

Technical comment on Evidence for a limit to human lifespan

Philipp Berens and Tom Wallis

October 11, 2016

Dong et al. claim to present statistical evidence in favor of an absolute limit to the human lifespan. Here we present a reanalysis of a central figure in their paper showing that in fact the data is uninformative with regards to the question whether there is a limit to human lifespan or not.

The model by the authors

The authors graph the maximum age reported at death (MRAD) for each year between 1968 and 2006. We acquired the data using WebPlotDigitizer and rounded the numbers to full years (which is what likely was the case for the original data). Originally the data came from the IDL Database.

Here is the raw data, as presented by the authors, fitting separate regression for years up to 1994 and after 1995.

The plot shows the raw data points in black and separate linear fits with 95%-CIs for years before and after 1995. It is not clear from the paper why the authors chose 1995 as a point to separate models.

We can also obtain the statistics for this model by fitting a linear model with the additional group-variable as predictor including interactions, allowing for a changes slope and offset for the data after 1995.

```
mdl1 <- lm(Age ~ Year * Group, tbl)
summary.lm(mdl1)

##
## Call:
## lm(formula = Age ~ Year * Group, data = tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4359 -1.0757  0.0871  0.7616  6.1166
##
## Coefficients:
##              Estimate Std. Error
## (Intercept)   -191.0734    116.0246
## Year              0.1531      0.0585
## Group>=1995    858.5757    347.1723
## Year:Group>=1995 -0.4293      0.1737
```

```
##               t value Pr(>|t|)
## (Intercept)   -1.647   0.1104
## Year          2.617   0.0139 *
## Group>=1995   2.473   0.0195 *
## Year:Group>=1995 -2.472   0.0196 *
## ---
## Signif. codes:
##  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.956 on 29 degrees of freedom
## Multiple R-squared:  0.4163, Adjusted R-squared:  0.3559
## F-statistic: 6.893 on 3 and 29 DF,  p-value: 0.001209

# additionally show that you get the same
# model fit from this combined regression:
tbl2 <- tbl
tbl2$yhat <- predict mdl1

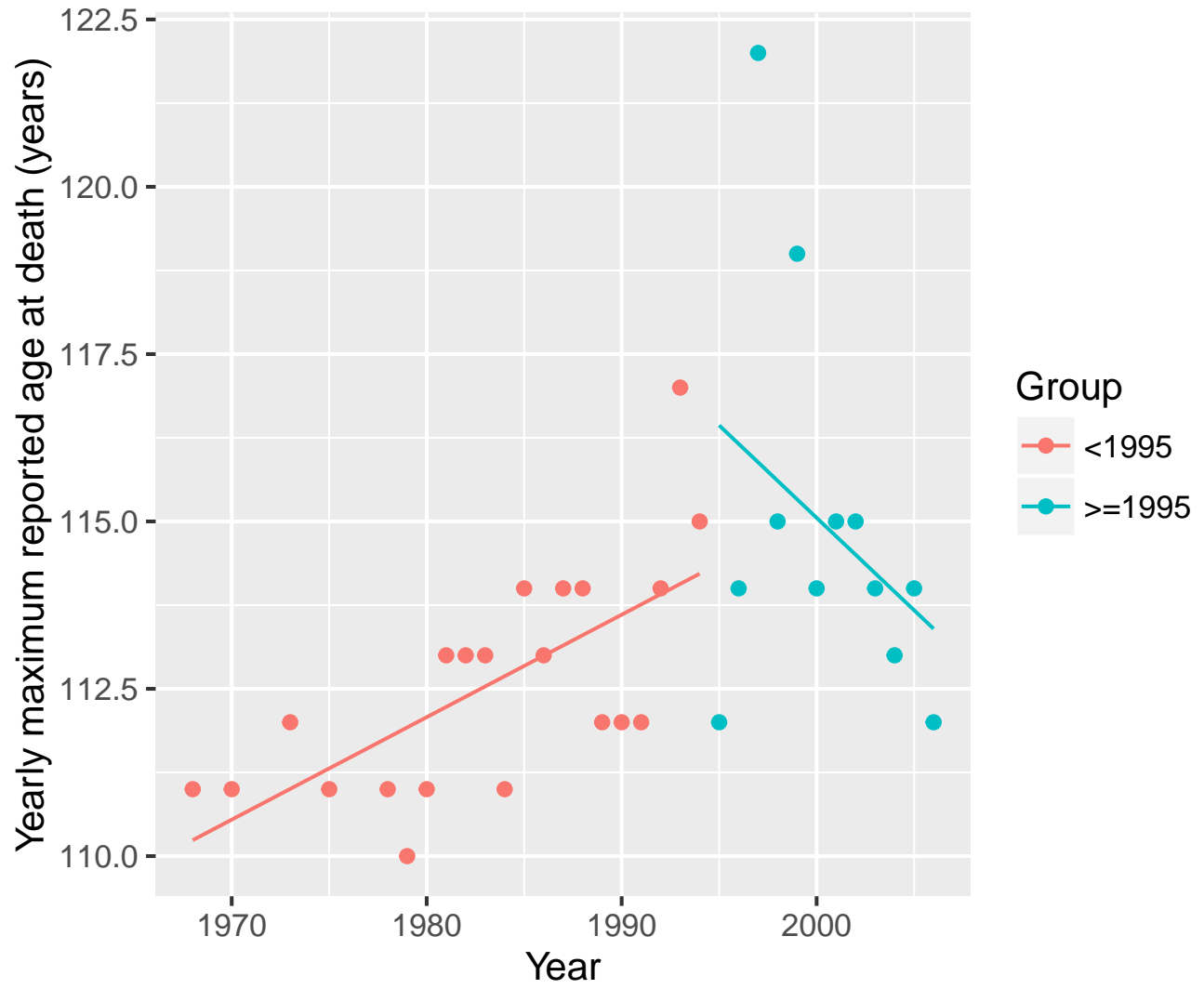
plt <- ggplot(tbl2, aes(x = Year, y = Age, colour = Group)) +
  geom_point() + geom_line(aes(y = yhat)) +
  ylab("Yearly maximum reported age at death (years)")

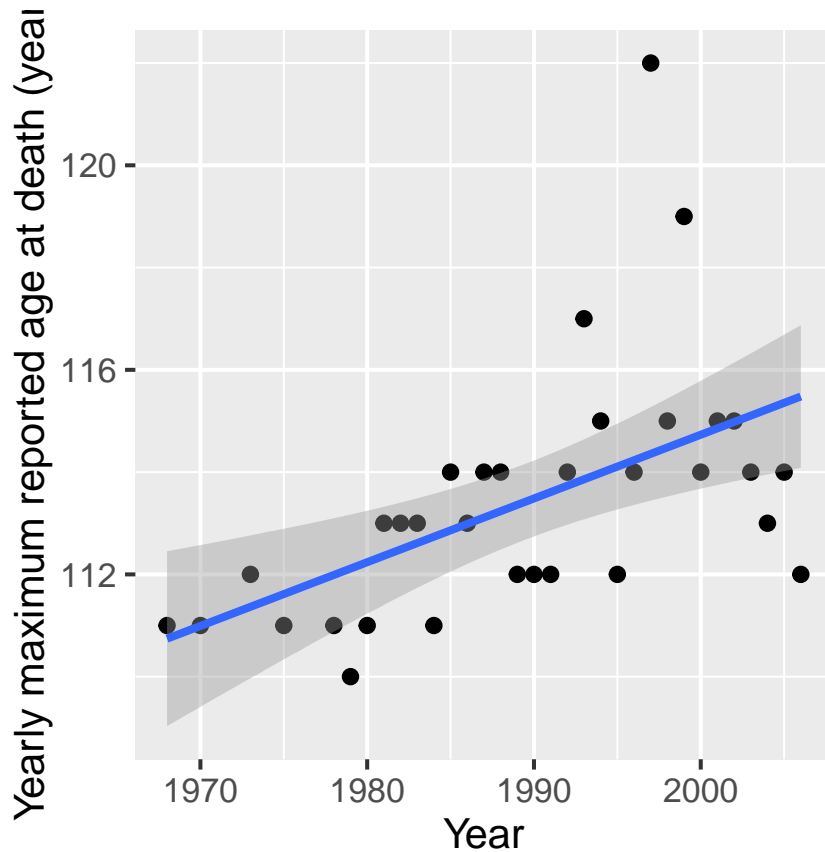
plt
```

