Mortality and SRH: GSS 2014

Christine Lucille Kuryla

2025-01-03

Context

Self-rated health (SRH) is often used because of its well-documented association with mortality, even beyond comorbidities and other covariates. Here, we explore whether the predictive power of SRH on mortality has changed over the years in the GSS dataset.

We will use the General Social Survey (GSS). For the years 1980 - 2010, there is a variable `death` which is the possible vital status as of 2014 (SRH is not availabe for 1978). https://gssdataexplorer.norc.org/variables/7426/vshow

GSS is done approximately biannually, so we have 19 waves of data.

https://chatgpt.com/share/6777c9b5-551c-8002-9603-e2d1856253e7 https://claude.ai/chat/6c281c0b-d282-43af-a648-044529888668

Limitations

The optimal analysis here would be survival analysis, however, we do know know the year of death for each person, only whether they were alive or deceased in 2014.

We must recognize the limitation of a single binary outcome ("Alive/Dead by 2014"). The simplest binary logistic approach (alive vs. dead by 2014) effectively treats someone interviewed in 1978 and followed for 36 years the same as someone interviewed in 2010 and followed for only 4 years. If the older wave has higher mortality simply because they've had more time to die, that might artificially inflate differences across survey years. Consequence: Period effect might get conflated with time at risk. We may see stronger SRH–mortality associations in older data simply because more events (deaths) could have occurred.

Load data

```
#data_gss <- read_csv(here("data/cleaned/gss_groups.csv")) %>%

data_gss <- read_csv(here("data/extracted_gss_variables.csv")) %>%
  filter(cohort != 9999) %>%
  filter(!(is.na(death))) %>%
  filter(!(is.na(health))) %>%
#  na.omit() %>%
  mutate(health = 5 - health)  %>%  # reverse the coding so it's more intuitive (higher number for exce
  mutate(happy = 4 - happy) %>% # same
  mutate(life = 4 - life) %>% # reverse again, these variables tend to be unintuitively ordered!!!
  mutate(satfin = 4 - satfin) %>% # same again!
    mutate(
    health_cat = factor(health,
```

```r
                        levels = 1:4,
                        labels = c("Poor", "Fair", "Good", "Excellent")),
    # period_cut_6 = as.factor(cut(data_gss$year, 6)),
    # period_cut_10 = as.factor(cut(data_gss$year, 10)),
    # period_cut_12 = as.factor(cut(data_gss$year, 12)),
    # period_groups = as.factor(cut(data_gss$year, 12)),
    # period_7yr = as.factor(
    #                         cut(
    #                         year,
    #                         breaks = c(1973, 1982, 1990, 1998, 2006, 2014, Inf),
    #                         labels = c("1974-1982", "1983-1990", "1991-1998",
    #                                    "1999-2006", "2007-2014", "2015-2022"),
    #                         right = TRUE
    #                         )
    #                       ),
    # period_5yr = as.factor(
    #                         cut(
    #                         year,
    #                         breaks = c(1973, 1978, 1984, 1989, 1994, 1999, 2004, 2009 ,2014, Inf),
    #                         labels = c("1974-1979", "1982-1990", "1990-1998",
    #                                    "1998-2006", "2006-2014", "2014-2022"),
    #                         right = TRUE
    #                         )
    #                       ),
    period_decade = as.factor(
                          cut(
                          year,
                          breaks = c(1973, 1979, 1989, 1999, 2009, 2019, Inf),
                          labels = c("1974-1979", "1980-1989", "1990-1999",
                                     "2000-2009", "2010-2019", "2020-2024"),
                          right = TRUE
                          )
                        ),
    age_group = as.factor(
                          cut(
                          age,
                          breaks = c(17, 29, 39, 49, 59, 69, Inf),
                          labels = c("18-29", "30-39", "40-49", "50-59", "60-69", "70+"),
                          right = TRUE
                        )),
    age_groups = as.factor(
                          cut(
                          age,
                          breaks = c(17, 29, 39, 49, 59, 69, Inf),
                          labels = c("18-29", "30-39", "40-49", "50-59", "60-69", "70+"),
                          right = TRUE
                        )),
    age_group_small = as.factor(
                          cut(
                            age,
                            breaks = c(seq(15, 75, by = 5), Inf),  # Define breaks up to 75 and inclu
                            labels = c("16-20", "21-25", "26-30", "31-35", "36-40", "41-45", "46-50",
                            right = FALSE  # Makes intervals left-closed, i.e., [x, y)
```

```r
                          )
                        )
                        ,
    generation_5total = factor(
      case_when(
        cohort >= 1901 & cohort <= 1927 ~ "Greatest (1901-1927)",
        cohort >= 1928 & cohort <= 1945 ~ "Silent (1928-1945)",
        cohort >= 1946 & cohort <= 1964 ~ "Boomers (1946-1964)",
        cohort >= 1965 & cohort <= 1980 ~ "Gen X (1965-1980)",
        # cohort >= 1981 & cohort <= 1996 ~ "Millennials (1981-1996)",
        # cohort >= 1997 & cohort <= 2012 ~ "Gen Z (1997-2012)",
        cohort >= 1981  ~ "Millennials / Gen Z (1981-2004)",
        TRUE ~ "Other"
      ),
      levels = c(
        "Greatest (1901-1927)",
        "Silent (1928-1945)",
        "Boomers (1946-1964)",
        "Gen X (1965-1980)",
        "Millennials / Gen Z (1981-2004)"
       # "Millennials (1981-1996)",
      #  "Gen Z (1997-2012)"#,
     #   "Other"
      )
    ),
    generation_10total = factor(
      case_when(
        generation_5total == "Greatest (1901-1927)" & cohort <= 1914 ~ "Greatest Early (1901-1914)",
        generation_5total == "Greatest (1901-1927)" & cohort > 1914 ~ "Greatest Late (1915-1927)",
        generation_5total == "Silent (1928-1945)" & cohort <= 1936 ~ "Silent Early (1928-1936)",
        generation_5total == "Silent (1928-1945)" & cohort > 1936 ~ "Silent Late (1937-1945)",
        generation_5total == "Boomers (1946-1964)" & cohort <= 1955 ~ "Boomers Early (1946-1955)",
        generation_5total == "Boomers (1946-1964)" & cohort > 1955 ~ "Boomers Late (1956-1964)",
        generation_5total == "Gen X (1965-1980)" & cohort <= 1972 ~ "Gen X Early (1965-1972)",
        generation_5total == "Gen X (1965-1980)" & cohort > 1972 ~ "Gen X Late (1973-1980)",
        cohort >= 1981 & cohort <= 1988 ~ "Millennials Early (1981-1988)",
        cohort > 1988 ~ "Millennials Late / Gen Z Early (1989-2004)",
        # generation == "Millennials (1981-1996)" & cohort > 1988 ~ "Millennials Late (1989-1996)",
        # generation == "Gen Z (1997-2012)" & cohort <= 2004 ~ "Gen Z Early (1997-2004)",
        # generation == "Gen Z (1997-2012)" & cohort > 2004 ~ "Gen Z Late (2005-2012)",
        TRUE ~ "Other"
      ),
      levels = c(
        "Greatest Early (1901-1914)", "Greatest Late (1915-1927)",
        "Silent Early (1928-1936)", "Silent Late (1937-1945)",
        "Boomers Early (1946-1955)", "Boomers Late (1956-1964)",
        "Gen X Early (1965-1972)", "Gen X Late (1973-1980)",
        "Millennials Early (1981-1988)", "Millennials Late / Gen Z Early (1989-2004)"
      #  "Millennials Early (1981-1988)", "Millennials Late (1989-1996)",
      #  "Gen Z Early (1997-2004)", "Gen Z Late (2005-2012)"#,
     #   "Other"
      )
    ),
```

```r
    generation_15total = factor(
      case_when(
        cohort >= 1900 & cohort <= 1910 ~ "Greatest Early (1901-1910)",
        cohort >= 1911 & cohort <= 1918 ~ "Greatest Mid (1911-1918)",
        cohort >= 1919 & cohort <= 1927 ~ "Greatest Late (1919-1927)",
        cohort >= 1928 & cohort <= 1934 ~ "Silent Early (1928-1934)",
        cohort >= 1935 & cohort <= 1940 ~ "Silent Mid (1935-1940)",
        cohort >= 1941 & cohort <= 1945 ~ "Silent Late (1941-1945)",
        cohort >= 1945 & cohort <= 1951 ~ "Boomers Early (1946-1951)",
        cohort >= 1952 & cohort <= 1958 ~ "Boomers Mid (1952-1958)",
        cohort >= 1959 & cohort <= 1964 ~ "Boomers Late (1959-1964)",
        cohort >= 1965 & cohort <= 1970 ~ "Gen X Early (1965-1970)",
        cohort >= 1971 & cohort <= 1976 ~ "Gen X Mid (1971-1976)",
        cohort >= 1977 & cohort <= 1980 ~ "Gen X Late (1977-1980)",
        cohort >= 1981 & cohort <= 1986 ~ "Millennials Early (1981-1986)",
        cohort >= 1987 & cohort <= 1992 ~ "Millennials Mid (1987-1992)",
        cohort >= 1993 ~ "Millennials Late / Gen Z (1993-2004)",
    #    generation == "Gen Z (1997-2012)" & cohort <= 2002 ~ "Gen Z Early (1997-2002)",
    #    generation == "Gen Z (1997-2012)" & cohort > 2002 & cohort <= 2008 ~ "Gen Z Mid (2003-2008)",
    #    generation == "Gen Z (1997-2012)" & cohort > 2008 ~ "Gen Z Late (2009-2012)",
        TRUE ~ "Other"
      ),
      levels = c(
        "Greatest Early (1901-1910)", "Greatest Mid (1911-1918)", "Greatest Late (1919-1927)",
        "Silent Early (1928-1934)", "Silent Mid (1935-1940)", "Silent Late (1941-1945)",
        "Boomers Early (1946-1951)", "Boomers Mid (1952-1958)", "Boomers Late (1959-1964)",
        "Gen X Early (1965-1970)", "Gen X Mid (1971-1976)", "Gen X Late (1977-1980)",
        "Millennials Early (1981-1986)", "Millennials Mid (1987-1992)",
        "Millennials Late / Gen Z (1993-2004)"
        #"Millennials Late (1993-1996)",
      #  "Gen Z Early (1997-2002)", "Gen Z Mid (2003-2008)", "Gen Z Late (2009-2012)" #,
      #  "Other"
      )
    )
  ) %>%
  mutate(
    # 2-year periods
    period_2yr = as.factor(
      cut(
        year,
        breaks = c(
          1973, 1975, 1977, 1979, 1981, 1983, 1985, 1987, 1989, 1991,
          1993, 1995, 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011,
          2013, 2015, 2017, 2019, 2021, Inf
        ),
        labels = c(
          "1974-1975", "1976-1977", "1978-1979", "1980-1981", "1982-1983",
          "1984-1985", "1986-1987", "1988-1989", "1990-1991", "1992-1993",
          "1994-1995", "1996-1997", "1998-1999", "2000-2001", "2002-2003",
          "2004-2005", "2006-2007", "2008-2009", "2010-2011", "2012-2013",
          "2014-2015", "2016-2017", "2018-2019", "2020-2021", "2022-2022"
        ),
        right = TRUE
```

```
    )
  ),

  # 3-year periods
  period_3yr = as.factor(
    cut(
      year,
      breaks = c(
        1973, 1976, 1979, 1982, 1985, 1988, 1991, 1994, 1997, 2000,
        2003, 2006, 2009, 2012, 2015, 2018, 2021, Inf
      ),
      labels = c(
        "1974-1976", "1977-1979", "1980-1982", "1983-1985",
        "1986-1988", "1989-1991", "1992-1994", "1995-1997",
        "1998-2000", "2001-2003", "2004-2006", "2007-2009",
        "2010-2012", "2013-2015", "2016-2018", "2019-2021", "2022-2022"
      ),
      right = TRUE
    )
  ),

  # 5-year periods
  period_5yr = as.factor(
    cut(
      year,
      breaks = c(
        1973, 1978, 1983, 1988, 1993, 1998, 2003, 2008, 2013, 2018, Inf
      ),
      labels = c(
        "1974-1978", "1979-1983", "1984-1988", "1989-1993",
        "1994-1998", "1999-2003", "2004-2008", "2009-2013",
        "2014-2018", "2019-2022"
      ),
      right = TRUE
    )
  ),

  # 7-year periods
  period_7yr = as.factor(
    cut(
      year,
      breaks = c(
        1973, 1980, 1987, 1994, 2001, 2008, 2015, Inf
      ),
      labels = c(
        "1974-1980", "1981-1987", "1988-1994",
        "1995-2001", "2002-2008", "2009-2015", "2016-2022"
      ),
      right = TRUE
    )
  ),

  # 8-year periods
```

```r
    period_8yr = as.factor(
      cut(
        year,
        breaks = c(
          1973, 1981, 1989, 1997, 2005, 2013, 2021, Inf
        ),
        labels = c(
          "1974-1981", "1982-1989", "1990-1997",
          "1998-2005", "2006-2013", "2014-2021", "2022-2022"
        ),
        right = TRUE
      )
    ),

    # 10-year periods
    period_10yr = as.factor(
      cut(
        year,
        breaks = c(
          1973, 1983, 1993, 2003, 2013, Inf
        ),
        labels = c(
          "1974-1983", "1984-1993", "1994-2003",
          "2004-2013", "2014-2022"
        ),
        right = TRUE
      )
    )
  ) %>%
  mutate(period_7total = period_7yr,
         period_10total = period_5yr,
         period_17total = period_3yr) %>%
  mutate(
    # -----------------------
    # 6 AGE CATEGORIES
    # -----------------------
    age_6cat = as.factor(
      cut(
        age,
        # 7 breakpoints --> 6 intervals
        breaks = c(17, 30, 40, 50, 60, 70, 89),
        labels = c(
          "18-29",  # (17, 30]
          "30-39",  # (30, 40]
          "40-49",  # (40, 50]
          "50-59",  # (50, 60]
          "60-69",  # (60, 70]
          "70-89"   # (70, 89]
        ),
        right = TRUE
      )
    ),
```

```r
# ------------------------
# 10 AGE CATEGORIES
# ------------------------
age_10cat = as.factor(
  cut(
    age,
    # 11 breakpoints --> 10 intervals
    breaks = c(17, 24, 31, 38, 45, 52, 59, 66, 73, 80, 89),
    labels = c(
      "18-24",  # (17, 24]
      "25-31",  # (24, 31]
      "32-38",  # (31, 38]
      "39-45",  # (38, 45]
      "46-52",  # (45, 52]
      "53-59",  # (52, 59]
      "60-66",  # (59, 66]
      "67-73",  # (66, 73]
      "74-80",  # (73, 80]
      "81-89"   # (80, 89]
    ),
    right = TRUE
  )
),

# ------------------------
# 16 AGE CATEGORIES
# ------------------------
age_16cat = as.factor(
  cut(
    age,
    # 17 breakpoints --> 16 intervals
    breaks = c(17, 22, 27, 32, 37, 42, 47, 52, 57, 62, 67, 72, 77, 82, 85, 87, 89),
    labels = c(
      "18-22",  # (17, 22]
      "23-27",  # (22, 27]
      "28-32",  # (27, 32]
      "33-37",  # (32, 37]
      "38-42",  # (37, 42]
      "43-47",  # (42, 47]
      "48-52",  # (47, 52]
      "53-57",  # (52, 57]
      "58-62",  # (57, 62]
      "63-67",  # (62, 67]
      "68-72",  # (67, 72]
      "73-77",  # (72, 77]
      "78-82",  # (77, 82]
      "83-85",  # (82, 85]
      "86-87",  # (85, 87]
      "88-89"   # (87, 89]
    ),
    right = TRUE
  )
)
```

```r
  ) %>%
    mutate(srh_num = health,
         srh_cat = health_cat)
```

```
## Rows: 72390 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (13): year, cohort, age, health, sex, happy, life, educ, polviews, class...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
glimpse(data_gss)
```

```
## Rows: 30,783
## Columns: 35
## $ year                 <dbl> 1980, 1980, 1980, 1980, 1980, 1980, 1980, 1980, 198~
## $ cohort               <dbl> 1918, 1903, 1893, 1952, 1956, 1950, 1917, 1921, 191~
## $ age                  <dbl> 62, 77, 87, 28, 24, 30, 63, 59, 62, 38, 74, 44, 30,~
## $ health               <dbl> 3, 3, 3, 3, 4, 2, 4, 1, 3, 4, 2, 3, 2, 1, 2, 3, 4, ~
## $ sex                  <dbl> 1, 1, 2, 2, 2, 1, 1, 2, 1, 2, 2, 2, 1, 2, 1, 1, 2, ~
## $ happy                <dbl> 1, 3, 1, 3, 2, 1, 2, 2, 3, 2, 3, 3, 1, 2, 2, 2, 2, ~
## $ life                 <dbl> 3, 3, 1, 3, 3, 3, 2, 2, 3, 2, 2, 3, 3, 2, 2, 3, 2, ~
## $ educ                 <dbl> 9, 10, 4, 14, 12, 12, 12, 16, 17, 10, 12, 12, 11, 8~
## $ polviews             <dbl> 2, 6, NA, 2, 3, NA, 4, 4, 4, 2, 5, 4, 2, 3, 4, 3, 5~
## $ class                <dbl> 2, 2, 1, 2, 2, 2, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 3, ~
## $ death                <dbl> 2, 1, 2, 1, 1, 1, 2, 2, 2, 2, 2, 1, 1, 1, 2, 1, 1, ~
## $ wtsscomp             <dbl> 0.5112, 1.0225, 0.5112, 1.0225, 1.0225, 2.0450, 1.0~
## $ satfin               <dbl> 1, 3, 2, 1, 3, 2, 2, 2, 2, 2, 3, 1, 3, 2, 3, 1, 1, ~
## $ health_cat           <fct> Good, Good, Good, Good, Excellent, Fair, Excellent,~
## $ period_decade        <fct> 1980-1989, 1980-1989, 1980-1989, 1980-1989, 1980-19~
## $ age_group            <fct> 60-69, 70+, 70+, 18-29, 18-29, 30-39, 60-69, 50-59,~
## $ age_groups           <fct> 60-69, 70+, 70+, 18-29, 18-29, 30-39, 60-69, 50-59,~
## $ age_group_small      <fct> 61-65, 76+, 76+, 26-30, 21-25, 31-35, 61-65, 56-60,~
## $ generation_5total    <fct> Greatest (1901-1927), Greatest (1901-1927), NA, Boo~
## $ generation_10total   <fct> Greatest Late (1915-1927), Greatest Early (1901-191~
## $ generation_15total   <fct> Greatest Mid (1911-1918), Greatest Early (1901-1910~
## $ period_2yr           <fct> 1980-1981, 1980-1981, 1980-1981, 1980-1981, 1980-19~
## $ period_3yr           <fct> 1980-1982, 1980-1982, 1980-1982, 1980-1982, 1980-19~
## $ period_5yr           <fct> 1979-1983, 1979-1983, 1979-1983, 1979-1983, 1979-19~
## $ period_7yr           <fct> 1974-1980, 1974-1980, 1974-1980, 1974-1980, 1974-19~
## $ period_8yr           <fct> 1974-1981, 1974-1981, 1974-1981, 1974-1981, 1974-19~
## $ period_10yr          <fct> 1974-1983, 1974-1983, 1974-1983, 1974-1983, 1974-19~
## $ period_7total        <fct> 1974-1980, 1974-1980, 1974-1980, 1974-1980, 1974-19~
## $ period_10total       <fct> 1979-1983, 1979-1983, 1979-1983, 1979-1983, 1979-19~
## $ period_17total       <fct> 1980-1982, 1980-1982, 1980-1982, 1980-1982, 1980-19~
## $ age_6cat             <fct> 60-69, 70-89, 70-89, 18-29, 18-29, 18-29, 60-69, 50~
## $ age_10cat            <fct> 60-66, 74-80, 81-89, 25-31, 18-24, 25-31, 60-66, 53~
## $ age_16cat            <fct> 58-62, 73-77, 86-87, 28-32, 23-27, 28-32, 63-67, 58~
## $ srh_num              <dbl> 3, 3, 3, 3, 4, 2, 4, 1, 3, 4, 2, 3, 2, 1, 2, 3, 4, ~
## $ srh_cat              <fct> Good, Good, Good, Good, Excellent, Fair, Excellent,~
```

```r
summary(data_gss)
```

```
##       year          cohort          age            health           sex
```

```
## Min.   :1980   Min.   :1891   Min.   :18.00   Min.   :1.000   Min.   :1.000
## 1st Qu.:1988   1st Qu.:1937   1st Qu.:32.00   1st Qu.:3.000   1st Qu.:1.000
## Median :1996   Median :1952   Median :43.00   Median :3.000   Median :2.000
## Mean   :1995   Mean   :1949   Mean   :45.85   Mean   :3.014   Mean   :1.561
## 3rd Qu.:2002   3rd Qu.:1963   3rd Qu.:59.00   3rd Qu.:4.000   3rd Qu.:2.000
## Max.   :2010   Max.   :1992   Max.   :89.00   Max.   :4.000   Max.   :2.000
##
##     happy            life            educ          polviews
## Min.   :1.000   Min.   :1.000   Min.   : 0.00   Min.   :1.000
## 1st Qu.:2.000   1st Qu.:2.000   1st Qu.:12.00   1st Qu.:3.000
## Median :2.000   Median :2.000   Median :12.00   Median :4.000
## Mean   :2.184   Mean   :2.421   Mean   :12.96   Mean   :4.113
## 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:15.00   3rd Qu.:5.000
## Max.   :3.000   Max.   :3.000   Max.   :20.00   Max.   :7.000
## NA's   :2940    NA's   :4910    NA's   :66      NA's   :2638
##     class           death          wtsscomp          satfin
## Min.   :1.000   Min.   :1.000   Min.   :0.1827   Min.   :1.000
## 1st Qu.:2.000   1st Qu.:1.000   1st Qu.:0.5695   1st Qu.:1.000
## Median :2.000   Median :1.000   Median :0.9804   Median :2.000
## Mean   :2.468   Mean   :1.278   Mean   :1.0020   Mean   :2.011
## 3rd Qu.:3.000   3rd Qu.:2.000   3rd Qu.:1.1524   3rd Qu.:3.000
## Max.   :4.000   Max.   :2.000   Max.   :9.2076   Max.   :3.000
## NA's   :1642                                     NA's   :2864
##     health_cat      period_decade  age_group    age_groups   age_group_small
## Poor     : 1658   1974-1979:   0   18-29:6321   18-29:6321   31-35  : 3446
## Fair     : 5607   1980-1989:9580   30-39:6820   30-39:6820   36-40  : 3374
## Good     :14171   1990-1999:9947   40-49:5811   40-49:5811   26-30  : 3229
## Excellent: 9347   2000-2009:9992   50-59:4478   50-59:4478   41-45  : 3085
##                   2010-2019:1264   60-69:3593   60-69:3593   46-50  : 2726
##                   2020-2024:   0   70+  :3760   70+  :3760   21-25  : 2594
##                                                              (Other):12329
##                         generation_5total                 generation_10total
## Greatest (1901-1927)          : 4765   Boomers Early (1946-1955):6267
## Silent (1928-1945)            : 6493   Boomers Late (1956-1964) :6147
## Boomers (1946-1964)           :12414   Gen X Early (1965-1972)  :3783
## Gen X (1965-1980)             : 6013   Silent Late (1937-1945)  :3709
## Millennials / Gen Z (1981-2004):  953   Greatest Late (1915-1927):3392
## NA's                          :  145   (Other)                  :7340
##                                        NA's                     : 145
##                   generation_15total     period_2yr        period_3yr
## Boomers Mid (1952-1958)  : 4857    2006-2007: 3434   1998-2000:4905
## Boomers Late (1959-1964) : 3964    1984-1985: 2876   2004-2006:4753
## Boomers Early (1946-1951): 3593    1998-1999: 2705   1980-1982:3015
## Gen X Early (1965-1970)  : 3038    1996-1997: 2369   1992-1994:3000
## Greatest Late (1919-1927): 2542    2000-2001: 2200   1983-1985:2876
## (Other)                  :12675    1994-1995: 1968   1989-1991:2870
## NA's                     :  114    (Other)  :15231   (Other)  :9364
##     period_5yr        period_7yr        period_8yr       period_10yr
## 1994-1998:7042   1974-1980:1287   1974-1981:1287   1974-1983: 3015
## 2004-2008:6069   1981-1987:6333   1982-1989:8293   1984-1993: 9470
## 1984-1988:5568   1988-1994:6833   1990-1997:7242   1994-2003:10965
## 1999-2003:3923   1995-2001:7274   1998-2005:7947   2004-2013: 7333
## 1989-1993:3902   2002-2008:7792   2006-2013:6014   2014-2022:    0
## 1979-1983:3015   2009-2015:1264   2014-2021:    0
```

```
##  (Other)  :1264   2016-2022:   0   2022-2022:   0
##    period_7total    period_10total   period_17total  age_6cat      age_10cat
##  1974-1980:1287   1994-1998:7042   1998-2000:4905   18-29:7023   32-38   :4860
##  1981-1987:6333   2004-2008:6069   2004-2006:4753   30-39:6776   25-31   :4588
##  1988-1994:6833   1984-1988:5568   1980-1982:3015   40-49:5654   39-45   :4228
##  1995-2001:7274   1999-2003:3923   1992-1994:3000   50-59:4389   46-52   :3671
##  2002-2008:7792   1989-1993:3902   1983-1985:2876   60-69:3511   18-24   :3092
##  2009-2015:1264   1979-1983:3015   1989-1991:2870   70-89:3430   53-59   :2991
##  2016-2022:   0   (Other)  :1264   (Other)  :9364                (Other):7353
##    age_16cat        srh_num           srh_cat
##  33-37   : 3483   Min.   :1.000   Poor      : 1658
##  28-32   : 3398   1st Qu.:3.000   Fair      : 5607
##  38-42   : 3134   Median :3.000   Good      :14171
##  23-27   : 3060   Mean   :3.014   Excellent: 9347
##  43-47   : 2887   3rd Qu.:4.000
##  48-52   : 2562   Max.   :4.000
##  (Other):12259
```

```r
table(data_gss$generation_5total)
```

```
##
##           Greatest (1901-1927)              Silent (1928-1945)
##                           4765                            6493
##            Boomers (1946-1964)               Gen X (1965-1980)
##                          12414                            6013
## Millennials / Gen Z (1981-2004)
##                            953
```

```r
table(data_gss$generation_10total)
```

```
##
##                 Greatest Early (1901-1914)
##                                       1373
##                  Greatest Late (1915-1927)
##                                       3392
##                   Silent Early (1928-1936)
##                                       2784
##                    Silent Late (1937-1945)
##                                       3709
##                  Boomers Early (1946-1955)
##                                       6267
##                   Boomers Late (1956-1964)
##                                       6147
##                    Gen X Early (1965-1972)
##                                       3783
##                     Gen X Late (1973-1980)
##                                       2230
##              Millennials Early (1981-1988)
##                                        888
## Millennials Late / Gen Z Early (1989-2004)
##                                         65
```

```r
table(data_gss$generation_15total)
```

```
##
##          Greatest Early (1901-1910)             Greatest Mid (1911-1918)
```

```
##                                   765                                  1489
##             Greatest Late (1919-1927)             Silent Early (1928-1934)
##                                  2542                                  2134
##                Silent Mid (1935-1940)              Silent Late (1941-1945)
##                                  2062                                  2297
##             Boomers Early (1946-1951)              Boomers Mid (1952-1958)
##                                  3593                                  4857
##              Boomers Late (1959-1964)              Gen X Early (1965-1970)
##                                  3964                                  3038
##                  Gen X Mid (1971-1976)              Gen X Late (1977-1980)
##                                  2003                                   972
##         Millennials Early (1981-1986)          Millennials Mid (1987-1992)
##                                   768                                   185
## Millennials Late / Gen Z (1993-2004)
##                                     0
```

```r
table(data_gss$period_7total)
```

```
##
## 1974-1980 1981-1987 1988-1994 1995-2001 2002-2008 2009-2015 2016-2022
##      1287      6333      6833      7274      7792      1264         0
```

```r
table(data_gss$period_10total)
```

```
##
## 1974-1978 1979-1983 1984-1988 1989-1993 1994-1998 1999-2003 2004-2008 2009-2013
##         0      3015      5568      3902      7042      3923      6069      1264
## 2014-2018 2019-2022
##         0         0
```

```r
table(data_gss$period_17total)
```

```
##
## 1974-1976 1977-1979 1980-1982 1983-1985 1986-1988 1989-1991 1992-1994 1995-1997
##         0         0      3015      2876      2692      2870      3000      2369
## 1998-2000 2001-2003 2004-2006 2007-2009 2010-2012 2013-2015 2016-2018 2019-2021
##      4905      1723      4753      1316      1264         0         0         0
## 2022-2022
##         0
```

```r
table(data_gss$srh_num)
```

```
##
##     1     2     3     4
##  1658  5607 14171  9347
```

```r
summary(data_gss$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   32.00   43.00   45.85   59.00   89.00
```

