HWS death rates for 2009–2011, by poverty and educational attainment

Eric Ossiander
March 22, 2013

We computed death rates for Washington residents who died during 2009–2011, categorized by the poverty level and the educational attainment level of the census tracts where they lived.

We used census data to categorize each census tract into one of 4 poverty categories (Table 1), according to the percent of the population living in poverty in the census tract. The percent in poverty data came from table B17001 in the 2007–2011 ACS summary file data. We also categorized each census tract into one of 5 educational attainment categories, according to the percent of college graduates among people age 25 or older living in the census tract (Table 1). The educational attainment data came from table B15002 in the 2007–2011 ACS summary file data.

Table 1: Census tracts were labeled by these percent poverty and educational attainment categories.

Poverty categories
(Percent in poverty)
0-4.9
5-9.9
10-19.9
20+
Educational attainment categories
(Percent college graduates)
0-14.9
15-24.9
25-34.9
35-44.9
45+

We computed age-adjusted death rates for residents of each census tract category. We computed death rates for all causes, and for each of the causes listed in Table 2. The NCHS 10-year age groups were used for age-adjustment, except as for colorectal cancer, for which we also computed age-adjusted rates with the NCI 5-year age groups. All age-adjustment was done to the 2000 US standard population. We obtained population by census tract from the 2010 US Census summary files.

Numbers of deaths used in the analysis

Some geocoded death records did not link to the census data (see Table 3). Apparently, this happens because the geocode field in the geocoded death files does not include the full geocode. Census tract codes are unique only within counties, but county code is not included in the geocode field. Therefore we have to combine the tract code with the “reported county” field to link to census data. Apparently, the reported county does not agree with the geocoded county in some cases.

Table 2: Cause-of-death categories for which we computed age-adjusted death rates. Rates were adjusted to the US 2000 standard population. The NCHS 10-year age-groupings were used for adjustment, unless otherwise noted.

Cause-of-death group	ICD-10 codes	notes
All Causes	A00–Z99	
All Disease Causes	A00–R99	
All Cancers	C00–C97	
Colorectal CA, NCI	C18, C19, C20, C26.0	used NCI 5-year age groups for age-adjustment
Colorectal CA, HP2010	C18, C19, C20, C21	
Skin Cancer	C43	
Breast Cancer	C50	
Cervical Cancer	C53	
Diabetes	E10–E14	
Coronary Heart Disease	I20–I25	
Acute MI	I21, I22	
Stroke	I60–I69	
Drowning	W65–W74, V90, V92	
Suicide	X60–X84, Y87.0	

Table 3: Number of deaths used for computing rates, by year

	Year			
	2009	2010	2011	Total
deaths to Washington residents	48,202	47,981	49,386	145,569
Exclusions				
missing age	1	1	1	3
geocode matchscore too low	666	1,085	818	2,569
geocode did not link to census	28	11	15	54
deaths used to compute rates	47,507	46,884	48,552	142,943

SAS program

This is the SAS program we used to merge the death geocodes with the standard death files, link to the census tract poverty and education data, and omcpute the age-adjusted weights.