Consistent with the paper, the fitted model has a slope of 0.153 years for years before 1995 and one of -0.276 for years afterwards (compare their Figure 2a). We also plot the regression line from this combined model to show that it is the same as the authors' separate regression fits.

A linear model

A simple alternative hypothesis to the claim of the authors would be that MRAD actually keeps increasing and therefore, that there is no limit to human lifespan. To model this idea, we fit a simple linear model to the data:





The plots shows the raw data points again, with a linear regression with 95% CIs fitted to all the data.

```
mdl2 <- lm(Age ~ Year, tbl)
summary.lm(mdl2)

##
## Call:
## lm(formula = Age ~ Year, data = tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4791 -1.3544  0.0139  0.6343  7.6428
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept) -134.56615   69.84837  -1.927
## Year          0.12465    0.03511   3.550
##              Pr(>|t|)
## (Intercept)  0.06325 .
## Year         0.00125 **
## ---
```

```
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.088 on 31 degrees of freedom
## Multiple R-squared: 0.2891, Adjusted R-squared: 0.2662
## F-statistic: 12.61 on 1 and 31 DF, p-value: 0.001251
```

In this case, MRAD increases slightly by 0.12 years per year.

Showing the models side-by-side

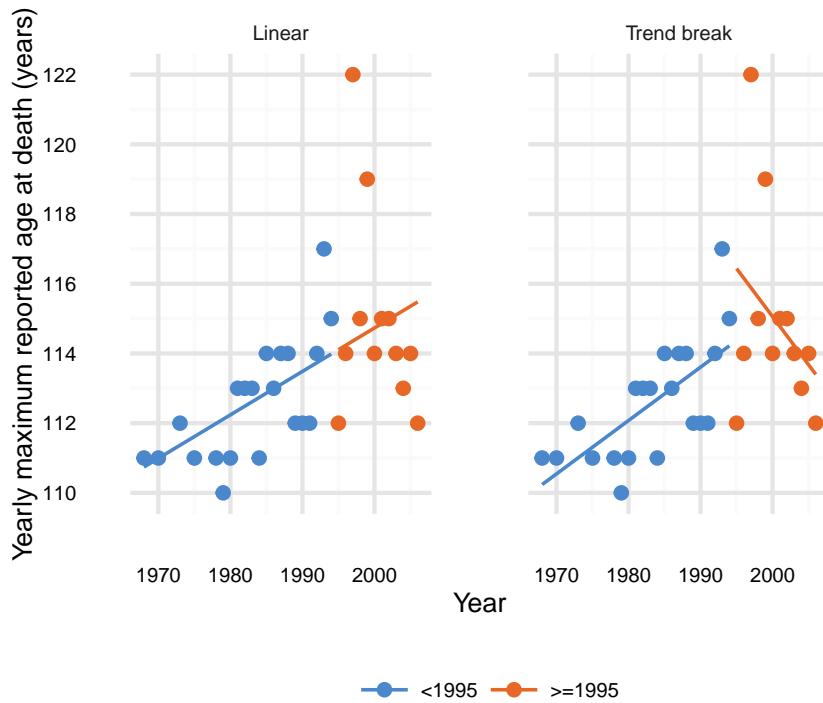
```
# prepare data:
tmp1 <- tbl
tmp2 <- tbl
tmp1$model <- "Trend break"
tmp1$yhat <- predict mdl1
tmp2$model <- "Linear"
tmp2$yhat <- predict mdl2

combined_data <- rbind(tmp1, tmp2)

plt <- ggplot(combined_data, aes(x = Year, y = Age,
  colour = Group)) + facet_grid(~model) + geom_point() +
  geom_line(aes(y = yhat))

# appearance:
plt <- plt + scale_y_continuous(name = "Yearly maximum reported age at death (years)",
  breaks = seq(110, 126, by = 2)) + ylab("Yearly maximum reported age at death (years)") +
  theme_minimal(base_size = 8) + theme(panel.grid.major = element_line(colour = "grey90",
  size = 0.6)) + theme(panel.margin = unit(2,
  "lines")) + scale_color_manual(values = c("#4A87CB",
  "#E76826"), name = "") + theme(legend.position = "bottom")

plt
```



Model comparison

Which model is better? In the paper, the authors fail to provide evidence for their model, they seem to argue that the data looks like there is a saturation effect or a decline in MRAD after 1995.

One can do better and objectively compare the two fitted models. If we look at the output of the models above, the model by the authors explains a little more variance in the data than the linear model (0.42 vs. 0.29). On the other hand, the model also uses four parameters to do so, compared to only two in the linear model.

We can therefore ask if the increase in explained variance is “worth” the additional parameters. A number of model comparison metrics exist; in general these weigh the tradeoff between model fit and complexity differently. We present the results of several classical model comparison metrics below.

Classical ANOVA

First, we consider the two models as a nested set and compare them using classical ANOVA.

```
anova mdl2, mdl1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: Age ~ Year
## Model 2: Age ~ Year * Group
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      31 135.12
## 2      29 110.95  2    24.171 3.159 0.05739 .
## ---
## Signif. codes:
##  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The two extra degrees of freedom in the Dong et al model does not lead to a statistically-significant improvement in residual error over the simple linear model (following the traditional $p < .05$ cutoff for significance).

AIC

Another comparison metric with appealing theoretical links to information theory (see Burnham & Anderson, 2002) is the so-called Akaike Information Criterion (AIC). In the AIC, smaller values denote better models.

```
AIC(md11)
## [1] 143.6636
AIC(md12)
## [1] 146.1678
```

Here, the model of Dong et al has the lower AIC and is thus the preferred model. However, following the heuristics suggested by Burnham & Anderson (2002, p.70), an AIC difference of -2.5 indicates that the data provide “substantial” support for the simpler linear model.

BIC

A related model comparison metric is the Bayesian Information Criterion (BIC), which is more conservative than AIC because it additionally penalises models with more parameters.

```
BIC(md11)
## [1] 151.1461
BIC(md12)
## [1] 150.6574
```

Following Raftery (1995), a BIC difference of 0.49 is not worth mentioning, providing no evidence of one versus the other model.

Bayes Factors

Finally, Bayes Factors (the ratio of the posterior model evidence), can be easily computed to compare simple linear models using the BayesFactor package (Morey & Rouder, 2015).