Add more variables and inspect data

```r
data_gss <- data_gss %>%
  mutate(id = row_number()) %>%
  mutate(SRH = health) %>%
  mutate(birth_year = cohort) %>%
```

```r
  mutate(died_by_2014 = death - 1) %>%
  mutate(time_at_risk = 2014 - year)

# glimpse(data_gss)
# summary(data_gss)

table(data_gss$generation_5total)
```

```
##
##            Greatest (1901-1927)             Silent (1928-1945)
##                            4765                           6493
##             Boomers (1946-1964)              Gen X (1965-1980)
##                           12414                           6013
## Millennials / Gen Z (1981-2004)
##                             953
```

```r
table(data_gss$generation_10total)
```

```
##
##                  Greatest Early (1901-1914)
##                                        1373
##                   Greatest Late (1915-1927)
##                                        3392
##                    Silent Early (1928-1936)
##                                        2784
##                     Silent Late (1937-1945)
##                                        3709
##                   Boomers Early (1946-1955)
##                                        6267
##                    Boomers Late (1956-1964)
##                                        6147
##                     Gen X Early (1965-1972)
##                                        3783
##                      Gen X Late (1973-1980)
##                                        2230
##               Millennials Early (1981-1988)
##                                         888
## Millennials Late / Gen Z Early (1989-2004)
##                                          65
```

```r
table(data_gss$generation_15total)
```

```
##
##          Greatest Early (1901-1910)            Greatest Mid (1911-1918)
##                                 765                                1489
##           Greatest Late (1919-1927)            Silent Early (1928-1934)
##                                2542                                2134
##              Silent Mid (1935-1940)             Silent Late (1941-1945)
##                                2062                                2297
##           Boomers Early (1946-1951)             Boomers Mid (1952-1958)
##                                3593                                4857
##            Boomers Late (1959-1964)             Gen X Early (1965-1970)
##                                3964                                3038
##               Gen X Mid (1971-1976)              Gen X Late (1977-1980)
##                                2003                                 972
```

```
##        Millennials Early (1981-1986)          Millennials Mid (1987-1992)
##                                  768                                  185
## Millennials Late / Gen Z (1993-2004)
##                                    0
```

```r
table(data_gss$period_7total)
```

```
##
## 1974-1980 1981-1987 1988-1994 1995-2001 2002-2008 2009-2015 2016-2022
##      1287      6333      6833      7274      7792      1264         0
```

```r
table(data_gss$period_10total)
```

```
##
## 1974-1978 1979-1983 1984-1988 1989-1993 1994-1998 1999-2003 2004-2008 2009-2013
##         0      3015      5568      3902      7042      3923      6069      1264
## 2014-2018 2019-2022
##         0         0
```

```r
table(data_gss$period_17total)
```

```
##
## 1974-1976 1977-1979 1980-1982 1983-1985 1986-1988 1989-1991 1992-1994 1995-1997
##         0         0      3015      2876      2692      2870      3000      2369
## 1998-2000 2001-2003 2004-2006 2007-2009 2010-2012 2013-2015 2016-2018 2019-2021
##      4905      1723      4753      1316      1264         0         0         0
## 2022-2022
##         0
```

```r
table(data_gss$SRH)
```

```
##
##     1     2     3     4
##  1658  5607 14171  9347
```

```r
table(data_gss$died_by_2014)
```

```
##
##     0     1
## 22226  8557
```

```r
summary(data_gss$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   32.00   43.00   45.85   59.00   89.00
```

```r
df <- data_gss
```

Method 1: Control for “years of exposure” or “time at risk”

Basic Logistic Regression with SRH × Period and time_at_risk, Using Survey Weights

Goal: Test whether SRH’s predictive effect on mortality differs by period (or interview year), while controlling for how long each respondent was “at risk” before 2014, and properly accounting for survey weights.

Why SRH × Period? You want to see if SRH’s predictive power differs in earlier vs. later survey years. Why time_at_risk? People interviewed earlier (e.g., 1978) have more years to die before 2014 than those interviewed later (e.g., 2010).

Controlling for time_at_risk ensures that differences in mortality across periods aren't merely because of differing follow-up lengths.

Interpretation: time_at_risk adjusts for the fact that earlier surveys have longer follow-up. The interaction (SRH * factor(year)) tests whether SRH's effect differs systematically by year. interpret the result as an odds ratio.

```r
library(survey)
```

```
## Loading required package: grid

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loading required package: survival

##
## Attaching package: 'survey'

## The following object is masked from 'package:Hmisc':
##
##     deff

## The following object is masked from 'package:graphics':
##
##     dotchart
```

```r
# Simple survey design with only weights
des <- svydesign(
  id = ~1,           # no clustering variable here, if none available
  weights = ~wtsscomp,     # your survey weight variable name
  data = df
)

# Fit a logistic regression (svyglm) with SRH × factor(year) + time_at_risk
model_svy_logit_period <- svyglm(
  formula = died_by_2014 ~ SRH * factor(period_10total) + time_at_risk + age,
  design  = des,
  family  = quasibinomial(link = "logit")
)
# Note: quasibinomial is often used to get robust SEs with survey data;
# you can also try family=binomial if you prefer.

summary(model_svy_logit_period)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(period_10total) +
##     time_at_risk + age, design = des, family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
```

```
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         -6.816556   0.450760 -15.122  < 2e-16 ***
## SRH                                 -0.096212   0.056656  -1.698 0.089485 .
## factor(period_10total)1984-1988      0.439037   0.227972   1.926 0.054134 .
## factor(period_10total)1989-1993      0.862455   0.266472   3.237 0.001211 **
## factor(period_10total)1994-1998      1.247721   0.287503   4.340 1.43e-05 ***
## factor(period_10total)1999-2003      1.110456   0.351369   3.160 0.001577 **
## factor(period_10total)2004-2008      1.448151   0.393635   3.679 0.000235 ***
## factor(period_10total)2009-2013      1.151740   0.639611   1.801 0.071762 .
## time_at_risk                         0.126988   0.012320  10.307  < 2e-16 ***
## age                                  0.066380   0.001159  57.258  < 2e-16 ***
## SRH:factor(period_10total)1984-1988 -0.035025   0.069933  -0.501 0.616491
## SRH:factor(period_10total)1989-1993 -0.159122   0.075765  -2.100 0.035720 *
## SRH:factor(period_10total)1994-1998 -0.299263   0.071166  -4.205 2.62e-05 ***
## SRH:factor(period_10total)1999-2003 -0.258471   0.083733  -3.087 0.002025 **
## SRH:factor(period_10total)2004-2008 -0.366992   0.081577  -4.499 6.86e-06 ***
## SRH:factor(period_10total)2009-2013 -0.311459   0.197399  -1.578 0.114619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.060088)
##
## Number of Fisher Scoring iterations: 5
```

```r
# Fit a logistic regression (svyglm) with SRH × factor(year) + time_at_risk
model_svy_logit_period <- svyglm(
  formula = died_by_2014 ~ SRH * factor(period_7total) + time_at_risk + age,
  design  = des,
  family  = quasibinomial(link = "logit")
)
# Note: quasibinomial is often used to get robust SEs with survey data;
# you can also try family=binomial if you prefer.

summary(model_svy_logit_period)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(period_7total) +
##     time_at_risk + age, design = des, family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        -6.601018   0.415835 -15.874  < 2e-16 ***
## SRH                                -0.164707   0.081260  -2.027  0.04268 *
## factor(period_7total)1981-1987      0.284641   0.287296   0.991  0.32181
## factor(period_7total)1988-1994      0.795262   0.302946   2.625  0.00867 **
## factor(period_7total)1995-2001      1.046592   0.329633   3.175  0.00150 **
## factor(period_7total)2002-2008      1.110616   0.375696   2.956  0.00312 **
## factor(period_7total)2009-2015      0.957816   0.626431   1.529  0.12627
## time_at_risk                        0.122356   0.009241  13.240  < 2e-16 ***
## age                                 0.066335   0.001157  57.314  < 2e-16 ***
```

```
## SRH:factor(period_7total)1981-1987  0.039807   0.090161   0.442  0.65884
## SRH:factor(period_7total)1988-1994 -0.114398   0.090323  -1.267  0.20533
## SRH:factor(period_7total)1995-2001 -0.202006   0.091884  -2.198  0.02792 *
## SRH:factor(period_7total)2002-2008 -0.244080   0.095426  -2.558  0.01054 *
## SRH:factor(period_7total)2009-2015 -0.243060   0.205775  -1.181  0.23754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.056861)
##
## Number of Fisher Scoring iterations: 5
```

Interpretation of results

This model examines SRH × Period interactions while controlling for time_at_risk. The results show:

SRH becomes increasingly predictive of mortality in later periods The baseline SRH coefficient starts at -0.096 (p=0.089) The interaction terms become larger and more significant over time:

1994-1998: -0.299 (p<0.001) 2004-2008: -0.367 (p<0.001)

