

Utah Health Improvement Index Methodology

Indicators

The 2015-2017 Utah Health Improvement Index (HII) was based on nine indicators from the Area Deprivation Index (ADI)¹ that were available in Utah's Behavioral Risk Factor Surveillance System (BRFSS) in 2015-2017. These nine indicators are:

- Factor 1 - <8th grade education
- Factor 2 - HS diploma
- Factor 4 - Median Income
- Factor 5 - Income disparity
- Factor 9 - Owner Occupied Housing
- Factor 10 - Unemployment Rate (age 16+)
- Factor 11 - Below poverty level
- Factor 12 - Below 150% of poverty level
- Factor 13 - Single Family Households with Children

For the 2022 HII, we reviewed data for the nine HII indicators from the BRFSS and the American Community Survey (ACS). Several models with various combinations of sources were constructed for the nine indicators. The indicators and sources listed below offered the highest level of precision with the most recent data:

2016-2020 ACS file for

- a) Factor 1 - <8th grade education
- b) Factor 2 - HS diploma
- c) Factor 4 - Median Income
- d) Factor 5 - Income disparity
- e) Factor 10 - Unemployment Rate (age 16+)

2016-2020 BRFSS five-year dataset for

- a) Factor 9 - Owner Occupied Housing
- b) Factor 11 - Below poverty level
- c) Factor 12 - Below 150% of poverty level
- d) Factor 13 - Single Family Households with Children

Use of these datasets for the nine indicators yielded no missing values for 99 Utah small health areas. The 2022 HII includes the following additional methodological changes. In the 2015-2017 HII, factor 5 was computed as $\log(100 * (\% \text{ with income} < 10k / \% \text{ with income} > 50k))$. In the 2022 HII the Gini coefficient from the ACS was used for income disparity. Another slight change refers to the unemployment rate. When using the 2015-2017 BRFSS data this indicator was for ages 18 and above. The ACS data for unemployment used for the 2022 HII is based on ages 16+.

Five year ACS PUMS data sets contain Zip Code Tabulation Areas (ZCTAs) instead of zip codes. ZCTAs closely approximate zip codes and were combined to produce rates for each small area for the five indicators from the ACS. R code to aggregate the ACS data into small area estimates (developed by UDOH CSTE Fellow Erica Bennion) is included in the appendix

Factor Analysis

After the nine indicators were computed for all 99 small areas, a factor analysis was performed using a principal component analysis with no priors and at least two factors. The factor scoring coefficients for the first factor were used as weights for the nine indicators to compute the HII. The SAS code is included in the appendix.

The following table shows the weights for the factors. The 2000 Singh coefficients are published coefficients from all census tracts in the United States using the 2000 census. The Singh coefficients are derived from more than nine factors and are only presented to show the relative consistency of the weights across data and years. As shown, the coefficients did not change significantly between 2017 and 2020. The negative factors stay negative and the positive factors stay positive. The magnitude of the factors is relatively similar from 2017 to 2022. Factor 5 was coded differently (Gini coefficient rather than log ratio of incomes) so the magnitude of that coefficient would necessarily change.

	Singh 2000 Coefficient	2017 HII Coefficient	2020 HII Coefficient
Factor 1 - <8th grade education	0.0969	0.1136	0.1650
Factor 2 - HS diploma	-0.1090	-0.1629	-0.1868
Factor 4 - Median Income	-0.1082	-0.1769	-0.1803
Factor 5 - Income disparity	0.0823	0.1613	0.0321
Factor 9 - Owner Occupied Housing	-0.0708	-0.1008	-0.1188
Factor 10 - Unemployment Rate (age 16+)	0.0826	0.1124	0.1454
Factor 11 - Below poverty level	0.1074	0.1830	0.1885
Factor 12 - Below 150% of poverty level	0.1157	0.1842	0.1994
Factor 13 - Single Family Households with Children	0.0810	0.0880	0.0940

Construction of the Index

The coefficients were multiplied by the indicators and summed for each small area to create a raw HII score.

Raw HII = coef1*factor1 + coef2* factor 2+...+coef13* factor 13.

