Income and Health Disparities in Aging

Student ID: 11484265

2024-11-07

Contents

Introduction

The relationship between income trajectories and self-reported health has been well-documented, with positive associations observed in prior research (Akanni et al., 2022). Understanding this link over multiple time points is crucial for examining the cyclical or cumulative effects of socioeconomic factors on health as individuals age. Income not only influences access to essential resources such as healthcare, nutritious food, and stable housing, but also affects mental well-being and stress levels—all of which are vital for healthy aging. By tracking this relationship longitudinally, we can discern whether changes in income, such as those resulting from retirement or financial strain, directly affect health outcomes, or if deteriorating health itself limits income potential (e.g., through medical expenses or reduced work capacity).

To examine the relationship between income and self-reported health outcomes in the aging population, cross-lagged models will be applied to four waves of data from the English Longitudinal Study of Ageing (Banks et al., 2024). While cross-sectional data offer only a snapshot, longitudinal data enables the modeling of temporal order between changes in income and health. This approach allows for the exploration of causal inferences, though it is important to note that cross-lagged models do not establish causation definitively. By analyzing data across multiple waves, the natural dynamics of aging can be better captured, providing a more accurate and comprehensive understanding of these processes over time.

Data and Methods

The English Longitudinal Study of Ageing (ELSA) is a comprehensive dataset that tracks individuals aged 50 and older in England, gathering information across a wide range of domains, including socioeconomic status, health, and psychological well-being. This analysis focuses on data from waves 7 to 10, providing a four-wave longitudinal perspective on the relationship between income and health over time.

Variables

- Income: This variable represents the equivalised total income at the benefit unit level, adjusted for unit size. Total income includes employment income, self-employment income, state benefits, state pensions, private pensions, asset income, and other sources. Each member of a benefit unit is assigned the unit's total income, providing a measure of financial resources adjusted for household composition. Missing income data are excluded from the analysis. The income variable is standardized prior to modelling to allow for consistent comparison.
- Self-Reported Health: Self-reported health status is used as an indicator of overall well-being, with previous research showing it as a reliable predictor of mortality and morbidity. The Health and Retirement Study (HRS) version is utilized due to incomplete data from the Health Survey for England (HSE) version. The original responses range from 1 (excellent health) to 5 (poor health), which are reversed and standardized for clarity before modelling. Records with missing health data are excluded from the analysis.
- IDs: Individual and household IDs are used to accurately link data points across waves, ensuring proper tracking of participants over time for longitudinal analysis.

Descriptive statistics

Income is somewhat right-skewed, with a median of £30,000 per annum, and is characterized by an extremely high maximum value. This indicates that while most individuals earn around the median, a small number of individuals have significantly higher incomes, pulling the distribution to the right.

Combined with the summary statistics, the frequency table for self-reported health shows a slight right skewness, suggesting that more individuals perceive themselves as healthier than neutral.

Table 1: Statistics on annual income (hundreds of pounds) and self-reported health.

	Income	Health_HRS
Min.	0.0000	1.000000
1st Qu.	250.0517	2.000000
Median	366.9548	3.000000
Mean	446.2101	3.201678
3rd Qu.	536.5778	4.000000
Max.	19028.3809	5.000000

Table 2: Frequency of response on self-reported health.

Response	Freq
1	2206
2	5485
3	9936
4	8858
5	3548

Model Sequence

A series of cross-lagged panel models will be used to explore the relationships between income and health, progressively adding constraints and random intercepts to improve model fit.

1. Unconstrained Cross-Lagged Model: This initial model estimates the relationships between income and health without imposing any constraints across the waves. It serves as a baseline for understanding how these variables influence each other over time, with all parameters freely estimated.

2. Equality-Constrained Models: Constraints will be introduced gradually, one wave at a time. After each addition, a Chi-Square difference test (ANOVA) will be performed to compare the new model with the previous best-fitting model, helping to determine whether the added constraint improves the model fit. This step identifies whether certain effects are stable across time.

3. Random Intercept Cross-Lagged Model: Random intercepts will be incorporated for both income and health to account for individual differences that may consistently affect these variables across waves (e.g., baseline levels of income or health that vary between individuals). ANOVA will again be used to assess whether the inclusion of random intercepts significantly enhances the model fit.

Analysis

Model 1 (Base)