```
bf1 <- lmBF(Age ~ Year * Group, tbl)
bf2 <- lmBF(Age ~ Year, tbl)

bf1/bf2

## Bayes factor analysis
## -----
## [1] Year * Group : 1.240718 ±0.53%
##
## Against denominator:
##   Age ~ Year
## ---
## Bayes factor type: BFlinearModel, JZS
```

Under these default priors (assuming for example that both models are equally likely *a priori*), the models receive approximately equal support from the data (the Dong et al model is favoured with an odds ratio of 1.24-to-1).

Changing the priors of BayesFactor from “medium” to “ultrawide” on the standardised effect size scale did not appreciably affect these conclusions; with ultrawide effect-size priors the simpler model is now preferred with odds of 1.09-to-1.

Bayesian estimation of model parameters

Here we take a Bayesian approach to model estimation, and fit the full linear model including interaction terms. We employ a Student-t prior with a mean of zero, standard deviation of 2.5 and five degrees of freedom, which yields modest shrinkage of the coefficients towards zero, i.e. enforcing some conservatism in inference.

We fit the models using the package `rstanarm`, which allows relatively straightforward use of Bayesian methods.

```
bmdl <- stan_glm(Age ~ Year * Group, tbl, prior = student_t(5,
  0, 2.5), prior_intercept = student_t(5, 0,
  50), family = gaussian(), adapt_delta = 0.99)

##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 1).
##
## Chain 1, Iteration:   1 / 2000 [ 0%] (Warmup)
```



```

## Chain 1, Iteration: 200 / 2000 [ 10%] (Warmup)
## Chain 1, Iteration: 400 / 2000 [ 20%] (Warmup)
## Chain 1, Iteration: 600 / 2000 [ 30%] (Warmup)
## Chain 1, Iteration: 800 / 2000 [ 40%] (Warmup)
## Chain 1, Iteration: 1000 / 2000 [ 50%] (Warmup)
## Chain 1, Iteration: 1001 / 2000 [ 50%] (Sampling)
## Chain 1, Iteration: 1200 / 2000 [ 60%] (Sampling)
## Chain 1, Iteration: 1400 / 2000 [ 70%] (Sampling)
## Chain 1, Iteration: 1600 / 2000 [ 80%] (Sampling)
## Chain 1, Iteration: 1800 / 2000 [ 90%] (Sampling)
## Chain 1, Iteration: 2000 / 2000 [100%] (Sampling)
## Elapsed Time: 1.23667 seconds (Warm-up)
##                2.31164 seconds (Sampling)
##                3.54831 seconds (Total)
##
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 2).
##
## Chain 2, Iteration: 1 / 2000 [ 0%] (Warmup)
## Chain 2, Iteration: 200 / 2000 [ 10%] (Warmup)
## Chain 2, Iteration: 400 / 2000 [ 20%] (Warmup)
## Chain 2, Iteration: 600 / 2000 [ 30%] (Warmup)
## Chain 2, Iteration: 800 / 2000 [ 40%] (Warmup)
## Chain 2, Iteration: 1000 / 2000 [ 50%] (Warmup)
## Chain 2, Iteration: 1001 / 2000 [ 50%] (Sampling)
## Chain 2, Iteration: 1200 / 2000 [ 60%] (Sampling)
## Chain 2, Iteration: 1400 / 2000 [ 70%] (Sampling)
## Chain 2, Iteration: 1600 / 2000 [ 80%] (Sampling)
## Chain 2, Iteration: 1800 / 2000 [ 90%] (Sampling)
## Chain 2, Iteration: 2000 / 2000 [100%] (Sampling)
## Elapsed Time: 1.01169 seconds (Warm-up)
##                1.01493 seconds (Sampling)
##                2.02662 seconds (Total)
##
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 3).
##
## Chain 3, Iteration: 1 / 2000 [ 0%] (Warmup)
## Chain 3, Iteration: 200 / 2000 [ 10%] (Warmup)
## Chain 3, Iteration: 400 / 2000 [ 20%] (Warmup)
## Chain 3, Iteration: 600 / 2000 [ 30%] (Warmup)
## Chain 3, Iteration: 800 / 2000 [ 40%] (Warmup)
## Chain 3, Iteration: 1000 / 2000 [ 50%] (Warmup)
## Chain 3, Iteration: 1001 / 2000 [ 50%] (Sampling)

```

