

Estimating Bayes Factors for a Slowdown in life expectancy since 2012

Introduction

Though it is commonly accepted that life expectancy gains have slowed down, or been ‘stalling’, in the UK since around 2012, this determination has been made largely on the basis of descriptive statistics and figures, rather than any form of statistical hypothesis test. This determination appears appropriate given that, in the UK, there have been substantial changes in the economy and labour market, in how social security has been administered, in the funding for adult social care, and in fiscal policy (‘austerity’) within the UK since the Global Financial Crisis (GFC) of 2008, and the Conservative-led Coalition of 2010, and subsequent Conservative and Conservative-led governments in the years since. It is particularly noteworthy that the economic growth rate since 2008 has been the slowest, and post-recession recovery the weakest, in living memory. It would seem implausible to assume that largely unprecedented declines in economic growth rates, experienced for over a decade, would *not* have some adverse effect on population health. There are clear reasons to assume that fiscal and economic factors will have a causal and adverse role in population health, as measured through mortality measures including age-standardised mortality rates and life expectancy at birth and from other ages. However, to date the question of *whether, how likely, and by what degree* a slowdown in life expectancy has occurred has not been addressed in most of the public health literature within the framework of formal statistical inference and hypothesis testing. This short paper addresses this gap in the literature.

The Classical framework for statistical inference and hypothesis testing is the Neyman-Pearson Hypothesis Testing (NPHT) framework. This involves producing a statistical model definition of the Null Hypothesis, a model in which the outcome *is not* affected by the exposure, and comparing observed outcomes against the range of outcomes generated by the Null Model. If the observed values are more extreme than the vast majority of the values that could be expected if the Null Hypothesis were true, then the Null Hypothesis can be described as ‘rejected’, and so there is some reason to believe the Alternative Hypothesis, that the relation between exposure and outcome is as hypothesised, is correct. By convention, the observed values must be more extreme than at least 5% of the values expected under the Null, so even if the evidence suggests that the Alternative Hypothesis is *more likely* than the Null Hypothesis, the Null Hypothesis can still be ‘failed to be rejected’.

The Classical framework for hypothesis testing, with its critical threshold of 5% or lower, can be likened to a criminal court, in which one of the sides in a dispute - the prosecution - has a higher burden of proof than the other side - the defence. The Null Hypothesis must be disproven ‘beyond reasonable doubt’ in order for the judgement to be made in favour of the Alternative Hypothesis. Because there are so many potential exposure-outcome relationships that could be tested for, and because the cost of producing more observations to test is often low, even higher burdens of proof - critical thresholds of 1% or lower - are increasingly being favoured. The critical threshold used to test for the presence of the Higgs boson, for example, was ‘five sigma’, a probability that the values observed, if the Standard Model were correct, of less than around 1 in 3.5 million. This threshold was chosen both because the evidence in favour of the Higgs Boson needed to be so strong that it would be unrealistic to assume it would ever be undermined in the whole of Human civilisation, and also because the marginal cost of producing new observations - new readings from the Large Hadron Collider - was miniscule, with the LHC capable of generating around 600 million new observations per second.

By contrast, and by definition, a country’s population can produce one new value of observed annual life expectancy per year, and often takes over a year to finalise life expectancy values. If we propose that life expectancy has fallen since 2012, and that the last available life expectancy value for the UK is for 2017, we therefore have just five observations - five observed annual changes - to weigh up. These five observations are unlikely to tip the scales heavily towards either the side of the defence - that life expectancies improvements

have not slowed - or the prosecution - that the improvements *have* slowed - but we can still weigh them to make better determinations and decisions. And weigh them we should, for each of these observations potentially carries within it tens of thousands of additional deaths, lives not lived, preventable deaths that may have been caused by deliberate and intentional government policy. To wait until the weight of evidence tips the scales of Classical hypothesis testing beyond an arbitrary 5% critical threshold will likely mean waiting, and not acting, for a generation or more, decades in which we will have not acted, but have had some evidence throughout to suggest we should have acted, and as a result of not acting the total excess death toll may have passed from five, to six, to seven figures. To not act according to the best possible interpretation of the only available evidence is negligent. A more pragmatic approach to weighing the evidence in favour of different hypotheses about the slowdown is needed.

The Bayes Factor provides this more pragmatic approach. If the Classical framework works like a Criminal court, with the burden of proof on the prosecution, so the Bayes Factor operates more like a Civil court, requiring only that the 'balance of probabilities' favour the prosecution's case over the defence. The Bayes Factor is defined as the ratio of two likelihoods: The likelihood of the model of Alternative Hypothesis (the prosecution) divided by the likelihood of the model of the Null Hypothesis (the defence). A Bayes Factor of 2.0 means the evidence favours the prosecution twice as strongly as the defence, and a Bayes factor of 0.5 (1/2.0) means the evidence favours the defence twice as strongly as the prosecution. Importantly, a Bayes Factor above 1 means the evidence favours the prosecution, and a Bayes Factor below 1 means the evidence favours the defence. As new evidence is introduced, the Bayes Factor can be updated, with the scales tipping either ever more towards the the side already favoured, or becoming more equivocal and balanced.