Listing 1: Deathrates2009-2011.sas

```
*/
/*
Deathrates2009-2011.sas Compute death rates for the 2012 HWS report.

I am to compute age-adjusted death rates for 2009-2011 for these categories:

All Causes                A00-R99
All Cancers               C00-C97
Colorectal CA, NCI        C18,C19,C20, C26.0
Colorectal CA, HP2010     C18,C19,C20, C21
Skin Cancer               C43
Breast Cancer             C50
Cervical Cancer           C53
Diabetes                  E10-E14
Coronary Heart Disease    I20 - I25
Acute MI                  I21,I22
Stroke                    I60 - I69
Drowning                  W65-W74, V90,V92
Suicide                   X60-X84,Y87.0  Add U03

I am to compute rates by educational attainment and poverty levels, by
computing rates for groups of census tracts according to what education
and poverty categories they fall into. I use the NCHS 10-year age groups for
age-adjustment, except as noted above for colorectal cancer.

The census tract education and poverty data and population by age
are in these files:
c:\user\projects\HWS\pop_by_tract\poveduc_10year_2011.sas7bdat
c:\user\projects\HWS\pop_by_tract\poveduc_5year_2011.sas7bdat

The categories are:

 collegecat  (percent college grads)
 ------------
  0-14.9
 15-24.9
 25-34.9
 35-44.9
 45+


 povcat (percent in poverty)
 ----------
  0-4.9
  5-9.9
 10-19.9
 20+


The 19 NCI 5-year age groups are:
under 1
1-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
```

```
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85+

The 11 NCHS 10-year age groups are:
under 1
1-4
5-14
15-24
25-34
35-44
45-54
55-64
65-74
75-84
85+

This is a list of the sources for the geocode match (with frequencies for 2009).
I will exclude records with sources CITY CENTROID, POST OFFICE CENTROID, and
ZIP CENTROID. These sources all have values for matchscore of 40 or lower, so I
will use that criteria to exclude these records (this also excludes records
with missing geocodes).

                              The FREQ Procedure

 source                    Frequency     Percent
 ------------------------------------------------
 CENTRUS STREETS               492        1.02
 CITY CENTROID                 547        1.14
 ESRI                           12        0.02
 HWY                            11        0.02
 NAVSTREETS                      1        0.00
 PARCELS                      1268        2.63
 POST OFFICE CENTROID           61        0.13
 ROADS                         317        0.66
 ST1000                         20        0.04
 STREET CENTROID               196        0.41
 TIGER 2006                      2        0.00
 TIGER06                      2306        4.79
 TIGER10                     42840       88.93
 ZIP CENTROID                   29        0.06
 ZIP PLUS 4 CENTROID            70        0.15

                       Frequency Missing = 29
*/
libname hwspop    'c:\user\projects\HWS\pop_by_tract';

/*
Steps:
1. Read death geocode files, and merge them with the standard death files
2. Merge death and census tract files to add education/poverty categories to
   the death files.
3. Extract qualifying deaths (eg colorectal cancers) from death files
4. Tabulate deaths by census tract education or poverty category and age
5. Tabulate population by census tract education or poverty category and age
6. Merge death and population and compute age-adjusted rates
*/

/*
Create a file that has the census tract education and poverty categories
*/
proc sort data=hwspop.poveduc_10year_2011 out=poveduc(keep=county tractID
      collegecat povcat) nodupkey;
```

```
   by tractID;
run;
data poveduc2;
   length tract2010 $ 7;
   set poveduc;
   tract2010 = substr(tractID,8,4)||'.'||substr(tractID,12,2);
run;

%macro deathgeo(year);
/*
Step 1. Read death geocodes and merge with death file
*/
data geo&year;
   infile "c:\data\death\Geocoded data\geod&year..los";
   input
      @1 certno       $char10.
      @11 county      $char2.
      @70 source      $char20.
      @90 matchscore  3.
      @93 tract2000   $char7.
      @118 tract2010  $char7.
      ;
run;
proc freq data=death.dea&year(where=('01' le cnty_res le '39'));
   tables age;
run;
