

Online Technical Appendix for Fenton, Wyper, McCartney, Minton, 'Socioeconomic inequality in recent mortality trends in Scotland'

Jon Minton

Introduction

This document describes all data, methods and code used to generate the results within the above-listed article.

All data are publically available, and you are free to interrogate and use the code in your own analyses.

All data preparation, analyses, figures and tables were generated using the R programming language, with the exception of some initial data preparation done within Microsoft Excel. This data preparation only involved changing the structure of existing data within an Excel sheet, rather than the contents of such data.

Data

The data used are from this National Records of Scotland (NRS) webpage. The dataset 'Latest tables based on 2013 ESP' (European Standardized Population) was downloaded in Excel format.

All analyses presented in the paper were based on table 7 of this Excel Workbook.

The contents of table 7 of the Workbook were rearranged so as to conform with data structure recommended by (Wickham, (2014) 'Tidy Data')[<https://www.jstatsoft.org/article/view/v059i10/v59i10.pdf>], producing a new datasheet, called `flat_data`, which was imported directly into R. All further processing and analyses were conducted within R.

Workflow and results

The following code chunks perform the analyses which are presented in the paper.

Where additional information are required as to the methods, they are described adjacent to the code chunks.

Loading the required packages

```
pacman::p_load(tidyverse, readxl, cowplot, kableExtra)
```

Note: `pacman` is an R package for managing R packages. It has to be installed once using `install.packages("pacman")`, but once installed will either install or load other packages as required.

Load data

```
dta <- read_excel("data/ASMR_SIMD_2001_2017_indexed_trends.xlsx", sheet = "flat_data")
```

Tidying the data

```
names(dta) <- c("year", "overall", "q1", "q2", "q3", "q4", "q5", "gender")
```

```
dta_tidy <- dta %>%
  gather(key = "simd", value = "asmr", -year, -gender) %>%
  mutate(SIMD = factor(simd,
    levels = c("q1", "q2", "q3", "q4", "q5", "overall"),
    labels = c("Q.1 (Most deprived)",
      "Q.2", "Q.3", "Q.4", "Q.5 (Least deprived)", "Overall")
  )
)
```

