 0653 | 1 | 2 | 1 | S | 1 | 047 | 30 | 02 | 005 | 35047113 | |
| 1601031231 | L8N2Z3 | 3525005 | 537 | 033.00 | 044701 | 43246956079851089 | 000 | A9F111191 | R4A | 4809 | 0837 | 2 | 1 | 1 | S | 1 | 029 | 50 | 01 | 005 | 35032002 | |
| 1601032031 | K8A7W4 | 3547064 | 515 | 000.00 | 004912 | 45817759077093184 | 000 | A9F112181 | R4A | 5256 | 0157 | 4 | 5 | 3 | S | 1 | 070 | 15 | 04 | 075 | 35068254 | |
| 1601033332 | R2K0K5 | 4611040 | 602 | 134.00 | 071204 | 49930495097093590 | 000 | A9F111191 | R4A | 6221 | 10 | 2 | 3 | 1 | S | 1 | 014 | 50 | 09 | 040 | 46014208 | |
| 1601035633 | R2C5B2 | 4611040 | 602 | 120.02 | 085503 | 49900542096969280 | 000 | A9F111191 | R4A | 6221 | 10 | 2 | 4 | 1 | S | 1 | 014 | 50 | 09 | 040 | 46014003 | |

Sample printout from the GEOPROB dataset (file .PRB) PCCF+ VERSION 4

PARTIAL PRINT OF GEOPROB FILE (ERRORS & WARNINGS, BUT NO NOTES)