data death&year(keep=certno underly underly3 cnty_res age sex);
   set death.dea&year(where=('01' le cnty_res le '39' and age not in (.,999)));
run;
proc sort data=geo&year;
   by certno;
run;
proc sort data=death&year;
   by certno;
run;
data deathgeo&year;
   merge death&year(in=indeath) geo&year(in=ingeo);
   by certno;
   if indeath and ingeo;
run;
/*
Step 2. Merge death and census tract files to add education/poverty categories
to the death files.
*/
proc sort data=deathgeo&year;
   by county tract2010;
run;
proc sort data=poveduc2;
   by county tract2010;
run;
data deathgeo2_&year nolink&year;
   merge deathgeo&year(in=indeath) poveduc2(in=incensus);
   by county tract2010;
   if indeath and incensus and matchscore ge 50 then output deathgeo2_&year;
   else output nolink&year;
run;
proc freq data=nolink&year;
   where certno ne '';
   tables matchscore/missing;
run;
%mend;

%deathgeo(2009);
%deathgeo(2010);
%deathgeo(2011);

%macro rates_10year(sex,cause,causetit,sheet);
/*
```

```
Step 3. Extract qualifying deaths from deathgeo file
*/
data deathfil;
   set deathgeo2_2009 deathgeo2_2010 deathgeo2_2011;
   where &cause;
   select;
      when(age = 0)          agegrp = ' 0    ';
      when( 1 le age le  4)  agegrp = ' 1-4  ';
      when( 5 le age le 14)  agegrp = ' 5-14';
      when(15 le age le 24)  agegrp = '15-24';
      when(25 le age le 34)  agegrp = '25-34';
      when(35 le age le 44)  agegrp = '35-44';
      when(45 le age le 54)  agegrp = '45-54';
      when(55 le age le 64)  agegrp = '55-64';
      when(65 le age le 74)  agegrp = '65-74';
      when(75 le age le 84)  agegrp = '75-84';
      when(85 le age le 120) agegrp = '85+   ';
      end;
run;
/*
Step 4. Tabulate deaths by census tract education or poverty category and age
*/
proc summary data=deathfil nway;
   class collegecat agegrp;
   output out=deathcoll(rename=(_freq_=deaths));
run;
proc summary data=deathfil nway;
   class povcat agegrp;
   output out=deathpov(rename=(_freq_=deaths));
run;
/*
Step 5. Tabulate population by census tract education or poverty category and age
*/
proc summary data=hwspop.poveduc_10year_2011(where=(sex = "&sex")) nway;
   class agegrp collegecat;
   var pop;
   output out=popcoll sum=pop;
run;
proc summary data=hwspop.poveduc_10year_2011(where=(sex = "&sex")) nway;
   class agegrp povcat;
   var pop;
   output out=poppov sum=pop;
run;
/*
Step 6. Merge death and population and compute age-adjusted rates
*/
data proport(keep=agegrp prop);
   infile "c:\data\popln\2000prop.dat";
   input age 9-10 prop 21-28;
   select(age);
      when(0 ) agegrp = ' 0    ';
      when(1 ) agegrp = ' 1-4  ';
      when(5 ) agegrp = ' 5-14';
      when(15) agegrp = '15-24';
      when(25) agegrp = '25-34';
      when(35) agegrp = '35-44';
      when(45) agegrp = '45-54';
      when(55) agegrp = '55-64';
      when(65) agegrp = '65-74';
      when(75) agegrp = '75-84';
      when(85) agegrp = '85+   ';
      end;
run;
/*
First, add standard population weights to the population file
*/
data popcoll2;
   merge popcoll proport;
```

```
   by agegrp;
run;
proc sort data=popcoll2;
   by collegecat agegrp;
run;
data poppov2;
   merge poppov proport;
   by agegrp;
run;
proc sort data=poppov2;
   by povcat agegrp;
run;

data ratescoll(keep=totdeath totpop rate stderr glower gupper collegecat);
   length collegecat $ 7 totpop totdeath rate stderr glower gupper 8;
   retain rate maxwt;
   merge popcoll2 deathcoll;
   by collegecat agegrp;
   if first.collegecat then do;
      rate=0;
      totpop=0;
      totdeath=0;
      var2=0;
      maxwt=0;
      end;
   if deaths = . then deaths = 0;
/*
multiply population by 3 because we have 3 years of death data and 1 year pop
estimates, but divide by 3 again before the output statement.
*/
   pop = 3*pop;
   totpop + pop;
   totdeath+deaths;
   agerate=(deaths/pop);
   rate=rate+(agerate*prop*100000);
   var2+(deaths * (prop/pop)**2);
   if (prop/pop) > maxwt then maxwt = prop/pop;
   if last.collegecat then do;
      var2     = var2*(100000**2);
      stderr   = sqrt(var2);
      maxwt    = maxwt*100000;
      glower   = (var2/(2*rate))*cinv(0.025,2*(rate**2)/var2);