```

## Chain 3, Iteration: 1200 / 2000 [ 60%] (Sampling)
## Chain 3, Iteration: 1400 / 2000 [ 70%] (Sampling)
## Chain 3, Iteration: 1600 / 2000 [ 80%] (Sampling)
## Chain 3, Iteration: 1800 / 2000 [ 90%] (Sampling)
## Chain 3, Iteration: 2000 / 2000 [100%] (Sampling)
## Elapsed Time: 1.04679 seconds (Warm-up)
##               1.18861 seconds (Sampling)
##               2.2354 seconds (Total)
##
##
## SAMPLING FOR MODEL 'continuous' NOW (CHAIN 4).
##
## Chain 4, Iteration:    1 / 2000 [  0%] (Warmup)
## Chain 4, Iteration:  200 / 2000 [ 10%] (Warmup)
## Chain 4, Iteration:  400 / 2000 [ 20%] (Warmup)
## Chain 4, Iteration:  600 / 2000 [ 30%] (Warmup)
## Chain 4, Iteration:  800 / 2000 [ 40%] (Warmup)
## Chain 4, Iteration: 1000 / 2000 [ 50%] (Warmup)
## Chain 4, Iteration: 1001 / 2000 [ 50%] (Sampling)
## Chain 4, Iteration: 1200 / 2000 [ 60%] (Sampling)
## Chain 4, Iteration: 1400 / 2000 [ 70%] (Sampling)
## Chain 4, Iteration: 1600 / 2000 [ 80%] (Sampling)
## Chain 4, Iteration: 1800 / 2000 [ 90%] (Sampling)
## Chain 4, Iteration: 2000 / 2000 [100%] (Sampling)
## Elapsed Time: 0.937473 seconds (Warm-up)
##               0.944194 seconds (Sampling)
##               1.88167 seconds (Total)

```

We can summarize the fitted model and plot the posterior density over the parameters:

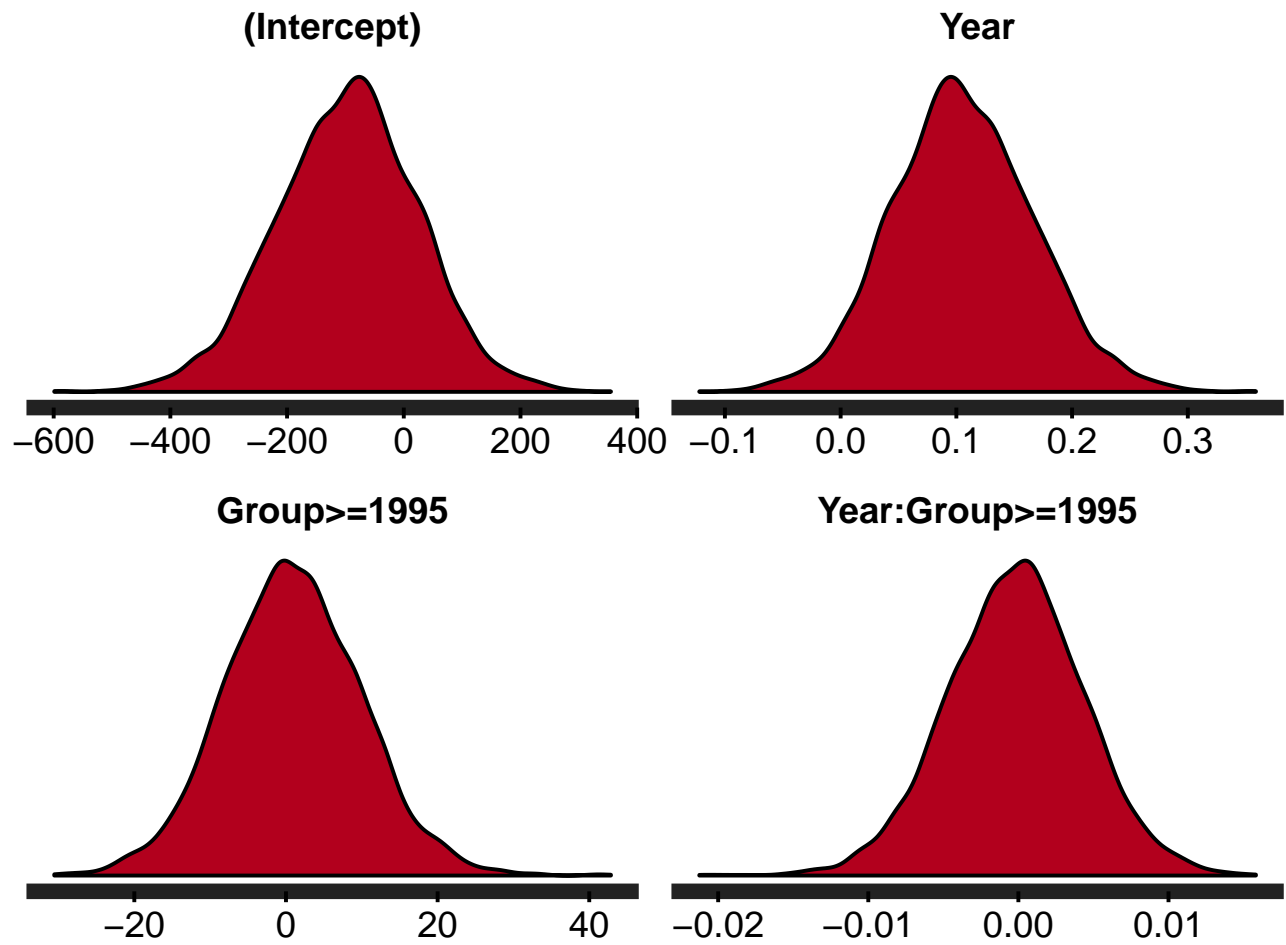
```
plot(bmdl, "dens")
```