| ID | PCODE | PRCDCSD | CMA | CT | DABLK | LL | HRSUB | DPL | DIAG | BLDG NAME,ADR(CPCOMM:CMA/DPL) | :CDNAME | CDTYP | CSDNAME | TY |
|---|--------|----------|-----|--------|---------|------|---------|-----|------------|--|------------------|-------|------------|----|
| 0 ERROR: NO MATCH TO PCCF---CHECK PCODE/ADDRESS &OR CODE MANUALLY | | | | | | | | | | | | | | |
| 1202050810 | AIK5J7 | 1001485 | 001 | 301.02 | 013501 | 4705 | 01 | 000 | 90I31994. | St. John's CMA | :Avalon Peninsul | DIV | CONCEPTIT* | |
| 1201026310 | B2M5B3 | 1200999 | 999 | 999.99 | 999900 | 4506 | 99 | 999 | 902..892. | | : | | * | |
| 1302025710 | GOK2K0 | 2410005 | 000 | 000.00 | 007009 | 4806 | 01 | 000 | 90I949949 | NOT CMACA | :Rimouski-Neiget | MRC | ESPRIT-SM* | |
| 1301031010 | H9G3X9 | 2466140 | 462 | 521.01 | 235801 | 4507 | 06 | 000 | 90I31994. | Montréal CMA | :Montréal | CU | DOLLARD-V* | |
| 1602451310 | K7K2T0 | 3510010 | 521 | 008.00 | 018405 | 4407 | 0241 | 000 | 90I11994. | Kingston CMA | :Frontenac | CTY | KINGSTONC* | |
| 1604153110 | M3Y4A1 | 3520005 | 535 | 999.99 | 999900 | 4307 | 99999 | 999 | 902..892. | Toronto CMA | :Toronto | DIV | TORONTO C* | |
| 1604305110 | R3N3L2 | 4611040 | 602 | 008.00 | 038001 | 4909 | 10 | 000 | 90I11994. | Winnipeg CMA | :Winnipeg | DIV | WINNIPEGC* | |
| 1802106710 | V1S4X1 | 5933042 | 925 | 006.00 | 004302 | 5012 | 14 | 000 | 90I21994. | Kamloops CA1 | :Thompson-Nicola | RD | KAMLOOPSC* | |
| 1802068310 | V4T4J5 | 5935027 | 915 | 102.02 | 015502 | 4911 | 13 | 175 | 90I41994. | Kelowna CA1:Westbank (UNP) | :Central Okanaga | RD | CENTRAL RD | |
| 1803049810 | V9C5T3 | 5917044 | 935 | 154.02 | 048004 | 4812 | 41 | 000 | 90I51994. | Victoria CMA | :Capital | RD | LANGFORDDM | |
| 1 ERROR: LINKED TO PO GEOG---CODE MANUALLY IF RESID ADD AVAILABLE | | | | | | | | | | | | | | |
| 1604055531 | R4J1A1 | 4611999 | 602 | 999.99 | 999900 | 4909 | 99 | 000 | JZ1I22824. | HEADINGLEY:Winnipeg CMA | :Winnipeg | DIV | * | |
| 1201059710 | AIK4G9 | 1001999 | 001 | 999.99 | 999900 | 4705 | 99 | 000 | K1I318341 | BOX 18001:18060 STN MAIN UPPER GULLIES | | | * | |
| 2 WARNING: NON-RESIDENTIAL PCODE---CHECK PCODE/ADDRESS (LEGIT RES?) | | | | | | | | | | | | | | |
| 1304154932 | H3L1B9 | -2400999 | 462 | 999.99 | 999900 | . | . 99 | 999 | E2F119191 | CENTRE MEDICAL HENRI-BOURASSA 222 HENRI-BOURA MONT | | | * | |
| 1603422510 | L4C9S7 | -3500999 | 535 | 999.99 | 999900 | . | . 99999 | 999 | E2F119191 | BUSINESS BUILDING 120 NEWKIRK RD RICHMOND HILL | | | * | |
| 1602226510 | T2S2T6 | -4800999 | 825 | 999.99 | 999900 | . | . 99 | 999 | E2F119191 | FOODVALE OFFICE COMPLEX 5005 ELBOW DR SW CALGARY | | | * | |
| 1601088310 | T5N4A3 | -4800999 | 835 | 999.99 | 999900 | . | . 99 | 999 | E2F119191 | PEOPLES TRUST PLAZA 10216 124 ST NW EDMONTON | | | * | |
| 1302161110 | H3N2Y1 | -2400999 | 462 | 999.99 | 999900 | . | . 99 | 999 | G2F119191 | VIDEOTRON LTEE 405 OGILVY AV 200 MONTREAL | | | * | |
| 1804030033 | V2A5A9 | -5900999 | 913 | 000.00 | 999900 | . | . 99 | 999 | G2D119171 | CITY OF PENTICTON 171 MAIN ST PENTICTON | | | * | |
| 3 WARNING: BUSINESS BLDG---CHECK PCODE/ADDRESS (LEGITIMATE RES?) | | | | | | | | | | | | | | |
| 1604118533 | L6Y2N4 | @3521010 | 535 | 572.05 | 020201 | 4307 | 0653 | 000 | E3F111191 | APARTMENT BLDG 430 MCMURCHY AVE S BRAMPTON | | | BRAMPTONC* | |
| 1604503732 | T5H4B9 | @4811061 | 835 | 046.00 | 020808 | 5311 | 25 | 000 | E3F111191 | HYS MEDICAL CENTRE 11010 101 ST NW EDMONTON | | | EDMONTONC* | |
| 4 WARNING: COMMERC/INSTITU---CHECK PCODE/ADDRESS (LEGITIMATE RES?) | | | | | | | | | | | | | | |
| 1801082533 | V5G4J3 | ?5915025 | 933 | 230.01 | 139201 | 4912 | 22 | 000 | BG4F111191 | BRITISH COLUMBIA INSTITUTE OF TECHNOLOGY 4200 BURN | | | BURNABY C* | |
| 1202190833 | A1B1S5 | @1001519 | 001 | 013.00 | 025301 | 4705 | 01 | 000 | G4F111191 | ST PATRICKS MERCY HOME 146 ELIZABETH AVE ST. JOHN' | ST. JOHN' | | JOHNC* | |
| 1202154133 | A2A2E1 | @1006017 | 010 | 000.00 | 003010 | 4805 | 03 | 000 | G4D112171 | CENTRAL NEWFOUNDLAND REGIONAL HEALTH CENTRE 5 GRAN | GRAND FAT* | | | |
| 1303089633 | H2C3H6 | @2466025 | 462 | 277.00 | 265801 | 4507 | 06 | 000 | G4F111191 | LES RESIDENCES LAURENDEAU,LEGARE,LOUVAIN 1725 MONT | MONTRÉALV* | | | |
| 1603169333 | MIH3A1 | @3520005 | 535 | 356.00 | 361001 | 4307 | 0495N | 000 | G4F111191 | CEDARBROOK LODGE 520 MARKHAM RD SCARBOROUGH | | | TORONTO C* | |
| 1602154410 | M9W4L3 | @3520005 | 535 | 246.00 | 184101 | 4307 | 0495A | 000 | G4F111191 | KIPLING ACRES HOME FOR THE AGED 2233 KIPLING ETOBI | | | TORONTO C* | |
| 1604515931 | N2L3G1 | @3530016 | 541 | 106.01 | 029605 | 4308 | 0765 | 000 | G4F111191 | UNIVERSITY OF WATERLOO 200 UNIVERSITY AVE W WATERL | WATERLOOC* | | | |
| 1604443433 | R1N3V4 | @4609029 | 607 | 000.00 | 001414H | 4909 | 40 | 000 | G4F112181 | LION'S PRAIRIE MANOR 24 9TH ST SE PORTAGE LA PRAIR | PORTAGE C* | | | |
| 1603468632 | R3N1V9 | @4611040 | 602 | 510.02 | 036601 | 4909 | 10 | 000 | G4F111191 | CANADIAN FORCES BASE WINNIPEG, KAPYONG BARRAC WINN | WINNIPEGC* | | | |
| 1601086332 | R7N1R7 | @4617050 | 000 | 000.00 | 001114 | 5110 | 60 | 000 | G4F111191 | DAUPHIN GENERAL HOSPITAL 625 3RD ST SW DAUPHIN | | | DAUPHIN C* | |
| 1603548732 | S4S3B4 | @4706027 | 705 | 002.02 | 049002 | 5010 | 04 | 000 | G4F111191 | EXTENDICARE/PARKSIDE 4540 RAE ST REGINA | | | REGINA C* | |
| 1602539533 | T5K0L4 | @4811061 | 835 | 032.02 | 015604H | 5311 | 25 | 000 | G4F111191 | GENERAL HOSPITAL 11111 JASPER AVE NW EDMONTON | | | EDMONTONC* | |
| 1803100131 | V6T1K2 | @5915020 | 933 | 069.00 | 094705 | 4912 | 32 | 000 | G4D111171 | WALTER GAGE RESIDENCE (UBC) 5959 STUDENT UN VANC | | | GREATER RD | |