The evidence and the prosecution's argument

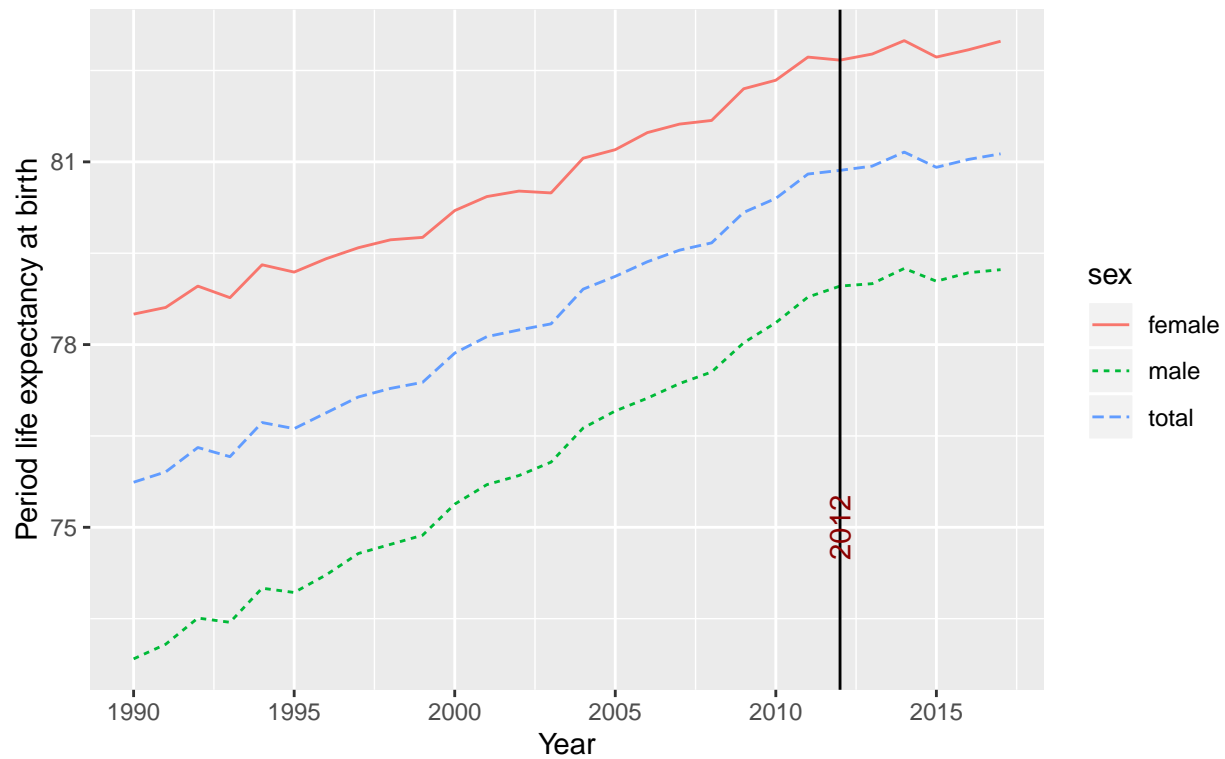
```
source("scripts/load_packages_and_functions.R")

e0_uk <- read_csv("data/e0_uk.csv")

## Parsed with column specification:
## cols(
##   year = col_double(),
##   sex = col_character(),
##   e0 = col_double()
## )

e0_uk %>%
  filter(year >= 1990) %>%
  ggplot(aes(x = year, y = e0, group = sex, colour = sex, linetype = sex)) +
  geom_line() +
  labs(x = "Year", y = "Period life expectancy at birth",
       title = "Life expectancy in the UK since 1990",
       caption = "Sources: HMD; ONS for 2017") +
  geom_vline(xintercept = 2012) +
  annotate(geom = "text", x = 2012, y = 75, label = "2012", colour = "darkred", angle = 90) +
  scale_x_continuous(breaks = seq(1990, 2015, by = 5))
```