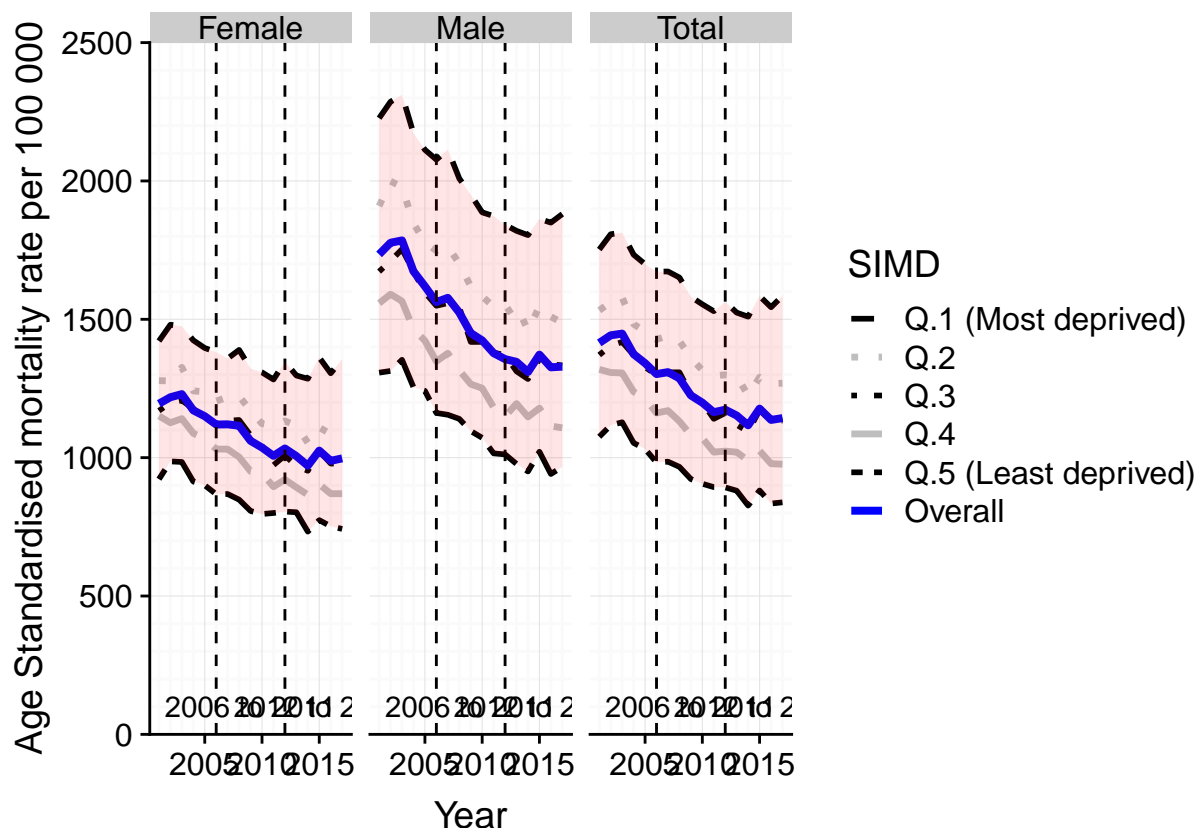
Visualising

This is one part of the first figure. The R package `ggplot2` is used to produce this visualisation.

```
p1 <- dta_tidy %>%
  ggplot(aes(x = year )) +

  facet_wrap(~gender) +
  geom_line(aes(y = asmr, group = SIMD, linetype = SIMD, size = SIMD, color = SIMD)) +
  scale_size_manual(values = c(1, 1.2, 1, 1.2, 1, 1.4)) +
  scale_linetype_manual(values = c(2,3,4,5, 6, 1)) +
  scale_color_manual(values = c("black", "grey", "black", "grey", "black", "blue")) +
  labs(x = "Year", y = "Age Standardised mortality rate per 100 000") +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 2500), minor_breaks = seq(0, 2500, by = 100)) +
  scale_x_continuous(minor_breaks = 2001:2017) +
  geom_vline(xintercept = 2012, linetype = "dashed") +
  geom_vline(xintercept = 2006, linetype = "dashed") +
  annotate("text", y = 100, x = 2006 + (2012 - 2006) / 2, label = "2006 to 2011") +
  annotate("text", y = 100, x = 2012 + (2018 - 2012) / 2, label = "2012 to 2017") +
  geom_ribbon(
    aes(x = year, ymin = q5, ymax = q1),
    alpha = 0.1, fill = "red",
    data = dta_tidy %>% filter(simd %in% c("q1", "q5")) %>% select(-SIMD) %>% spread(simd, asmr)
  ) +
  background_grid(major = "xy", minor = "xy")
```

p1



The following chunk calculates the percentage change, from the earlier to latter period, in ASMR by gender and SIMD quintile.

```
percent_changes <- dta_tidy %>%
  mutate(period = case_when(
    between(year, 2012, 2017) ~ "2012-2017",
    between(year, 2006, 2011) ~ "2006-2011",
    TRUE ~ NA_character_) %>% factor(levels = c("2012-2017", "2006-2011"))) %>%
  group_by(gender, simd, period) %>%
  filter(year == min(year) | year == max(year)) %>%
  filter(!is.na(period)) %>%
  group_by(gender, SIMD, period) %>%
  summarise(percent_change = - 100 * (1 - asmr[year == max(year)] / asmr[year == min(year)])) %>%
  ungroup()
```

```
## Warning: Factor `period` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

Note: The contents of `percent_changes` look as follows:

```
percent_changes

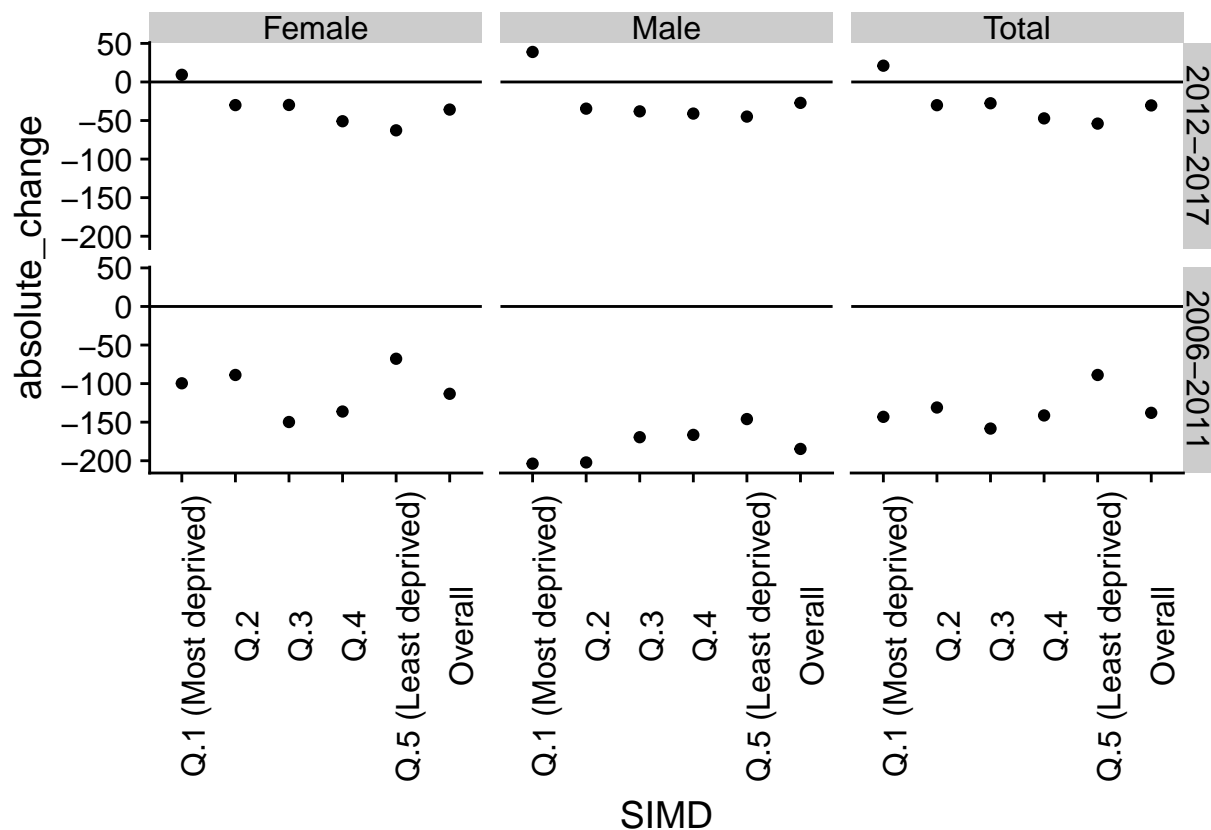
## # A tibble: 36 x 4
##   gender SIMD          period percent_change
##   <chr> <fct>          <fct>          <dbl>
## 1 Female Q.1 (Most deprived) 2012-2017         0.690
## 2 Female Q.1 (Most deprived) 2006-2011        -7.21
## 3 Female Q.2                2012-2017        -2.65
```