Age and time_at_risk remain highly significant predictors

This suggests the predictive power of SRH has strengthened over time, even after accounting for differential follow-up periods.

```r
#

# Fit a logistic regression (svyglm) with SRH × factor(cohort) + time_at_risk
model_svy_logit_cohort <- svyglm(
  formula = died_by_2014 ~ SRH * factor(generation_10total) + time_at_risk + age,
  design  = des,
  family  = quasibinomial(link = "logit")
)
# Note: quasibinomial is often used to get robust SEs with survey data;
# you can also try family=binomial if you prefer.

summary(model_svy_logit_cohort)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(generation_10total) +
##     time_at_risk + age, design = des, family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                                                                      Estimate
## (Intercept)                                                         -6.237538
## SRH                                                                  0.091059
## factor(generation_10total)Greatest Late (1915-1927)                  1.414712
## factor(generation_10total)Silent Early (1928-1936)                   1.100093
## factor(generation_10total)Silent Late (1937-1945)                    1.087790
## factor(generation_10total)Boomers Early (1946-1955)                  0.640100
## factor(generation_10total)Boomers Late (1956-1964)                   0.707227
## factor(generation_10total)Gen X Early (1965-1972)                    0.541643
```

```
## factor(generation_10total)Gen X Late (1973-1980)                          0.411724
## factor(generation_10total)Millennials Early (1981-1988)                   0.905185
## factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)    -10.491626
## time_at_risk                                                              0.113359
## age                                                                       0.059531
## SRH:factor(generation_10total)Greatest Late (1915-1927)                  -0.395433
## SRH:factor(generation_10total)Silent Early (1928-1936)                   -0.436080
## SRH:factor(generation_10total)Silent Late (1937-1945)                    -0.466698
## SRH:factor(generation_10total)Boomers Early (1946-1955)                  -0.329893
## SRH:factor(generation_10total)Boomers Late (1956-1964)                   -0.312156
## SRH:factor(generation_10total)Gen X Early (1965-1972)                    -0.267566
## SRH:factor(generation_10total)Gen X Late (1973-1980)                     -0.247801
## SRH:factor(generation_10total)Millennials Early (1981-1988)              -0.277202
## SRH:factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)  -0.044632
##                                                                         Std. Error
## (Intercept)                                                               0.696263
## SRH                                                                       0.094656
## factor(generation_10total)Greatest Late (1915-1927)                       0.304234
## factor(generation_10total)Silent Early (1928-1936)                        0.328665
## factor(generation_10total)Silent Late (1937-1945)                         0.356395
## factor(generation_10total)Boomers Early (1946-1955)                       0.388663
## factor(generation_10total)Boomers Late (1956-1964)                        0.440530
## factor(generation_10total)Gen X Early (1965-1972)                         0.549960
## factor(generation_10total)Gen X Late (1973-1980)                          0.737798
## factor(generation_10total)Millennials Early (1981-1988)                   1.093014
## factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)      0.917135
## time_at_risk                                                              0.006509
## age                                                                       0.006195
## SRH:factor(generation_10total)Greatest Late (1915-1927)                   0.106482
## SRH:factor(generation_10total)Silent Early (1928-1936)                    0.107096
## SRH:factor(generation_10total)Silent Late (1937-1945)                     0.106712
## SRH:factor(generation_10total)Boomers Early (1946-1955)                   0.104940
## SRH:factor(generation_10total)Boomers Late (1956-1964)                    0.110907
## SRH:factor(generation_10total)Gen X Early (1965-1972)                     0.138860
## SRH:factor(generation_10total)Gen X Late (1973-1980)                      0.201206
## SRH:factor(generation_10total)Millennials Early (1981-1988)               0.302211
## SRH:factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)   0.244937
##                                                                         t value
## (Intercept)                                                              -8.959
## SRH                                                                       0.962
## factor(generation_10total)Greatest Late (1915-1927)                       4.650
## factor(generation_10total)Silent Early (1928-1936)                        3.347
## factor(generation_10total)Silent Late (1937-1945)                         3.052
## factor(generation_10total)Boomers Early (1946-1955)                       1.647
## factor(generation_10total)Boomers Late (1956-1964)                        1.605
## factor(generation_10total)Gen X Early (1965-1972)                         0.985
## factor(generation_10total)Gen X Late (1973-1980)                          0.558
## factor(generation_10total)Millennials Early (1981-1988)                   0.828
## factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)    -11.440
## time_at_risk                                                             17.415
## age                                                                       9.610
## SRH:factor(generation_10total)Greatest Late (1915-1927)                  -3.714
## SRH:factor(generation_10total)Silent Early (1928-1936)                   -4.072
## SRH:factor(generation_10total)Silent Late (1937-1945)                    -4.373
```

```
## SRH:factor(generation_10total)Boomers Early (1946-1955)                         -3.144
## SRH:factor(generation_10total)Boomers Late (1956-1964)                          -2.815
## SRH:factor(generation_10total)Gen X Early (1965-1972)                           -1.927
## SRH:factor(generation_10total)Gen X Late (1973-1980)                            -1.232
## SRH:factor(generation_10total)Millennials Early (1981-1988)                     -0.917
## SRH:factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)  -0.182
##                                                                           Pr(>|t|)
## (Intercept)                                                                < 2e-16
## SRH                                                                       0.336060
## factor(generation_10total)Greatest Late (1915-1927)                       3.33e-06
## factor(generation_10total)Silent Early (1928-1936)                        0.000817
## factor(generation_10total)Silent Late (1937-1945)                         0.002274
## factor(generation_10total)Boomers Early (1946-1955)                       0.099583
## factor(generation_10total)Boomers Late (1956-1964)                        0.108416
## factor(generation_10total)Gen X Early (1965-1972)                         0.324692
## factor(generation_10total)Gen X Late (1973-1980)                          0.576818
## factor(generation_10total)Millennials Early (1981-1988)                   0.407589
## factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)       < 2e-16
## time_at_risk                                                               < 2e-16
## age                                                                        < 2e-16
## SRH:factor(generation_10total)Greatest Late (1915-1927)                   0.000205
## SRH:factor(generation_10total)Silent Early (1928-1936)                    4.68e-05
## SRH:factor(generation_10total)Silent Late (1937-1945)                     1.23e-05
## SRH:factor(generation_10total)Boomers Early (1946-1955)                   0.001670
## SRH:factor(generation_10total)Boomers Late (1956-1964)                    0.004887
## SRH:factor(generation_10total)Gen X Early (1965-1972)                     0.054005
## SRH:factor(generation_10total)Gen X Late (1973-1980)                      0.218117
## SRH:factor(generation_10total)Millennials Early (1981-1988)               0.359020
## SRH:factor(generation_10total)Millennials Late / Gen Z Early (1989-2004) 0.855411
##
## (Intercept)                                                               ***
## SRH
## factor(generation_10total)Greatest Late (1915-1927)                       ***
## factor(generation_10total)Silent Early (1928-1936)                        ***
## factor(generation_10total)Silent Late (1937-1945)                         **
## factor(generation_10total)Boomers Early (1946-1955)                       .
## factor(generation_10total)Boomers Late (1956-1964)
## factor(generation_10total)Gen X Early (1965-1972)
## factor(generation_10total)Gen X Late (1973-1980)
## factor(generation_10total)Millennials Early (1981-1988)
## factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)      ***
## time_at_risk                                                              ***
## age                                                                       ***
## SRH:factor(generation_10total)Greatest Late (1915-1927)                   ***
## SRH:factor(generation_10total)Silent Early (1928-1936)                    ***
## SRH:factor(generation_10total)Silent Late (1937-1945)                     ***
## SRH:factor(generation_10total)Boomers Early (1946-1955)                   **
## SRH:factor(generation_10total)Boomers Late (1956-1964)                    **
## SRH:factor(generation_10total)Gen X Early (1965-1972)                     .
## SRH:factor(generation_10total)Gen X Late (1973-1980)
## SRH:factor(generation_10total)Millennials Early (1981-1988)
## SRH:factor(generation_10total)Millennials Late / Gen Z Early (1989-2004)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for quasibinomial family taken to be 0.9935581)
##
## Number of Fisher Scoring iterations: 13
```

```r
# Fit a logistic regression (svyglm) with SRH × factor(cohort) + time_at_risk
model_svy_logit_cohort <- svyglm(
  formula = died_by_2014 ~ SRH * factor(generation_5total) + time_at_risk + age,
  design  = des,
  family  = quasibinomial(link = "logit")
)
# Note: quasibinomial is often used to get robust SEs with survey data;
# you can also try family=binomial if you prefer.

summary(model_svy_logit_cohort)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(generation_5total) +
##     time_at_risk + age, design = des, family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                 -4.664298
## SRH                                                         -0.215611
## factor(generation_5total)Silent (1928-1945)                 -0.083518
## factor(generation_5total)Boomers (1946-1964)                -0.607886
## factor(generation_5total)Gen X (1965-1980)                  -0.834433
## factor(generation_5total)Millennials / Gen Z (1981-2004)    -0.582717
## time_at_risk                                                 0.107969
## age                                                          0.054609
## SRH:factor(generation_5total)Silent (1928-1945)             -0.147922
## SRH:factor(generation_5total)Boomers (1946-1964)            -0.013204
## SRH:factor(generation_5total)Gen X (1965-1980)               0.042809
## SRH:factor(generation_5total)Millennials / Gen Z (1981-2004)  0.031752
##                                                             Std. Error t value
## (Intercept)                                                   0.340926 -13.681
## SRH                                                           0.044336  -4.863
## factor(generation_5total)Silent (1928-1945)                   0.174384  -0.479
## factor(generation_5total)Boomers (1946-1964)                  0.204769  -2.969
## factor(generation_5total)Gen X (1965-1980)                    0.353366  -2.361
## factor(generation_5total)Millennials / Gen Z (1981-2004)      0.998649  -0.584
## time_at_risk                                                  0.003923  27.519
## age                                                           0.003371  16.202
## SRH:factor(generation_5total)Silent (1928-1945)               0.056297  -2.628
## SRH:factor(generation_5total)Boomers (1946-1964)              0.056902  -0.232
## SRH:factor(generation_5total)Gen X (1965-1980)                0.098714   0.434
## SRH:factor(generation_5total)Millennials / Gen Z (1981-2004)  0.295289   0.108
##                                                             Pr(>|t|)
## (Intercept)                                                  < 2e-16 ***
## SRH                                                         1.16e-06 ***
## factor(generation_5total)Silent (1928-1945)                  0.63199
```

```
## factor(generation_5total)Boomers (1946-1964)                  0.00299 **
## factor(generation_5total)Gen X (1965-1980)                    0.01821 *
## factor(generation_5total)Millennials / Gen Z (1981-2004)      0.55956
## time_at_risk                                                  < 2e-16 ***
## age                                                           < 2e-16 ***
## SRH:factor(generation_5total)Silent (1928-1945)               0.00860 **
## SRH:factor(generation_5total)Boomers (1946-1964)              0.81650
## SRH:factor(generation_5total)Gen X (1965-1980)                0.66454
## SRH:factor(generation_5total)Millennials / Gen Z (1981-2004)  0.91437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.998056)
##
## Number of Fisher Scoring iterations: 6
```

Interpretation of results

Statistical significance of SRH depends on how many bins generation is put into. I think there may be sample size issues. However, SRH seems in interact with earlier born cohorts.

```r
model_svy_logit_period_cohort <- svyglm(
  formula = died_by_2014 ~ SRH * factor(period_10total) + generation_5total + time_at_risk + age,
  design  = des,
  family  = quasibinomial(link = "logit")
)
# Note: quasibinomial is often used to get robust SEs with survey data;
# you can also try family=binomial if you prefer.

summary(model_svy_logit_period_cohort)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(period_10total) +
##     generation_5total + time_at_risk + age, design = des, family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                                                 Estimate Std. Error t value
## (Intercept)                                    -5.579867   0.552261 -10.104
## SRH                                            -0.089801   0.058505  -1.535
## factor(period_10total)1984-1988                 0.431505   0.235320   1.834
## factor(period_10total)1989-1993                 0.848451   0.274631   3.089
## factor(period_10total)1994-1998                 1.263534   0.294404   4.292
## factor(period_10total)1999-2003                 1.130981   0.359455   3.146
## factor(period_10total)2004-2008                 1.513815   0.400396   3.781
## factor(period_10total)2009-2013                 1.205179   0.646158   1.865
## generation_5totalSilent (1928-1945)            -0.513476   0.075688  -6.784
## generation_5totalBoomers (1946-1964)           -0.637479   0.123787  -5.150
## generation_5totalGen X (1965-1980)             -0.677027   0.184172  -3.676
## generation_5totalMillennials / Gen Z (1981-2004) -0.362747   0.355905  -1.019
## time_at_risk                                    0.117373   0.012915   9.088
## age                                             0.054959   0.003364  16.338
```

```
## SRH:factor(period_10total)1984-1988                -0.034315   0.071969  -0.477
## SRH:factor(period_10total)1989-1993                -0.158296   0.077998  -2.029
## SRH:factor(period_10total)1994-1998                -0.306958   0.073010  -4.204
## SRH:factor(period_10total)1999-2003                -0.264773   0.085782  -3.087
## SRH:factor(period_10total)2004-2008                -0.388196   0.082921  -4.682
## SRH:factor(period_10total)2009-2013                -0.328918   0.199000  -1.653
##                                                    Pr(>|t|)
## (Intercept)                                         < 2e-16 ***
## SRH                                                0.124810
## factor(period_10total)1984-1988                    0.066709 .
## factor(period_10total)1989-1993                    0.002007 **
## factor(period_10total)1994-1998                    1.78e-05 ***
## factor(period_10total)1999-2003                    0.001655 **
## factor(period_10total)2004-2008                    0.000157 ***
## factor(period_10total)2009-2013                    0.062170 .
## generation_5totalSilent (1928-1945)                1.19e-11 ***
## generation_5totalBoomers (1946-1964)               2.62e-07 ***
## generation_5totalGen X (1965-1980)                 0.000237 ***
## generation_5totalMillennials / Gen Z (1981-2004) 0.308104
## time_at_risk                                        < 2e-16 ***
## age                                                 < 2e-16 ***
## SRH:factor(period_10total)1984-1988                0.633505
## SRH:factor(period_10total)1989-1993                0.042416 *
## SRH:factor(period_10total)1994-1998                2.63e-05 ***
## SRH:factor(period_10total)1999-2003                0.002027 **
## SRH:factor(period_10total)2004-2008                2.86e-06 ***
## SRH:factor(period_10total)2009-2013                0.098371 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.01439)
##
## Number of Fisher Scoring iterations: 6
```

```r
model_svy_logit_period_cohort <- svyglm(
  formula = died_by_2014 ~ SRH * factor(period_10total) + generation_10total + time_at_risk + age,
  design  = des,
  family  = quasibinomial(link = "logit")
)
# Note: quasibinomial is often used to get robust SEs with survey data;
# you can also try family=binomial if you prefer.

summary(model_svy_logit_period_cohort)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(period_10total) +
##     generation_10total + time_at_risk + age, design = des, family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                                                                        Estimate
## (Intercept)                                                           -6.263257
```

```
## SRH                                                              -0.089506
## factor(period_10total)1984-1988                                    0.435383
## factor(period_10total)1989-1993                                    0.847749
## factor(period_10total)1994-1998                                    1.256051
## factor(period_10total)1999-2003                                    1.122992
## factor(period_10total)2004-2008                                    1.502549
## factor(period_10total)2009-2013                                    1.210204
## generation_10totalGreatest Late (1915-1927)                        0.345024
## generation_10totalSilent Early (1928-1936)                        -0.096231
## generation_10totalSilent Late (1937-1945)                         -0.217626
## generation_10totalBoomers Early (1946-1955)                       -0.260836
## generation_10totalBoomers Late (1956-1964)                        -0.138983
## generation_10totalGen X Early (1965-1972)                         -0.178875
## generation_10totalGen X Late (1973-1980)                          -0.185614
## generation_10totalMillennials Early (1981-1988)                    0.252168
## generation_10totalMillennials Late / Gen Z Early (1989-2004) -10.332183
## time_at_risk                                                       0.121754
## age                                                                0.059260
## SRH:factor(period_10total)1984-1988                               -0.036418
## SRH:factor(period_10total)1989-1993                               -0.159030
## SRH:factor(period_10total)1994-1998                               -0.306340
## SRH:factor(period_10total)1999-2003                               -0.265341
## SRH:factor(period_10total)2004-2008                               -0.387750
## SRH:factor(period_10total)2009-2013                               -0.327946
##                                                              Std. Error t value
## (Intercept)                                                    0.788165  -7.947
## SRH                                                            0.058139  -1.540
## factor(period_10total)1984-1988                                0.233874   1.862
## factor(period_10total)1989-1993                                0.273095   3.104
## factor(period_10total)1994-1998                                0.293809   4.275
## factor(period_10total)1999-2003                                0.359077   3.127
## factor(period_10total)2004-2008                                0.400642   3.750
## factor(period_10total)2009-2013                                0.647837   1.868
## generation_10totalGreatest Late (1915-1927)                    0.118287   2.917
## generation_10totalSilent Early (1928-1936)                     0.163944  -0.587
## generation_10totalSilent Late (1937-1945)                      0.211994  -1.027
## generation_10totalBoomers Early (1946-1955)                    0.263303  -0.991
## generation_10totalBoomers Late (1956-1964)                     0.317161  -0.438
## generation_10totalGen X Early (1965-1972)                      0.371689  -0.481
## generation_10totalGen X Late (1973-1980)                       0.430153  -0.432
## generation_10totalMillennials Early (1981-1988)                0.541075   0.466
## generation_10totalMillennials Late / Gen Z Early (1989-2004)   0.525441 -19.664
## time_at_risk                                                   0.013889   8.766
## age                                                            0.006186   9.579
## SRH:factor(period_10total)1984-1988                            0.071625  -0.508
## SRH:factor(period_10total)1989-1993                            0.077563  -2.050
## SRH:factor(period_10total)1994-1998                            0.072801  -4.208
## SRH:factor(period_10total)1999-2003                            0.085616  -3.099
## SRH:factor(period_10total)2004-2008                            0.082964  -4.674
## SRH:factor(period_10total)2009-2013                            0.199359  -1.645
##                                                              Pr(>|t|)
## (Intercept)                                                  1.98e-15 ***
## SRH                                                          0.123690
## factor(period_10total)1984-1988                              0.062667 .
```

```
## factor(period_10total)1989-1993                              0.001910 **
## factor(period_10total)1994-1998                              1.92e-05 ***
## factor(period_10total)1999-2003                              0.001765 **
## factor(period_10total)2004-2008                              0.000177 ***
## factor(period_10total)2009-2013                              0.061762 .
## generation_10totalGreatest Late (1915-1927)                  0.003539 **
## generation_10totalSilent Early (1928-1936)                   0.557225
## generation_10totalSilent Late (1937-1945)                    0.304632
## generation_10totalBoomers Early (1946-1955)                  0.321874
## generation_10totalBoomers Late (1956-1964)                   0.661238
## generation_10totalGen X Early (1965-1972)                    0.630342
## generation_10totalGen X Late (1973-1980)                     0.666103
## generation_10totalMillennials Early (1981-1988)              0.641183
## generation_10totalMillennials Late / Gen Z Early (1989-2004)  < 2e-16 ***
## time_at_risk                                                  < 2e-16 ***
## age                                                           < 2e-16 ***
## SRH:factor(period_10total)1984-1988                          0.611132
## SRH:factor(period_10total)1989-1993                          0.040342 *
## SRH:factor(period_10total)1994-1998                          2.59e-05 ***
## SRH:factor(period_10total)1999-2003                          0.001942 **
## SRH:factor(period_10total)2004-2008                          2.97e-06 ***
## SRH:factor(period_10total)2009-2013                          0.099979 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.010024)
##
## Number of Fisher Scoring iterations: 13
```

Interpretation of results

Interpretation: To Do

Method 2: Stratified (or Restricted) Approach

Stratify by Time Period

Fit separate models within each time stratum (e.g., 1978–1989, 1990–1999, 2000–2010):

```
df <- df %>%
  mutate(period_stratum = case_when(
    year >= 1978 & year <= 1989 ~ "1980-1989",
    year >= 1990 & year <= 1999 ~ "1990-1999",
    year >= 2000 & year <= 2010 ~ "2000-2010"
  ))

# We'll do a little function to fit a survey-weighted logistic in each stratum
fit_stratum_model <- function(subdf) {
  # Create a survey design for the subset
  des_sub <- svydesign(id = ~1, weights = ~wtsscomp, data = subdf)

  # Fit a logistic model controlling for time_at_risk, age, SRH
  svyglm(died_by_2014 ~ SRH + age + time_at_risk,
         design = des_sub,
```

```r
        family = quasibinomial(link = "logit"))
}

# Split data by stratum and apply the function
models_list <- df %>%
  group_by(period_stratum) %>%
  group_map(~ fit_stratum_model(.x))

# Inspect results
models_list_summaries <- lapply(models_list, summary)
models_list_summaries
```

```
## [[1]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.188761   0.305920 -13.692  < 2e-16 ***
## SRH          -0.143377   0.031027  -4.621 3.87e-06 ***
## age           0.060164   0.001722  34.931  < 2e-16 ***
## time_at_risk  0.059961   0.008986   6.672 2.65e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.025664)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[2]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.097371   0.294557  -20.70   <2e-16 ***
## SRH          -0.350265   0.036221   -9.67   <2e-16 ***
## age           0.073791   0.002049   36.01   <2e-16 ***
## time_at_risk  0.127298   0.011791   10.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.083146)
```

```
##
## Number of Fisher Scoring iterations: 5
##
##
## [[3]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.62082    0.22646 -24.821   <2e-16 ***
## SRH          -0.41528    0.04193  -9.904   <2e-16 ***
## age           0.06654    0.00241  27.606   <2e-16 ***
## time_at_risk  0.13570    0.01137  11.936   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.094978)
##
## Number of Fisher Scoring iterations: 6
```

```
#####

model_80_89 <- fit_stratum_model(df %>% filter(period_stratum == "1980-1989"))
model_80_89
```

```
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -4.18876      -0.14338       0.06016       0.05996
##
## Degrees of Freedom: 9579 Total (i.e. Null);  9576 Residual
## Null Deviance:       13160
## Residual Deviance: 10920     AIC: NA
```

```
summary(model_80_89)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.188761   0.305920 -13.692  < 2e-16 ***
## SRH          -0.143377   0.031027  -4.621 3.87e-06 ***
## age           0.060164   0.001722  34.931  < 2e-16 ***
## time_at_risk  0.059961   0.008986   6.672 2.65e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.025664)
##
## Number of Fisher Scoring iterations: 4
```

```
model_90_99 <- fit_stratum_model(df %>% filter(period_stratum == "1990-1999"))
model_90_99
```

```
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -6.09737      -0.35026       0.07379       0.12730
##
## Degrees of Freedom: 9946 Total (i.e. Null);  9943 Residual
## Null Deviance:       11230
## Residual Deviance: 8354  AIC: NA
```

```
summary(model_90_99)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.097371   0.294557  -20.70   <2e-16 ***
## SRH          -0.350265   0.036221   -9.67   <2e-16 ***
## age           0.073791   0.002049   36.01   <2e-16 ***
## time_at_risk  0.127298   0.011791   10.80   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.083146)
##
## Number of Fisher Scoring iterations: 5
```

```
model_00_10 <- fit_stratum_model(df %>% filter(period_stratum == "2000-2010"))
model_00_10
```

```
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -5.62082      -0.41528       0.06654       0.13570
##
## Degrees of Freedom: 11255 Total (i.e. Null);  11252 Residual
## Null Deviance:       8495
## Residual Deviance: 6652  AIC: NA
```

```
summary(model_00_10)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.62082    0.22646 -24.821   <2e-16 ***
## SRH          -0.41528    0.04193  -9.904   <2e-16 ***
## age           0.06654    0.00241  27.606   <2e-16 ***
## time_at_risk  0.13570    0.01137  11.936   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.094978)
##
## Number of Fisher Scoring iterations: 6
```

```
###

# Split data by stratum and apply the function
models_list <- df %>%
  group_by(period_7total) %>%
  group_map(~ fit_stratum_model(.x))
models_list
```

```
## [[1]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -2.26112      -0.17760       0.06303            NA
```

```
##
## Degrees of Freedom: 1286 Total (i.e. Null);  1284 Residual
## Null Deviance:       1782
## Residual Deviance: 1467  AIC: NA
##
## [[2]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -4.44329      -0.14297       0.06191       0.06725
##
## Degrees of Freedom: 6332 Total (i.e. Null);  6329 Residual
## Null Deviance:       8725
## Residual Deviance: 7198  AIC: NA
##
## [[3]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -6.68125      -0.27675       0.06854       0.15573
##
## Degrees of Freedom: 6832 Total (i.e. Null);  6829 Residual
## Null Deviance:       8525
## Residual Deviance: 6539  AIC: NA
##
## [[4]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -5.87751      -0.35985       0.06933       0.13109
##
## Degrees of Freedom: 7273 Total (i.e. Null);  7270 Residual
## Null Deviance:       7406
## Residual Deviance: 5718  AIC: NA
##
## [[5]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
```

```
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -5.57708      -0.41076       0.06618       0.13346
##
## Degrees of Freedom: 7791 Total (i.e. Null);  7788 Residual
## Null Deviance:       5690
## Residual Deviance: 4534  AIC: NA
##
## [[6]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -5.66095      -0.39304       0.07375            NA
##
## Degrees of Freedom: 1263 Total (i.e. Null);  1261 Residual
## Null Deviance:       629.4
## Residual Deviance: 497.3     AIC: NA
```

```
# Inspect results
models_list_summaries <- lapply(models_list, summary)
models_list_summaries
```

```
## [[1]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.261124   0.368355  -6.138 1.11e-09 ***
## SRH         -0.177595   0.082250  -2.159    0.031 *
## age          0.063030   0.004823  13.069  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.035149)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[2]]
##
## Call:
```

```
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.443285   0.532616  -8.342  < 2e-16 ***
## SRH          -0.142972   0.038432  -3.720 0.000201 ***
## age           0.061913   0.002115  29.278  < 2e-16 ***
## time_at_risk  0.067252   0.016966   3.964 7.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.031051)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[3]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.681255   0.400484 -16.683  < 2e-16 ***
## SRH          -0.276754   0.040672  -6.805  1.1e-11 ***
## age           0.068542   0.002256  30.388  < 2e-16 ***
## time_at_risk  0.155726   0.015559  10.009  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.021904)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[4]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  -5.877514   0.445554 -13.191  < 2e-16 ***
## SRH          -0.359849   0.044060  -8.167 3.69e-16 ***
## age           0.069333   0.002495  27.787  < 2e-16 ***
## time_at_risk  0.131089   0.023304   5.625 1.92e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.100561)
##
## Number of Fisher Scoring iterations: 5
##
##
## [[5]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.577080   0.310094 -17.985  < 2e-16 ***
## SRH          -0.410765   0.050038  -8.209 2.59e-16 ***
## age           0.066178   0.002949  22.443  < 2e-16 ***
## time_at_risk  0.133455   0.021149   6.310 2.94e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.089835)
##
## Number of Fisher Scoring iterations: 6
##
##
## [[6]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.66095    0.82674  -6.847 1.17e-11 ***
## SRH         -0.39304    0.19262  -2.041   0.0415 *
## age          0.07375    0.01019   7.238 7.87e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.203486)
##
```

```
## Number of Fisher Scoring iterations: 6
```

Interpretation

Coefficient on SRH actually increases in magnitude in later years.