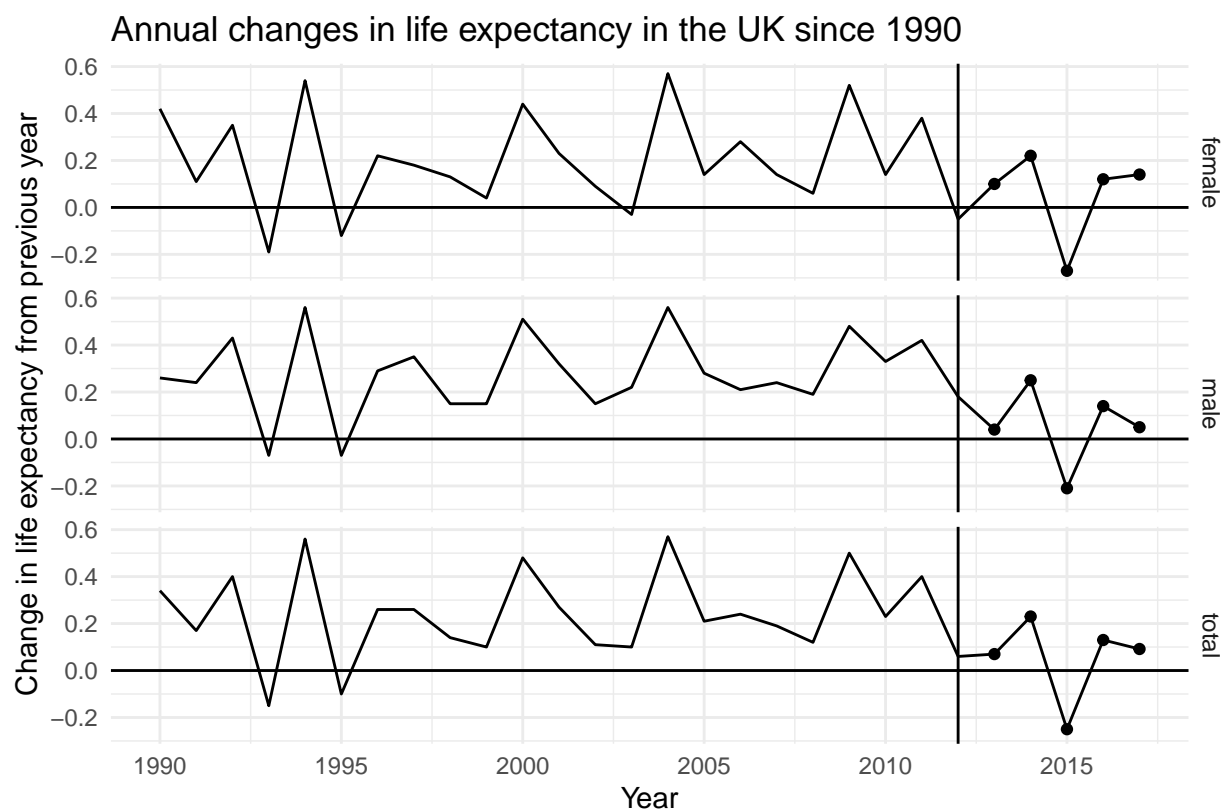
Life expectancy in the UK since 1990



Sources: HMD; ONS for 2017

Another way of presenting this, which makes for an easier model of the Null Hypothesis, is to look at annual changes in life expectancy between successive years. This is shown as follows

```
e0_uk %>%
  group_by(sex) %>%
  arrange(year) %>%
  mutate(ch_e0 = e0 - lag(e0)) %>%
  filter(year >= 1990) %>%
  ggplot(aes(x = year, y = ch_e0)) +
  geom_line() +
  geom_point(aes(x = year, y = ch_e0), inherit.aes = FALSE,
             data = . %>% filter(year > 2012))
  ) +
  labs(x = "Year", y = "Change in life expectancy from previous year",
       title = "Annual changes in life expectancy in the UK since 1990",
       caption = "Sources: HMD; ONS for 2017") +
  geom_vline(xintercept = 2012) +
  scale_x_continuous(breaks = seq(1990, 2015, by = 5)) +
  geom_hline(yintercept = 0) +
  facet_grid(sex ~ .) +
  theme_minimal()
```



Sources: HMD; ONS for 2017

The prosecution's argument is therefore that the average rate of annual improvement since 2012, the five points to the right of the vertical line, has been smaller than the rate of improvement observed in the earlier period. The defence's argument is therefore that the average annual rate of annual improvement is *not* lower than in the earlier period. As there is some degree of variability in annual improvement rates this is not completely implausible: If each annual improvement rate is a random draw from a statistical distribution and, being random, a small series of 'bad draws' are possible. The likelihood of a run of 'bad luck', successive draws from the lower end of the improvement distribution, decreases with each observation, however. We will therefore see how the Bayes Factor, favouring the assumption of some degree of slowdown in life expectancy gains, changes as each of the five observations seen after 2012 are added in sequence. This will give a sense of whether and by what extent the evidence in favour of a slowdown has been increasing, an exercise that should be repeated as each new observation becomes available.

Defining the models

The annual improvement in life expectancy for a year t can be modelled as a draw from a Normal distribution with parameters μ , the expectation of the model, and σ^2 , its variability. For the Null model, the proposition is as follows:

The annual changes in life expectancy after 2012 are draws from Normal distribution with μ equal to the sample mean observed between 1990 and 2012 inclusive, and σ^2 equal to the variance in these observations. These results are the following by gender and in total:

```
e0_uk %>%
  group_by(sex) %>%
  arrange(year) %>%
  mutate(ch_e0 = e0 - lag(e0)) %>%
  filter(between(year, 1990, 2012)) %>%
```

```
summarise(
  mu = mean(ch_e0), var = var(ch_e0)
)
```

```
## # A tibble: 3 x 3
##   sex      mu    var
##   <chr> <dbl> <dbl>
## 1 female 0.200 0.0443
## 2 male   0.277 0.0285
## 3 total  0.237 0.0358
```

i.e. the Null model proposes that the average annual rate of improvement per year remains at 0.20 years/year for females, 0.28 years per year for males, and 0.24 years per year for both genders combined. The variance is assumed to be higher for females than for males.