```
bmdl
```

```

## stan_glm(formula = Age ~ Year * Group, family = gaussian(), data = tbl,
##          prior = student_t(5, 0, 2.5), prior_intercept = student_t(5,
##          0, 50), adapt_delta = 0.99)
##
## Estimates:
##           Median MAD_SD
## (Intercept)   -92.4  119.2
## Year           0.1    0.1
## Group>=1995    0.7    8.9
## Year:Group>=1995 0.0    0.0

```



```
## sigma          2.2    0.3
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##           Median MAD_SD
## mean_PPD 113.4    0.5
##
## Observations: 33  Number of unconstrained parameters: 5
```

Comparing the fitted model to the frequentist models above shows that the posterior median of the linear effect of Year (0.103) is similar to the estimated value above (0.153), but shrunk towards zero by the prior. The posterior density on the interaction term is centered around zero during inference, arguing that there is little evidence of a different slope after 1995. There is a small effect of the interaction term on the y-intercept, increasing the estimated y-intercept by 900. This is likely an artefact of the model parametrization.

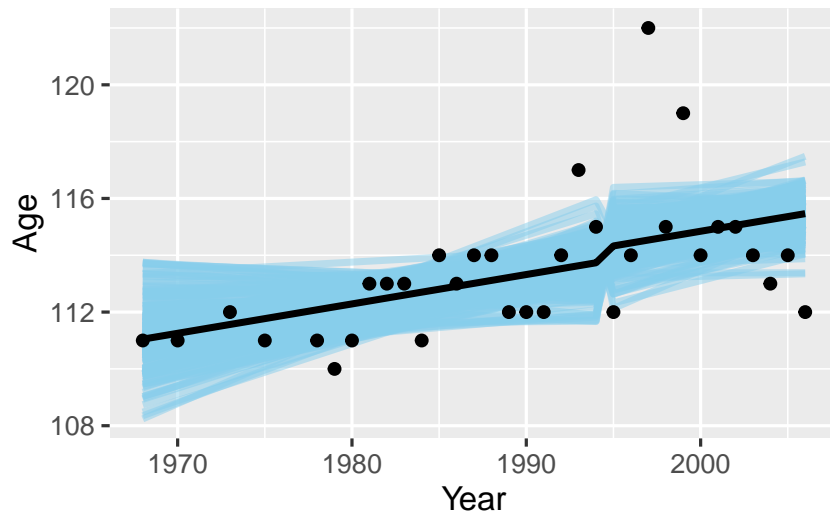
```
draws <- as.data.frame(as.matrix(bmdl))
X <- draws[1:200, 1:4]
foomdl <- mdl1
tbl2 <- tbl["Year"]

base <- ggplot(tbl, aes(x = Year, y = Age))
for (i in 1:200) {
  foomdl$coefficients <- c(X[i, 1], X[i, 2],
    X[i, 3], X[i, 4])
  tbl2["Pred"] <- predict.lm(foomdl, tbl)

  base <- base + geom_line(data = tbl2, mapping = aes(x = Year,
    y = Pred), color = "skyblue", alpha = 0.5,
    size = 1.1)
}

X = coef(bmdl)
foomdl$coefficients <- c(X[1], X[2], X[3], X[4])
tbl2["Pred"] <- predict.lm(foomdl, tbl)

base + geom_point() + geom_line(data = tbl2, mapping = aes(x = Year,
  y = Pred), size = 1.1)
```



We check certain properties of the Bayesian fitting procedures graphically:

```
stan_diag(bmdl)
```

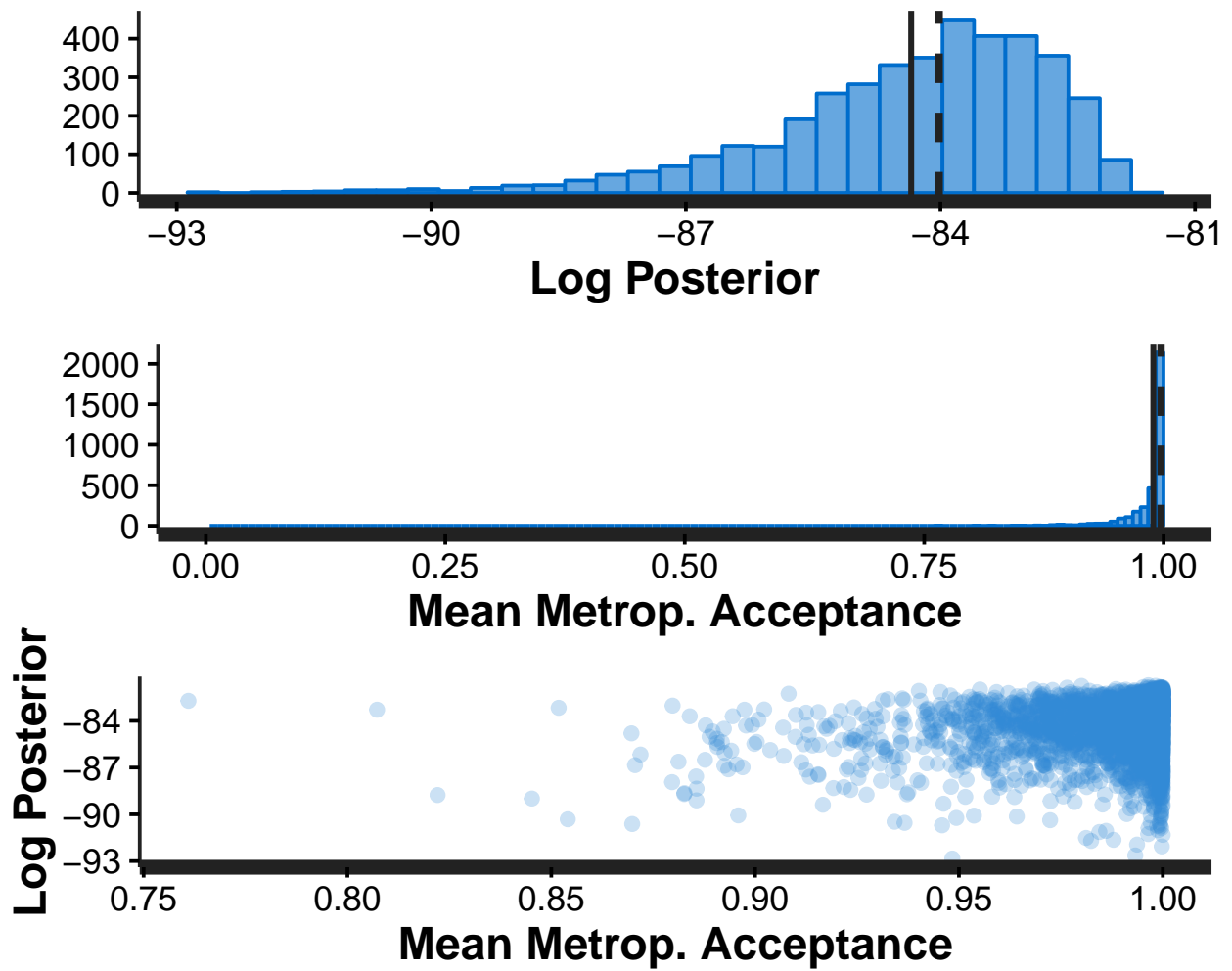
```
## Warning: Removed 1 rows containing missing values
## (geom_bar).
```

Extended data figure

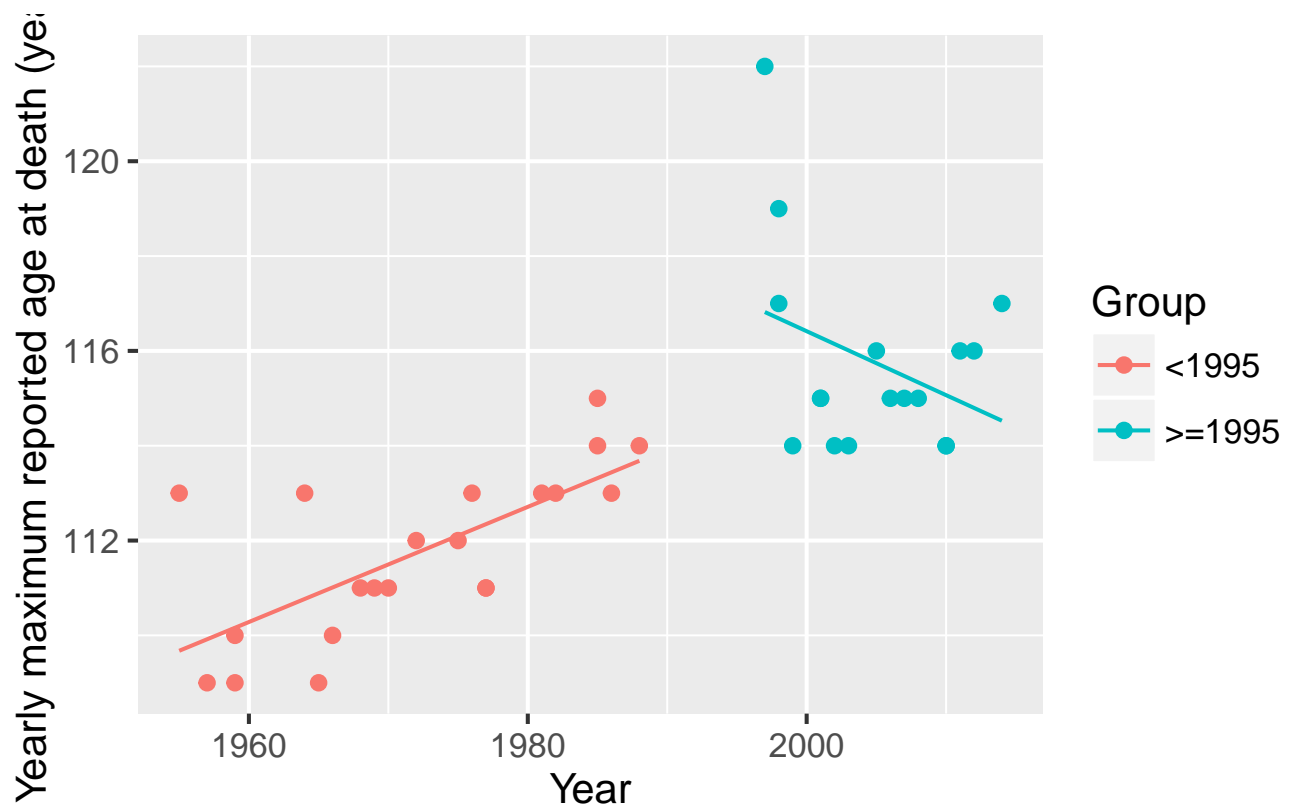
The authors present a similar dataset from an independent source, the Gerontological Research Group, in Extended Data Figure 6. They find similar results to those reported for the main analysis, and they argue that this provides independent evidence for their central analysis.

We again acquired these data using WebPlotDigitizer, and present them below.