Model 1 is the base model, incorporating both auto-regressive and cross-lagged paths for income and health without any constraints. This model serves as a starting point to examine the dynamic relationships between income and health over time, with all parameters freely estimated, providing an initial understanding of the interactions between these variables.

```
##
##   # Cross-lagged paths
##   Income_W8 ~ 1 + Income_W7 + Health_HRS_W7
##   Income_W9 ~ 1 + Income_W8 + Health_HRS_W8
##   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
##
##   Health_HRS_W8 ~ 1 + Health_HRS_W7 + Income_W7
##   Health_HRS_W9 ~ 1 + Health_HRS_W8 + Income_W8
##   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
##
##   # correlation
##   Health_HRS_W7 ~~ Income_W7
##   Health_HRS_W8 ~~ Income_W8
##   Health_HRS_W9 ~~ Income_W9
##   Health_HRS_W10 ~~ Income_W10
```

```
## lavaan 0.6-19 ended normally after 14 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        32
##
##                                                   Used       Total
##   Number of observations                          3781       11700
##
## Model Test User Model:
##
##   Test statistic                              1423.225
##   Degrees of freedom                                12
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W8 ~
##     Income_W7         0.567    0.013   44.350    0.000
##     Health_HRS_W7     0.104    0.014    7.339    0.000
##   Income_W9 ~
##     Income_W8         0.530    0.016   33.281    0.000
```

```
##     Health_HRS_W8     0.042    0.017    2.455    0.014
##   Income_W10 ~
##     Income_W9         0.283    0.015   18.455    0.000
##     Health_HRS_W9     0.055    0.018    2.992    0.003
##   Health_HRS_W8 ~
##     Health_HRS_W7     0.668    0.012   54.174    0.000
##     Income_W7         0.029    0.011    2.654    0.008
##   Health_HRS_W9 ~
##     Health_HRS_W8     0.651    0.012   52.448    0.000
##     Income_W8         0.060    0.012    5.190    0.000
##   Health_HRS_W10 ~
##     Health_HRS_W9     0.650    0.013   49.844    0.000
##     Income_W9         0.028    0.011    2.543    0.011
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W7 ~~
##     Health_HRS_W7     0.174    0.016   10.883    0.000
##  .Income_W8 ~~
##    .Health_HRS_W8     0.025    0.009    2.771    0.006
##  .Income_W9 ~~
##    .Health_HRS_W9    -0.003    0.011   -0.226    0.821
##  .Income_W10 ~~
##    .Health_HRS_W10    0.042    0.013    3.312    0.001
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Income_W8        -0.007    0.013   -0.500    0.617
##    .Income_W9        -0.014    0.016   -0.837    0.402
##    .Income_W10       -0.033    0.017   -1.941    0.052
##    .Health_HRS_W8     0.060    0.012    5.154    0.000
##    .Health_HRS_W9    -0.036    0.012   -3.083    0.002
##    .Health_HRS_W10   -0.117    0.012   -9.487    0.000
##     Income_W7         0.136    0.017    8.039    0.000
##     Health_HRS_W7     0.204    0.015   13.439    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Income_W8         0.646    0.015   43.480    0.000
##    .Income_W9         0.951    0.022   43.480    0.000
##    .Income_W10        1.100    0.025   43.480    0.000
##    .Health_HRS_W8     0.485    0.011   43.480    0.000
##    .Health_HRS_W9     0.496    0.011   43.480    0.000
##    .Health_HRS_W10    0.564    0.013   43.480    0.000
##     Income_W7         1.080    0.025   43.480    0.000
##     Health_HRS_W7     0.872    0.020   43.480    0.000
```

The results indicate a significant positive effect of previous income on subsequent income across waves, although the effect size diminishes over time. Health has a smaller, yet still positive, effect on income in all waves. Health demonstrates strong stability over time, with previous health significantly predicting future health, while income has a much smaller positive effect on health in each wave.

The covariances between income and health are generally positive but small, and only significant in Waves 7, 8, and 10. This suggests that while there is a positive relationship between income and health, it is not consistently strong across all waves.

Model 2-4

Equality constraints are applied to the cross-lagged paths between income and health, one wave at a time, to test if the relationships between these variables remain consistent across the waves. This allows for a more refined understanding of whether the effects of income on health, and vice versa, vary or stay stable over time.

```
##
##   # Cross-lagged paths
##   Income_W8 ~ 1 + Income_W7 + a*Health_HRS_W7
##   Income_W9 ~ 1 + Income_W8 + Health_HRS_W8
##   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
##
##   Health_HRS_W8 ~ 1 + Health_HRS_W7 + a*Income_W7
##   Health_HRS_W9 ~ 1 + Health_HRS_W8 + Income_W8
##   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
##
##   # correlation
##   Health_HRS_W7 ~~ Income_W7
##   Health_HRS_W8 ~~ Income_W8
##   Health_HRS_W9 ~~ Income_W9
##   Health_HRS_W10 ~~ Income_W10
```

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff    RMSEA Df diff Pr(>Chisq)
## fit_1 12 76494 76694 1423.2
## fit_2 13 76509 76703 1440.4     17.161 0.065377       1  3.435e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When equality constraints are applied to wave 8, there is a significant improvement in model fit compared to the base model. However, the constrained model has a slightly higher AIC and BIC, indicating a trade-off between simplicity and fit quality. Nevertheless, the fit indices suggest the model still provides a good overall fit despite the added constraint.

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff RMSEA Df diff Pr(>Chisq)
## fit_2 13 76509 76703 1440.4
## fit_3 14 76508 76695 1441.1    0.72466     0       1     0.3946
```

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff    RMSEA Df diff Pr(>Chisq)
## fit_2 13 76509 76703 1440.4
## fit_4 14 76509 76696 1441.9     1.5487 0.012046       1     0.2133
```