The Alternative Hypothesis has usually been expressed verbally as a ‘slowdown’. This implies that the expected life expectancy improvements after 2012 are below those previously observed, but are still positive (i.e. that there have been slower rates of improvement, rather than a complete halting of improvements). However a large number of potential model specifications are compatible with this hypothesis, each of which may be more, or less, likely than the Null Model. In order to represent this range of possible Alternative Model specifications compatible with the Slowdown Hypothesis, a grid search approach is used: The μ values for each gender, and total, based on the 1990-2012 observations are multiplied by various percentages ranging, by 1% at a time, from 0% to 100%. Multiplying μ by 0% is the hypothesis that mortality rates have slowed down completely, whereas multiplying μ by 100% is the hypothesis that there has been no slowdown whatsoever. For each of these % reduction hypotheses, compatible with the Slowdown Hypothesis, the Bayes Factor compared with the Null Model, no slowdown, is calculated. The schedules showing how the Bayes Factor changes with each proposed % reduction in annual improvement are shown first for the data point 2012-13, then for the series of data points {2012-13, 2013-14}, up to the full available sequence of five observations from 2012-13 through to 2016-17. This shows how the accumulated weight of evidence has shifted over time.

```
e0_uk_bf <-
  e0_uk %>%
  group_by(sex) %>%
  arrange(year) %>%
  mutate(ch_e0 = e0 - lag(e0)) %>%
  nest() %>%
  crossing(after_end = 2013:2017) %>%
  mutate(
    bayes_df = map2(
      after_end, data, ~calc_bayes_factors(after_period = c(2013, .x), before_period = c(1990, 2012), or
    )
  ) %>%
  select(sex, after_end, bayes_df) %>%
  unnest() %>%
  mutate(
    period = paste0("2012-", str_sub(after_end, 3,4))
  )
```

Results

The following shows the Bayes Factor based on the single observation 2012-13:

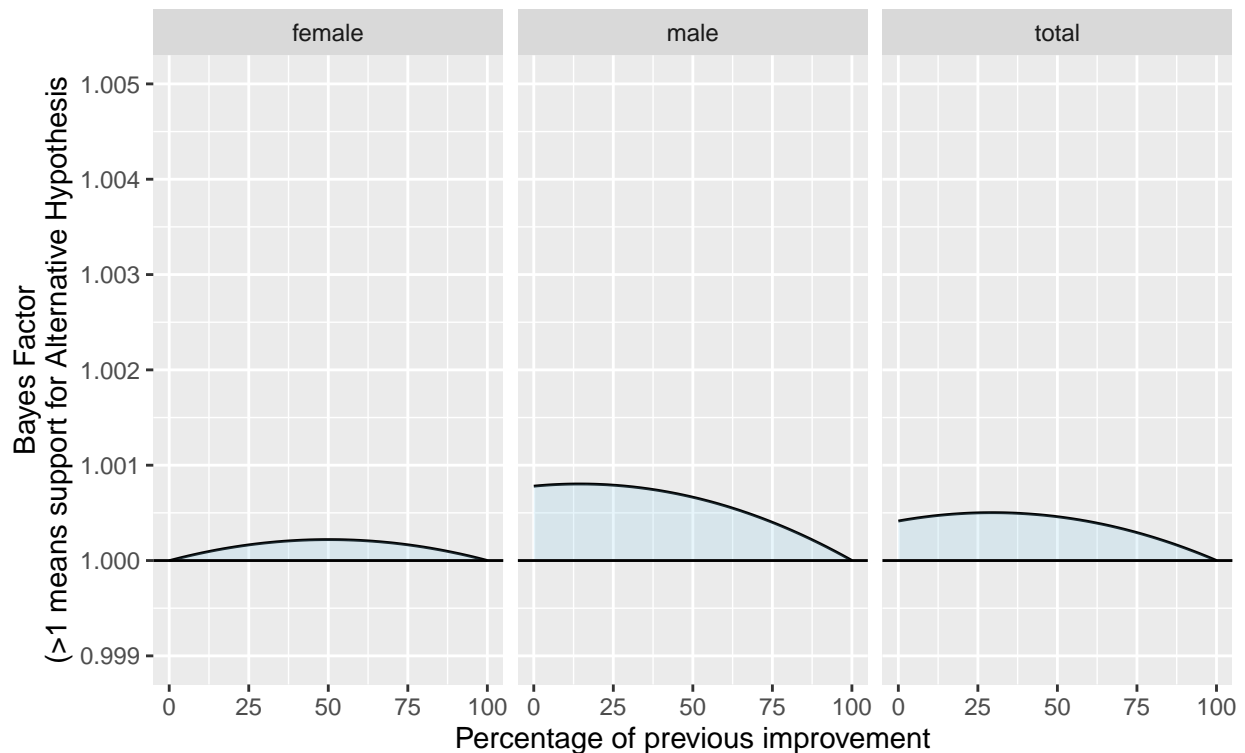
```
e0_uk_bf %>%
  filter(period == "2012-13") %>%
```

```

mutate(perc = 100 * perc) %>%
ggplot(aes(x = perc, y = bayes_factor)) +
geom_line() +
geom_ribbon(aes(ymin = 1, x = perc, ymax = bayes_factor),
            data = e0_uk_bf %>%
              filter(period == "2012-13") %>%
              mutate(perc = 100 * perc),
            inherit.aes = FALSE, fill = "skyblue", alpha = 0.2) +
facet_wrap(~sex) +
scale_y_continuous(limits = c(0.999, 1.005),
                   breaks = seq(0.999, 1.0050, by = 0.001)
                   ) +
geom_hline(yintercept = 1) +
labs(
  x = "Percentage of previous improvement",
  y = "Bayes Factor\n(>1 means support for Alternative Hypothesis",
  title = "Bayes Factor for various proposed levels of slowdown",
  subtitle = "Based on 2012-13 data alone"
)