```
## 4 Female Q.2          2006-2011      -7.37
## 5 Female Q.3          2012-2017      -2.96
## 6 Female Q.3          2006-2011     -13.3
## 7 Female Q.4          2012-2017     -5.53
## 8 Female Q.4          2006-2011     -13.2
## 9 Female Q.5 (Least deprived) 2012-2017 -7.79
## 10 Female Q.5 (Least deprived) 2006-2011 -7.81
## # ... with 26 more rows
```

Note: The code chunk above can be readily adapted to showing absolute change (for example), as follows

```
dta_tidy %>%
  mutate(period = case_when(
    between(year, 2012, 2017) ~ "2012-2017",
    between(year, 2006, 2011) ~ "2006-2011",
    TRUE ~ NA_character_) %>% factor(levels = c("2012-2017", "2006-2011"))) %>%
  group_by(gender, simd, period) %>%
  filter(year == min(year) | year == max(year)) %>%
  filter(!is.na(period)) %>%
  group_by(gender, SIMD, period) %>%
  summarise(
    percent_change = - 100 * (1 - asmr[year == max(year)] / asmr[year == min(year)]),
    absolute_change = asmr[year==max(year)] - asmr[year == min(year)] # additional line
  ) %>%
  ungroup() %>% # indicative visualisation
  ggplot(aes(x = SIMD, y = absolute_change)) +
  geom_point() +
  facet_grid(period~gender) +
  geom_hline(yintercept = 0) +
  theme(axis.text.x = element_text(angle = 90))
```

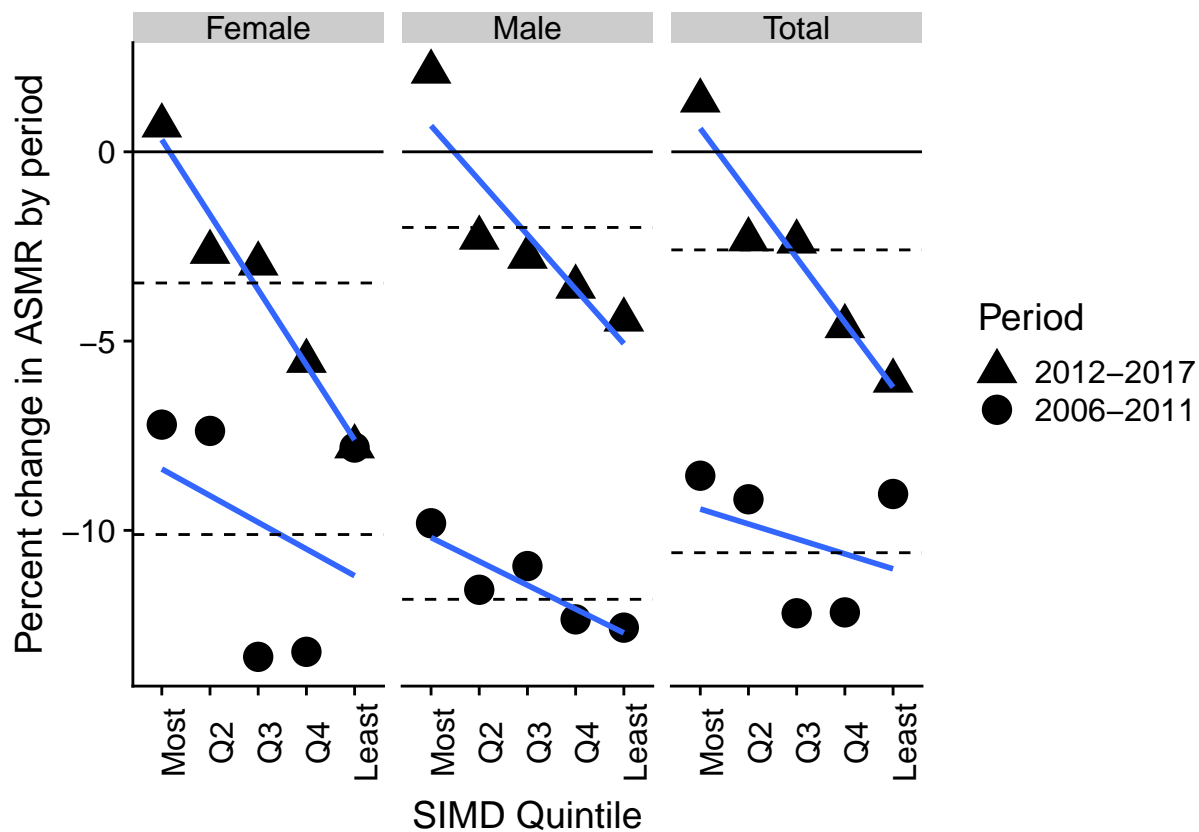
```
## Warning: Factor `period` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```