The models with additional constraints on the cross-lagged paths for waves 9 and 10 (fit_3 and fit_4) did not yield significant results, indicating that these models may be too restrictive. This suggests that the influence may be stronger in one direction. The coefficients in these models are very small; if pressed, one might conclude that income has a stronger impact on health in wave 9, while health has a stronger impact on income in wave 10.

```
## lavaan 0.6-19 ended normally after 15 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        32
##   Number of equality constraints                     1
##
##                                                   Used       Total
##   Number of observations                          3781       11700
##
## Model Test User Model:
##
##   Test statistic                              1440.385
##   Degrees of freedom                                13
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W8 ~
##     Income_W7         0.576    0.013   45.647    0.000
##     Hlt_HRS_W7 (a)    0.058    0.009    6.628    0.000
##   Income_W9 ~
##     Income_W8         0.530    0.016   33.292    0.000
##     Hlt_HRS_W8        0.042    0.017    2.461    0.014
##   Income_W10 ~
##     Income_W9         0.283    0.015   18.452    0.000
##     Hlt_HRS_W9        0.055    0.018    2.992    0.003
##   Health_HRS_W8 ~
##     Hlt_HRS_W7        0.660    0.012   54.102    0.000
##     Income_W7  (a)    0.058    0.009    6.628    0.000
##   Health_HRS_W9 ~
##     Hlt_HRS_W8        0.651    0.012   52.576    0.000
##     Income_W8         0.060    0.012    5.192    0.000
##   Health_HRS_W10 ~
##     Hlt_HRS_W9        0.650    0.013   49.856    0.000
##     Income_W9         0.028    0.011    2.542    0.011
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W7 ~~
##     Health_HRS_W7     0.174    0.016   10.883    0.000
##  .Income_W8 ~~
##    .Health_HRS_W8     0.026    0.009    2.801    0.005
##  .Income_W9 ~~
```

```
##     .Health_HRS_W9    -0.003    0.011   -0.226    0.821
##   .Income_W10 ~~
##     .Health_HRS_W10    0.042    0.013    3.312    0.001
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Income_W8         0.002    0.013    0.119    0.905
##    .Income_W9        -0.014    0.016   -0.837    0.403
##    .Income_W10       -0.033    0.017   -1.941    0.052
##    .Health_HRS_W8     0.058    0.012    4.955    0.000
##    .Health_HRS_W9    -0.036    0.012   -3.082    0.002
##    .Health_HRS_W10   -0.117    0.012   -9.487    0.000
##     Income_W7         0.136    0.017    8.039    0.000
##     Health_HRS_W7     0.204    0.015   13.439    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Income_W8         0.648    0.015   43.480    0.000
##    .Income_W9         0.951    0.022   43.480    0.000
##    .Income_W10        1.100    0.025   43.480    0.000
##    .Health_HRS_W8     0.486    0.011   43.480    0.000
##    .Health_HRS_W9     0.496    0.011   43.480    0.000
##    .Health_HRS_W10    0.564    0.013   43.480    0.000
##     Income_W7         1.080    0.025   43.480    0.000
##     Health_HRS_W7     0.872    0.020   43.480    0.000
```

Although only wave 7 has the constraint, the results of model 2 are largely consistent with those of model 1. The constrained coefficient for health in wave 7 predicting income in wave 8 decreased from 0.104 to 0.058. Otherwise, income and health continue to significantly predict each other over time, though with very small coefficients, and both cross-lagged effects remain positive. Given the small values, it's unclear whether this represents a meaningful difference in practice.

Model 5-7

Equality constraints were applied to the auto-regressive paths for each wave and compared to the previous model.

```
##
##      # Cross-lagged paths
##      Income_W8 ~ 1 + b*Income_W7 + a*Health_HRS_W7
##      Income_W9 ~ 1 + c*Income_W8 + Health_HRS_W8
##      Income_W10 ~ 1 + d*Income_W9 + Health_HRS_W9
##
##      Health_HRS_W8 ~ 1 + b*Health_HRS_W7 + a*Income_W7
##      Health_HRS_W9 ~ 1 + c*Health_HRS_W8 + Income_W8
##      Health_HRS_W10 ~ 1 + d*Health_HRS_W9 + Income_W9
##
##      # correlation
##      Health_HRS_W7 ~~ Income_W7
##      Health_HRS_W8 ~~ Income_W8
##      Health_HRS_W9 ~~ Income_W9
##      Health_HRS_W10 ~~ Income_W10
```

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_2 13 76509 76703 1440.4
## fit_5 14 76531 76718 1463.8     23.435 0.07703       1  1.292e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_5 14 76531 76718 1463.8
## fit_6 15 76564 76745 1499.2     35.383 0.09536       1  2.709e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_6 15 76564 76745 1499.2
## fit_7 16 76884 77059 1821.5     322.26 0.29149       1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Retaining all constraints results in a significant improvement in the model.

```
## lavaan 0.6-19 ended normally after 17 iterations
##
```

```
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        32
##   Number of equality constraints                     4
##
##                                                   Used       Total
##   Number of observations                          3781       11700
##
## Model Test User Model:
##
##   Test statistic                              1821.465
##   Degrees of freedom                                16
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W8 ~
##     Income_W7  (b)    0.620    0.009   69.945    0.000
##     Hlt_HRS_W7 (a)    0.059    0.009    6.749    0.000
##   Income_W9 ~
##     Income_W8  (c)    0.605    0.010   61.442    0.000
##     Hlt_HRS_W8        0.028    0.017    1.591    0.112
##   Income_W10 ~
##     Income_W9  (d)    0.498    0.010   49.048    0.000
##     Hlt_HRS_W9        0.014    0.019    0.739    0.460
##   Health_HRS_W8 ~
##     Hlt_HRS_W7 (b)    0.620    0.009   69.945    0.000
##     Income_W7  (a)    0.059    0.009    6.749    0.000
##   Health_HRS_W9 ~
##     Hlt_HRS_W8 (c)    0.605    0.010   61.442    0.000
##     Income_W8         0.067    0.011    6.023    0.000
##   Health_HRS_W10 ~
##     Hlt_HRS_W9 (d)    0.498    0.010   49.048    0.000
##     Income_W9         0.051    0.011    4.785    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W7 ~~
##     Health_HRS_W7     0.174    0.016   10.883    0.000
##  .Income_W8 ~~
##    .Health_HRS_W8     0.028    0.009    3.073    0.002
##  .Income_W9 ~~
##    .Health_HRS_W9    -0.002    0.011   -0.177    0.860
##  .Income_W10 ~~
##    .Health_HRS_W10    0.050    0.013    3.753    0.000
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
```

```
##    .Income_W8        -0.005    0.013   -0.345    0.730
##    .Income_W9        -0.018    0.016   -1.080    0.280
##    .Income_W10       -0.039    0.018   -2.193    0.028
##    .Health_HRS_W8     0.066    0.012    5.708    0.000
##    .Health_HRS_W9    -0.028    0.012   -2.380    0.017
##    .Health_HRS_W10   -0.102    0.012   -8.200    0.000
##     Income_W7         0.136    0.017    8.039    0.000
##     Health_HRS_W7     0.204    0.015   13.439    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Income_W8         0.650    0.015   43.480    0.000
##    .Income_W9         0.957    0.022   43.480    0.000
##    .Income_W10        1.157    0.027   43.480    0.000
##    .Health_HRS_W8     0.487    0.011   43.480    0.000
##    .Health_HRS_W9     0.498    0.011   43.480    0.000
##    .Health_HRS_W10    0.584    0.013   43.480    0.000
##     Income_W7         1.080    0.025   43.480    0.000
##     Health_HRS_W7     0.872    0.020   43.480    0.000
```