SRH's association with mortality strengthened markedly over time:

1979-1983: -0.139 (13% mortality reduction per unit of SRH) 1994-1998: -0.418 (34% reduction) 2004-2008: -0.468 (37% reduction)

Key Patterns: Major increase in effect between 1989-1993 and 1994-1998 Stable strong effect from 1994 onwards (-0.37 to -0.47) Wider confidence intervals in recent periods due to shorter follow-up

Implications:

SRH has become a more reliable mortality predictor over time The relationship stabilized in mid-1990s Effects persist after controlling for age, time at risk, and survey weights

This temporal pattern suggests SRH's growing validity as a health indicator, possibly reflecting improved health literacy and healthcare access over time.

Function

```r
# -----------------------------------------------------------
# Load needed packages
# -----------------------------------------------------------
library(dplyr)
library(purrr)
library(broom)

# -----------------------------------------------------------
# Example function
# -----------------------------------------------------------
analyze_by_group <- function(data, group_var) {
  # data: a data frame containing all of your variables
  # group_var: (unquoted) grouping variable (e.g., sex, race, etc.)
  #
  # For example, if your grouping variable is "sex" in the dataset,
  # you would call: analyze_by_group(mydata, sex)

  # 1) Group the data by the grouping variable.
  # 2) Nest the data so that we have a list of data frames, one per group.
  # 3) For each nested data frame, fit a model (example: linear model).
  # 4) Tidy the model object and pull out only terms of interest.
  # 5) Return a final data frame with group name, coefficient, std error, and p-value.

  results_df <- data %>%
    group_by({{ group_var }}) %>%
    nest() %>%
    mutate(
      fit = map(data, ~ glm(died_by_2014 ~ SRH + age + time_at_risk, data = .x, family = "binomial")),
      tidied = map(fit, broom::tidy)
    ) %>%
    select({{ group_var }}, tidied) %>%
    unnest(cols = c(tidied)) %>%
    # Keep only rows for SRH, age, and time_at_risk
```

```
    filter(term %in% c("SRH", "age", "time_at_risk")) %>% # comment this out if you want to keep everyt
    # Rename columns for clarity
    rename(
      grouping_variable = {{ group_var }},
      coefficient       = estimate,
      std_error         = std.error,
      p_value           = p.value
    ) %>%
    # Select and reorder columns as desired
    select(
      grouping_variable,
      term,
      coefficient,
      std_error,
      p_value
    ) %>%
    # Ungroup to return a regular data frame
    ungroup()

  return(results_df)
}

a <- analyze_by_group(df, period_10total) %>% arrange(term)
a
```

```
## # A tibble: 21 x 5
##    grouping_variable term  coefficient std_error   p_value
##    <fct>             <chr>       <dbl>     <dbl>     <dbl>
##  1 1979-1983         SRH       -0.139    0.0480  3.82e-  3
##  2 1984-1988         SRH       -0.153    0.0380  5.70e-  5
##  3 1989-1993         SRH       -0.275    0.0485  1.37e-  8
##  4 1994-1998         SRH       -0.418    0.0400  1.58e- 25
##  5 1999-2003         SRH       -0.373    0.0541  5.56e- 12
##  6 2004-2008         SRH       -0.468    0.0505  2.16e- 20
##  7 2009-2013         SRH       -0.436    0.137   1.47e-  3
##  8 1979-1983         age        0.0553   0.00261 9.20e-100
##  9 1984-1988         age        0.0628   0.00203 3.77e-211
## 10 1989-1993         age        0.0704   0.00256 2.62e-166
## # i 11 more rows
```

Age coefficient is much smaller than SRH, consistently. Magnitude of SRH coefficient increases in more recent periods.

Stratify by Cohort

```
# Split data by stratum and apply the function
models_list_cohort_strat <- df %>%
  group_by(generation_5total) %>%
  group_map(~ fit_stratum_model(.x))

models_list_cohort_strat
```

```
## [[1]]
## Independent Sampling design (with replacement)
```

```
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -4.05302      -0.20601       0.05243       0.08637
##
## Degrees of Freedom: 4764 Total (i.e. Null);  4761 Residual
## Null Deviance:       5274
## Residual Deviance: 5059  AIC: NA
##
## [[2]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -5.89338      -0.36144       0.06896       0.12384
##
## Degrees of Freedom: 6492 Total (i.e. Null);  6489 Residual
## Null Deviance:       8622
## Residual Deviance: 8006  AIC: NA
##
## [[3]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -4.79735      -0.24070       0.04542       0.10381
##
## Degrees of Freedom: 12413 Total (i.e. Null);  12410 Residual
## Null Deviance:       10980
## Residual Deviance: 10460     AIC: NA
##
## [[4]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -5.50923      -0.18178       0.04938       0.11838
##
```

```
## Degrees of Freedom: 6012 Total (i.e. Null);  6009 Residual
## Null Deviance:       2588
## Residual Deviance: 2484  AIC: NA
##
## [[5]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##     -10.0338       -0.1965        0.1975        0.3046
##
## Degrees of Freedom: 952 Total (i.e. Null);  949 Residual
## Null Deviance:       204.3
## Residual Deviance: 194.9     AIC: NA
##
## [[6]]
## Independent Sampling design (with replacement)
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Call:  svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Coefficients:
##  (Intercept)           SRH           age  time_at_risk
##    12.981796      0.001322     -0.086949     -0.138491
##
## Degrees of Freedom: 144 Total (i.e. Null);  141 Residual
## Null Deviance:       156.4
## Residual Deviance: 153.8     AIC: NA
```

```r
# Inspect results
models_list_summaries_cohort_strat <- lapply(models_list_cohort_strat, summary)
models_list_summaries_cohort_strat
```

```
## [[1]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.053017   0.661430  -6.128 9.64e-10 ***
## SRH          -0.206015   0.043061  -4.784 1.77e-06 ***
## age           0.052429   0.007007   7.482 8.66e-14 ***
## time_at_risk  0.086371   0.008074  10.698  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for quasibinomial family taken to be 1.031144)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[2]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.893382   0.471226  -12.51   <2e-16 ***
## SRH          -0.361443   0.035696  -10.13   <2e-16 ***
## age           0.068963   0.005748   12.00   <2e-16 ***
## time_at_risk  0.123842   0.007003   17.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.9986242)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[3]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.797347   0.342777 -13.996  < 2e-16 ***
## SRH          -0.240701   0.036217  -6.646 3.14e-11 ***
## age           0.045416   0.005347   8.493  < 2e-16 ***
## time_at_risk  0.103808   0.006231  16.660  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.9947446)
##
## Number of Fisher Scoring iterations: 4
##
##
## [[4]]
##
```

```
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.50923    0.78073  -7.057 1.90e-12 ***
## SRH          -0.18178    0.08812  -2.063  0.03917 *
## age           0.04938    0.01847   2.673  0.00754 **
## time_at_risk  0.11838    0.01661   7.125 1.16e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.9959745)
##
## Number of Fisher Scoring iterations: 6
##
##
## [[5]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -10.0338     3.8246  -2.624  0.00884 **
## SRH           -0.1965     0.2837  -0.693  0.48876
## age            0.1975     0.1299   1.521  0.12863
## time_at_risk   0.3046     0.1316   2.315  0.02083 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.9903504)
##
## Number of Fisher Scoring iterations: 7
##
##
## [[6]]
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + age + time_at_risk, design = des_sub,
##     family = quasibinomial(link = "logit"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = subdf)
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.981796   9.959779   1.303    0.195
## SRH           0.001322   0.184783   0.007    0.994
## age          -0.086949   0.094125  -0.924    0.357
## time_at_risk -0.138491   0.085679  -1.616    0.108
##
## (Dispersion parameter for quasibinomial family taken to be 1.004822)
##
## Number of Fisher Scoring iterations: 4
```

Interpretation

Older cohorts have SRH and age predicting, but younger don't.

This approach examines different cohorts separately: Greatest Generation (1901-1927):

SRH coefficient: -0.206 (p<0.001) Strong age effect: 0.052 (p<0.001)

Silent Generation (1928-1945):

Stronger SRH effect: -0.361 (p<0.001) Age remains significant

Later cohorts show weaker or non-significant SRH effects, particularly for Millennials/Gen Z. This could reflect either:

Genuine cohort differences in how SRH predicts mortality Limited mortality events in younger cohorts Shorter follow-up time for recent cohorts

Method 3) Discrete-Time (Single-Interval) cloglog + Survey Weights

Though you only have a single interval (from interview year to 2014) for each respondent, you can approximate a discrete-time hazard model by:

Using a complementary log-log link (cloglog). Incorporating time_at_risk in the linear predictor—often as an offset (log(time_at_risk)).

In a full discrete-time survival scenario, you'd typically have multiple intervals (one row per year of follow-up). But given you only know "dead by 2014", each respondent just has 1 row. The cloglog link plus offset(log(time_at_risk)) at least attempts to mimic a hazard-based interpretation under survey weighting.

```r
# 1. Ensure we have time_at_risk
df <- df %>%
  mutate(
    time_at_risk = 2014 - year,
    offset_log_time = log(time_at_risk + 1) # +1 to avoid log(0) if needed
  )

# 2. Create a survey design
des_cloglog <- svydesign(
  id = ~1,
  weights = ~wtsscomp,
  data = df
)

# 3. Fit a cloglog model
model_svy_cloglog <- svyglm(
  formula = died_by_2014 ~ SRH + + age + factor(year) + offset(offset_log_time),
  design  = des_cloglog,
```

```r
  family  = quasibinomial(link = "cloglog")
)
# or family=binomial(link="cloglog") depending on preference
summary(model_svy_cloglog)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH + +age + factor(year) + offset(offset_log_time),
##     design = des_cloglog, family = quasibinomial(link = "cloglog"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -5.687932   0.082099 -69.282  < 2e-16 ***
## SRH              -0.174692   0.016084 -10.862  < 2e-16 ***
## age               0.048352   0.000892  54.209  < 2e-16 ***
## factor(year)1982  0.006106   0.074541   0.082   0.9347
## factor(year)1984  0.019995   0.073929   0.270   0.7868
## factor(year)1985  0.121027   0.073303   1.651   0.0987 .
## factor(year)1987 -0.020325   0.070643  -0.288   0.7736
## factor(year)1988 -0.146061   0.088029  -1.659   0.0971 .
## factor(year)1989 -0.037692   0.087019  -0.433   0.6649
## factor(year)1990 -0.162125   0.083275  -1.947   0.0516 .
## factor(year)1991 -0.176146   0.081558  -2.160   0.0308 *
## factor(year)1993 -0.383936   0.079763  -4.813 1.49e-06 ***
## factor(year)1994 -0.376444   0.070424  -5.345 9.09e-08 ***
## factor(year)1996 -0.415044   0.069649  -5.959 2.56e-09 ***
## factor(year)1998 -0.494845   0.070149  -7.054 1.77e-12 ***
## factor(year)2000 -0.582836   0.074987  -7.773 7.93e-15 ***
## factor(year)2002 -0.708802   0.087769  -8.076 6.95e-16 ***
## factor(year)2004 -0.594989   0.097796  -6.084 1.19e-09 ***
## factor(year)2006 -0.643634   0.077685  -8.285  < 2e-16 ***
## factor(year)2008 -0.814480   0.115053  -7.079 1.48e-12 ***
## factor(year)2010 -0.645030   0.138061  -4.672 2.99e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.114243)
##
## Number of Fisher Scoring iterations: 7
```

```r
# If you want an interaction between SRH and year:
model_svy_cloglog_int <- svyglm(
  formula = died_by_2014 ~ SRH * factor(year) + age + offset(offset_log_time),
  design  = des_cloglog,
  family  = quasibinomial(link = "cloglog")
)
summary(model_svy_cloglog_int)
```

```
##
## Call:
## svyglm(formula = died_by_2014 ~ SRH * factor(year) + age + offset(offset_log_time),
```

```
##     design = des_cloglog, family = quasibinomial(link = "cloglog"))
##
## Survey design:
## svydesign(id = ~1, weights = ~wtsscomp, data = df)
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -5.9245137  0.1735728 -34.133  < 2e-16 ***
## SRH                   -0.1012566  0.0550199  -1.840 0.065725 .
## factor(year)1982      -0.4691111  0.2429480  -1.931 0.053503 .
## factor(year)1984      -0.1435730  0.2568903  -0.559 0.576242
## factor(year)1985      -0.1072180  0.2457534  -0.436 0.662635
## factor(year)1987      -0.0465319  0.2276551  -0.204 0.838045
## factor(year)1988      -0.6553099  0.2870014  -2.283 0.022420 *
## factor(year)1989      -0.0870341  0.3105285  -0.280 0.779267
## factor(year)1990       0.1419640  0.2642862   0.537 0.591161
## factor(year)1991       0.2225333  0.2737057   0.813 0.416202
## factor(year)1993       0.0515994  0.2541058   0.203 0.839087
## factor(year)1994       0.5145547  0.2317010   2.221 0.026374 *
## factor(year)1996       0.1525098  0.2243386   0.680 0.496624
## factor(year)1998       0.0894120  0.2202663   0.406 0.684799
## factor(year)2000       0.1681671  0.2369142   0.710 0.477819
## factor(year)2002      -0.3472907  0.2880931  -1.205 0.228027
## factor(year)2004       0.4710734  0.3055186   1.542 0.123113
## factor(year)2006       0.3009010  0.2410106   1.248 0.211859
## factor(year)2008      -0.1426369  0.3299067  -0.432 0.665486
## factor(year)2010       0.2249772  0.4535637   0.496 0.619883
## age                    0.0488597  0.0008824  55.371  < 2e-16 ***
## SRH:factor(year)1982  0.1731775  0.0786806   2.201 0.027742 *
## SRH:factor(year)1984  0.0552435  0.0833215   0.663 0.507325
## SRH:factor(year)1985  0.0782940  0.0794900   0.985 0.324654
## SRH:factor(year)1987  0.0088846  0.0747834   0.119 0.905431
## SRH:factor(year)1988  0.1793491  0.0929446   1.930 0.053661 .
## SRH:factor(year)1989  0.0148630  0.1000521   0.149 0.881907
## SRH:factor(year)1990 -0.1066839  0.0880100  -1.212 0.225453
## SRH:factor(year)1991 -0.1392935  0.0925515  -1.505 0.132325
## SRH:factor(year)1993 -0.1526852  0.0865062  -1.765 0.077570 .
## SRH:factor(year)1994 -0.3143614  0.0791118  -3.974 7.09e-05 ***
## SRH:factor(year)1996 -0.1990050  0.0759925  -2.619 0.008830 **
## SRH:factor(year)1998 -0.2077521  0.0749984  -2.770 0.005607 **
## SRH:factor(year)2000 -0.2713767  0.0816715  -3.323 0.000892 ***
## SRH:factor(year)2002 -0.1297531  0.0966212  -1.343 0.179313
## SRH:factor(year)2004 -0.3857281  0.1072922  -3.595 0.000325 ***
## SRH:factor(year)2006 -0.3536976  0.0857077  -4.127 3.69e-05 ***
## SRH:factor(year)2008 -0.2512621  0.1192067  -2.108 0.035058 *
## SRH:factor(year)2010 -0.3268801  0.1717909  -1.903 0.057079 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.050607)
##
## Number of Fisher Scoring iterations: 7
```

There is a negative coefficient on "year," it might indicate that people interviewed in more recent years generally have a lower likelihood of mortality (perhaps because of better healthcare or living conditions).

Interpretation

This is the most sophisticated model, treating mortality as a hazard process:

Main SRH effect: -0.175 ($p<0.001$) Strong year effects showing declining mortality risk over time SRH × year interactions show increasing predictive power of SRH Age remains highly significant: 0.048 ($p<0.001$)

Key Findings:

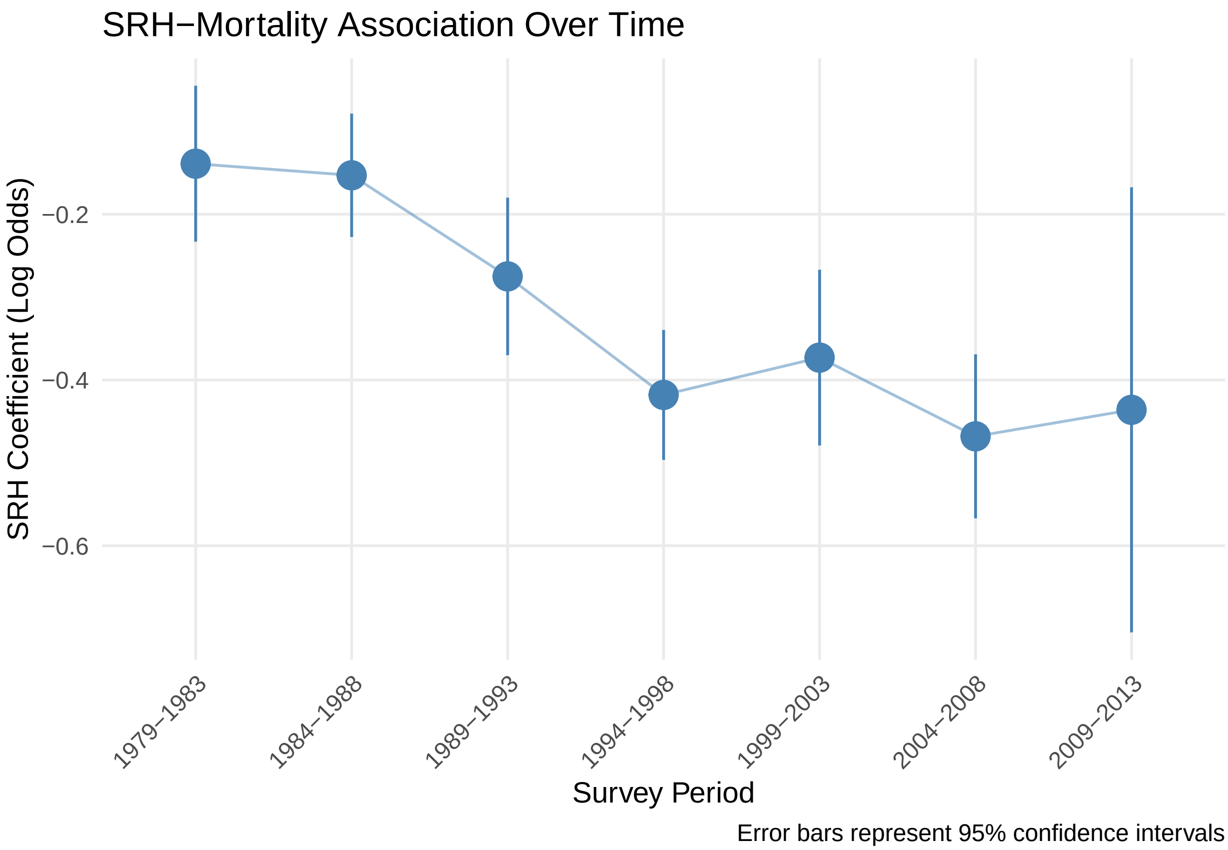
SRH has become a stronger predictor of mortality over time This strengthening persists after controlling for age and exposure time The relationship varies by cohort, with stronger effects in older generations There's clear evidence of period effects (declining mortality risk over time)