The following shows the second part of the figure in the paper

```
p2 <- percent_changes %>%
  mutate(SIMD = factor(SIMD,
    levels = c("Q.1 (Most deprived)",
      "Q.2", "Q.3", "Q.4", "Q.5 (Least deprived)", "Overall"),
    labels = c("Most", "Q2", "Q3", "Q4", "Least", "Overall")
  )
) %>%
  filter(SIMD != "Overall") %>%
  ggplot(aes(x = SIMD, y = percent_change, group = period, shape = period)) +
  facet_wrap(~ gender) +
  geom_point(size = 5) +
  stat_smooth(method = "lm", se = F) + # This produces the blue line with the regression slopes
  geom_hline(yintercept = 0) +
  geom_hline( # This adds the overall percent change
    aes(yintercept = percent_change, group = period),
    data = percent_changes %>% filter(SIMD == "Overall"),
    linetype = "dashed"
  ) +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(y = "Percent change in ASMR by period", x = "SIMD Quintile") +
  scale_shape_manual("Period", values = c(17, 16))
```

p2

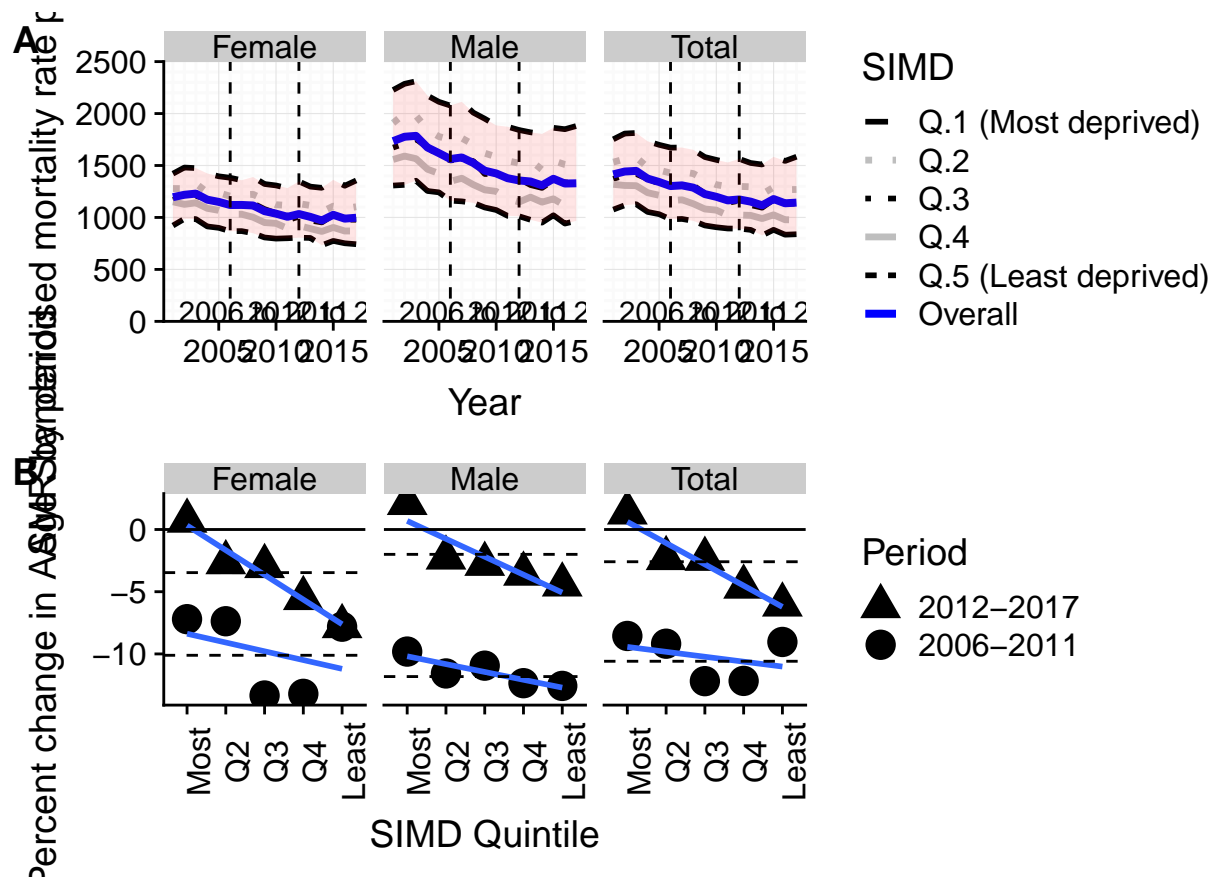


Combined figure

The following produces the combined figure comprising the two parts shown previously. The figure is rendered as a png format image at 300dpi, and placed in the directory 'figures'.

```
p_both <- plot_grid(p1, p2, labels = c("A", "B"), ncol = 1, align = "v")
```

```
p_both
```



```
ggsave("figures/combined_figure_2012.png", dpi = 300, units = "cm", height = 30, width = 30)
```

Table

The following shows how the contents of the table were produced.