```

Bayes Factor for various proposed levels of slowdown
Based on 2012-13 data alone



The following figure updates this to include the next year

```

e0_uk_bf %>%
  filter(period %in% c("2012-13", "2012-14")) %>%
  mutate(perc = 100 * perc) %>%

```

```

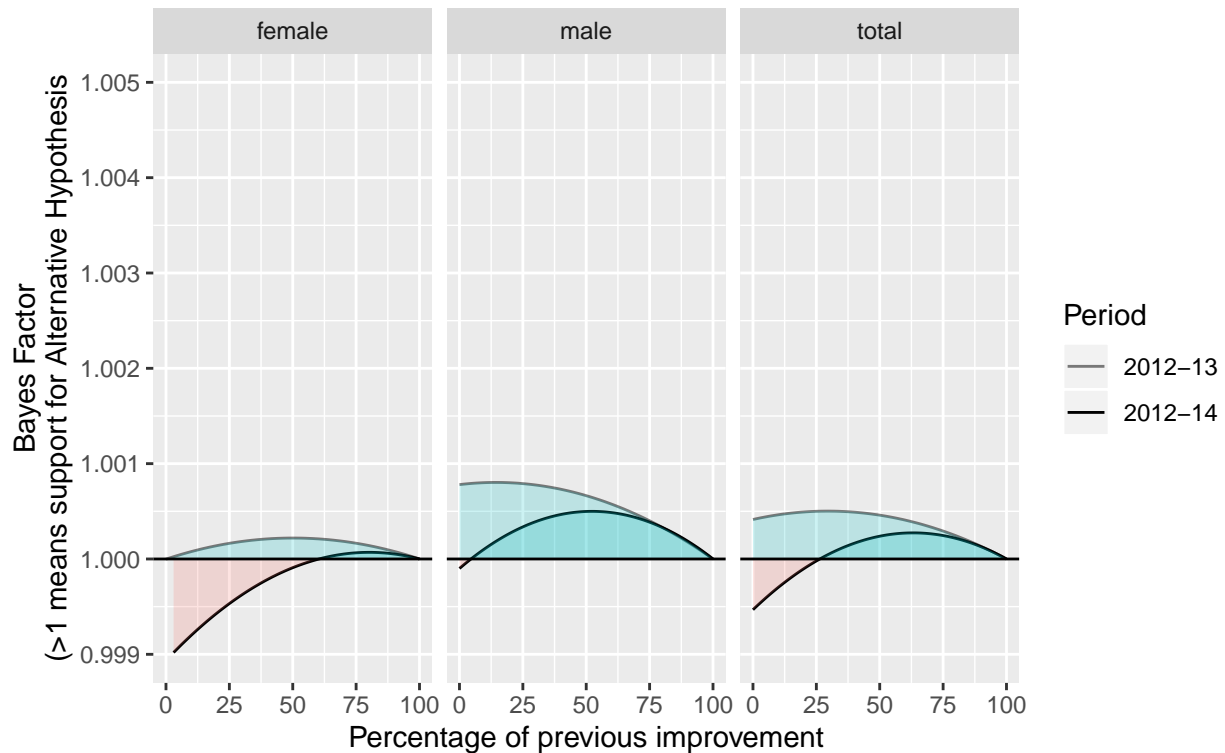
ggplot(aes(x = perc, y = bayes_factor, alpha = period)) +
  geom_line() +
  geom_ribbon(aes(
    ymin = ifelse(is_pos, 1, bayes_factor),
    ymax = ifelse(is_pos, bayes_factor, 1),
    x = perc, group = paste0(is_pos, period),
    fill = is_pos
  ),
    data = e0_uk_bf %>%
      filter(period %in% c("2012-13", "2012-14")) %>%
      mutate(perc = 100 * perc) %>%
      mutate(is_pos = bayes_factor > 1),
    inherit.aes = FALSE, alpha = 0.2) +
  facet_wrap(~sex) +
  scale_y_continuous(limits = c(0.999, 1.005),
    breaks = seq(0.999, 1.0050, by = 0.001)
  ) +
  scale_alpha_discrete("Period", range = c(0.5, 1), breaks = c("2012-13", "2012-14")) +
  geom_hline(yintercept = 1) +
  labs(
    x = "Percentage of previous improvement",
    y = "Bayes Factor\n(>1 means support for Alternative Hypothesis",
    title = "Bayes Factor for various proposed levels of slowdown",
    subtitle = "Based on periods 2012-13, and 2012-14"
  ) +
  guides(fill = FALSE)