      gupper   = ((var2 + maxwt**2)/(2*(rate+maxwt)))*
                 cinv(0.975,2*(rate+maxwt)**2/(var2+maxwt**2));
      totpop = totpop/3;
      output;
      end;
run;
data ratespov(keep=totdeath totpop rate stderr glower gupper povcat);
   length povcat $ 7 totpop totdeath rate stderr glower gupper 8;
   retain rate maxwt;
   merge poppov2 deathpov;
   by povcat agegrp;
   if first.povcat then do;
      rate=0;
      totpop=0;
      totdeath=0;
      var2=0;
      maxwt=0;
      end;
   if deaths = . then deaths = 0;
/*
multiply population by 3 because we have 3 years of death data and 1 year pop
estimates, but divide by 3 again before the output statement.
*/
   pop = 3*pop;
   totpop + pop;
   totdeath+deaths;
```

```
   agerate=(deaths/pop);
   rate=rate+(agerate*prop*100000);
   var2+(deaths * (prop/pop)**2);
   if (prop/pop) > maxwt then maxwt = prop/pop;
   if last.povcat then do;
      var2     = var2*(100000**2);
      stderr   = sqrt(var2);
      maxwt    = maxwt*100000;
      glower   = (var2/(2*rate))*cinv(0.025,2*(rate**2)/var2);
      gupper   = ((var2 + maxwt**2)/(2*(rate+maxwt)))*
                 cinv(0.975,2*(rate+maxwt)**2/(var2+maxwt**2));
      totpop = totpop/3;
      output;
      end;
run;
title "Age-adjusted rates for &causetit";
proc print data=ratespov;
run;
proc print data=ratescoll;
run;
title;
data &sheet;
   set ratespov ratescoll;
run;
proc export data=&sheet
   outfile = "c:\user\projects\HWS\Deathrates\rates_2009to2011.xls"
   dbms=excel
   replace
   ;
run;
%mend;

%rates_10year(T,%str('A00' le underly3 le 'Z99'),All causes (A00-Z99),
   All_cause);
%rates_10year(T,%str('A00' le underly3 le 'R99'),All disease causes (A00-R99),
   All_dx_causes);
%rates_10year(T,%str('C00' le underly3 le 'C97'),All cancers (C00-C97),
   All_cancer);
%rates_10year(T,%str('C18' le underly3 le 'C21'),
   Colorectal CA (C18-C21),Colorectal_Cancer);
%rates_10year(T,%str(underly3 = 'C43'),Skin cancer (C43),Skin_Cancer);

%rates_10year(F,%str(underly3 = 'C50' and sex = 'F'),
   Breast cancer (C50) (rate in females),Breast_cancer);
%rates_10year(F,%str(underly3 = 'C53' and sex = 'F'),
   Cervical cancer (C53) (rate in females),Cervical_cancer);

%rates_10year(T,%str('E10' le underly3 le 'E14'),Diabetes (E10-E14),Diabetes);
%rates_10year(T,%str('I20' le underly3 le 'I25'),Coronary Heart Disease (I20-I25),
   Coronary_heart_disease);
%rates_10year(T,%str('I21' le underly3 le 'I22'),Acute MI (I21-I22),Acute_MI);
%rates_10year(T,%str('I60' le underly3 le 'I69'),Stroke (I60-I69),Stroke);
%rates_10year(T,%str('W65' le underly3 le 'W74' or underly3 in ('V90','V92')),
   Drowning (W65-W74, V90, V92),Drowning);
%rates_10year(T,%str('X60' le underly3 le 'X84' or underly = 'Y870'),
   Suicide (X60-X84,Y87.0),Suicide);


%macro rates_5year(sex,cause,causetit,file);
/*
This macro is for the NCI age-adjustment with 5-year age groupings

Step 3. Extract qualifying deaths from deathgeo file
*/
data deathfil;
   set deathgeo2_2009 deathgeo2_2010 deathgeo2_2011;
   where &cause;
   select;
```

```
      when(age = 0)          agegrp = ' 0    ';
      when( 1 le age le  4) agegrp = ' 1-4  ';
      when( 5 le age le  9) agegrp = ' 5-9  ';
      when(10 le age le 14) agegrp = '10-14';
      when(15 le age le 19) agegrp = '15-19';
      when(20 le age le 24) agegrp = '20-24';
      when(25 le age le 29) agegrp = '25-29';
      when(30 le age le 34) agegrp = '30-34';
      when(35 le age le 39) agegrp = '35-39';
      when(40 le age le 44) agegrp = '40-44';
      when(45 le age le 49) agegrp = '45-49';
      when(50 le age le 54) agegrp = '50-54';
      when(55 le age le 59) agegrp = '55-59';
      when(60 le age le 64) agegrp = '60-64';
      when(65 le age le 69) agegrp = '65-69';
      when(70 le age le 74) agegrp = '70-74';
      when(75 le age le 79) agegrp = '75-79';
      when(80 le age le 84) agegrp = '80-84';
      when(85 le age)       agegrp = '85+   ';
      end;
run;
/*
Step 4. Tabulate deaths by census tract education or poverty category and age
*/
proc summary data=deathfil nway;
   class collegecat agegrp;
   output out=deathcoll(rename=(_freq_=deaths));
run;
proc summary data=deathfil nway;
   class povcat agegrp;