These summarise a series of univariate linear regressions of SIMD quintile against percentage change in ASMR within distinct periods.

The columns **intercept** and **gradient** present summary information about point estimates and 95% confidence intervals for the intercept and gradient of the the regressions, which are then presented in the formatted table.

A functional programming approach was adopted in order to produce the same analyses consistently for different gender and period combinations, using functions within the (**purrr** package)[<https://purrr.tidyverse.org/>].

```
get_ci <- function(x){
  tmp <- x %>% summary() %>% coefficients()

  return(
    list(
      lower = tmp[,1] - 1.96 * tmp[,2],
      upper = tmp[,1] + 1.96 * tmp[,2]
    )
  )
}
```

```

# Model parameters
tbl_1 <- percent_changes %>%
  filter(SIMD != "Overall") %>%
  mutate(qnt = unclass(SIMD) - 1) %>% # This is so the intercept refers to the 1st quintile (not the 'z
  select(gender, period, percent_change, qnt) %>%
  group_by(gender, period) %>%
  nest() %>%
  mutate(mdl = map(data, ~lm(percent_change ~ qnt, data = .x))) %>%
  mutate(`R. sq.` = map_dbl(mdl, ~summary(.x)[`r.squared`][1])) %>%
  mutate(gradient = map_dbl(mdl, ~coef(.x)[`qnt`])) %>%
  mutate(intercept = map_dbl(mdl, ~coef(.x)[`Intercept`])) %>%
  mutate(cis = map(mdl, get_ci)) %>%
  mutate(
    int_lower = map_dbl(cis, ~.[`lower`][1]),
    int_upper = map_dbl(cis, ~.[`upper`][1]),
    grd_lower = map_dbl(cis, ~.[`lower`][2]),
    grd_upper = map_dbl(cis, ~.[`upper`][2])
  ) %>%
  select(gender, period, `R. sq.`,
         gradient, grd_lower, grd_upper,
         intercept, int_lower, int_upper
  ) %>%
  mutate(
    gradient = paste0(
      format(round(gradient, 2), nsmall = 2),
      " (",
      format(round(grd_lower, 2), nsmall = 2),
      ", ",
      format(round(grd_upper, 2), nsmall = 2),
      ")"
    )
  ) %>%
  mutate(
    intercept = paste0(
      format(round(intercept, 2), nsmall = 2),
      " (",
      format(round(int_lower, 2), nsmall = 2),
      ", ",
      format(round(int_upper, 2), nsmall = 2),
      ")"
    )
  ) %>%
  select(-grd_lower, -grd_upper, -int_lower, -int_upper)

tbl_1

```

```

## # A tibble: 6 x 5
##   gender period   `R. sq.` gradient      intercept
##   <chr>   <fct>     <dbl> <chr>         <chr>
## 1 Female 2012-2017  0.960 -1.98 (-2.44, -1.53) " 0.32 ( -0.80, 1.44)"
## 2 Female 2006-2011  0.122 -0.70 (-2.84, 1.44) " -8.38 (-13.62, -3.14)"
## 3 Male   2012-2017  0.803 -1.44 (-2.25, -0.63) " 0.69 ( -1.29, 2.67)"
## 4 Male   2006-2011  0.790 -0.63 (-1.00, -0.26) -10.19 (-11.09, -9.29)
## 5 Total  2012-2017  0.928 -1.71 (-2.25, -1.17) " 0.62 ( -0.70, 1.93)"

```


Table 1: Percent change in ASMR by gender, SIMD quintile, and period

gender	period	Percentages						R. sq.	gradient
		Q.1 (Most deprived)	Q.2	Q.3	Q.4	Q.5 (Least deprived)	Overall		
Female	2006-2011	-7.21	-7.37	-13.34	-13.21	-7.81	-10.11	0.12	-0.70 (-2.39, 1.60)
Female	2012-2017	0.69	-2.65	-2.96	-5.53	-7.79	-3.46	0.96	-1.98 (-2.99, -0.97)
Male	2006-2011	-9.81	-11.57	-10.94	-12.35	-12.57	-11.82	0.79	-0.63 (-1.60, 0.81)
Male	2012-2017	2.11	-2.27	-2.78	-3.57	-4.44	-2.00	0.80	-1.44 (-2.45, -0.43)
Total	2006-2011	-8.56	-9.18	-12.19	-12.17	-9.04	-10.59	0.12	-0.40 (-1.60, 0.81)
Total	2012-2017	1.35	-2.32	-2.38	-4.61	-6.05	-2.59	0.93	-1.71 (-2.72, -0.70)

Note:

Overall: Whole of Scotland. R.Sq. : R-Squared for model. Gradient: Increase in % change per unit increase in quintile. In

```
## 6 Total 2006-2011 0.121 -0.40 (-1.60, 0.81) " -9.44 (-12.39, -6.48)"
```

```
tbl_2 <- percent_changes %>% spread(SIMD, percent_change)
```

```
tbl_both <- inner_join(tbl_2, tbl_1)
```

```
## Joining, by = c("gender", "period")
```

```
tbl_both
```

```
## # A tibble: 6 x 11
##   gender period `Q.1 (Most depr~` Q.2 Q.3 Q.4 `Q.5 (Least dep~
##   <chr> <fct>      <dbl> <dbl> <dbl> <dbl>      <dbl>
## 1 Female 2012-~      0.690 -2.65 -2.96 -5.53      -7.79
## 2 Female 2006-~     -7.21 -7.37 -13.3 -13.2      -7.81
## 3 Male 2012-~      2.11 -2.27 -2.78 -3.57      -4.44
## 4 Male 2006-~     -9.81 -11.6 -10.9 -12.4     -12.6
## 5 Total 2012-~      1.35 -2.32 -2.38 -4.61      -6.05
## 6 Total 2006-~     -8.56 -9.18 -12.2 -12.2     -9.04
## # ... with 4 more variables: Overall <dbl>, `R. sq.` <dbl>,
## #   gradient <chr>, intercept <chr>
```

The following converts the above table into a more neatly formatted table using the `kable` and `kableExtra` packages.

```
tbl_both %>%
  mutate(period = factor(period, levels = c("2006-2011", "2012-2017"))) %>%
  arrange(gender, period) %>%
  knitr::kable(
    digits = 2,
    caption = "Percent change in ASMR by gender, SIMD quintile, and period"
  ) %>%
  kableExtra::kable_styling() %>%
  kableExtra::add_header_above(c(" ", " ", "Percentages" = 6, "Model results" = 3)) %>%
  kableExtra::footnote("Overall: Whole of Scotland. R.Sq. : R-Squared for model. Gradient: Increase in % change per unit increase in quintile. In
```

Approach discussed in sensitivity analysis

The sensitivity analysis to the paper showed the effect of using the fitted values for the first and last year in each of the periods, rather than the values themselves. This approach can address any concern that the first and last year within either period were in any way anomalous or uncharacteristic of change within the period as a whole.

The function `broom::augment` was used to extract fitted values for each year within each period, gender and SIMD combination. The fitted values, `.fitted`, were then used in place of the observed values, `asmr`, as in the main analyses.

```
percent_changes_pred <- dta_tidy %>%
  mutate(period = case_when(
    between(year, 2012, 2017) ~ "2012-2017",
    between(year, 2006, 2011) ~ "2006-2011",
    TRUE ~ NA_character_) %>% factor(levels = c("2012-2017", "2006-2011"))) %>%
  group_by(gender, SIMD, simd, period) %>%
  nest() %>%
  mutate(mdl = map(data, ~lm(asmr ~ year, data = .x))) %>%
  mutate(aug = map(mdl, broom::augment)) %>%
  select(-data, -mdl) %>%
  unnest() %>%
  filter(!is.na(period)) %>%
  group_by(gender, SIMD, period) %>%
  filter(year == min(year) | year == max(year)) %>%
  summarise(percent_change = - 100 * (1 - .fitted[year == max(year)] / .fitted[year == min(year)])) %>%
  ungroup()
```

```
## Warning: Factor `period` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## Warning: Factor `period` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
percent_changes_pred_overall <- percent_changes_pred %>%
  group_by(gender, period) %>%
  summarise(percent_change = mean(percent_change)) %>%
  ungroup() %>%
  mutate(SIMD = "Mean of quintiles") %>%
  select(gender, SIMD, period, percent_change)

percent_changes_pred <- percent_changes_pred %>%
  bind_rows(percent_changes_pred_overall) %>%
  mutate(SIMD = factor(SIMD,
    levels = c("Q.1 (Most deprived)",
               "Q.2", "Q.3", "Q.4", "Q.5 (Least deprived)", "Overall", "Mean of quintiles"),
    labels = c("Most", "Q2", "Q3", "Q4", "Least", "Overall", "Mean of quintiles")
  )
)
```

```
## Warning in bind_rows(x, .id): binding factor and character vector,
## coercing into character vector
```

```
## Warning in bind_rows(x, .id): binding character and factor vector,
## coercing into character vector
```

The subfigure using this alternative stratgy is therefore produced as follows:

```
p2a <- percent_changes_pred %>%
  filter(!(SIMD %in% c("Overall", "Mean of quintiles"))) %>%
  ggplot(aes(x = SIMD, y = percent_change, group = period, shape = period)) +
  facet_wrap(~ gender) +
  geom_point(size = 5) +
  stat_smooth(method = "lm", se = F) +
```