In addition to prior observations, stronger auto-regressive effects for both income and health across the waves are now evident.

It is worth noting that adding constraints to the auto-regressive terms has led to the cross-lagged effects of health on income in waves 9 and 10 becoming non-significant. This suggests that the stability of each variable over time might be accounting for some of the variation in the relationships between income and health.

The covariances reveal significant positive correlations between income and health at waves 7, 8, and 10, indicating that individuals with higher income tend to report poorer health, and vice versa. However, the covariance in wave 9 is insignificant, suggesting a weaker or potentially absent association between income and health at that time point.

The intercept for Health_HRS_W7 is positive, indicating that the average health score at wave 7 is above average. In contrast, the intercepts for income in waves 8 and 9 are not significant, suggesting that the average income levels at these time points do not significantly differ from the overall mean, with a slight decline observed from wave 7, where income was notably higher than average.

Model 8

Random intercepts for both health and income across all four waves are added in this model to estimate the variation that remains stable over time. This allows for accounting for individual differences in baseline levels of income and health that do not change across waves, providing a more accurate representation of how these variables evolve relative to each individual's starting point.

```
##
##     # Cross-lagged paths
##     Income_W8 ~ 1 + b*Income_W7 + a*Health_HRS_W7
##     Income_W9 ~ 1 + c*Income_W8 + Health_HRS_W8
##     Income_W10 ~ 1 + d*Income_W9 + Health_HRS_W9
##
##     Health_HRS_W8 ~ 1 + b*Health_HRS_W7 + a*Income_W7
##     Health_HRS_W9 ~ 1 + c*Health_HRS_W8 + Income_W8
##     Health_HRS_W10 ~ 1 + d*Health_HRS_W9 + Income_W9
##
##     # correlation
##     Health_HRS_W7 ~~ Income_W7
##     Health_HRS_W8 ~~ Income_W8
##     Health_HRS_W9 ~~ Income_W9
##     Health_HRS_W10 ~~ Income_W10
##
##     rm =~ 1*Income_W7 + 1*Income_W8 + 1*Income_W9 + 1*Income_W10
##     rm =~ 1*Health_HRS_W7 + 1*Health_HRS_W8 + 1*Health_HRS_W9 + 1*Health_HRS_W10
```

```
##
## Chi-Squared Difference Test
##
##       Df   AIC   BIC   Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_8 15 75653 75833  587.53
## fit_7 16 76884 77059 1821.46     1233.9 0.57104       1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As seen in the ANOVA results, the inclusion of random intercepts significantly improves the model fit, indicating that accounting for stable individual differences in income and health over time provides a better representation of the data.

```
## lavaan 0.6-19 ended normally after 26 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of model parameters                        33
##   Number of equality constraints                     4
##
##                                                   Used       Total
##   Number of observations                          3781       11700
##
## Model Test User Model:
##
##   Test statistic                               587.534
##   Degrees of freedom                                15
```

```
##   P-value (Chi-square)                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   rm =~
##     Income_W7         1.000
##     Income_W8         1.000
##     Income_W9         1.000
##     Income_W10        1.000
##     Health_HRS_W7     1.000
##     Health_HRS_W8     1.000
##     Health_HRS_W9     1.000
##     Health_HRS_W10    1.000
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   Income_W8 ~
##     Income_W7  (b)    0.314    0.010   32.862    0.000
##     Hlt_HRS_W7 (a)   -0.222    0.009  -23.963    0.000
##   Income_W9 ~
##     Income_W8  (c)    0.269    0.010   25.759    0.000
##     Hlt_HRS_W8       -0.339    0.018  -19.047    0.000
##   Income_W10 ~
##     Income_W9  (d)    0.188    0.010   17.954    0.000
##     Hlt_HRS_W9       -0.349    0.019  -18.527    0.000
##   Health_HRS_W8 ~
##     Hlt_HRS_W7 (b)    0.314    0.010   32.862    0.000
##     Income_W7  (a)   -0.222    0.009  -23.963    0.000
##   Health_HRS_W9 ~
##     Hlt_HRS_W8 (c)    0.269    0.010   25.759    0.000
##     Income_W8        -0.201    0.012  -17.433    0.000
##   Health_HRS_W10 ~
##     Hlt_HRS_W9 (d)    0.188    0.010   17.954    0.000
##     Income_W9        -0.165    0.011  -15.231    0.000
##
## Covariances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##  .Income_W7 ~~
##    .Health_HRS_W7    -0.165    0.012  -14.288    0.000
##  .Income_W8 ~~
##    .Health_HRS_W8    -0.092    0.009  -10.839    0.000
##  .Income_W9 ~~
##    .Health_HRS_W9    -0.102    0.011   -9.626    0.000
##  .Income_W10 ~~
##    .Health_HRS_W10   -0.074    0.012   -6.164    0.000
##
## Intercepts:
##                    Estimate  Std.Err  z-value  P(>|z|)
```