```

```
## Warning: Using alpha for a discrete variable is not advised.
```

```
## Warning: Removed 3 rows containing missing values (geom_path).
```

Bayes Factor for various proposed levels of slowdown Based on periods 2012–13, and 2012–14



Whereas the Bayes Factors for all proposed percentages of slowdown were positive when the first observation (2012–13) was used (values above 1 at all proposed Percentage values), 2013–14 saw a higher rate of improvement than the previous year, and so the addition of this observation both reduced the maximum Bayes Factor, and found higher rates of slowdown to be less likely than no slowdown (the Null model).

Repeating this updating exercise for each of the subsequent years' observations leads to the following figure.

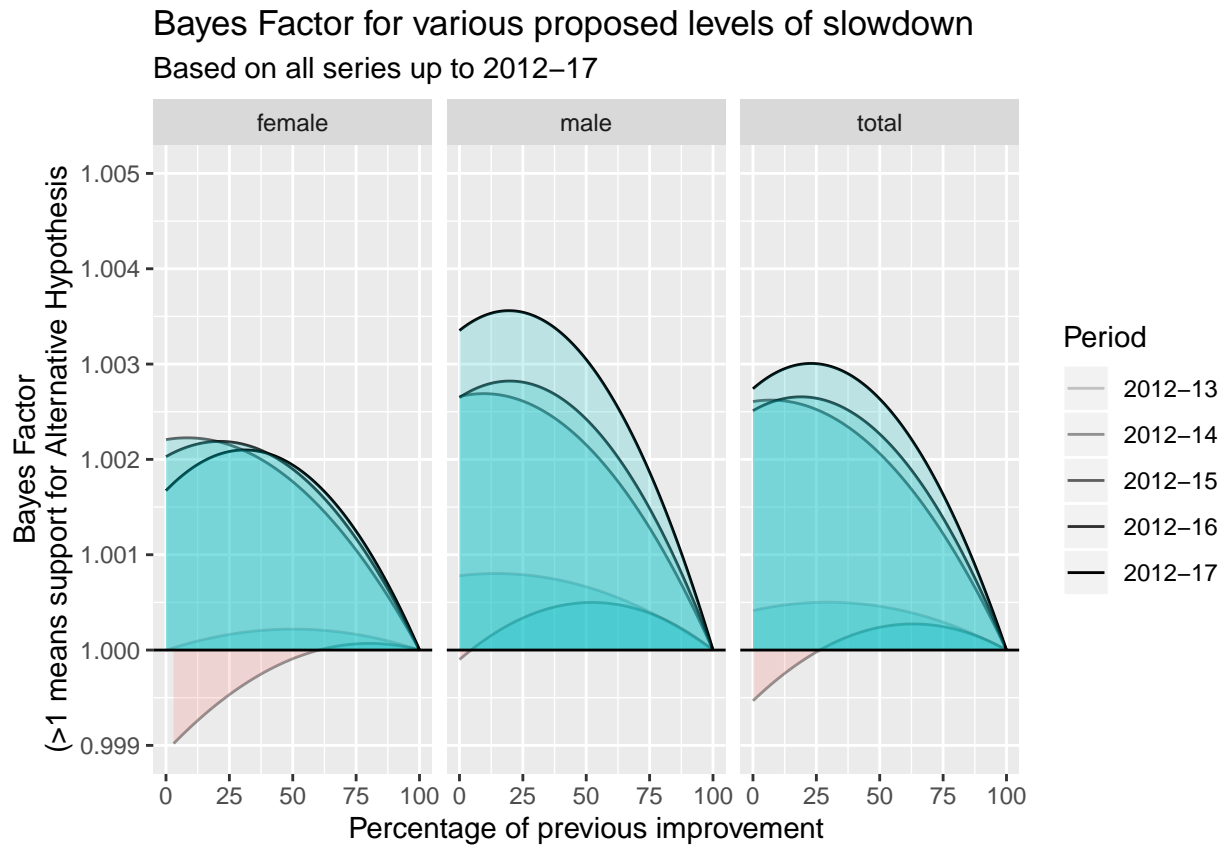
```
e0_uk_bf %>%
  mutate(perc = 100 * perc) %>%
  ggplot(aes(x = perc, y = bayes_factor, alpha = period)) +
  geom_line() +
  geom_ribbon(aes(
    ymin = ifelse(is_pos, 1, bayes_factor),
    ymax = ifelse(is_pos, bayes_factor, 1),
    x = perc, group = paste0(is_pos, period),
    fill = is_pos
  )),
  data = e0_uk_bf %>%
    mutate(perc = 100 * perc) %>%
    mutate(is_pos = bayes_factor > 1),
    inherit.aes = FALSE, alpha = 0.2) +
  facet_wrap(~sex) +
  scale_y_continuous(limits = c(0.999, 1.005),
    breaks = seq(0.999, 1.0050, by = 0.001)
  ) +
  scale_alpha_discrete("Period", range = c(0.2, 1), breaks = c("2012-13", "2012-14", "2012-15", "2012-16"))
```



```
geom_hline(yintercept = 1) +
labs(
  x = "Percentage of previous improvement",
  y = "Bayes Factor\n(>1 means support for Alternative Hypothesis",
  title = "Bayes Factor for various proposed levels of slowdown",
  subtitle = "Based on all series up to 2012-17"
) +
guides(fill = FALSE)
```

```
## Warning: Using alpha for a discrete variable is not advised.
```

```
## Warning: Removed 3 rows containing missing values (geom_path).
```