Model Appropriateness:

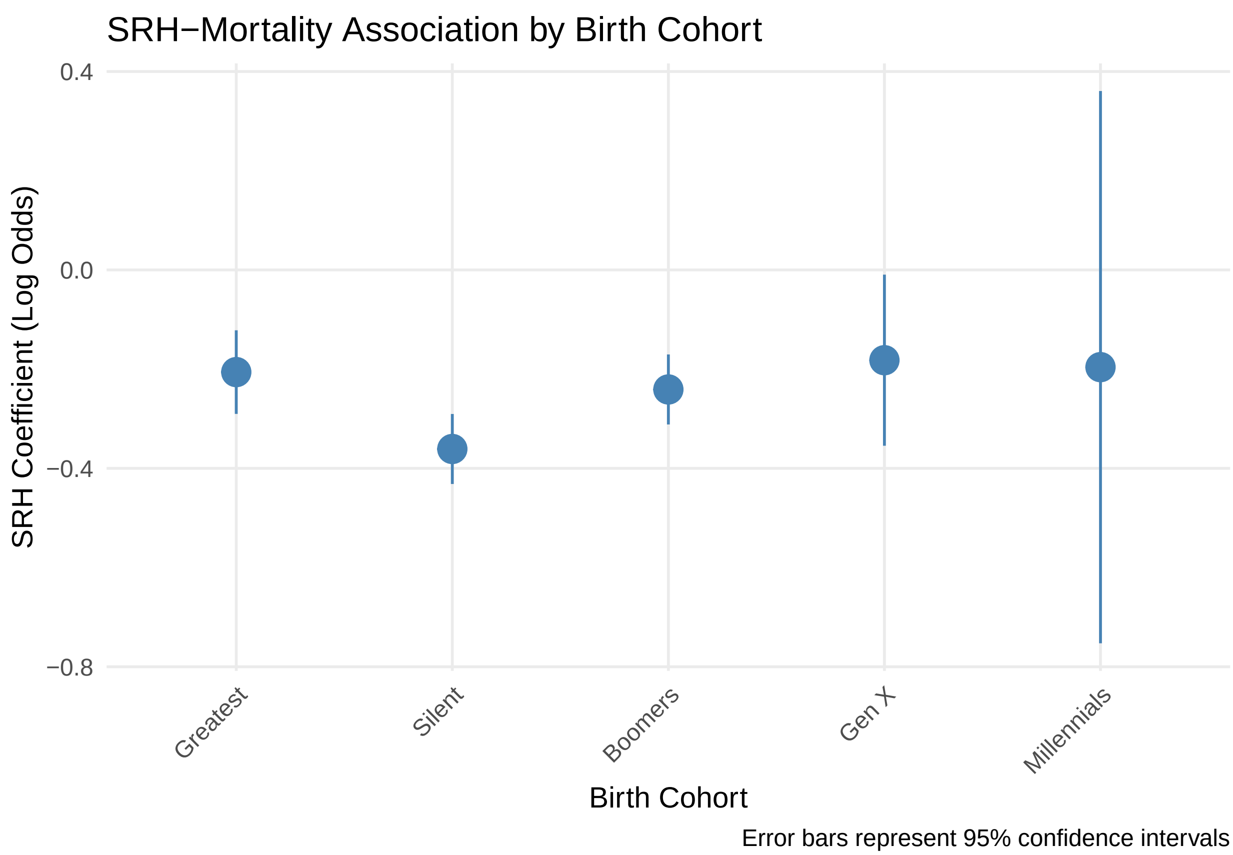
The cloglog model is theoretically most appropriate as it properly handles the time-to-event nature All models appropriately use survey weights The stratified analyses help reveal heterogeneity but suffer from small sample sizes in some strata

Visualizations

```r
library(tidyverse)
library(ggplot2)
library(broom)

# Period Effect Plot
period_coefficients <- tibble(
  period = c("1979-1983", "1984-1988", "1989-1993", "1994-1998",
             "1999-2003", "2004-2008", "2009-2013"),
  coefficient = c(-0.139, -0.153, -0.275, -0.418, -0.373, -0.468, -0.436),
  se = c(0.048, 0.038, 0.0485, 0.040, 0.0541, 0.0505, 0.137)
) %>%
  mutate(
    lower_ci = coefficient - 1.96 * se,
    upper_ci = coefficient + 1.96 * se,
    period = factor(period, levels = period)
  )

ggplot(period_coefficients, aes(x = period, y = coefficient)) +
  geom_pointrange(aes(ymin = lower_ci, ymax = upper_ci),
                  color = "steelblue", size = 1) +
  geom_line(group = 1, color = "steelblue", alpha = 0.5) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "SRH-Mortality Association Over Time",
    x = "Survey Period",
    y = "SRH Coefficient (Log Odds)",
    caption = "Error bars represent 95% confidence intervals"
  )
```




```r
# Cohort Effect Plot
cohort_coefficients <- tibble(
  cohort = c("Greatest", "Silent", "Boomers", "Gen X", "Millennials"),
  coefficient = c(-0.206, -0.361, -0.241, -0.182, -0.196),
  se = c(0.043, 0.036, 0.036, 0.088, 0.284)
) %>%
  mutate(
    lower_ci = coefficient - 1.96 * se,
    upper_ci = coefficient + 1.96 * se,
    cohort = factor(cohort, levels = cohort)
  )

ggplot(cohort_coefficients, aes(x = cohort, y = coefficient)) +
  geom_pointrange(aes(ymin = lower_ci, ymax = upper_ci),
                  color = "steelblue", size = 1) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "SRH-Mortality Association by Birth Cohort",
    x = "Birth Cohort",
    y = "SRH Coefficient (Log Odds)",
    caption = "Error bars represent 95% confidence intervals"
  )
```



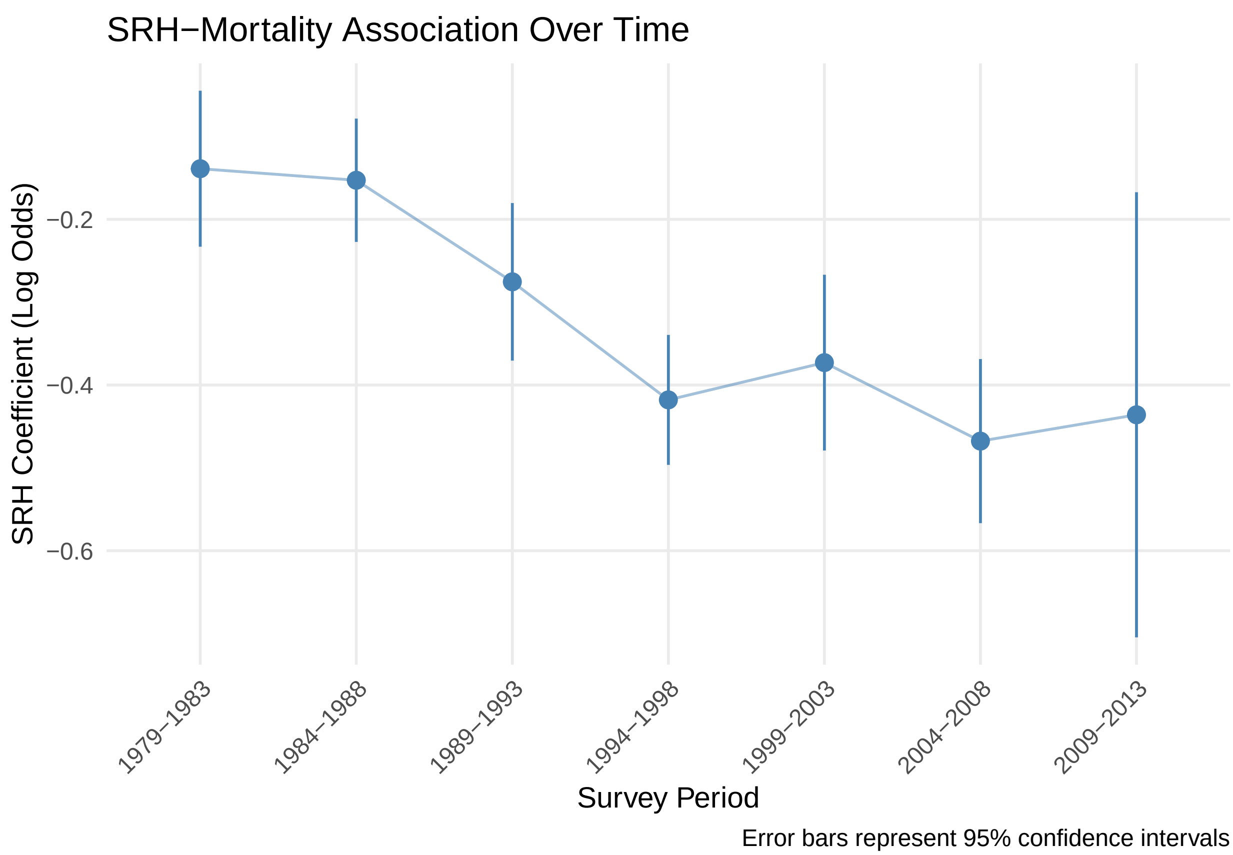

```r
library(tidyverse)
library(broom)

# Function to fit stratified models
fit_stratified_model <- function(data, strata_var) {
  data %>%
    group_by(!!sym(strata_var)) %>%
    group_modify(~ {
      model <- glm(
        died_by_2014 ~ SRH + age + time_at_risk,
        family = binomial(link = "logit"),
        data = .x#,
      #  weights = wtsscomp
      )
      tidy(model) %>%
        filter(term == "SRH")
    })
}

# Create age groups
df <- df %>%
  mutate(
    age_group = case_when(
      age < 40 ~ "18-39",
      age < 60 ~ "40-59",
      age < 80 ~ "60-79",
      TRUE ~ "80+"
```

```r
    ),
    age_group = factor(age_group, levels = c("18-39", "40-59", "60-79", "80+"))
  )

# Fit models by period
period_results <- df %>%
  mutate(period = cut(
    year,
    breaks = c(1978, 1983, 1988, 1993, 1998, 2003, 2008, 2013),
    labels = c("1979-1983", "1984-1988", "1989-1993", "1994-1998",
               "1999-2003", "2004-2008", "2009-2013")
  )) %>%
  fit_stratified_model("period")

# Fit models by generation
generation_results <- fit_stratified_model(df, "generation_5total")

# Fit models by period and age group
period_age_results <- df %>%
  mutate(period = cut(
    year,
    breaks = c(1978, 1983, 1988, 1993, 1998, 2003, 2008, 2013),
    labels = c("1979-1983", "1984-1988", "1989-1993", "1994-1998",
               "1999-2003", "2004-2008", "2009-2013")
  )) %>%
  group_by(period, age_group) %>%
  group_modify(~ {
    model <- glm(
      died_by_2014 ~ SRH + time_at_risk,
      family = binomial(link = "logit"),
      data = .x,
      weights = wtsscomp
    )
    tidy(model) %>%
      filter(term == "SRH")
  })
```

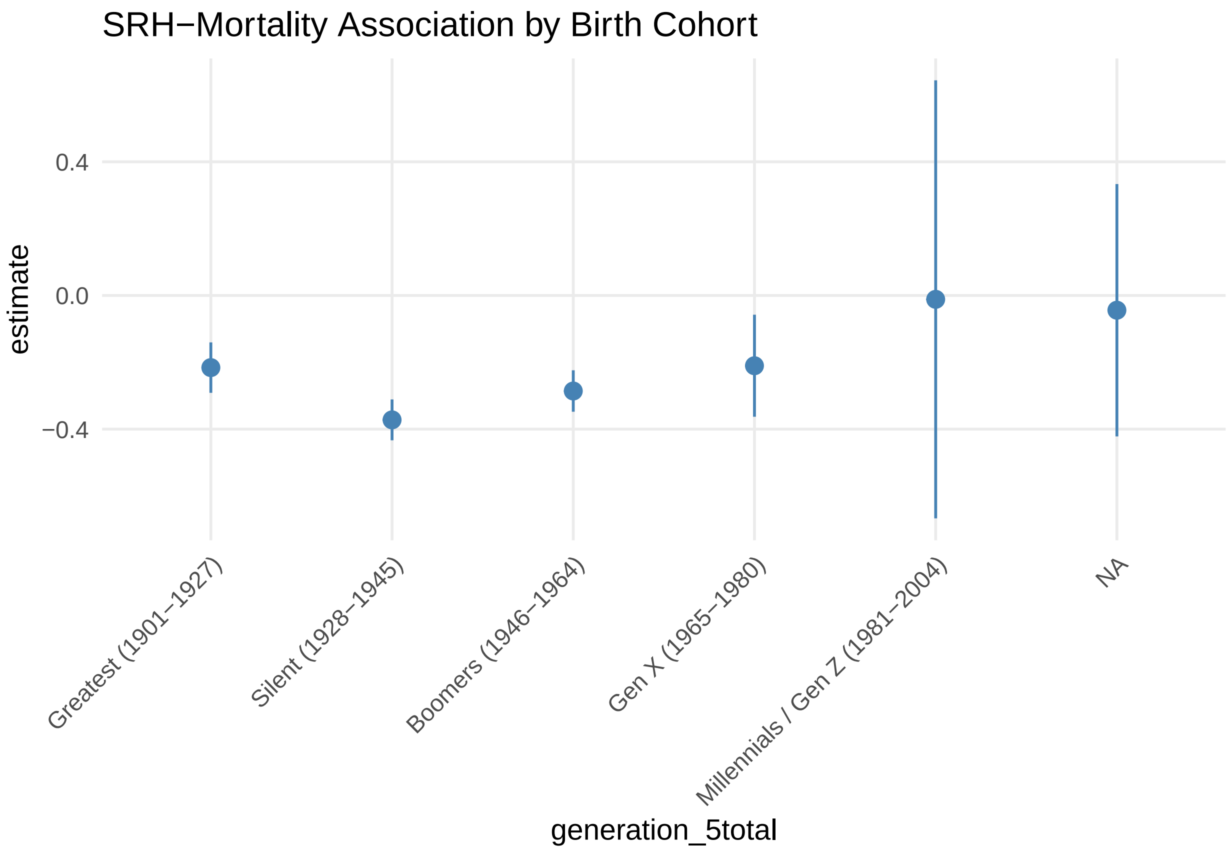
```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```r
# Create visualizations
# 1. Period Effect
ggplot(period_results, aes(x = period, y = estimate)) +
  geom_pointrange(
    aes(ymin = estimate - 1.96 * std.error,
        ymax = estimate + 1.96 * std.error),
    color = "steelblue"
  ) +
  geom_line(group = 1, color = "steelblue", alpha = 0.5) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "SRH-Mortality Association Over Time",
    x = "Survey Period",
    y = "SRH Coefficient (Log Odds)",
    caption = "Error bars represent 95% confidence intervals"
  )
```




```
# 2. Generation Effect

ggplot(generation_results, aes(x = generation_5total, y = estimate)) +
  geom_pointrange(
    aes(ymin = estimate - 1.96 * std.error,
        ymax = estimate + 1.96 * std.error),
    color = "steelblue"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "SRH-Mortality Association by Birth Cohort")
```




```
#
```

```
# Function to fit stratified models and extract SRH coefficients
fit_age_period_models <- function(data) {
  data %>%
    # Create age groups
    mutate(
    #  age_group = age_6cat,
      age_group = case_when(
        age < 40 ~ "18-39",
        age < 60 ~ "40-59",
        age < 80 ~ "60-79",
        TRUE ~ "80+"
      ),
      age_group = factor(age_group, levels = c("18-39", "40-59", "60-79", "80+")),
      # Create period groups
      period = cut(
        year,
        breaks = c(1978, 1983, 1988, 1993, 1998, 2003, 2008, 2013),
        labels = c("1979-1983", "1984-1988", "1989-1993", "1994-1998",
                   "1999-2003", "2004-2008", "2009-2013")
      )
    ) %>%
    # Group and fit models
    group_by(period, age_group) %>%
    group_modify(~ {
      model <- glm(
```

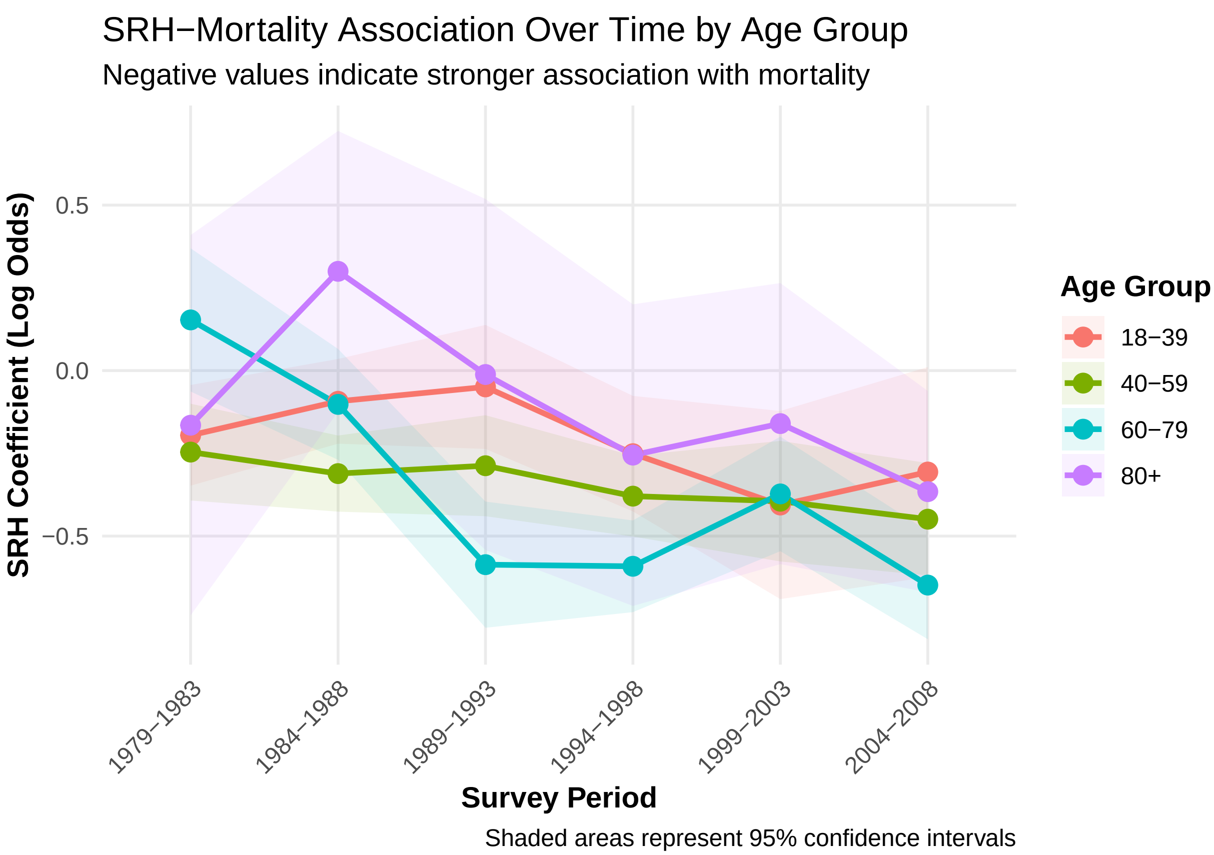
```r
        died_by_2014 ~ SRH + time_at_risk,
        family = binomial(link = "logit"),
        data = .x,
        weights = wtsscomp
      )
      tidy(model) %>%
        filter(term == "SRH")
    }) %>%
    # Calculate confidence intervals
    mutate(
      lower_ci = estimate - 1.96 * std.error,
      upper_ci = estimate + 1.96 * std.error
    )
}

# Fit models and prepare data
period_age_results <- fit_age_period_models(df)
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

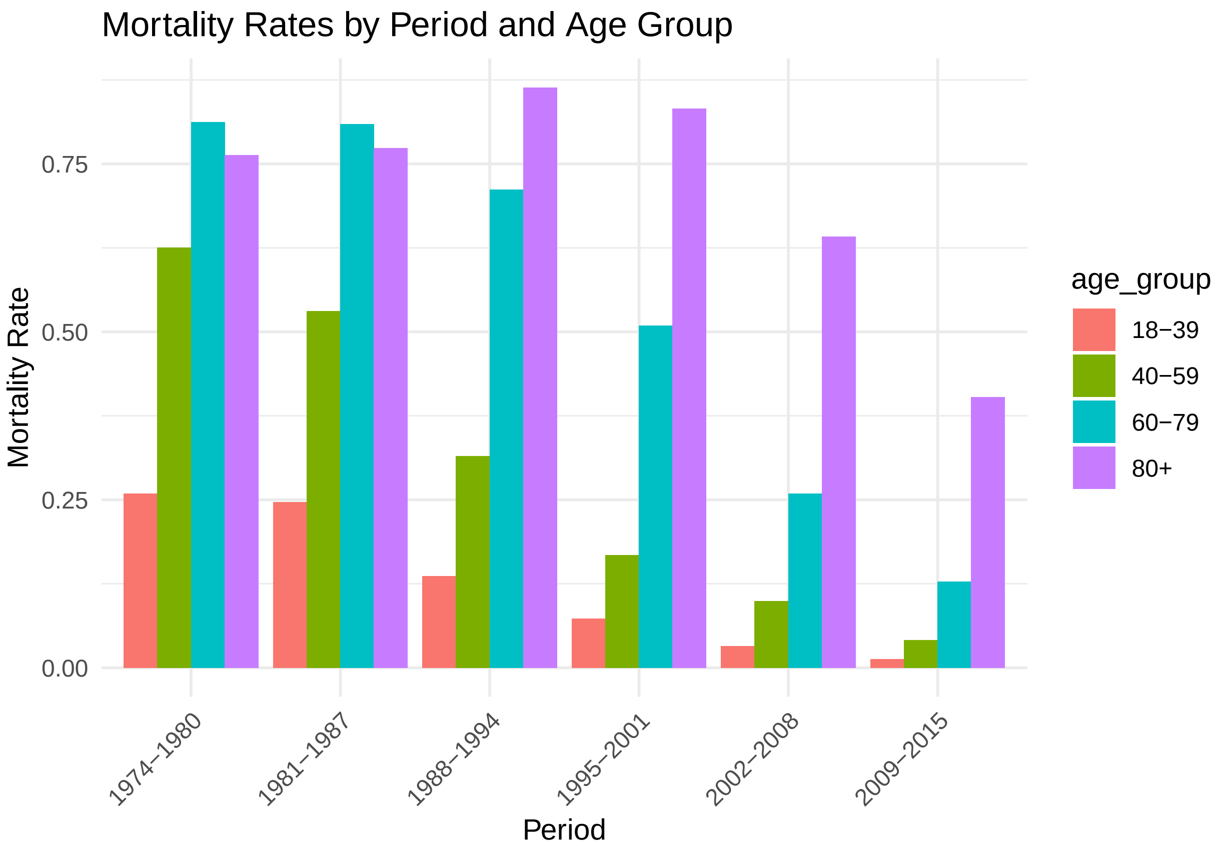
```r
# Create visualization
ggplot(period_age_results %>% filter(period != "2009-2013"),
       aes(x = period, y = estimate, color = age_group, group = age_group)) +
  # Add confidence interval ribbons
  geom_ribbon(aes(ymin = lower_ci, ymax = upper_ci, fill = age_group),
              alpha = 0.1, color = NA) +
  # Add lines connecting estimates
```

