GSA 2022 Abstract Code

Peter Sun

March 3, 2022

Contents

1 Load Packages and Data

```
library(tidyverse)
library(haven)
library(sjlabelled)
library(ggpubr)
library(kableExtra)
library(scales)

# Avoid select clashes
select <- dplyr::select
recode <- dplyr::recode
summarize <- dplyr::summarize
```

2 Re-Read Project-Level Environmental Variables

```
readRenviron(".Renviron")
```

3 Import Work, Age, Wave 14 (2018) Flag, and Nursing Home Status

```
# Import "randhrs1992_2018v1.dta"
rand.long <- read_dta(Sys.getenv("HRS_LONG"),
  col_select = c(hhid, pn, r14agey_e, inw14,
                 r14work, # Currently Working for Pay
                 s14hhidpn)) %>% # Spouse Identifier for Caregiving
  rename(worker = r14work) %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()

# Import nursing home status from "trk2018tr_r.dta"
tracker <- read_dta(Sys.getenv("HRS_TRACKER_2018_20"),
  col_select = c(hhid, pn, qnurshm))
```

4 Import Geography Data

```
# Import "HRSXREGION18.dta"
geo <- read_dta(Sys.getenv("HRS_REGION_2018_82"),
    col_select = c(hhid, pn, beale2013_18, region18)) %>%
  rename(rural = beale2013_18) %>%
  mutate(rural = recode(rural,
    `1` = "Urban",
    `2` = "Urban",
    `3` = "Rural",
    .default = NA_character_)) %>%
  mutate(region = recode(region18,
    `1` = "northeast",
    `2` = "northeast",
    `3` = "midwest",
    `4` = "midwest",
    `5` = "south",
    `6` = "south",
    `8` = "west",
    `9` = "west",
    .default = NA_character_)) %>%
  mutate(division = recode(region18,
    `1` = "northeast_new_england",
    `2` = "northeast_middle_atlantic",
    `3` = "midwest_east_north",
    `4` = "midwest_west_north",
    `5` = "south_south_atlantic",
    `6` = "south_east_south",
    `7` = "south_west_south",
    `8` = "west_mountain",
    `9` = "west_pacific", .default = NA_character_)) %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()
```

5 Import Volunteering Data

```
vol.18 <- read_dta(Sys.getenv("HRS_2018_FAT"),
    col_select = c(hhid, pn, qg086)) %>%
  rename(volunteer = qg086) %>%
  mutate(volunteer = recode(volunteer,
    `1` = 1,
    `5` = 0,
    .default = NA_real_))
table(vol.18$volunteer)
```

```
##
##     0     1
## 11328  5772
```

6 Import Caregiving Data

```
# Import 2018 RAND Fat File (File name: h04f1c.dta)
care18 <- read_dta(Sys.getenv("HRS_2018_FAT"),
                   col_select = c("hhidpn", "hhid", "pn",

                                  # adl helpers
                                  starts_with("qg033_"),

                                  # iadl helpers
                                  starts_with("qg055_"),

                                  # caregiving grandchildren
                                  "qe060",

                                  # caregiving parental personal
                                  "qf119",

                                  # caregiving parental errands
                                  "qf139")) %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()

# Identify participants who had an ADL or IADL helper
# spouse_helper_sum counts the number of 2s (spouse helper) in qg033 and qg055
# has_spouse_helper is 1 if there is at least one 2s, 0 if none
care18b <- care18 %>%
  mutate(spouse_helper_sum = rowSums(
    ifelse(
      select(., starts_with("qg033") | starts_with("qg055")) == 2, 1, 0
    ), na.rm = TRUE)) %>%
  mutate(has_spouse_helper = ifelse(spouse_helper_sum >= 1, 1, 0))

# Extract participants who have a spousal ADL/IADL caregiver
# Merge their spouse PN
# Then create a dataset with hhid and pn of spouse and an indication of
# whether or not they are a spousal caregiver
spousal_caregivers.18 <- care18b %>%
  filter(has_spouse_helper == 1) %>%
  select(hhid, pn, has_spouse_helper) %>%
  left_join(rand.long %>% select(hhid, pn, s14hhidpn),
            by = c("hhid", "pn")) %>%
  select(hhidpn = s14hhidpn, caregiver_spousal = has_spouse_helper)

# Merge the spousal_caregivers data back to the dataset
care18c <- care18b %>%
  left_join(spousal_caregivers.18, by = c("hhidpn"))

# Format parental/grandchildren caregivers
# Create caregiver_parental if either personal or errands == 1
care18d <- care18c %>%
  mutate(across(.cols = c(qe060, qf119, qf139),
```

```
    ~recode(., `1` = 1, `5` = 0, `8` = 0, `9` = 0,
            .default = NA_real_))) %>%
  rename(caregiver_grandchildren = qe060,
         caregiver_parental_personal = qf119,
         caregiver_parental_errands = qf139) %>%
  mutate(caregiver_parental =
    ifelse(caregiver_parental_personal == 1 | caregiver_parental_errands == 1,
           1, 0))

# Create general category of caregiving (Note: ignore zeroes)
care18e <- care18d %>%
  select(hhid, pn, cs = caregiver_spousal, cp = caregiver_parental,
         cg = caregiver_grandchildren) %>%
  mutate(Caregiver_Sum = rowSums(select(., cs:cg), na.rm = T),
         caregiver = ifelse(Caregiver_Sum >= 1, 1, 0)) %>%
  select(hhid, pn, caregiver)
```