The raw HII scores were then normalized to a mean of 100 and a standard deviation of 20 to make interpretation easier using the following formula.

$$\text{Standardized HII} = 100 + \left\{ \frac{20(\text{RawHII} - \text{meanHII})}{\text{StdHII}} \right\}$$

Where meanHII is the average of the raw HII scores across all small areas and

stdHII is the standard deviation of raw HII scores across all small areas

After the standardized HII was computed the small areas were grouped into five groups with the following formula:

Very Low = <80.00
Low = 80.00-94.99
Average = 95.00-104.99
High = 105.00-119.99
Very High = >=120.00

Appendix

Appendix 1 – R Code to aggregate American Community Survey Data into small areas (Erica Bennion)

```
#ZCTA to Small Area Aggregation -----
```

```
#This program creates a key that maps ZCTAs to Small Areas and includes HUD ratios to be applied later to modify population counts of ZCTAs that cross multiple Small Areas.
```

```
#Created: 1/12/21 by Erica Bennion
```

```
#Updated: 5/8/21 by Erica Bennion
```

```
#Load packages -----
```

```
library(tidyverse)
```

```
library(readxl)
```

```
#File locations -----
```

```
folder <- "//Utah/ddcp/users/BHP/CSTEFelloweb1/HII-Map/Data/"
```

```
#Load data -----
```

```
area_def <- read_excel(paste0(folder, "2018 Geographies MASTER.xlsx"))
```

```
# Received this spreadsheet of small areas, Zip codes, counties, cities, HUD ratios, etc. from Mike Friedrichs; spreadsheet was created by Kim Neerings
```

```
# These were the most recent data available as of January 2021
```

```
zcta_correct <- read_excel(paste0(folder, "Zip_to_zcta_complete.xlsx"))
```

This data set was downloaded from <https://udsmapper.org/zip-code-to-zcta-crosswalk/> in April 2021 and filtered to all Utah ZCTA/Zip codes that do not match

#Clean small area definition data

```
area_def <- area_def %>%
```

```
  mutate(HUD = ifelse(`HUD Res Ratio` == "--", 0.000, round(as.numeric(`HUD Res Ratio`), 3))) %>%  
#round HUD ratio to nearest 100th and set those marked "--" to 0
```

```
  rename("SmallArea" = `Small Area Name`) %>%
```

```
  select(Zip, SmallArea, HUD) %>% #select only the columns needed
```

#Note: this results in a warning about NA's introduced by coercion, but I don't see any NAs introduced so I think it's fine.

```
  mutate(Zip = as.character(Zip)) %>% #convert Zip to character to prepare for merge
```

```
  filter(!(SmallArea %in% c("Not Assigned", "Not assigned")), HUD > 0) #get rid of zip codes that aren't  
assigned to a small area and those that have a HUD ratio of 0 (if HUD = 0 then no one lives there, so  
there shouldn't be a row of data)
```

#Merge zip-code-to-small-area key with ZCTA-to-Zip-code key

```
area_def <- left_join(select(zcta_correct, ZIP_CODE, ZCTA), area_def, by = c("ZIP_CODE" = "Zip")) %>%
```

```
  select(-ZIP_CODE) %>% #get rid of Zip code column because we are now using ZCTA (used in ACS data)
```

```
  filter(!(is.na(SmallArea))) #get rid of rows missing a Small Area designation
```

#Get rid of duplicates

```
area_def <- area_def[-which(duplicated(area_def, by = c("ZCTA", "Small Area Name"))), ]
```

#Save data set

```
saveRDS(area_def, "C:/Users/CSTEFelloweb1/Documents/HII-Map/Data/ZCTA-HUD-Ratios.rds")
```

```
rm(zcta_correct, folder) #remove zip population count data set - no longer needed
```

#GINI or Income Inequality

```

#Packages

library(tidyverse)

#Load Income Inequality Data -----

gini <- read.csv("C:/Users/CSTEFelloweb1/Downloads/ACSDT5Y2020.B19083_2022-03-
22T171347/ACSDT5Y2020.B19083_data_with_overlays_2022-03-22T160012.csv")

gini$NAME <- gsub("ZCTA5 ", "", gini$NAME, fixed = T)

#Import ZCTA to Small Area crosswalk data

hud_ratios <- readRDS("C:/Users/CSTEFelloweb1/Documents/HII-Map/hii-map/Data/ZCTA-HUD-
Ratios.rds")

#Join ACS data to crosswalk data

gini <- left_join(hud_ratios, gini, by = c("ZCTA" = "NAME"))

rm(hud_ratios) #No longer need crosswalk data set

#Multiply everything by the HUD ratio

gini2 <- gini %>% #create temporary data set with only the columns that will be multiplied

  select(-c(ZCTA, SmallArea)) %>% #remove ZCTA and Small Area columns since they shouldn't be
multiplied

  mutate(across(.fns = ~ as.numeric(.) * gini$HUD)) #multiply HUD by all remaining columns

gini <- cbind(select(gini, SmallArea),select(gini2, -c(HUD, GEO_ID))) #add Small Area back in (don't need
ZCTA anymore), remove HUD column

rm(gini2)

#Separate estimates and MOEs into separate data sets

gini_est <- gini[ , 1:2]

gini_moe <- gini[ -1, c(1,3)]

rm(gini) #No longer need gini data set