   output out=deathpov(rename=(_freq_=deaths));
run;
/*
Step 5. Tabulate population by census tract education or poverty category and age
*/
proc summary data=hwspop.poveduc_5year_2011(where=(sex = "&sex")) nway;
   class agegrp collegecat;
   var pop;
   output out=popcoll sum=pop;
run;
proc summary data=hwspop.poveduc_5year_2011(where=(sex = "&sex")) nway;
   class agegrp povcat;
   var pop;
   output out=poppov sum=pop;
run;
/*
Step 6. Merge death and population and compute age-adjusted rates
*/

data proport(keep=agegrp prop);
   infile "c:\data\popln\2000_19g.dat";
   input age 9-10 prop 21-28;
   select(age);
      when(0 ) agegrp = ' 0    ';
      when(1 ) agegrp = ' 1-4  ';
      when(5 ) agegrp = ' 5-9  ';
      when(10) agegrp = '10-14';
      when(15) agegrp = '15-19';
      when(20) agegrp = '20-24';
      when(25) agegrp = '25-29';
      when(30) agegrp = '30-34';
      when(35) agegrp = '35-39';
      when(40) agegrp = '40-44';
      when(45) agegrp = '45-49';
      when(50) agegrp = '50-54';
      when(55) agegrp = '55-59';
      when(60) agegrp = '60-64';
      when(65) agegrp = '65-69';
```

```
      when(70) agegrp = '70-74';
      when(75) agegrp = '75-79';
      when(80) agegrp = '80-84';
      when(85) agegrp = '85+   ';
      end;
run;
/*
First, add standard population weights to the population file
*/
data popcoll2;
   merge popcoll proport;
   by agegrp;
run;
proc sort data=popcoll2;
   by collegecat agegrp;
run;
data poppov2;
   merge poppov proport;
   by agegrp;
run;
proc sort data=poppov2;
   by povcat agegrp;
run;

data ratescoll(keep=totdeath totpop rate stderr glower gupper collegecat);
   length collegecat $ 7 totpop totdeath rate stderr glower gupper 8;
   retain rate maxwt;
   merge popcoll2 deathcoll;
   by collegecat agegrp;
   if first.collegecat then do;
      rate=0;
      totpop=0;
      totdeath=0;
      var2=0;
      maxwt=0;
      end;
   if deaths = . then deaths = 0;
/*
multiply population by 3 because we have 3 years of death data and 1 year pop
estimates, but divide by 3 again before the output statement.
*/
   pop = 3*pop;
   totpop + pop;
   totdeath+deaths;
   agerate=(deaths/pop);
   rate=rate+(agerate*prop*100000);
   var2+(deaths * (prop/pop)**2);
   if (prop/pop) > maxwt then maxwt = prop/pop;
   if last.collegecat then do;
      var2     = var2*(100000**2);
      stderr   = sqrt(var2);
      maxwt    = maxwt*100000;
      glower   = (var2/(2*rate))*cinv(0.025,2*(rate**2)/var2);
      gupper   = ((var2 + maxwt**2)/(2*(rate+maxwt)))*
                 cinv(0.975,2*(rate+maxwt)**2/(var2+maxwt**2));
      totpop = totpop/3;
      output;
      end;
run;
data ratespov(keep=totdeath totpop rate stderr glower gupper povcat);
   length povcat $ 7 totpop totdeath rate stderr glower gupper 8;
   retain rate maxwt;
   merge poppov2 deathpov;
   by povcat agegrp;
   if first.povcat then do;
      rate=0;
      totpop=0;
      totdeath=0;
```

```
      var2=0;
      maxwt=0;
      end;
   if deaths = . then deaths = 0;
/*
multiply population by 3 because we have 3 years of death data and 1 year pop
estimates, but divide by 3 again before the output statement.
*/
   pop = 3*pop;
   totpop + pop;
   totdeath+deaths;
   agerate=(deaths/pop);
   rate=rate+(agerate*prop*100000);
   var2+(deaths * (prop/pop)**2);
   if (prop/pop) > maxwt then maxwt = prop/pop;
   if last.povcat then do;
      var2     = var2*(100000**2);
      stderr   = sqrt(var2);
      maxwt    = maxwt*100000;
      glower   = (var2/(2*rate))*cinv(0.025,2*(rate**2)/var2);
      gupper   = ((var2 + maxwt**2)/(2*(rate+maxwt)))*
                 cinv(0.975,2*(rate+maxwt)**2/(var2+maxwt**2));
      totpop = totpop/3;
      output;
      end;
run;
title "Age-adjusted rates for &causetit (NCI 5-year age groups)";
proc print data=ratespov;
run;
proc print data=ratescoll;
run;
title;

data &file;
   set ratespov ratescoll;
run;
proc export data=&file
   outfile = "c:\user\projects\HWS\Deathrates\rates&file..xls"
   dbms=excel
   replace
   ;
run;
%mend;

%rates_5year(T,%str('C18' le underly3 le 'C20' or underly = 'C260'),
   Colorectal ca (C18-C20; C26.0),NCI_colorectalCA_1999to2011);

*/
```