Strategies for using the results of PCCF+

- after the initial coding is done ...

Code your data only once, but analyse them many times

- **Be sure to correct all serious problems identified by the automated coding. It usually takes a couple of iterations to get the whole file clean.**
- **The importance of the problems identified by the diagnostic codes depends on the data set and on the analyses to be done. Retain the diagnostic codes!**
- **Once coded, the same dataset can be used for various kinds of studies (eg SES disparities, access to services, environmental health).**

“Linking” to other geographically-coded data using SAS merge; by GEOuid

- **Select a level: DA, CT, CMACA, etc.**
- **Choose the census/other variables of interest**
 - **census profiles available at all geographic levels***
- **Sort both files by GEOYYuid, then**
- **Merge by GEOYYuid**
- **Rewrite your file with the new variables appended**
- *Also for linking to geographically-oriented administrative data*

Dealing with different vintages of census geography

- **Example:** to compare age-specific death rates for 1985-87 vs 1995-97 vs 1999-2002
- **Problem:** Version 4 codes to 2001 + 1996 census geography; Version 3 codes to 1996; Version 2 codes to 1991; Ver 1 (NA) coded to 1986
- **Solution:** Code all with Vers 4, then use EA/DA conversion tables (“pretty good”, but not perfect).

Calculating distances

- Latitude & longitude for each record
- Sample distance calculation shown in GEORES.SAS (find “calculate dist”)
- Supplemental program DIST4D.SAS for many-to-many distances

Important geographic references

- **PCCF, PCCF+ documentation (.pdf files)**
- **Geographic relationships**
 - **GeoSuite 2001, 1996 (aka GeoRef 1996)**
 - **Geographic Attributes File 1991, 1986**
 - **Geography Tape File 1981, 1976, 1971**
- **Inter-censal EA/DA translations 1981-2001**
- **Neighbourhood income quintiles 1981-2001**
- **SGC translations (approximate)**

Other useful tools

- **Postal Code Viewer** (STC internal http://pcodeweb/web_apps/index.htm)
 - partial addresses fine, wildcards & truncation
 - reverse lookup with wildcards, truncation
- **GeoSearch 2001** (www.statcan.ca)
 - place names, street names, road intersections
 - postal code (formerly)
 - shows map layers incl BLK DA CT CMACA
- **Canada 411 – postal codes from phone numbers**
 - may refer to billing address vs service location
 - portability of numbers within prefixes => vintage impt

Pitfalls of automated coding: some examples (1)

- **Problem:** In a study of psychiatric problems among Manitoba children, dozens of children had the same downtown Winnipeg postal code.
- **Diagnosis:** Examination of the building name and address showed the postal code referred to the office of the provincial trustee responsible for minor children in provincial care. Use of the geography and neighbourhood characteristics associated with that postal code would have seriously biased the study results.
- **Solution:** Most non-residential postal codes including those for government and institutions can be identified by looking at the building / organization name and address in the problem output. Then either find the postal code for the true place of residence (if appropriate re study aims) or set geography to missing (as was done for this study).

Pitfalls of automated coding (2)

- **Problem:** In a study Quebec births, many births were for mothers with the same few urban postal codes. The delivery mode type of those postal codes was not B (for large apartment buildings).
- **Diagnosis:** It was determined that missing postal codes were being administratively assigned the postal code of the hospital of birth, so that health region could be assigned even though the mother's postal code was unknown. Use of the associated small-area geography and/or neighbourhood characteristics would have systematically biased the results.
- **Solution:** Identify postal codes for hospitals, which should not be accepted as place of residence of the mother. Then either use the address information (if available) to find the mother's own postal code or set geography to missing (as was done for this study).