```

```

#Sum estimates of ZCTAs within same small areas

gini_est <- gini_est %>%

  group_by(SmallArea) %>%

  summarise(across(everything(), ~ sum(., na.rm = T)))

#Sum squares of MOEs of ZCTAs within same small areas, i.e. sqrt(sum(x^2))

gini_se <- gini_moe %>%

  group_by(SmallArea) %>% #this group by is just a way of telling R to ignore the SmallArea column in
the function on the next line

  summarise(across(everything(), ~ . ^ 2)) %>% #square MOEs

  group_by(SmallArea) %>%

  summarise(across(everything(), ~ sqrt(sum(., na.rm = T))/1.645))

gini <- left_join(gini_est,gini_se)

colnames(gini) <- c("SmallArea", "gini", "gini_SE")

#Export data frame

saveRDS(gini, file = "C:/Users/CSTEFelloweb1/Documents/HII-Map/hii-map/Data/gini.rds")

#Clean environment

rm(convert_df)

```

Appendix 2 – SAS Code for Principal Component Analysis

```

libname HIIlib 'E:\Health Improvement Index\2017_2020 Updated HII';
proc factor data=HIIlib.adifactors99
  simple
  method=prin
  priors=one
  mineigen=1
  /* rotate=varimax */
  score
  /* round */
  plots=scree;

```

```

        *flag=desired-size-of-"significant"-factor-loadings;
        var factor1 factor2 factor4 factor5 factor9 factor10 factor11
factor12 factor13;
run;

data HIIfactors99;
    set HIIlib.adifactors99 (rename=(factor1=fact1 factor2=fact2
factor4=fact4 factor5=fact5 factor9=fact9
                                factor10=fact10 factor11=fact11
factor12=fact12 factor13=fact13)); *oldname=newname;
run;

proc factor data=work.HIIfactors99
    simple
    method=prin
    priors=one
    nfact=2
    /* rotate=varimax */
    /* round */
    score
    plots=scree
    out=adifacts
    outstat=coefs;
    *flag=desired-size-of-"significant"-factor-loadings;
    var fact1 fact2 fact4 fact5 fact9 fact10 fact11 fact12 fact13;
run;

data HIIcoefs (rename=(fact1=coef1 fact2=coef2 fact4=coef4 fact5=coef5
fact9=coef9 fact10=coef10 fact11=coef11 fact12=coef12 fact13=coef13));

    set coefs;
    if _type_='SCORE' and _name_='Factor1';
run;

proc contents data=HIIcoefs;run;
proc print data=HIIcoefs;run;

proc sort data=HIIfactors99;
    by sarea_r2;
run;
data HIIindex;
    merge HIIcoefs HIIfactors99 ;
    retain c1 c2 c4 c5 c9 c10 c11 c12 c13;
    drop coef1 coef2 coef4 coef5 coef9 coef10 coef11 coef12 coef13;
    if _n_=1 then do;
        c1=coef1;
        c2=coef2;
        c4=coef4;
        c5=coef5;
        c9=coef9;
        c10=coef10;
        c11=coef11;
        c12=coef12;
        c13=coef13;
    end;

```

```
BaseScore=(c1*fact1)+
           (c2*fact2)+
           (c4*fact4)+
           (c5*fact5)+
           (c9*fact9)+
           (c10*fact10)+
           (c11*fact11)+
           (c12*fact12)+
           (c13*fact13);
run;

proc print data=HIIIndex;
  var BaseScore;
  id sarea_r2;
run;

proc means data=HIIIndex;
  var BaseScore;
  output out=HIIstats;
run;

proc print data=HIIstats;run;
```

References

1. Area Deprivation and Widening Inequalities in US Mortality, 1969-1998. July 2003, Vol 93, No. 7 | American Journal of Public Health