```
##    .Income_W8          0.094    0.015    6.113    0.000
##    .Income_W9          0.087    0.018    4.758    0.000
##    .Income_W10         0.011    0.019    0.588    0.556
##    .Health_HRS_W8      0.166    0.014   12.140    0.000
##    .Health_HRS_W9      0.064    0.014    4.605    0.000
##    .Health_HRS_W10    -0.062    0.014   -4.283    0.000
##    .Income_W7          0.136    0.017    7.884    0.000
##    .Health_HRS_W7      0.204    0.015   13.874    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .Income_W7          0.788    0.020   39.616    0.000
##    .Income_W8          0.542    0.014   37.918    0.000
##    .Income_W9          0.870    0.022   40.272    0.000
##    .Income_W10         0.992    0.024   41.147    0.000
##    .Health_HRS_W7      0.484    0.013   37.949    0.000
##    .Health_HRS_W8      0.354    0.010   35.670    0.000
##    .Health_HRS_W9      0.379    0.010   36.978    0.000
##    .Health_HRS_W10     0.456    0.012   38.657    0.000
##     rm                 0.334    0.011   29.487    0.000
```

Now that the between-person variation (0.334) has been accounted for, the cross-lagged effects have reversed. The difference in cross-lagged effects between this model and the previous ones is likely due to the inclusion of random intercepts. These random intercepts capture stable, individual-level differences in income and health, isolating within-person variability over time. Without random intercepts, the model combines both between-person and within-person effects, which can distort or inflate the observed relationships. By adjusting for baseline levels of income and health, the model with random intercepts focuses on time-specific changes, resulting in reversed cross-lagged effects and providing a clearer and more accurate representation of the dynamic relationship between income and health.

Model Assumptions

Sample Size

Klein recommends a ratio of 20 cases per variable measured, and this guideline is met in the current model. With a total of 78,443 observations and 33 model parameters, the average number of observations per parameter is 2,377, which comfortably exceeds the recommended ratio, ensuring sufficient statistical power for the analysis.

Variable Distributions

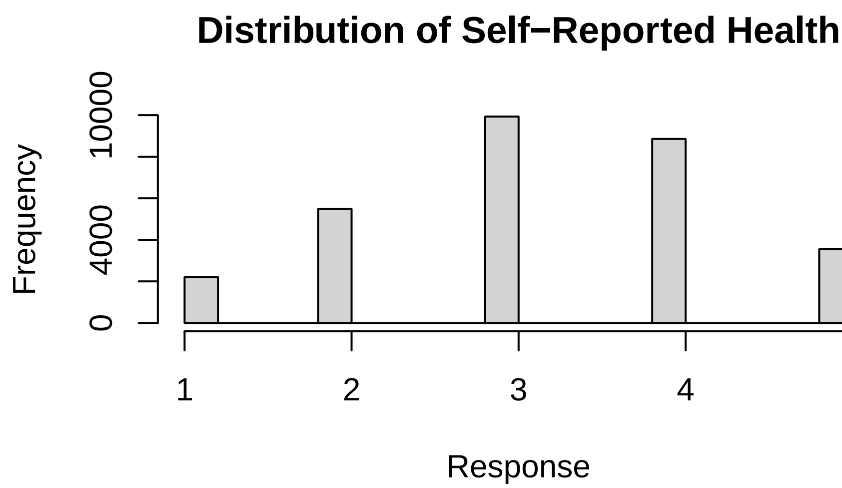
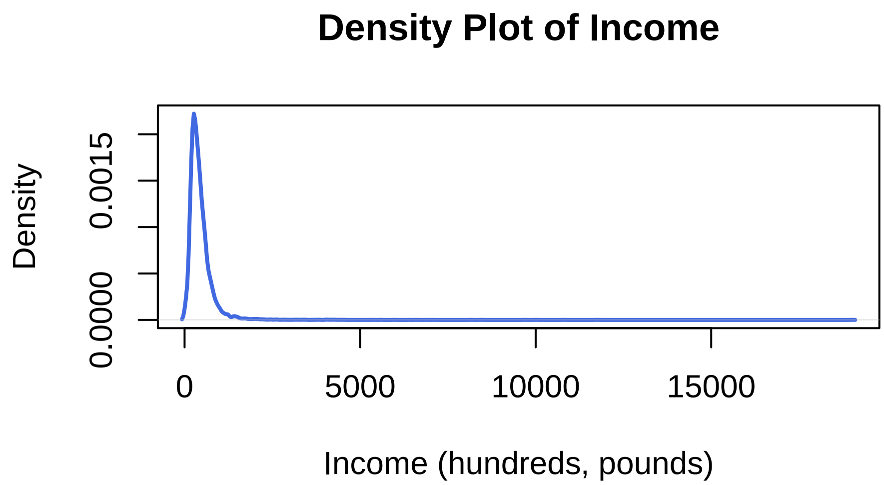
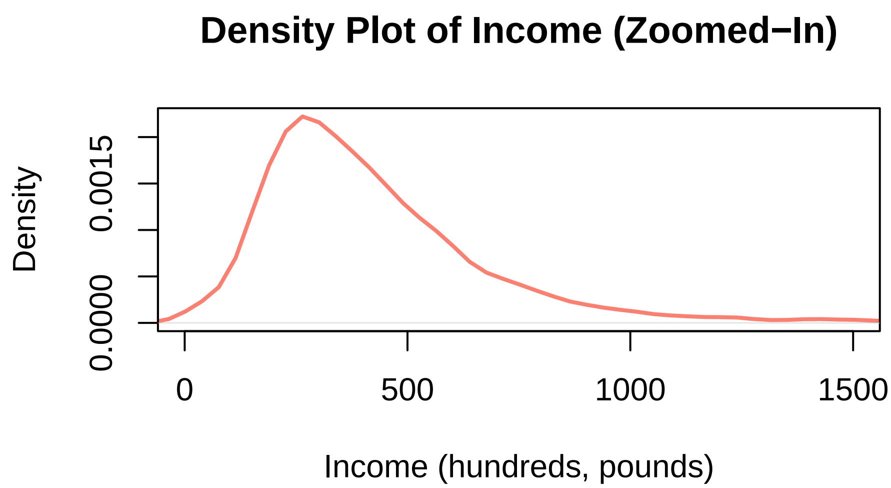


Figure 1: Self-reported health, 1=Poor, 5=Excellent.

Health is fairly evenly distributed across the sample, satisfying the requirement for a balanced distribution of the variable in the analysis. This distribution allows for a more robust and generalizable model, as extreme skewness or heavy clustering in one category could otherwise distort the relationships being tested.





While there are extremely high outliers for income, the majority of the data points, with income under £150,000, exhibit an acceptable level of skewness. This suggests that while the data contains some extreme values, the distribution of income for the majority of individuals is relatively normal, which supports the validity of the analysis and the robustness of the results for most of the sample.

Temporal Order

The flow of influence in the model is unidirectional, consistent with the structure of panel studies conducted in waves. In such studies, events that occur later in time cannot influence those that have already occurred. This temporal ordering ensures that the analysis reflects a logical causal direction, where earlier events potentially influence later outcomes, but not vice versa.

Conclusions

This study aimed to explore the relationship between income and health over time, specifically examining how each influences the other across multiple waves of data collection. A cross-lagged model was constructed to assess the bidirectional influences of income and health. The model included lagged paths that tested the predictive power of income on health outcomes and vice versa, while controlling for autoregressive effects within each variable. The results suggest a somewhat stronger influence of income on health, particularly in wave 9, although the relationship turned negative after accounting for between-person variation. This finding contrasts with previous research. However, due to the limitations of the model, causal relationships cannot be definitively established or inferred.

Longitudinal data analysis was crucial for answering the research question, as it enabled the observation of changes and relationships over time, rather than relying on a snapshot at a single point. This approach allowed for a more nuanced understanding of how income and health influence each other across different waves, capturing the dynamic nature of these variables.

The analysis encountered several limitations. Self-reported health data are subject to measurement errors, and incorporating objective health indicators could improve the accuracy of the findings. While autoregressive effects account for some stability, including additional variables, such as education, household wealth, and other socio-demographic factors, would provide a more comprehensive understanding of the income-health relationship. Furthermore, the study's limited timeframe, spanning only waves 7 to 10, may not capture long-term dynamics between income and health. Finally, the UK-based sample limits the generalizability of the findings to other populations and settings.

References

Akanni, L., Lenhart, O., & Morton, A. (2022). Income trajectories and self-rated health status in the UK. SSM - Population Health, 17, 101035. https://doi.org/10.1016/j.ssmph.2022.101035

Banks, J., Batty, G. David, Breedvelt, J., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., Zaninotto, P. (2024). English Longitudinal Study of Ageing: Waves 0-10, 1998-2023. [data collection]. 40th Edition. UK Data Service. SN: 5050, DOI: http://doi.org/10.5255/UKDA-SN-5050-27

Klein, R. B. (1998). Principles and practice of structural equation modeling. New York: Guilford Press.

Appendix A - R code: data cleaning