7 Merge Datasets and Create Multiple Productive Activities

```
df <- rand.long %>%
  left_join(tracker, by = c("hhid", "pn")) %>%
  filter(inw14 == 1) %>% # in wave 14
  filter(qnurshm %in% c(5, 6, 7)) %>% # community-dwelling
  filter(r14agey_e >= 65) %>% # age 65+
  left_join(geo, by = c("hhid", "pn")) %>%
  left_join(vol.18, by = c("hhid", "pn")) %>%
  left_join(care18e, by = c("hhid", "pn")) %>%
  mutate(multi = ifelse(volunteer == 1 | caregiver == 1 | worker == 1, 1, 0))
```

7.1 Study Sample Size

```
nrow(df) # 2018 HRS sample of age 65+ community-dwelling individuals
```

```
## [1] 8728
```

8 Helper Functions

```
# Function for contingency table
get_kab <- function(data, geo, iv) {
  data %>%
    count({{ geo }}, rural, {{ iv }}) %>%
    group_by({{ geo }}, rural) %>%
    filter(!is.na({{ geo }})) %>%
    filter(!is.na(rural)) %>%
    filter(!is.na({{ iv }})) %>%
    mutate(pct = scales::percent(n / sum(n), accuracy = .1)) %>%
    ungroup() %>%
    filter({{ iv }} == 1) %>%
    select({{ geo }}, rural, pct)
}

# Function for two proportion z-test
get_prop <- function(data, geo, iv) {
  data %>%
    count({{ geo }}, rural, {{ iv }}) %>%
    group_by({{ geo }}, rural) %>%
    filter(!is.na({{ geo }})) %>%
    filter(!is.na(rural)) %>%
    filter(!is.na({{ iv }})) %>%
    mutate(sum = sum(n)) %>%
    ungroup() %>%
    filter({{ iv }} == 1) %>%
    pivot_wider(names_from = "rural", values_from = n:sum) %>%
    rowwise() %>%
    mutate(p = prop.test(x = c(n_Rural, n_Urban),
                         n = c(sum_Rural, sum_Urban))$p.value,
           rural_prop = prop.test(x = c(n_Rural, n_Urban),
                                  n = c(sum_Rural, sum_Urban))$estimate[1],
           urban_prop = prop.test(x = c(n_Rural, n_Urban),
                                  n = c(sum_Rural, sum_Urban))$estimate[2]) %>%
    mutate(rural_prop = percent(rural_prop, accuracy = .1),
           urban_prop = percent(urban_prop, accuracy = .1)) %>%
    kbl(booktabs = T, linesep = "", digits = 1) %>%
    kable_styling(position = "center") %>%
    kable_styling(latex_options = c("striped", "hold_position"))
}
```

