Data and Replication materials

Mirjam Allik

11 July 2019

This document provides information on the data and some examples of replication of the results for the article *Differences in Ill Health and in Socioeconomic Inequalities in Health by Ethnic Groups: A cross-sectional study using 2011 Scottish Census* Ethnicity & Health by Mirjam Allik, Denise Brown, Ruth Dundas and Alastair H Leyland DOI:10.1080/13557858.2019.1643009

The data set *SRH_by_ethnicity_age_DZ_2011.csv* gives the number of people by Self-rated health (SRH) status, age group and ethnicity for each datazone in Scotland in 2011. It also gives a number of deprivation measures for each datazone. The data uses 2001 datazones. The data sets *st_rates_by_ethnicity_age_dep_2011.csv* and *RII_SII_by_ethnicity_age_dep_2011.csv* are derived based on the above mentioned data and provide, respectively, standardized rates and health inequality measures by age group and ethnicity. They provide the results presented in the article and also some additional results not presented in the article.

Data Sources

The data were put together using a number of different sources. The self-rated health are available from the Scottish Census commissioned tables website under section *Identity* as a zipped folder *CT_0229a_2011_to_CT_0229b_2011_csv.zip*. https://www.scotlandscensus.gov.uk/ods-web/data-warehouse.html#additionaltab

The data on the census based deprivation measure is available from Allik, M., Brown, D., Dundas, R. and Leyland, A. H. (2016) Small-area deprivation measure data sets for Scotland, 2001 and 2011. Data In Brief, 7, pp. 1682-1686. (doi:10.1016/j.dib.2016.04.060)

Data on the Carstairs score comes from Brown, D., Allik, M., Dundas, R. and Leyland, A. H. (2014) Carstairs Scores for Scottish Postcode Sectors, Datazones and Output Areas from the 2011 Census. Technical Report. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow.

SIMD data is from The Scottish Government. (2012) "SIMD 2012 Background Data - Part 2 Overall Ranks and Domain Ranks." http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis/Background-Data-2012/Background2SIMD2012

Ethnic groups are labelled in the data sets as in the Scottish Census data. Labels that were abbreviated are given in Table 1.

Table 1: Ethnic group labels and census categories

Ethnic group label	Census category
White (total)	White Scottish, White British, White Irish, White Polish, Other White
Asian (total)	Pakistani, Chinese, Indian and Other Asian
Pakistani	Pakistani, Pakistani Scottish or Pakistani British
Indian	Indian, Indian Scottish or Indian British
Chinese	Chinese, Chinese Scottish or Chinese British
Caribbean	Caribbean or Black
Mixed	Mixed or multiple ethnic groups

Variable descriptions

Table 2: Variables and their descriptions for data set *SRH_by_ethnicity_age_DZ_2011.csv*

Variable name	Description
datazone	2001 Datazone
age	Age group, coded as follows: 1 - 0-9; 2 - zero values; 3 - 10-14; 4 - 15-19; 5 - 20-24; 6 - 25-29; 7 - 30-34; 8 - 35-39; 9 - 40-44; 10 - 45-49; 11 - 50-54; 12 - 55-59; 13 - 60-64; 14 - 65-69; 15 - 70-74; 16 - 75-79; 17 - 80-84; 18 - 85 and above; 19 - all ages
ethnicity	Ethnic group as labelled, see also Table 1.
bad	Number of people who stated their health was “bad”" or “very bad”"
fg	Number of people who stated their health was “fairly good”
good	Number of people who stated their health was “good or”very good"
limited	Number of people who stated their day-to-day activities were limited a little or a lot
not.limited	Number of people who stated their day-to-day activities were not limited
pop	Number of people in the age and ethnic group in the datazone.
total.pop	Total number of people in the datazone
all.hh	Total number of people in households in the datazone
all.com	Total number of people in communal establishments in the datazone
Q.carstairs	Quintile of the Carstairs score (1 - least deprived; 5 - most deprived)
D.carstairs	Decile of the Carstairs score (1 - least deprived; 10 - most deprived)
carstairs	Carstairs score
SIMD.rank	SIMD rank
income.rank	SIMD income domain rank
dep.measure	Deprivation score, from Allik, M et al (2016) listed above
D.dep.measure	Decile of the deprivation measure (1 - least deprived; 10 - most deprived)
Q.dep.measure	Quintile of the deprivation measure (1 - least deprived; 5 - most deprived)
D.income.rank	Decile of the SIMD income domain (1 - least deprived; 10 - most deprived)
Q.income.rank	Quintile of the SIMD income domain (1 - least deprived; 5 - most deprived)
UR2FOLD	2-fold Scottish Government urban-rural classification
UR6FOLD	6-fold Scottish Government urban-rural classification
UR8FOLD	8-fold Scottish Government urban-rural classification
HB	Health Board
council	Council area

Table 3: Variables and their descriptions for data set *st_rates_by_ethnicity_age_dep_2011.csv*