```
> library(tidyverse)

> library(data.table)

> # Store data file paths
> tbl_fin <-
+   list.files(path = "./data/financial/",
+              pattern = "\\.tab$",
+              full.names = T)

> tbl_ifs <-
+   list.files(path = "./data/ifs/",
+              pattern = "\\.tab$",
+              full.names = T)

> # Initialize financial and individual and family survey (ifs) data df
> fin <- data.frame()

> ifs <- data.frame()

> # Combine financial and ifs data into respective dfs
> for (i in 1:10){
+
+   # read wave data
+   fin0 <- read.delim(tbl_fin[i])
+
+   # find and add wave number based on file name
+   no <- str_extract(tbl_fin[i], "\\d+")
+   fin0$Wave <- no
+
+   # find column with wave-specific household id
+   wave_house <- colnames(fin0)[grep("^idahhw", colnames(fin0))]
+
+   # add columns of interest to overall financial data
+   fin0 <- fin0 %>% select(c("Wave","idauniq", wave_house,
+                                     "eqtotinc_bu_s","nettotw_bu_s"))
+   colnames(fin0) <- c("Wave", "ID_Ind", "ID_Household", "Income", "Wealth")
+
+   fin <- rbind(fin, fin0)
+
+ }

> # Ifs
> for (i in 1:10){
+
+   # read wave data
+   ifs0 <- read.delim(tbl_ifs[i])
+
+   # ifs and add wave number based on file name
+   no <- str_extract(tbl_ifs[i], "\\d+")
+   ifs0$Wave <- no
```

```
+
+   # ifs column with wave-specific household id
+   wave_house <- colnames(ifs0)[grep("^idahhw", colnames(ifs0))]
+
+   # add columns of interest to overall ifs data
+   ifs0 <- ifs0 %>% select(c("Wave","idauniq", wave_house,
+                              "edqual", "srh_hrs", "srh_hse"
+                              ))
+
+   colnames(ifs0) <- c("Wave", "ID_Ind", "ID_Household", "Education",
+                        "Health_HRS", "Health_HSE")
+
+   ifs <- rbind(ifs, ifs0)
+
+ }