9 Results

9.1 Census Region Statistics

```
r1 <- get_kab(df, region, worker) %>% rename(worker = pct)
r2 <- get_kab(df, region, volunteer) %>% rename(volunteer = pct)
r3 <- get_kab(df, region, caregiver) %>% rename(caregiver = pct)
r4 <- get_kab(df, region, multi) %>% rename(multiple = pct)
r1 %>%
  left_join(r2, by = c("region", "rural")) %>%
  left_join(r3, by = c("region", "rural")) %>%
  left_join(r4, by = c("region", "rural")) %>%
  kbl(booktabs = T, linesep = "", digits = 1) %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

region	rural	worker	volunteer	caregiver	multiple
midwest	Rural	20.9%	42.0%	27.2%	62.4%
midwest	Urban	14.8%	37.7%	26.5%	57.9%
northeast	Rural	16.4%	29.5%	19.8%	50.3%
northeast	Urban	19.3%	31.5%	25.5%	57.2%
south	Rural	18.3%	31.0%	29.0%	56.0%
south	Urban	19.3%	33.1%	25.4%	56.4%
west	Rural	19.3%	37.0%	25.1%	59.7%
west	Urban	20.6%	30.2%	24.9%	56.3%

9.2 Census Division Statistics

```
d1 <- get_kab(df, division, worker) %>% rename(worker = pct)
d2 <- get_kab(df, division, volunteer) %>% rename(volunteer = pct)
d3 <- get_kab(df, division, caregiver) %>% rename(caregiver = pct)
d4 <- get_kab(df, division, multi) %>% rename(multiple = pct)
d1 %>%
  left_join(d2, by = c("division", "rural")) %>%
  left_join(d3, by = c("division", "rural")) %>%
  left_join(d4, by = c("division", "rural")) %>%
  kbl(booktabs = T, linesep = "", digits = 1) %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

division	rural	worker	volunteer	caregiver	multiple
midwest_east_north	Rural	21.3%	39.0%	28.2%	60.8%
midwest_east_north	Urban	14.9%	36.9%	26.6%	58.4%
midwest_west_north	Rural	20.3%	46.8%	25.6%	64.9%
midwest_west_north	Urban	14.4%	39.5%	26.3%	56.6%
northeast_middle_atlantic	Rural	15.1%	26.8%	20.3%	47.6%
northeast_middle_atlantic	Urban	19.6%	32.5%	26.9%	58.6%
northeast_new_england	Rural	20.5%	38.5%	17.9%	59.0%
northeast_new_england	Urban	18.4%	28.7%	21.5%	53.3%
south_east_south	Rural	18.9%	28.8%	35.1%	60.4%
south_east_south	Urban	20.7%	32.1%	26.2%	56.0%
south_south_atlantic	Rural	18.1%	31.9%	26.5%	54.2%
south_south_atlantic	Urban	19.0%	33.3%	25.3%	56.4%
south_west_south	Rural	14.7%	28.8%	25.1%	51.8%
south_west_south	Urban	17.9%	24.6%	25.7%	50.9%
west_mountain	Rural	23.6%	35.6%	24.0%	60.3%
west_mountain	Urban	16.9%	31.3%	23.5%	53.3%
west_pacific	Rural	7.8%	40.6%	28.1%	57.8%
west_pacific	Urban	21.9%	29.8%	25.4%	57.5%

9.3 Rural-Urban Comparisons (Two-Proportion Z Tests)

9.3.1 Region: Worker, Volunteer, Caregiver, and Multiple

```
get_prop(df, region, worker)
```

region	worker	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest	1	159	171	760	1158	0.0	20.9%	14.8%
northeast	1	27	203	165	1053	0.4	16.4%	19.3%
south	1	140	374	764	1941	0.6	18.3%	19.3%
west	1	46	298	238	1449	0.7	19.3%	20.6%

```
get_prop(df, region, volunteer)
```

region	volunteer	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest	1	320	436	761	1157	0.1	42.0%	37.7%
northeast	1	49	332	166	1054	0.7	29.5%	31.5%
south	1	237	643	765	1944	0.3	31.0%	33.1%
west	1	88	438	238	1448	0.0	37.0%	30.2%

```
get_prop(df, region, caregiver)
```