We can see from this that, with the addition of the annual change from 2014-15, the Bayes Factor in support of some degree of life expectancy improvement slowdown increased sharply, with more support for higher rates of slowdown (around 25% of previous levels) than smaller rates of slowdown (around 75% of previous levels). A simplified figure, using only the full available series, 2012-17, is shown below. The figures have been annotated to show the proposed extent of the slowdown, as a percentage of previous improvement rates, at which the Bayes Factor is maximised for each gender.

```
e0_uk_bf %>%
  mutate(perc = 100 * perc) %>%
  filter(period == "2012-17") %>%
  ggplot(aes(x = perc, y = bayes_factor)) +
  geom_line() +
  geom_ribbon(aes(
    ymin = ifelse(is_pos, 1, bayes_factor),
```

```

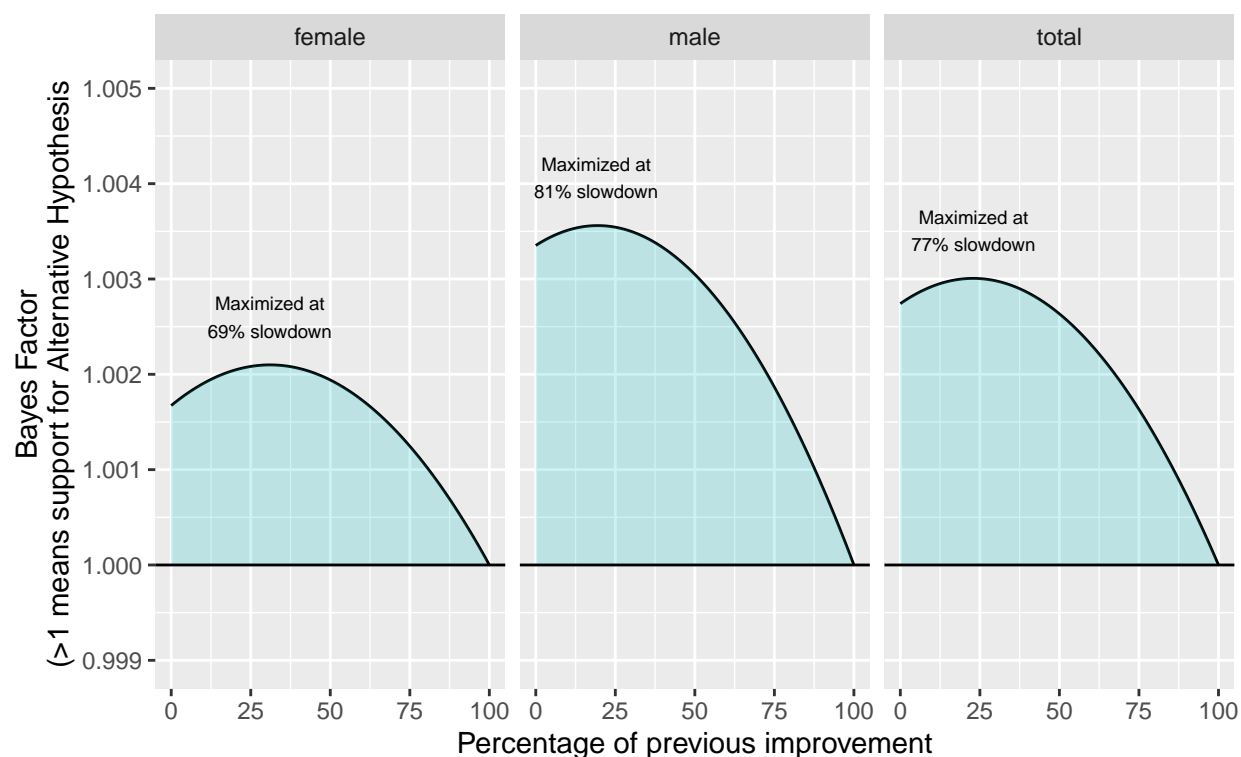
    ymax = ifelse(is_pos, bayes_factor, 1),
    x = perc, group = is_pos,
    fill = is_pos
  ),
  data = e0_uk_bf %>%
    mutate(perc = 100 * perc) %>%
    mutate(is_pos = bayes_factor > 1) %>%
    filter(period == "2012-17") ,
    inherit.aes = FALSE, alpha = 0.2) +
  facet_wrap(~sex) +
  scale_y_continuous(limits = c(0.999, 1.005),
    breaks = seq(0.999, 1.0050, by = 0.001)

  ) +
  geom_hline(yintercept = 1) +
  labs(
    x = "Percentage of previous improvement",
    y = "Bayes Factor\n(>1 means support for Alternative Hypothesis",
    title = "Bayes Factor for various proposed levels of slowdown",
    subtitle = "Based on complete series up to 2012-17"
  ) +
  guides(fill = FALSE) +
  geom_text(
    data = e0_uk_bf %>%
      mutate(perc = 100 * perc) %>%
      filter(after_end == 2017) %>%
      group_by(sex) %>%
      filter(bayes_factor == max(bayes_factor)) %>%
      mutate(text = paste0("Maximized at\n", 100 - round(perc, 0), "% slowdown")),
    mapping = aes(x = perc, y = bayes_factor + 0.0005, label = text),
    inherit.aes = FALSE, size = 2.5
  )