> # Join the 2 datasets by wave number, ind id, and wave-specific household id
> data <- inner_join(fin, ifs, by = c("Wave", "ID_Ind","ID_Household"))

> data0 <- data #backup

> # Check proportions of response rate
> table(data$Health_HSE)

> table(data$Health_HRS)

> # only keep HRS since many respondents weren't asked HSE version in many waves (code -3)
> data <- data %>% select(c(-Health_HSE))

> # remove records where any field is missing
> data <- data %>% filter(if_all(everything(), ~ . >= 0))

> num_vars <- c("Wave", "ID_Ind", "ID_Household", "Education",
+               "Health_HRS",  "Wealth", "Income")

> data[num_vars] <- lapply(data[num_vars], as.numeric)

> data <- data %>% filter(Wave > 6)

> # create a copy to store temporary values
> temp <- data$Health_HRS

> # apply simultaneous recoding
> data$Health_HRS[temp == 1] <- 5

> data$Health_HRS[temp == 2] <- 4

> data$Health_HRS[temp == 4] <- 2

> data$Health_HRS[temp == 5] <- 1

> # pivot to wide format
> data_w <- data %>% pivot_wider(id_cols = ID_Ind, names_from = Wave,
+     values_from = c(Wealth, Income, Education, Health_HRS),
```

```
+     names_sep = "_W"
+   )

> # standardize data
> data_w_std <- data_w %>%
+   mutate(across(starts_with("Education"), scale),
+          across(starts_with("Wealth"), scale),
+          across(starts_with("Health_HRS"), scale),
+          across(starts_with("Income"), scale))
```

Appendix B - R code: modelling

```
> library(lavaan)

> library(tidyverse)

> ##########################
> model_1 <- '
+   # Cross-lagged paths
+   Income_W8 ~ 1 + Income_W7 + Health_HRS_W7
+   Income_W9 ~ 1 + Income_W8 + Health_HRS_W8
+   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
+
+   Health_HRS_W8 ~ 1 + Health_HRS_W7 + Income_W7
+   Health_HRS_W9 ~ 1 + Health_HRS_W8 + Income_W8
+   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
+
+   # correlation
+   Health_HRS_W7 ~~ Income_W7
+   Health_HRS_W8 ~~ Income_W8
+   Health_HRS_W9 ~~ Income_W9
+   Health_HRS_W10 ~~ Income_W10
+ '

> # model 2: equal cross-effects wave 7
> model_2 <- '
+   # Cross-lagged paths
+   Income_W8 ~ 1 + Income_W7 + a*Health_HRS_W7
+   Income_W9 ~ 1 + Income_W8 + Health_HRS_W8
+   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
+
+   Health_HRS_W8 ~ 1 + Health_HRS_W7 + a*Income_W7
+   Health_HRS_W9 ~ 1 + Health_HRS_W8 + Income_W8
+   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
+
+   # correlation
+   Health_HRS_W7 ~~ Income_W7
+   Health_HRS_W8 ~~ Income_W8
+   Health_HRS_W9 ~~ Income_W9
+   Health_HRS_W10 ~~ Income_W10
+ '

> # model 3: equal cross-effects wave 8
> model_3 <- '
+   # Cross-lagged paths
+   Income_W8 ~ 1 + Income_W7 + a*Health_HRS_W7
+   Income_W9 ~ 1 + Income_W8 + b*Health_HRS_W8
+   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
+
+   Health_HRS_W8 ~ 1 + Health_HRS_W7 + a*Income_W7
+   Health_HRS_W9 ~ 1 + Health_HRS_W8 + b*Income_W8
+   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
+
```

```
+   # correlation
+   Health_HRS_W7 ~~ Income_W7
+   Health_HRS_W8 ~~ Income_W8
+   Health_HRS_W9 ~~ Income_W9
+   Health_HRS_W10 ~~ Income_W10
+ '

> # model 4: equal cross-effects wave 9
> model_4 <- '
+   # Cross-lagged paths
+   Income_W8 ~ 1 + Income_W7 + a*Health_HRS_W7
+   Income_W9 ~ 1 + Income_W8 + Health_HRS_W8
+   Income_W10 ~ 1 + Income_W9 + b*Health_HRS_W9
+
+   Health_HRS_W8 ~ 1 + Health_HRS_W7 + a*Income_W7
+   Health_HRS_W9 ~ 1 + Health_HRS_W8 + Income_W8
+   Health_HRS_W10 ~ 1 + Health_HRS_W9 + b*Income_W9
+
+   # correlation
+   Health_HRS_W7 ~~ Income_W7
+   Health_HRS_W8 ~~ Income_W8
+   Health_HRS_W9 ~~ Income_W9
+   Health_HRS_W10 ~~ Income_W10
+ '

> # model 5 : stability on income
> model_5 <- '
+   # Cross-lagged paths
+   Income_W8 ~ 1 + b*Income_W7 + a*Health_HRS_W7
+   Income_W9 ~ 1 + Income_W8 + Health_HRS_W8
+   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
+
+   Health_HRS_W8 ~ 1 + b*Health_HRS_W7 + a*Income_W7
+   Health_HRS_W9 ~ 1 + Health_HRS_W8 + Income_W8
+   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
+
+   # correlation
+   Health_HRS_W7 ~~ Income_W7
+   Health_HRS_W8 ~~ Income_W8
+   Health_HRS_W9 ~~ Income_W9
+   Health_HRS_W10 ~~ Income_W10
+ '