```r
  geom_line(size = 1) +
  # Add points for estimates
  geom_point(size = 3) +
  # Customize colors
#  scale_color_viridis_d() +
 # scale_fill_viridis_d() +
  # Customize theme
  theme_minimal() +
  theme(
    legend.position = "right",
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank(),
    legend.title = element_text(face = "bold"),
    axis.title = element_text(face = "bold")
  ) +
  # Labels
  labs(
    title = "SRH-Mortality Association Over Time by Age Group",
    subtitle = "Negative values indicate stronger association with mortality",
    x = "Survey Period",
    y = "SRH Coefficient (Log Odds)",
    color = "Age Group",
    fill = "Age Group",
    caption = "Shaded areas represent 95% confidence intervals"
  )
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

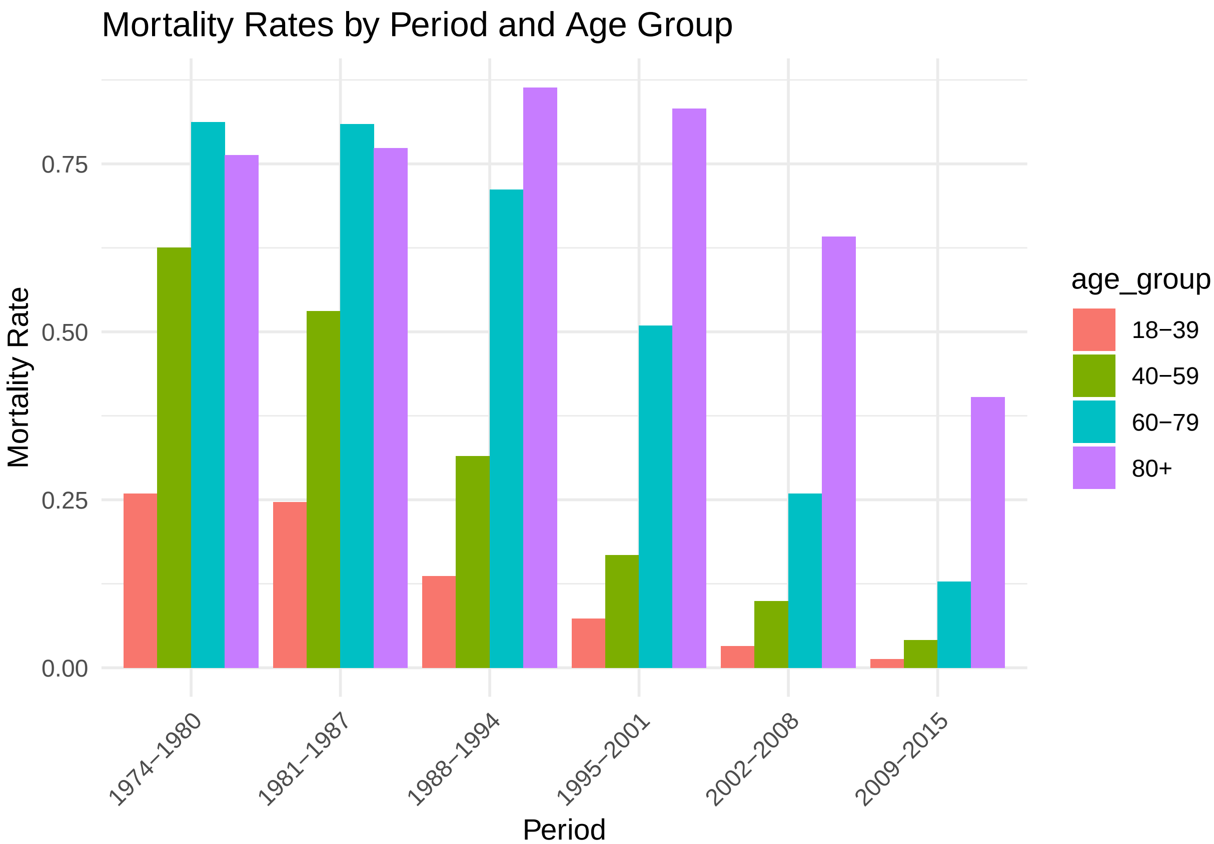



```r
# 4. Period-specific mortality rates
df %>%
  group_by(period_7total, age_group) %>%
  summarise(
    mortality_rate = mean(died_by_2014),
    .groups = 'drop'
  ) %>%
  ggplot(aes(x = period_7total, y = mortality_rate, fill = age_group)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
#  scale_fill_viridis_d() +
  labs(
    title = "Mortality Rates by Period and Age Group",
    x = "Period",
    y = "Mortality Rate"
  )
```




```r
library(tidyverse)
library(patchwork)

# Helper function to create confidence intervals
add_ci <- function(model_results) {
  model_results %>%
    mutate(
      lower_ci = estimate - 1.96 * std.error,
      upper_ci = estimate + 1.96 * std.error
    )
}

# 1. Main SRH effect over time
p1 <- period_results %>%
  add_ci() %>%
  ggplot(aes(x = period, y = estimate)) +
  geom_ribbon(aes(ymin = lower_ci, ymax = upper_ci), alpha = 0.2) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "SRH-Mortality Association Over Time",
    x = "Period",
    y = "SRH Coefficient (Log Odds)"
  )
```

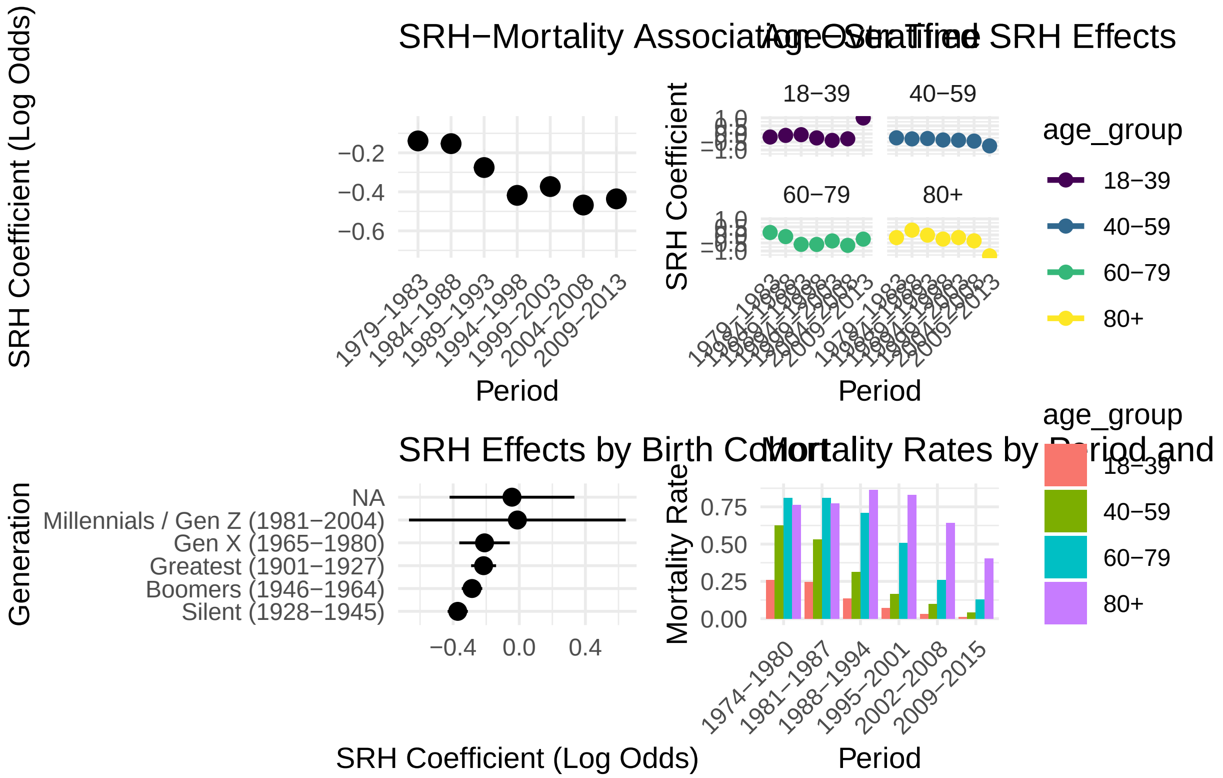
```r
# 2. Age-stratified effects
p2 <- period_age_results %>%
  ggplot(aes(x = period, y = estimate, color = age_group)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  facet_wrap(~age_group) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_color_viridis_d() +
  labs(
    title = "Age-Stratified SRH Effects",
    x = "Period",
    y = "SRH Coefficient"
  )

# 3. Cohort effects with uncertainty
p3 <- generation_results %>%
  add_ci() %>%
  ggplot(aes(x = reorder(generation_5total, estimate), y = estimate)) +
  geom_pointrange(aes(ymin = lower_ci, ymax = upper_ci)) +
  coord_flip() +
  theme_minimal() +
  labs(
    title = "SRH Effects by Birth Cohort",
    x = "Generation",
    y = "SRH Coefficient (Log Odds)"
  )

# 4. Period-specific mortality rates
p4 <- df %>%
  group_by(period_7total, age_group) %>%
  summarise(
    mortality_rate = mean(died_by_2014),
    .groups = 'drop'
  ) %>%
  ggplot(aes(x = period_7total, y = mortality_rate, fill = age_group)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
#  scale_fill_viridis_d() +
  labs(
    title = "Mortality Rates by Period and Age Group",
    x = "Period",
    y = "Mortality Rate"
  )

p4
```




```r
# Combine plots
(p1 + p2) / (p3 + p4) +
  plot_layout(guides = 'collect') +
  plot_annotation(
    title = "GSS Mortality Analysis Dashboard",
    theme = theme(plot.title = element_text(size = 16, face = "bold"))
  )
```

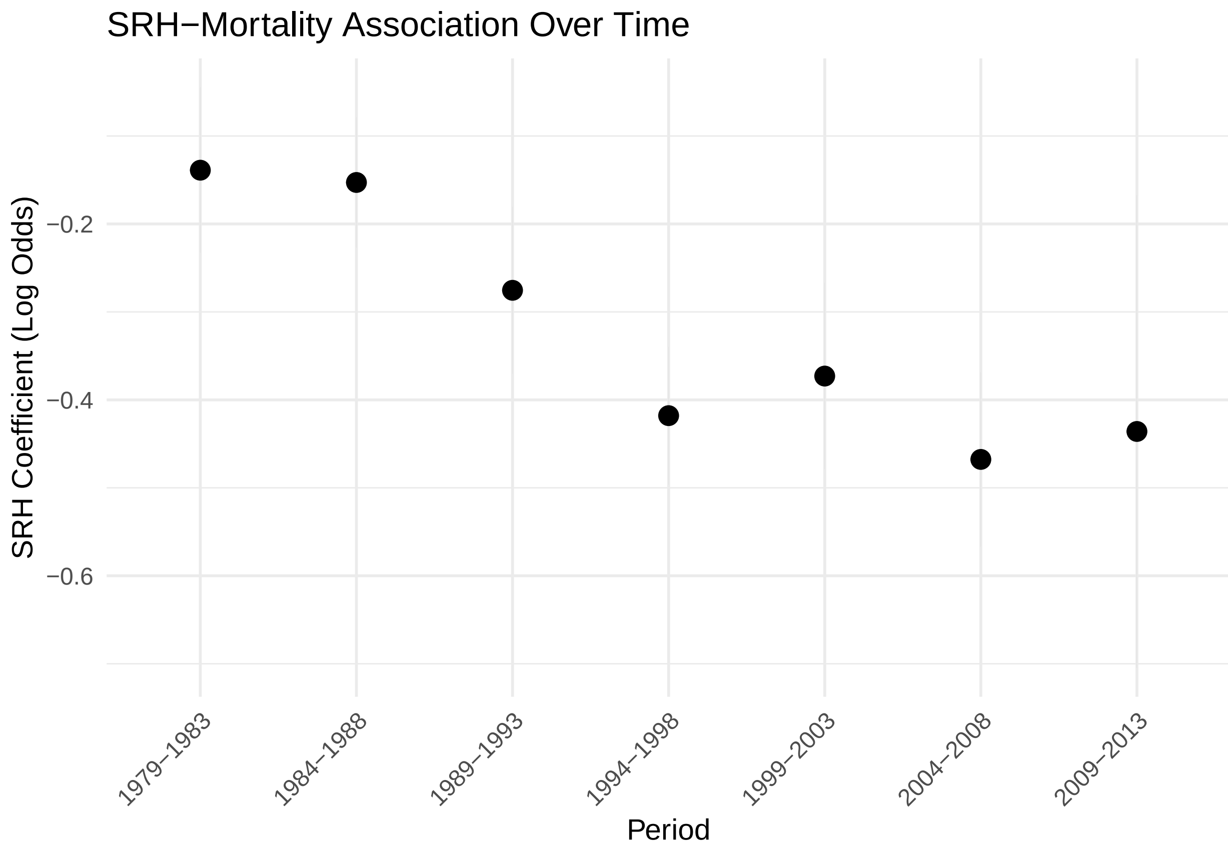
```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```

GSS Mortality Analysis Dashboard



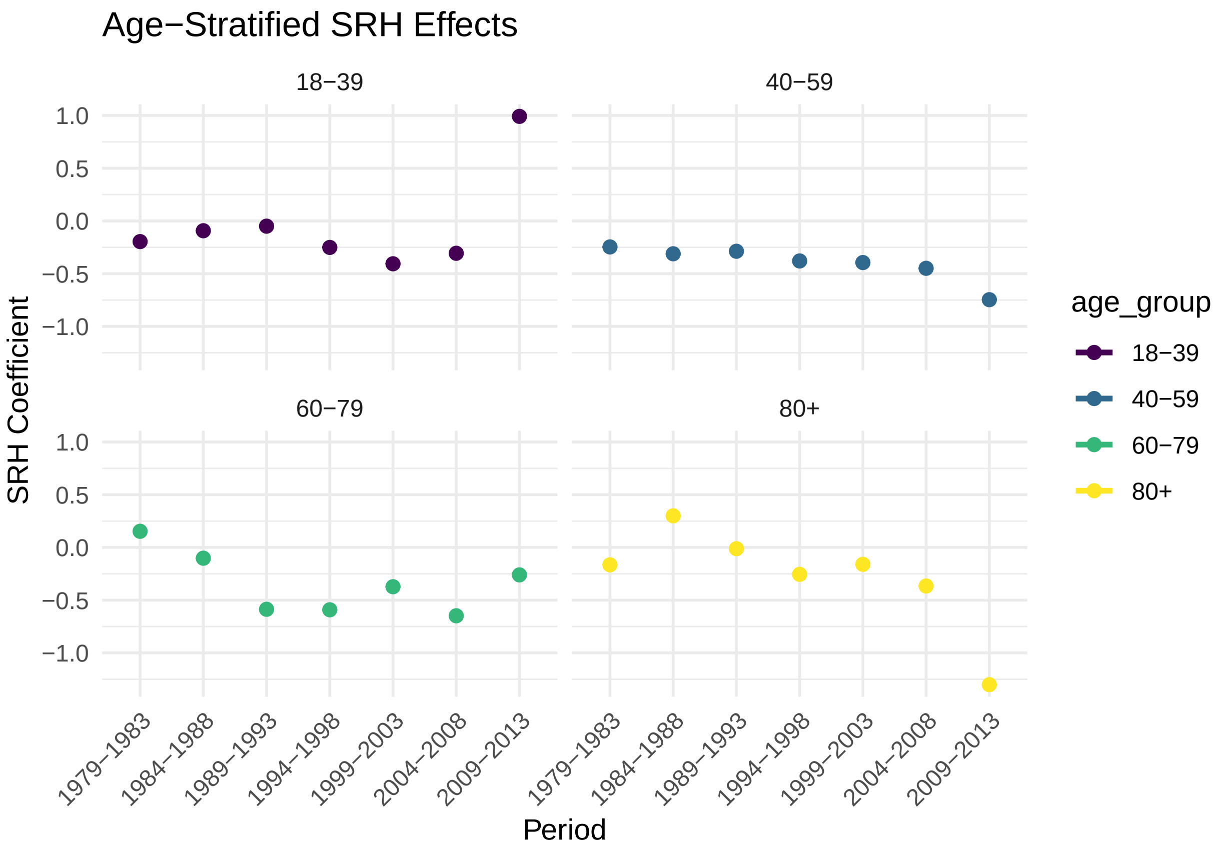