```

Bayes Factor for various proposed levels of slowdown

Based on complete series up to 2012–17



Discussion/Conclusion

This figure indicates that there is a higher likelihood of a severe slowdown, to around a fifth of previous values in males, and around a quarter of previous values in females, than of a more modest slowdown. There is also support for the hypothesis of a complete slowdown in life expectancy (0% of previous improvement rates). Although negative percentage values, i.e. hypotheses proposing a fall in life expectancy, have not been calculated, the observation that the Bayes Factors are above 1 at 0% improvement, and curve smoothly around their peak values, suggests that the recent data are more consistent with falling life expectancies than with life expectancies continuing to improve at the rates observed from 1990 to 2012.

Implications for research

It is straightforward to update the Bayes Factor when future years' life expectancy values become available. We recommend that this approach be incorporated with standard releases of life expectancy covering the UK and its constituent nations, with both the most up-to-date Bayes Factor schedule, and the impact of the new observation, being reported and discussed with each new release. This will provide a simple way of continuing to monitor the extent of the recent life expectancy slowdown and the accumulation of evidence in support of it. As with the addition of 2013-14 to the series, the impact on the Bayes Factor on an additional observation may reduce rather than increase the relative likelihood of a slowdown/fall in life expectancy compared with no slowdown, and by integrating and automating the inclusion of the Bayes Factor into new statistical releases the risk of selective reporting - whether to support the Slowdown or the No Slowdown Hypothesis - is minimised. Recent interim analysis of weekly deaths data for the year 2019 reported relatively few deaths in England & Wales compared with the 2009-2018 average in the first 12 months of the year; in 2018 more deaths occurred than average at the start of the year, but somewhat fewer towards the end. Given most of the UK's population are in England or Wales, when the UK's period life expectancies for

these two years become available, the support for the Slowdown Hypothesis may therefore reduce rather than increase. However, the analysis so far show more evidence in support of either Slowdown or decline in life expectancy than for improvements to increase as they did for at least the previous 22 years.

Implications for practice

Public Health needs to make timely decisions based on the best available evidence, even if that evidence is limited and may not meet the Criminal Court threshold of evidence implicit in Classical hypothesis testing frameworks. The Civil Court threshold of evidence of the Bayes Factor is perhaps closer to what most people, in a personal or professional capacity, do most of the time. We argue that, though the terminology may be unfamiliar, the core ideas that the Bayes Factor embody - of weighing the evidence, deciding to go with whichever option is better supported, and being willing to revise our decisions as more evidence comes to light - is simply common sense, and that not using this approach too often represents deciding to be indecisive, and to fail to act as best as we can now.

Technical appendix

Likelihood and Log Likelihood of the Normal Distribution

For computational reasons it is more common to calculate the log likelihood of a function rather than the likelihood itself. Defining $X = \{x_1, x_2, \dots, x_n\}$ as a series of n observations, the Log Likelihood of the Normal Distribution is as follows:

$$\log L(\mu, \sigma^2 | X = \{x_1, x_2, \dots, x_n\}) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

This is implemented as a function in R as follows:

```
get_ll <- function(x, mu, sig_sq){
  sig <- sqrt(sig_sq)
  n <- length(x)

  - n * log(sig) - (n/2) * log(2 * pi) - (1 / 2 * sig_sq) * sum((x - mu)^2)
}
```

The Bayes Factor is defined as ratio of Likelihoods of two models. In the general case, if $g(\theta)$ refers to a model with parameters θ , and θ_{null} and θ_{alt} to two different candidate parameters, then the Bayes Factor is

$$\frac{L(g(\theta_{alt})|X)}{L(g(\theta_{null})|X)}$$