Pitfalls of automated coding (3)

- **Problem:** In an early study using BC vital statistics data with nearly 100% presence of full postal codes, we were coding many deaths as residents of Montreal, Quebec, although the decedents had been born in other provinces or countries, and the provincial municipal coding showed BC place of residence.
- **Diagnosis:** The non-existent postal code H0H0H0 (ho-ho-ho!) was being assigned when no postal code was reported. PCCF+ imputed geography from partial postal codes, although error codes were also assigned.
- **Solution:** The full address was used to find a real postal code, or to assign geography manually if no postal code could be found.

Pitfalls of automated coding (4)

- **Problem:** The usual place of residence on vital statistics mortality files may legitimately include institutional addresses. How can we know when that is the case?
- **Diagnosis:** Systematically identify such cases by postal code (where unique) and by postal code and address (when not unique). In our studies of mortality by income, up to 15% of deaths are typically for residents of chronic care hospitals and other long-term health care facilities.
- **Solution:** Remove institutional residents from both deaths and population at risk (numerator and denominator). More of a problem for hospital separation data.

Pitfalls of automated coding (5)

- **Problem:** In a study set in the Kingston area, many health events were for a relatively few postal codes, which were not known to be hospitals or long-term health care facilities.
- **Diagnosis:** Closer examination showed them to be for prisons and university residences.
- **Solution:** Systematically identify such cases, and depending on the purposes of the study, decide whether or not to use such cases in the analysis. (Note: The smaller the study area, the greater the potential impact of such problems.)

Pitfalls of automated coding (6)

- **Problem:** In various studies, postal codes for businesses keep appearing in the field for place of residence, apparently not due to keying errors.
- **Diagnosis:** Likely a small but non-negligible proportion of persons either prefer to receive correspondence at their place of work, or mistakenly report the wrong postal code.
- **Solution:** Systematically identify postal codes for non-residential addresses. Try to recode based on street address or postal code reported on other records for the same person.

Pitfalls of automated coding (7)

- **Problem:** In a Nova Scotia study of socio-economic differentials in mental health based on person-oriented hospital data, the neighbourhood SES of the mentally ill, as determined from their current postal code, tended to decline over time.
- **Diagnosis:** Use of current postal code to assign neighbourhood SES would risk confusing cause with effect.
- **Solution:** In person-oriented analysis, assign neighbourhood SES based on postal code at *initial* hospitalization or diagnosis.

Pitfalls of automated coding (8)

- **Problem:** Some studies require geographic coding of business and industrial locations, including mines, manufacturing establishments and dumpsites.
- **Diagnosis:** The locations of such sites could be anywhere in the service area of postal code, unrelated to population distribution.
- **Solution:** The population-based assumptions on which resolution of multiple matches are made using PCCF+ are simply *not appropriate* for coding in such cases. Consider alternate coding methods based on nearest road intersection, retrieval of latitude and longitude information from other files, or use of GPS.

Access to *PCCF+*

- **STC data suppliers**
 - **Canadian Cancer Registries**
 - **Canadian Vital Statistics Registries**
 - **Canadian Inst for Health Info (CIHI)**
- **Canadian universities via DLI**
- **STC internal users via GeoDepot**
- **PCCF licensees (no additional cost)**
 - **Public Health Agency of Canada***
 - **Ontario Ministry of Health and LTC***

***PCCF+* via the
Data Liberation Initiative
(DLI)**

Most Canadian universities included

For info, see www.statcan.ca

then Learning Resources / Post-secondary /
Data Liberation Initiative (DLI)

- **Internal STC users:**

- [//geodepot/ftp/](#)

- http://geodepot.statcan.ca/Diss/Products/Products_e.cfm

- **External users:**

- **DLI at universities (& alternate source for STC internal use)**

- **STC regional offices (free to PCCF licensees)**

SPSS users

- **No plans to port PCCF+ to SPSS**
- **No need to learn SAS**
 - **just use it yourself as a black box (Steps 1-4)**
 - **get someone else to run PCCF+ for you, and send you the results**

Where to get help

- **DLI--See STC Website (Education, DLI) for list of contact persons at each university.**
- ***PCCF+* (Vers 4D): various persons, depending on nature of problem, etc. See *User's Guide*, p. 16.**
- **For other aspects of adding socioeconomic and spatial dimensions to health data, contact Russell Wilkins at STC**

Russell Wilkins

Health Analysis and Measurement Group

Statistics Canada, RHC-24A

Ottawa ON K1A 0T6

Tel: 1-613-951-5305

Fax: 1-613-951-3959

Email: wilkrus@statcan.ca

Thank you