Variable name	Description
ses	Small-area deprivation quintile (1 - least deprived; 5 - most deprived; overall - across all deprivation quintiles)
age	Age group as labelled (0-14, 15-29, 30-44, 45-59, 60-74, 75, 0-64, all)
rate	Standardized rate of ill health using the 2013 European Standard Population
CI_low	Lower 95% confidence interval for the rate of ill health
CI_high	Upper 95% confidence interval for the rate of ill health
ethnicity	Ethnic group as labelled, see also Table 1.
outcome	Health outcome: poor GH - poor general health; LLTI - limiting long-term illness

Table 4: Variables and their descriptions for data set *RII_SII_by_ethnicity_age_dep_2011.csv*

Variable name	Description
age	Age group as labelled (0-14, 15-29, 30-44, 45-59, 60-74, 75, 0-64, all)
estimate	Estimated health inequality
ci_low	Lower 95% confidence interval for estimated health inequality
ci_high	Upper 95% confidence interval for estimated health inequality
ethnicity	Ethnic group as labelled, see also Table 1.
measure	Measure used to estimate health inequalities: RII - Relative Index of Inequality; SII - Slope Index of Inequality
outcome	Health outcome: poor GH - poor general health; LLTI - limiting long-term illness
deprivation.var	Deprivation quintile used for the calculations: dep.quintile - Deprivation quintiles from Allik et al (2016); SIMD - SIMD income domain quintiles, Carstairs - Carstairs score quintiles

Replication

To calculate the standardized rates and the RII/SII the article used the R package SocEpi for health inequalities research available from https://github.com/m-allik/SocEpi.

Use the `st_rate()` function to calculate standardized rates (shown below). Note, the original data from the Scottish Census combined the age groups 0-4 and 5-9 into a single group 0-9. To use the SocEpi package the data was re-coded such that all counts for the age group 0-9 were coded as age group 0-4 (coded 1 in the data) and then the data was padded to create an additional age group 5-9 (coded 2 in the data). The padded age group 5-9 has zero counts for all health and population variables. In addition, the 2013 European Standard Population was accordingly adjusted and the age group 0-4 was given the weight 0.105 (0.050 + 0.055) and the age group 5-9 was given the weight zero.

```
library(SocEpi) # load library

d <- read.csv("SRH_by_ethnicity_age_DZ_2011.csv") # load data

# new standard population for padded data
pop_18 <- c(0.105, 0, 0.055, 0.055, 0.06, 0.06, 0.065, 0.07, 0.07, 0.07, 0.07,
            0.065, 0.06, 0.055, 0.05, 0.04, 0.025, 0.025)
sum(pop_18) # sums up to 1
```

```
## [1] 1
```

```
d <- d[d$age != 19, ] # remove all ages

# Standardized rates for White Scottish
st_rate(d, bad, pop, Q.dep.measure, age,
        ethnicity == "White Scottish", st_pop = pop_18)
```

```
## # A tibble: 48 x 5
##    ses   age     rate CI_low CI_high
##    <fct> <chr>  <dbl>  <dbl>   <dbl>
##  1 1     0-14    2.30   2.05    2.55
##  2 1     15-29   6.79   6.34    7.24
##  3 1     30-44  13.8   13.2    14.4
##  4 1     45-59  27.3   26.6    28.1
##  5 1     60-74  48.7   47.5    49.8
```

```
##  6 1     75    110.   107.    112.
##  7 2     0-14    2.93   2.65    3.22
##  8 2     15-29   8.63   8.15    9.10
##  9 2     30-44  21.3   20.6    22.0
## 10 2     45-59  45.8   44.9    46.8
## # ... with 38 more rows
```

To calculate RII, use the `rii()` function as shown below.

```
rii(d, bad, pop, Q.dep.measure, age,
    ethnicity == "White Scottish", st_pop = pop_18, RII = TRUE, W = T)
```

```
## # A tibble: 8 x 4
##   age     rii ci_low ci_high
##   <chr> <dbl>  <dbl>   <dbl>
## 1 0-14  1.30   1.18    1.41
## 2 15-29 1.25   1.18    1.31
## 3 30-44 1.98   1.95    2.01
## 4 45-59 2.05   2.02    2.07
## 5 60-74 1.68   1.65    1.70
## 6 75    0.934  0.905   0.963
## 7 0-64  1.95   1.94    1.97
## 8 all   1.63   1.62    1.65
```

To calculate SII, use the same `rii()` function as shown below, setting the argument `RII = FALSE`.

```
rii(d, bad, pop, Q.dep.measure, age,
    ethnicity == "White Scottish", st_pop = pop_18, RII = FALSE, W = T)
```

```
## # A tibble: 8 x 4
##   age      sii ci_low ci_high
##   <chr>  <dbl>  <dbl>   <dbl>
## 1 0-14    5.50   5.03    5.97
## 2 15-29  15.3   14.5    16.1
## 3 30-44  76.5   75.2    77.8
## 4 45-59 166.   164.    168.
## 5 60-74 192.   189.    194.
## 6 75    154.   149.    158.
## 7 0-64   82.3   81.6    83.0
## 8 all    99.8   99.0   101.
```