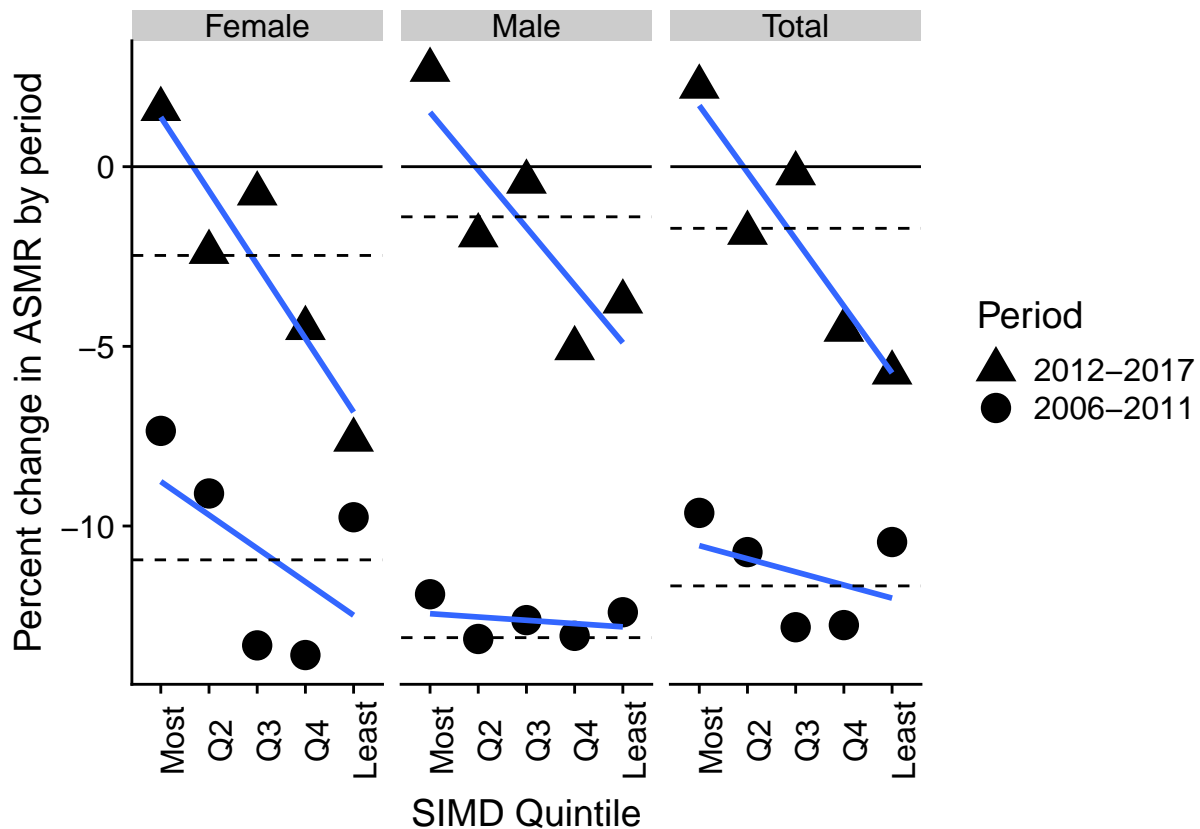
```

geom_hline(yintercept = 0) +
geom_hline(
  aes(yintercept = percent_change, group = period),
  data = percent_changes_pred %>% filter(SIMD == "Overall"),
  linetype = "dashed"
) +

theme(axis.text.x = element_text(angle = 90)) +
labs(y = "Percent change in ASMR by period", x = "SIMD Quintile") +
scale_shape_manual("Period", values = c(17, 16))

```

p2a



```

ggsave("figures/fig1a_using_alt_method_2012.png", dpi = 300, units = "cm", height = 16, width = 30)

```

The table of the above, using this alternative modelling strategy, is produced using the code chunks below

```

get_ci <- function(x){
  tmp <- x %>% summary() %>% coefficients()

  return(
    list(
      lower = tmp[,1] - 1.96 * tmp[,2],
      upper = tmp[,1] + 1.96 * tmp[,2]
    )
  )
}

```

```

}

# Model parameters
tbl_1a <- percent_changes_pred %>%
  filter(!(SIMD %in% c("Overall", "Mean of quintiles"))) %>%
  mutate(qnt = unclass(SIMD) - 1) %>% # This is so the intercept refers to the 1st quintile (not the 'z
  select(gender, period, percent_change, qnt) %>%
  group_by(gender, period) %>%
  nest() %>%
  mutate(mdl = map(data, ~lm(percent_change ~ qnt, data = .x))) %>%
  mutate(`R. sq.` = map_dbl(mdl, ~summary(.x)[`r.squared`][1])) %>%
  mutate(gradient = map_dbl(mdl, ~coef(.x)[`qnt`])) %>%
  mutate(intercept = map_dbl(mdl, ~coef(.x)[`Intercept`])) %>%
  mutate(cis = map(mdl, get_ci)) %>%
  mutate(
    int_lower = map_dbl(cis, ~.[`lower`][1]),
    int_upper = map_dbl(cis, ~.[`upper`][1]),
    grd_lower = map_dbl(cis, ~.[`lower`][2]),
    grd_upper = map_dbl(cis, ~.[`upper`][2])
  ) %>%
  select(gender, period, `R. sq.`,
         gradient, grd_lower, grd_upper,
         intercept, int_lower, int_upper
  ) %>%
  mutate(
    gradient = paste0(
      format(round(gradient, 2), nsmall = 2),
      " (",
      format(round(grd_lower, 2), nsmall = 2),
      ", ",
      format(round(grd_upper, 2), nsmall = 2),
      ")"
    )
  ) %>%
  mutate(
    intercept = paste0(
      format(round(intercept, 2), nsmall = 2),
      " (",
      format(round(int_lower, 2), nsmall = 2),
      ", ",
      format(round(int_upper, 2), nsmall = 2),
      ")"
    )
  ) %>%
  select(-grd_lower, -grd_upper, -int_lower, -int_upper)

tbl_1a