```
##
## Call:
## lm(formula = Age ~ Year * Group, data = ext_tbl)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5499 -1.0667 -0.2948  0.5918  5.1804
##
## Coefficients:
##              Estimate Std. Error
## (Intercept)  -127.67958   71.86639
## Year           0.12141    0.03644
## Group>=1995   513.76598  169.13487
```



```
## Year:Group>=1995    -0.25624    0.08462
##                      t value Pr(>|t|)
## (Intercept)         -1.777  0.08458 .
## Year                 3.332  0.00209 **
## Group>=1995          3.038  0.00456 **
## Year:Group>=1995    -3.028  0.00467 **
## ---
## Signif. codes:
##  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.66 on 34 degrees of freedom
## Multiple R-squared:  0.6694, Adjusted R-squared:  0.6402
## F-statistic: 22.95 on 3 and 34 DF, p-value: 2.654e-08
```



The fitted model has a slope of 0.121 years for years before 1995 (their slope = 0.1194) and one of -0.135 (their slope = -0.1367) for years afterwards (compare their Extended Data Figure 6). Differences to their model fit likely reflect data uncertainty from the digitisation process (owing to the poorer resolution of this figure in the paper).

```
## Analysis of Variance Table
##
```

```

## Model 1: Age ~ Year
## Model 2: Age ~ Year * Group
##   Res.Df    RSS Df Sum of Sq    F
## 1      36 123.789
## 2      34  93.689  2      30.1 5.4617
##   Pr(>F)
## 1
## 2 0.008772 **
## ---
## Signif. codes:
##  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Bayes factor analysis
## -----
## [1] Year * Group : 8.685352 ±0.93%
##
## Against denominator:
##   Age ~ Year
## ---
## Bayes factor type: BFlinearModel, JZS

```

Here, the ANOVA analysis does provide support for the Dong et al model. What about the other model comparison metrics? The AIC difference between the models is -6.6, which corresponds to “considerably less” support for the simple linear model relative to the authors’ model. The BIC difference is -3.3, which provides “positive evidence” (Raftery, 1995) in support of the authors’ model. Finally, the Bayes factor computed as above is about 8.8, which also provides positive support for the authors’ model.

Conclusion

These model comparison metrics, using both Frequentist, information theoretic and Bayesian approaches, yield the same conclusion: the central data reported in the paper do not support the argument by Dong et al that there is a limit to human lifespan. A simple linear model showing a positive relationship between year and lifespan is just as plausible given the data from Figure 2a.

The MRAD data from an independent source provided in Extended Data Figure 6 *does* provide some evidence in favour of the Dong et al “trend break” model, with one important caveat: data for the years 1989–1996 are missing from the authors’ plot. Since this data spans the key years of the “break”, they could have an important influence on these conclusions. The Dong et al paper does not clarify why this data is missing from the authors’ plot (compare to

data table online here, assuming that this is the source of the data the authors use).

References

- Burnham, K. P., & Anderson, D. R. (2002): Model selection and multimodel inference a practical information-theoretic approach. New York: Springer.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.
- Morey and Rouder (2015). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-2. <https://CRAN.R-project.org/package=BayesFactor>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.
- Vehtari, Gelman and Gabry (2016): Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, [arxiv link](#)