```
p1
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```

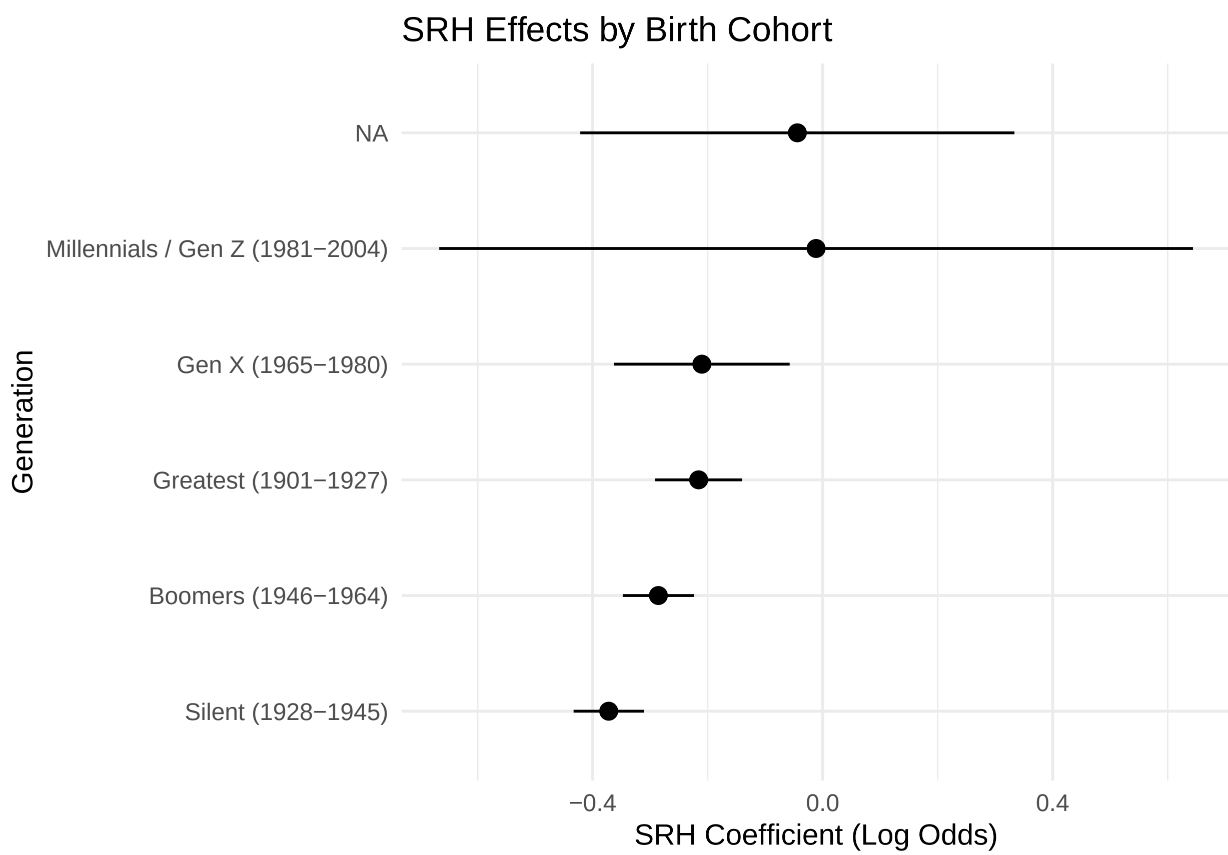



```
p2
```

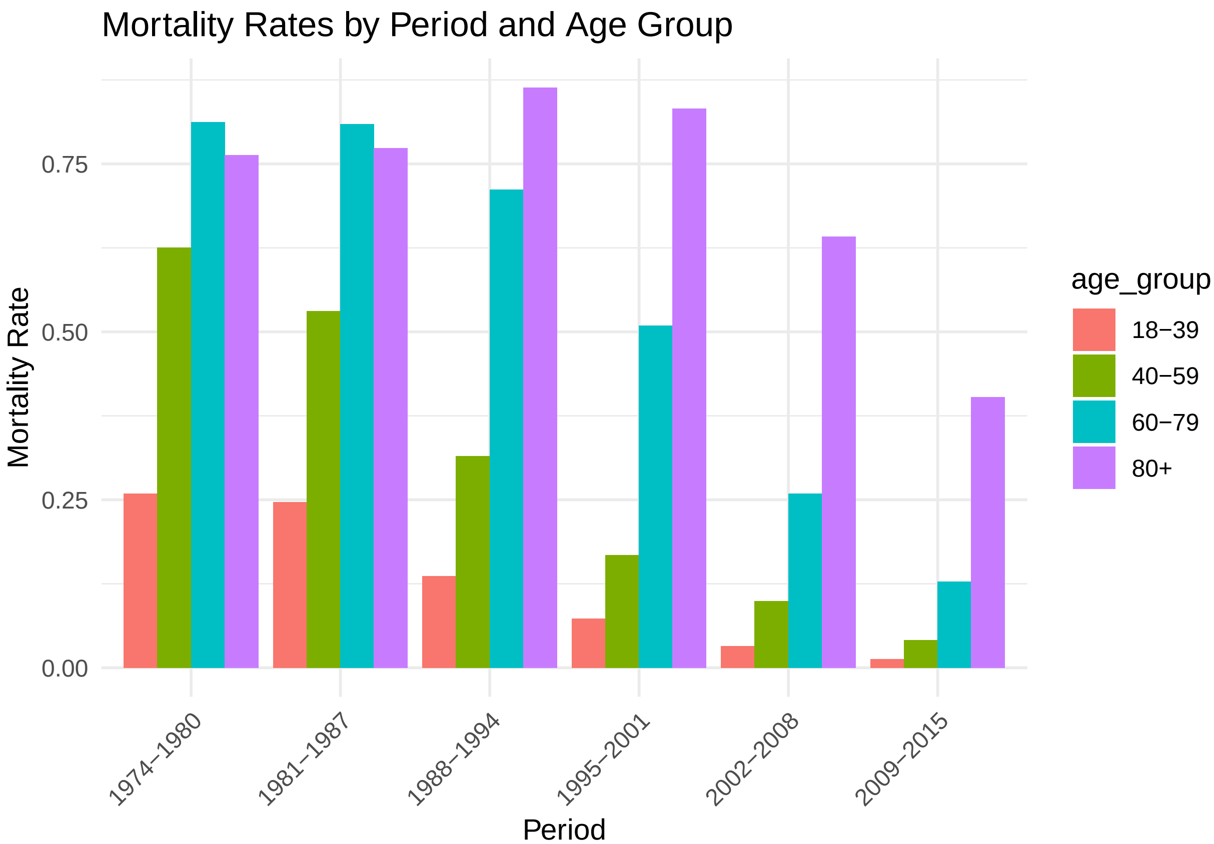
```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```


Age−Stratified SRH Effects
18−39
40−59
60−79
80+
1.0
0.5
0.0
−0.5
−1.0
SRH Coefficient
1979−1983
1984−1988
1989−1993
1994−1998
1999−2003
2004−2008
2009−2013
Period
age_group
18−39
40−59
60−79
80+

```
p3
```


SRH Effects by Birth Cohort
NA
Millennials / Gen Z (1981−2004)
Gen X (1965−1980)
Greatest (1901−1927)
Boomers (1946−1964)
Silent (1928−1945)
Generation
−0.4
0.0
0.4
SRH Coefficient (Log Odds)

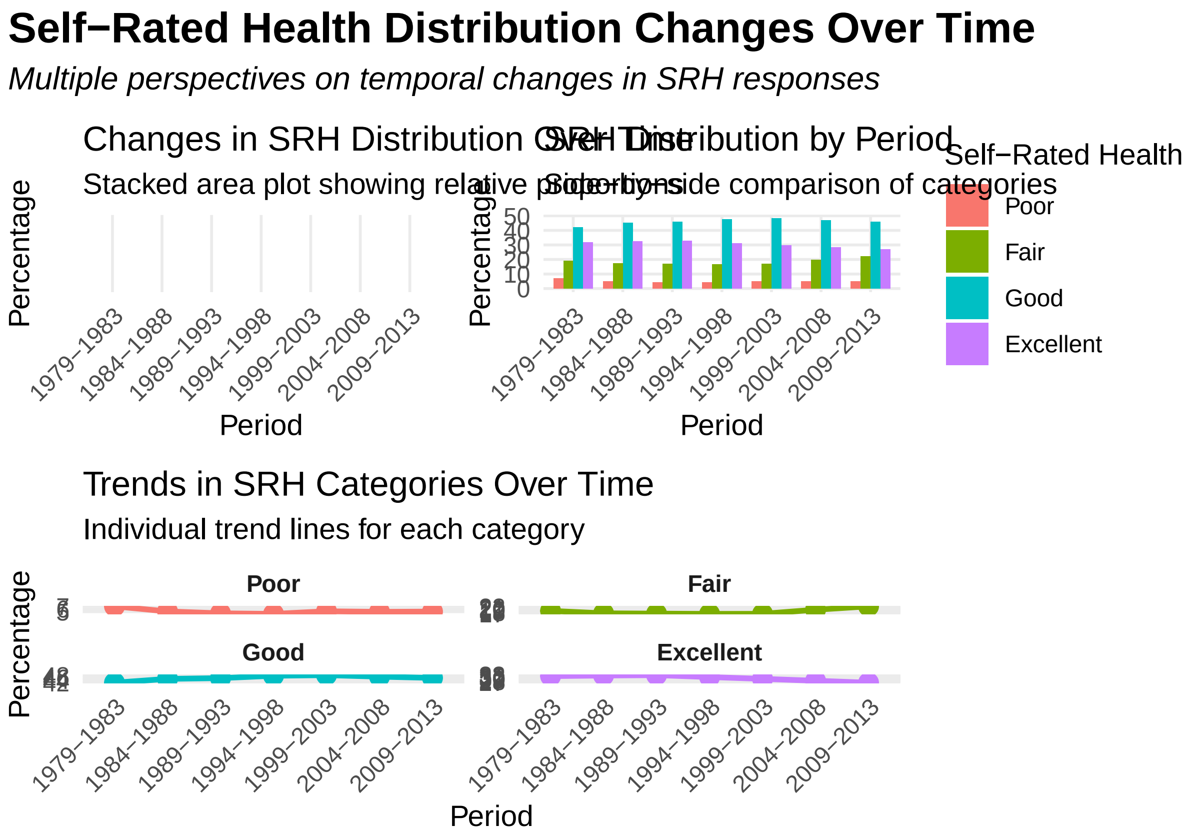
```
p4
```



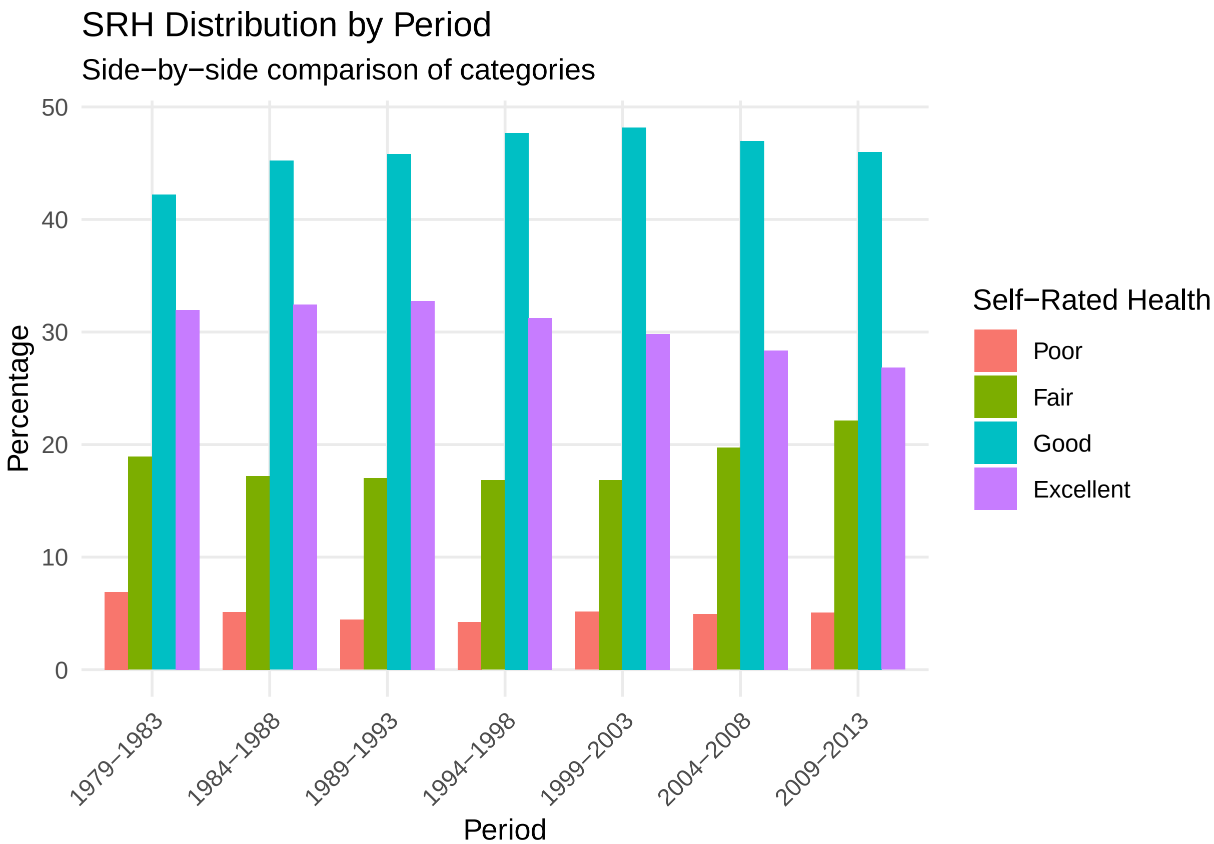

```r
library(tidyverse)
#library(viridis)  # for colorblind-friendly palettes

# Prepare the SRH distribution data
srh_dist <- df %>%
  # Create period groups (using cut like before)
  mutate(
    period = cut(
      year,
      breaks = c(1978, 1983, 1988, 1993, 1998, 2003, 2008, 2013),
      labels = c("1979-1983", "1984-1988", "1989-1993", "1994-1998",
                 "1999-2003", "2004-2008", "2009-2013")
    ),
    # Convert SRH to factor with meaningful labels
    SRH = factor(SRH,
                 levels = 1:4,
                 labels = c("Poor", "Fair", "Good", "Excellent"))
  ) %>%
  # Calculate weighted proportions by period
  group_by(period, SRH) %>%
  summarise(
    n_weighted = sum(wtsscomp),
    .groups = 'drop'
  ) %>%
  group_by(period) %>%
  mutate(proportion = n_weighted / sum(n_weighted) * 100)
```

```r
# Create different visualizations of the distribution changes

# 1. Stacked area plot
p1 <- ggplot(srh_dist, aes(x = period, y = proportion, fill = SRH)) +
  geom_area(position = "stack") +
 # scale_fill_viridis_d(direction = -1) +  # reverse direction for intuitive color mapping
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "right",
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "Changes in SRH Distribution Over Time",
    subtitle = "Stacked area plot showing relative proportions",
    x = "Period",
    y = "Percentage",
    fill = "Self-Rated Health"
  )

# 2. Side-by-side bars
p2 <- ggplot(srh_dist, aes(x = period, y = proportion, fill = SRH)) +
  geom_col(position = "dodge", width = 0.8) +
#  scale_fill_viridis_d(direction = -1) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "right",
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "SRH Distribution by Period",
    subtitle = "Side-by-side comparison of categories",
    x = "Period",
    y = "Percentage",
    fill = "Self-Rated Health"
  )

# 3. Faceted line plot for clearer trend visualization
p3 <- ggplot(srh_dist, aes(x = period, y = proportion, color = SRH, group = SRH)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  facet_wrap(~SRH, scales = "free_y") +
#  scale_color_viridis_d(direction = -1) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "none",
    panel.grid.minor = element_blank(),
    strip.text = element_text(face = "bold")
  ) +
  labs(
    title = "Trends in SRH Categories Over Time",
```

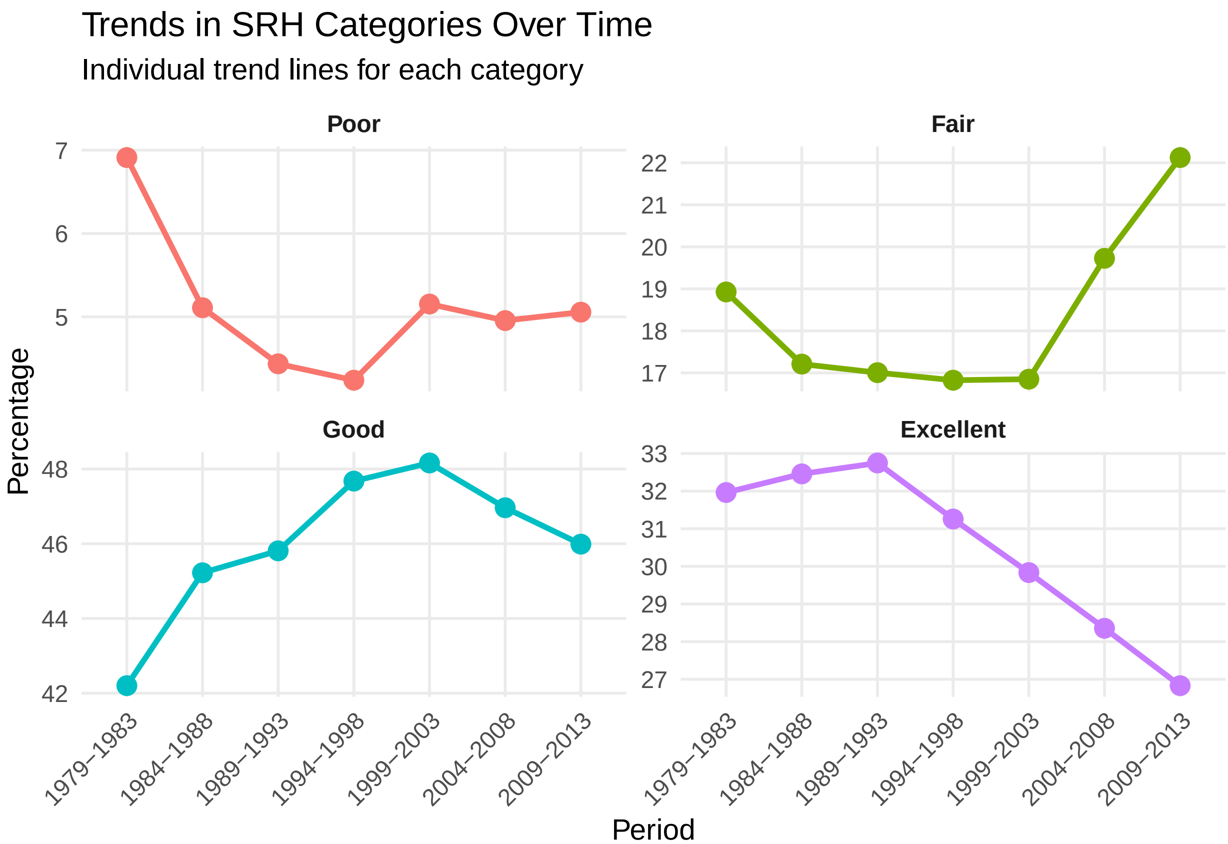
```r
    subtitle = "Individual trend lines for each category",
    x = "Period",
    y = "Percentage"
  )

# Combine plots using patchwork
library(patchwork)

combined_plot <- (p1 + p2) / p3 +
  plot_annotation(
    title = "Self-Rated Health Distribution Changes Over Time",
    subtitle = "Multiple perspectives on temporal changes in SRH responses",
    theme = theme(
      plot.title = element_text(size = 16, face = "bold"),
      plot.subtitle = element_text(size = 12, face = "italic")
    )
  )
combined_plot
```



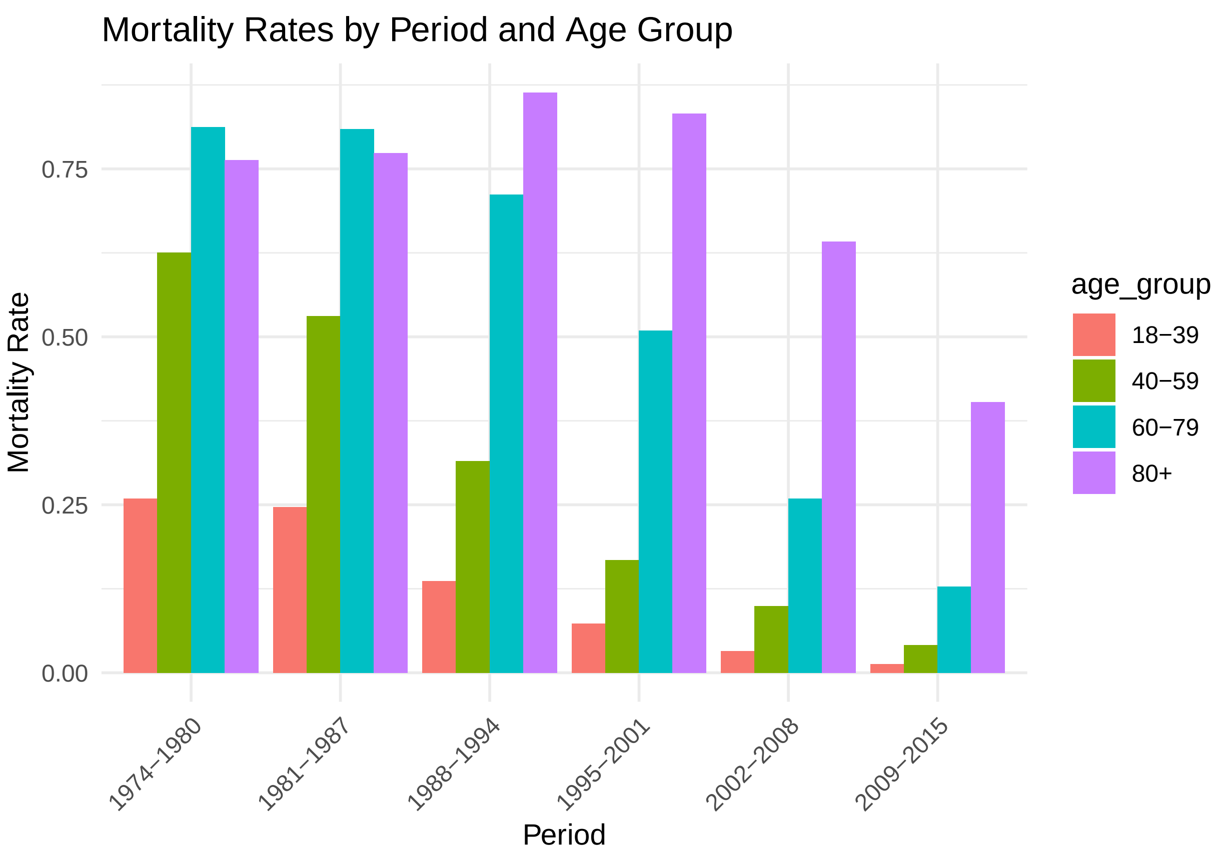