```

```

## # A tibble: 6 x 5
##   gender period   `R. sq.` gradient      intercept
##   <chr>   <fct>     <dbl> <chr>         <chr>
## 1 Female 2012-2017  0.848 -2.05 (-3.03, -1.07) " 1.38 ( -1.03,  3.78)"
## 2 Female 2006-2011  0.290 -0.93 (-2.58,  0.72) " -8.76 (-12.80, -4.72)"
## 3 Male   2012-2017  0.705 -1.60 (-2.78, -0.43) " 1.51 ( -1.36,  4.39)"

```

Table 2: Percent change in ASMR by gender, SIMD quintile, and period. (Alternative method)

gender	period	Percentages								
		Most	Q2	Q3	Q4	Least	Overall	R. sq.	gradient	intercept
Female	2006-2011	-7.35	-9.09	-13.32	-13.59	-9.76	-10.94	0.29	-0.93 (-2.58, 0.72)	-8.76 (-12.80, -4.72)
Female	2012-2017	1.60	-2.38	-0.75	-4.50	-7.60	-2.47	0.85	-2.05 (-3.03, -1.07)	1.38 (-1.03, 3.77)
Male	2006-2011	-11.90	-13.14	-12.62	-13.06	-12.40	-13.11	0.08	-0.09 (-0.44, 0.26)	-12.44 (-13.29, -11.59)
Male	2012-2017	2.69	-1.92	-0.42	-5.06	-3.76	-1.40	0.71	-1.60 (-2.78, -0.43)	1.51 (-1.36, 4.37)
Total	2006-2011	-9.63	-10.73	-12.82	-12.76	-10.45	-11.67	0.16	-0.37 (-1.31, 0.58)	-10.55 (-12.85, -8.24)
Total	2012-2017	2.22	-1.85	-0.19	-4.54	-5.73	-1.72	0.83	-1.86 (-2.80, -0.92)	1.70 (-0.60, 4.01)

Note:

Overall: Whole of Scotland. R.Sq. : R-Squared for model. Gradient: Increase in % change per unit increase in quintile. Intercept: Overall intercept.

```
## 4 Male    2006-2011    0.0813 -0.09 (-0.44,  0.26) -12.44 (-13.29, -11.59)
## 5 Total   2012-2017    0.834  -1.86 (-2.80, -0.92) "  1.70 ( -0.60,  4.01)"
## 6 Total   2006-2011    0.162  -0.37 (-1.31,  0.58) -10.55 (-12.85, -8.24)
```

```
tbl_2a <- percent_changes_pred %>% filter(SIMD != "Mean of quintiles") %>% spread(SIMD, percent_change)
```

```
tbl_both_a <- inner_join(tbl_2a, tbl_1a)
```

```
## Joining, by = c("gender", "period")
```

```
tbl_both_a
```

```
## # A tibble: 6 x 11
##   gender period    Most      Q2      Q3      Q4    Least Overall `R. sq.`
##   <chr>   <fct>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 Female 2012-~    1.60   -2.38  -0.751  -4.50   -7.60   -2.47   0.848
## 2 Female 2006-~   -7.35  -9.09 -13.3    -13.6   -9.76   -10.9    0.290
## 3 Male   2012-~    2.69  -1.92  -0.424  -5.06   -3.76   -1.40    0.705
## 4 Male   2006-~  -11.9 -13.1  -12.6   -13.1  -12.4   -13.1    0.0813
## 5 Total  2012-~    2.22  -1.85  -0.190  -4.54   -5.73   -1.72    0.834
## 6 Total  2006-~   -9.63 -10.7  -12.8   -12.8  -10.4   -11.7    0.162
## # ... with 2 more variables: gradient <chr>, intercept <chr>
```

And the table using the alternative approach

```
tbl_both_a %>%
  mutate(period = factor(period, levels = c("2006-2011", "2012-2017"))) %>%
  arrange(gender, period) %>%
  knitr::kable(
    digits = 2,
    caption = "Percent change in ASMR by gender, SIMD quintile, and period. (Alternative method)"
  ) %>%
  kableExtra::kable_styling() %>%
  kableExtra::add_header_above(c(" ", " ", "Percentages" = 6, "Model results" = 3)) %>%
  kableExtra::footnote("Overall: Whole of Scotland. R.Sq. : R-Squared for model. Gradient: Increase in % change per unit increase in quintile. Intercept: Overall intercept.")
```

Discussion

This document has provided descriptions of the data used, the processing performed on the data, and the code used to perform all analyses, visualisations, and tabulations. Our hope is this addresses any methodological concerns from viewers and reviewers, and will make it much more straightforward for anyone who wants to replicate and advance on our analyses to do so.