region	caregiver	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest	1	207	308	761	1162	0.8	27.2%	26.5%
northeast	1	33	270	167	1057	0.1	19.8%	25.5%
south	1	222	495	765	1947	0.1	29.0%	25.4%
west	1	60	362	239	1454	1.0	25.1%	24.9%

```
get_prop(df, region, multi)
```

region	multi	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest	1	474	669	760	1156	0.1	62.4%	57.9%
northeast	1	83	603	165	1054	0.1	50.3%	57.2%
south	1	428	1094	764	1941	0.9	56.0%	56.4%
west	1	142	816	238	1449	0.4	59.7%	56.3%

9.3.2 Division: Worker and Volunteer

```
get_prop(df, division, worker)
```

division	worker	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest_east_north	1	99	122	464	817	0.0	21.3%	14.9%
midwest_west_north	1	60	49	296	341	0.1	20.3%	14.4%
northeast_middle_atlantic	1	19	153	126	781	0.3	15.1%	19.6%
northeast_new_england	1	8	50	39	272	0.9	20.5%	18.4%
south_east_south	1	42	68	222	328	0.7	18.9%	20.7%
south_south_atlantic	1	98	306	542	1613	0.7	18.1%	19.0%
south_west_south	1	45	118	307	661	0.3	14.7%	17.9%
west_mountain	1	41	67	174	396	0.1	23.6%	16.9%
west_pacific	1	5	231	64	1053	0.0	7.8%	21.9%

```
get_prop(df, division, volunteer)
```

division	volunteer	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest_east_north	1	181	302	464	818	0.5	39.0%	36.9%
midwest_west_north	1	139	134	297	339	0.1	46.8%	39.5%
northeast_middle_atlantic	1	34	254	127	782	0.2	26.8%	32.5%
northeast_new_england	1	15	78	39	272	0.3	38.5%	28.7%
south_east_south	1	64	105	222	327	0.5	28.8%	32.1%
south_south_atlantic	1	173	538	543	1617	0.6	31.9%	33.3%
south_west_south	1	89	163	309	662	0.2	28.8%	24.6%
west_mountain	1	62	124	174	396	0.4	35.6%	31.3%
west_pacific	1	26	314	64	1052	0.1	40.6%	29.8%

9.3.3 Division: Caregiver and Multiple

```
get_prop(df, division, caregiver)
```

division	caregiver	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest_east_north	1	131	218	464	820	0.6	28.2%	26.6%
midwest_west_north	1	76	90	297	342	0.9	25.6%	26.3%
northeast_middle_atlantic	1	26	211	128	783	0.1	20.3%	26.9%
northeast_new_england	1	7	59	39	274	0.8	17.9%	21.5%
south_east_south	1	78	86	222	328	0.0	35.1%	26.2%
south_south_atlantic	1	144	409	543	1619	0.6	26.5%	25.3%
south_west_south	1	78	171	311	665	0.9	25.1%	25.7%
west_mountain	1	42	93	175	396	1.0	24.0%	23.5%
west_pacific	1	18	269	64	1058	0.7	28.1%	25.4%

```
get_prop(df, division, multi)
```

division	multi	n_Rural	n_Urban	sum_Rural	sum_Urban	p	rural_prop	urban_prop
midwest_east_north	1	282	477	464	817	0.4	60.8%	58.4%
midwest_west_north	1	192	192	296	339	0.0	64.9%	56.6%
northeast_middle_atlantic	1	60	458	126	782	0.0	47.6%	58.6%
northeast_new_england	1	23	145	39	272	0.6	59.0%	53.3%
south_east_south	1	134	183	222	327	0.3	60.4%	56.0%
south_south_atlantic	1	294	911	542	1614	0.4	54.2%	56.4%
south_west_south	1	159	337	307	662	0.9	51.8%	50.9%
west_mountain	1	105	211	174	396	0.1	60.3%	53.3%
west_pacific	1	37	605	64	1053	1.0	57.8%	57.5%