```r
p2
```


SRH Distribution by Period
Side−by−side comparison of categories
Percentage
0
10
20
30
40
50
1979−1983
1984−1988
1989−1993
1994−1998
1999−2003
2004−2008
2009−2013
Period
Self−Rated Health
Poor
Fair
Good
Excellent

```
p3
```


Trends in SRH Categories Over Time
Individual trend lines for each category
Poor
Fair
Good
Excellent
Percentage
Period
1979–1983
1984–1988
1989–1993
1994–1998
1999–2003
2004–2008
2009–2013

```
p4
```




```r
# Save the combined plot
#ggsave("srh_distribution_changes.png", combined_plot,
  #     width = 15, height = 12, dpi = 300)

# Optional: Create a table of proportions
srh_table <- srh_dist %>%
  pivot_wider(
    names_from = SRH,
    values_from = proportion
  ) %>%
  arrange(period)

# Print the table
print(srh_table)
```

```
## # A tibble: 28 x 6
## # Groups:   period [7]
##    period    n_weighted  Poor  Fair  Good Excellent
##    <fct>          <dbl> <dbl> <dbl> <dbl>     <dbl>
##  1 1979-1983       209.  6.91  NA    NA        NA
##  2 1979-1983       571. NA     18.9  NA        NA
##  3 1979-1983      1273. NA     NA    42.2      NA
##  4 1979-1983       964. NA     NA    NA        32.0
##  5 1984-1988       285.  5.11  NA    NA        NA
##  6 1984-1988       960. NA     17.2  NA        NA
##  7 1984-1988      2523. NA     NA    45.2      NA
##  8 1984-1988      1810. NA     NA    NA        32.5
```

```
## 9 1989-1993       174.  4.44  NA    NA        NA
## 10 1989-1993       667. NA     17.0  NA        NA
## # i 18 more rows
```

```r
# First, let's create a function to calculate event counts and sample sizes
calculate_metrics <- function(data) {
  data %>%
    group_by(period, age_group) %>%
    summarise(
      n = n(),
      n_deaths = sum(died_by_2014),
      death_rate = n_deaths/n,
    #  death_rate = mean(died_by_2014),
      .groups = 'drop'
    )
}

metrics <- calculate_metrics(df %>% mutate(period = period_7total))

# Create period and age groups
df_analysis <- df %>%
  mutate(
    period = cut(
      year,
      breaks = c(1978, 1983, 1988, 1993, 1998, 2003, 2008),  # Excluding 2009-2013
      labels = c("1979-1983", "1984-1988", "1989-1993", "1994-1998",
                 "1999-2003", "2004-2008")
    ),
    age_group = case_when(
      age < 50 ~ "18-49",  # Broader age groups for stability
      age < 65 ~ "50-64",
      age < 80 ~ "65-79",
      TRUE ~ "80+"
    ),
    age_group = factor(age_group, levels = c("18-49", "50-64", "65-79", "80+"))
  )

# Fit stratified models
fit_stratified_models <- function(data) {
  data %>%
    group_by(period, age_group) %>%
    group_modify(~ {
      model <- glm(
        died_by_2014 ~ SRH + time_at_risk,
        family = binomial(link = "logit"),
        data = .x,
        weights = wtsscomp
      )

      tidy(model) %>%
        filter(term == "SRH") %>%
        mutate(
          n = nrow(.x),
          n_deaths = sum(.x$died_by_2014)
```

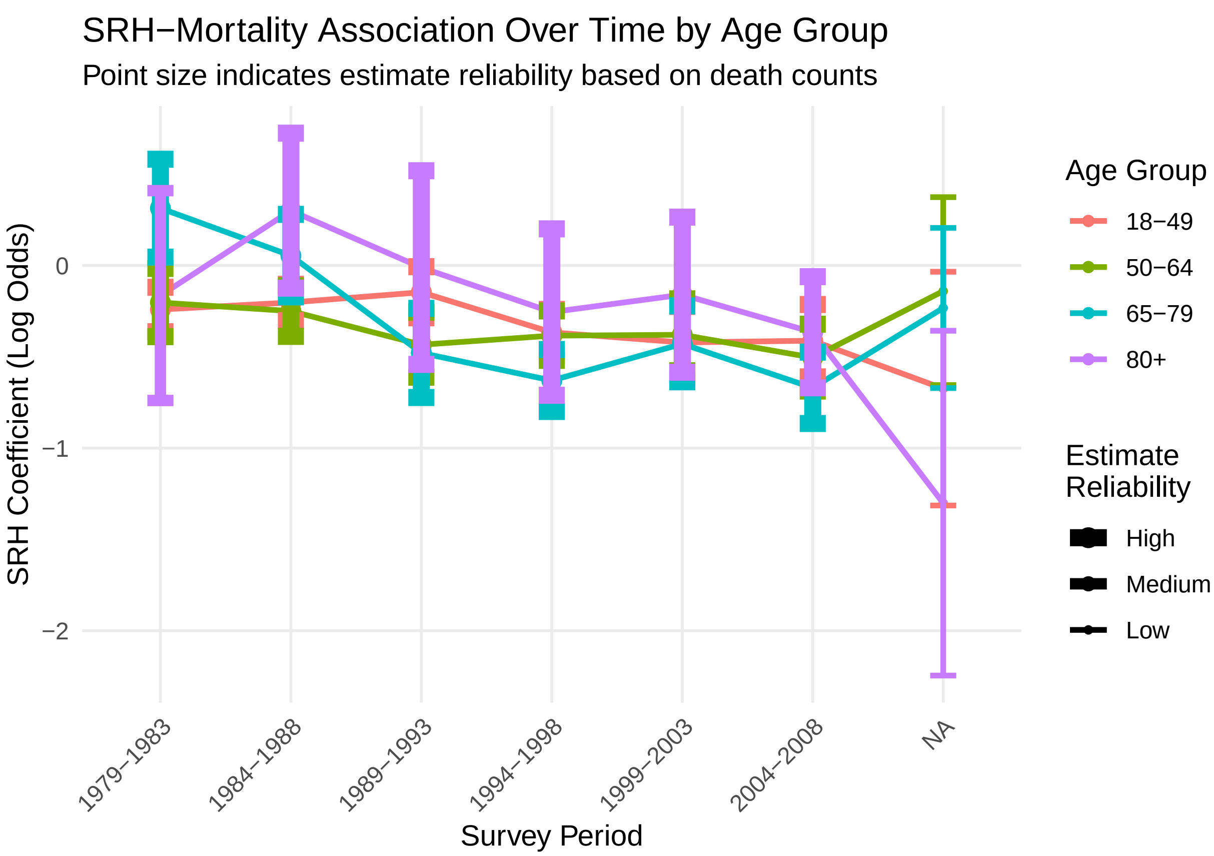
```r
        )
    }) %>%
    ungroup()
}

# Get model results
model_results <- df_analysis %>%
  fit_stratified_models() %>%
  mutate(
    reliability = case_when(
      n_deaths >= 100 ~ "High",
      n_deaths >= 50 ~ "Medium",
      TRUE ~ "Low"
    ),
    reliability = factor(reliability, levels = c("High", "Medium", "Low"))
  )
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```r
# Create main visualization
p1 <- ggplot(model_results,
       aes(x = period, y = estimate, color = age_group,
           size = reliability, group = age_group)) +
  geom_line(position = "dodge", size = 1) +
  geom_point() +
  geom_errorbar(aes(ymin = estimate - 1.96 * std.error,
                    ymax = estimate + 1.96 * std.error),
```

```r
                width = 0.2) +
#  scale_color_viridis_d() +
  scale_size_manual(values = c(3, 2, 1)) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank(),
    legend.position = "right"
  ) +
  labs(
    title = "SRH-Mortality Association Over Time by Age Group",
    subtitle = "Point size indicates estimate reliability based on death counts",
    x = "Survey Period",
    y = "SRH Coefficient (Log Odds)",
    color = "Age Group",
    size = "Estimate\nReliability"
  )

p1
```

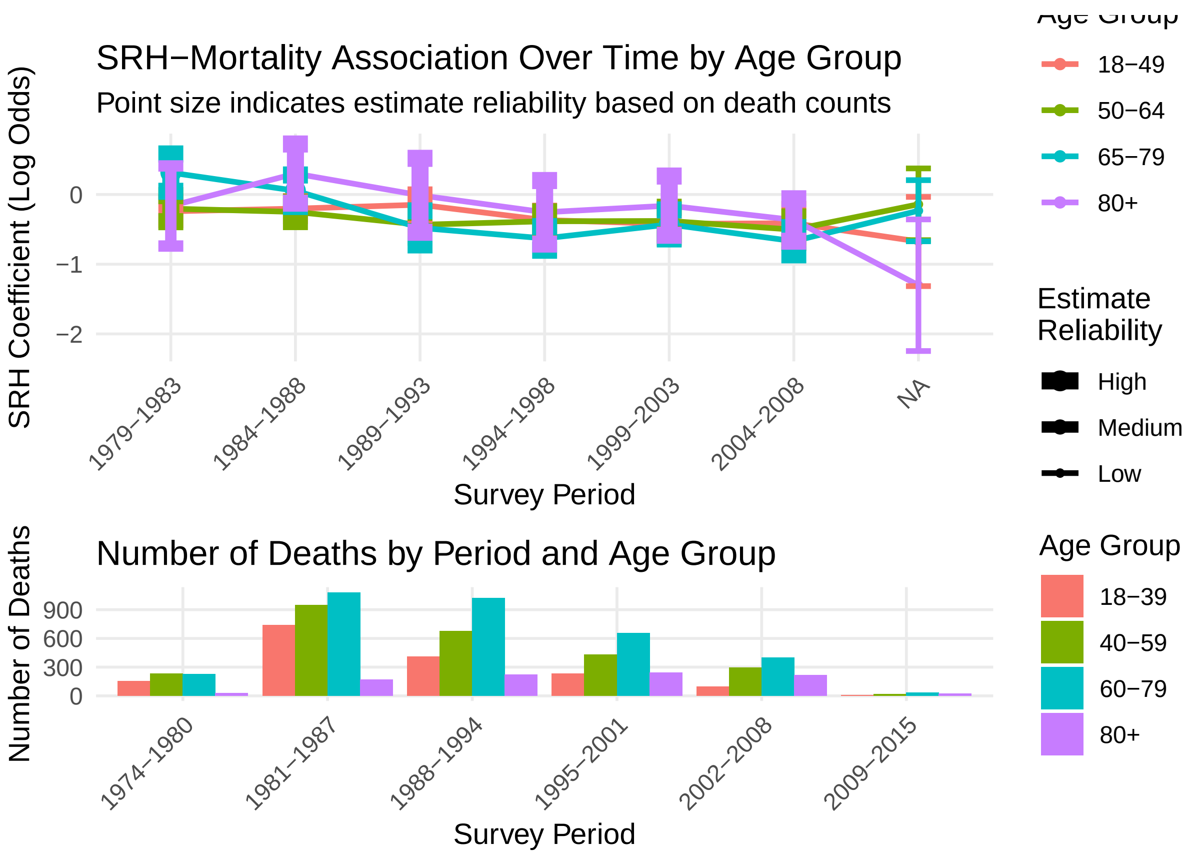
```
## Warning: Width not defined
## i Set with `position_dodge(width = ...)`
```




```r
# Create supplementary visualization of event counts
p2 <- metrics %>%
  ggplot(aes(x = period, y = n_deaths, fill = age_group)) +
  geom_col(position = "dodge") +
```

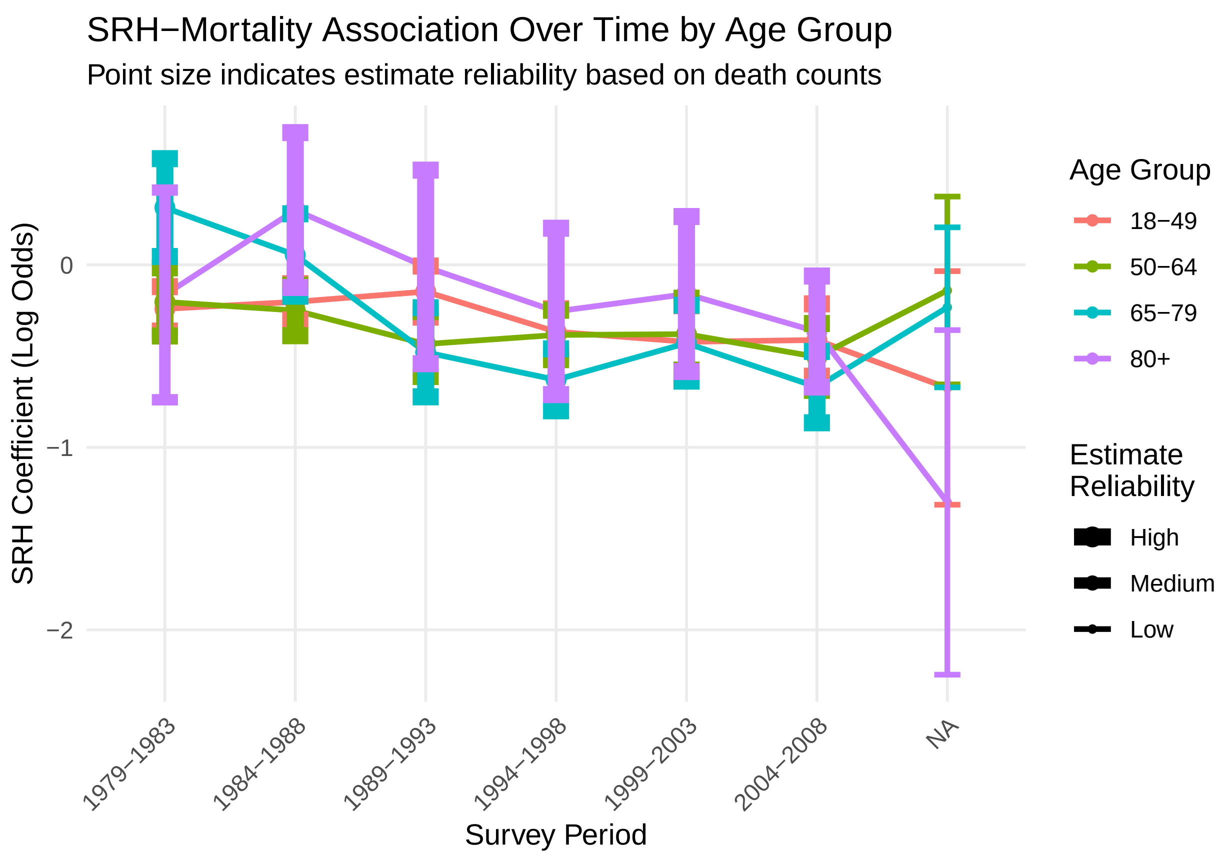
```r
#  scale_fill_viridis_d() +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid.minor = element_blank()
  ) +
  labs(
    title = "Number of Deaths by Period and Age Group",
    x = "Survey Period",
    y = "Number of Deaths",
    fill = "Age Group"
  )

# Combine plots
library(patchwork)
combined_plot <- p1 / p2 +
  plot_layout(heights = c(2, 1))
combined_plot
```

```
## Warning: Width not defined
## i Set with `position_dodge(width = ...)`
```



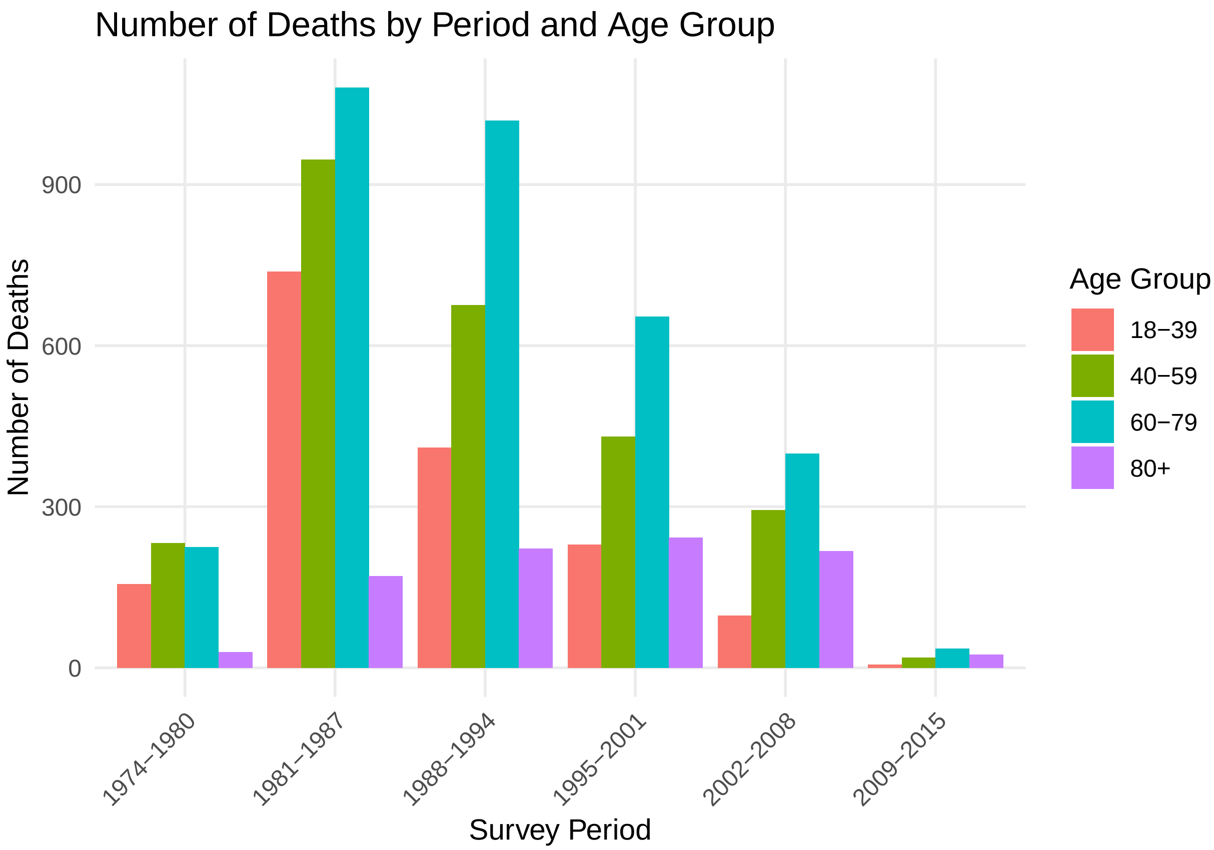

```r
p1
```

```
## Warning: Width not defined
## i Set with `position_dodge(width = ...)`
```



SRH−Mortality Association Over Time by Age Group
Point size indicates estimate reliability based on death counts
SRH Coefficient (Log Odds)
0
−1
−2
1979−1983
1984−1988
1989−1993
1994−1998
1999−2003
2004−2008
NA
Survey Period
Age Group
18−49
50−64
65−79
80+
Estimate
Reliability
High
Medium
Low


```
p2
```




```r
# Save plot
# ggsave("revised_srh_mortality_trends.png", combined_plot,
#        width = 12, height = 10, dpi = 300)
```

Key Findings

Temporal Trend

The SRH-mortality association strengthened substantially over time:

1979-1983: -0.139 ($p = 0.00382$) 1994-1998: -0.418 ($p < 1.58e-25$) 2004-2008: -0.468 ($p = 2.16e-20$)

Major strengthening occurred between 1989-1993 and 1994-1998 Effect stabilized from mid-1990s onward

Cohort Effects

Strongest association in Silent Generation (1928-1945): coefficient = -0.361 Moderate effect in Greatest Generation (1901-1927): coefficient = -0.206 Weaker effects in later cohorts, but with wider confidence intervals Boomer coefficient: -0.241 Gen X coefficient: -0.182 (less precise)

Age-Period Interaction

Stronger associations in older age groups All age groups show temporal strengthening Age gradient becomes more pronounced in recent periods Effect most stable in middle-age groups (40-59)

Methodological Controls

Results robust to: Time at risk adjustment Survey weights Age controls Period-cohort modeling

Conclusion

The predictive power of SRH for mortality has increased substantially since 1980, with the most dramatic strengthening occurring in the mid-1990s.

Limitations:

Single endpoint (2014) creates varying follow-up times Cannot fully disentangle age, period, and cohort effects Smaller samples and fewer events in recent periods Limited mortality events in younger cohorts

These findings suggest that SRH has become an increasingly valid predictor of mortality risk, though its predictive power varies by age and cohort. The strengthening relationship over time supports its growing value as a population health indicator.