> model_6 <- '
+   # Cross-lagged paths
+   Income_W8 ~ 1 + b*Income_W7 + a*Health_HRS_W7
+   Income_W9 ~ 1 + c*Income_W8 + Health_HRS_W8
+   Income_W10 ~ 1 + Income_W9 + Health_HRS_W9
+
+   Health_HRS_W8 ~ 1 + b*Health_HRS_W7 + a*Income_W7
+   Health_HRS_W9 ~ 1 + c*Health_HRS_W8 + Income_W8
+   Health_HRS_W10 ~ 1 + Health_HRS_W9 + Income_W9
+
+   # correlation
```

```
+   Health_HRS_W7 ~~ Income_W7
+   Health_HRS_W8 ~~ Income_W8
+   Health_HRS_W9 ~~ Income_W9
+   Health_HRS_W10 ~~ Income_W10
+ '

> model_7 <- '
+     # Cross-lagged paths
+     Income_W8 ~ 1 + b*Income_W7 + a*Health_HRS_W7
+     Income_W9 ~ 1 + c*Income_W8 + Health_HRS_W8
+     Income_W10 ~ 1 + d*Income_W9 + Health_HRS_W9
+
+     Health_HRS_W8 ~ 1 + b*Health_HRS_W7 + a*Income_W7
+     Health_HRS_W9 ~ 1 + c*Health_HRS_W8 + Income_W8
+     Health_HRS_W10 ~ 1 + d*Health_HRS_W9 + Income_W9
+
+     # correlation
+     Health_HRS_W7 ~~ Income_W7
+     Health_HRS_W8 ~~ Income_W8
+     Health_HRS_W9 ~~ Income_W9
+     Health_HRS_W10 ~~ Income_W10
+ '

> # model 8 : cross-lagged model with random intercept for Income
>
> model_8 <- '
+     # Cross-lagged paths
+     Income_W8 ~ 1 + b*Income_W7 + a*Health_HRS_W7
+     Income_W9 ~ 1 + c*Income_W8 + Health_HRS_W8
+     Income_W10 ~ 1 + d*Income_W9 + Health_HRS_W9
+
+     Health_HRS_W8 ~ 1 + b*Health_HRS_W7 + a*Income_W7
+     Health_HRS_W9 ~ 1 + c*Health_HRS_W8 + Income_W8
+     Health_HRS_W10 ~ 1 + d*Health_HRS_W9 + Income_W9
+
+     # correlation
+     Health_HRS_W7 ~~ Income_W7
+     Health_HRS_W8 ~~ Income_W8
+     Health_HRS_W9 ~~ Income_W9
+     Health_HRS_W10 ~~ Income_W10
+
+     rm =~ 1*Income_W7 + 1*Income_W8 + 1*Income_W9 + 1*Income_W10
+     rm =~ 1*Health_HRS_W7 + 1*Health_HRS_W8 + 1*Health_HRS_W9 + 1*Health_HRS_W10
+ '

> # Fit all models specs
> fit_1 <- sem(model_1, data = data_w_std)

> fit_2 <- sem(model_2, data = data_w_std)

> fit_3 <- sem(model_3, data = data_w_std)

> fit_4 <- sem(model_4, data = data_w_std)
```

```
> fit_5 <- sem(model_5, data = data_w_std)

> fit_6 <- sem(model_6, data = data_w_std)

> fit_7 <- sem(model_7, data = data_w_std)

> fit_8 <- sem(model_8, data = data_w_std)

> # Compare models progressively
> anova(fit_1, fit_2) # m2 is better

> anova(fit_2, fit_3) # m3 is better

> anova(fit_2, fit_4) # m2 is better

> anova(fit_2, fit_5) # m5 is better

> anova(fit_5, fit_6) # m6 is better

> anova(fit_6, fit_7) # m7 is better

> anova(fit_7, fit_8) # m8 is better
```

Appendix C - R code: visualization

```
> # Descriptive
> tabAll <- as.data.frame(apply(data, 2, summary))[,c(4,7)]

> tabHRS <- as.data.frame(table(data$Health_HRS))

> colnames(tabHRS) <- c("Response", "Freq")

> # Model summaries
> sumM1 <- summary(fit_1)

> sumM2 <- summary(fit_2)

> anvM12 <- anova(fit_1, fit_2)

> anvM23 <- anova(fit_2, fit_3)

> anvM24 <- anova(fit_2, fit_4)

> sumM5 <- summary(fit_5)

> anvM25 <- anova(fit_2, fit_5)

> sumM6 <- summary(fit_6)

> anvM56 <- anova(fit_5, fit_6)

> sumM7 <- summary(fit_7)

> anvM67 <- anova(fit_6, fit_7)

> sumM8 <- summary(fit_8)

> anvM78 <- anova(fit_7, fit_8)
```

Appendix D - R code: rmarkdown chunks

```
source(file="clean.R")
source(file="models.R")
source(file="vis.R")

kable(tabAll,align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))


kable(tabHRS,align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))

cat(model_1)
sumM1

cat(model_2)
anvM12


anvM23
anvM24


sumM2


cat(model_7)


anvM25
anvM56
anvM67


sumM7


cat(model_8)
anvM78


sumM8


hist(data$Health_HRS,
     main="Distribution of Self-Reported Health",
     xlab="Response",
     ylab="Frequency", )
```

```
plot(density(data$Income, na.rm=TRUE),
     main="Density Plot of Income",
     xlab="Income (hundreds, pounds)",
     ylab="Density",
     col="royalblue",
     lwd=2,
     xlim=c(min(data$Income), max(data$Income)))

plot(density(data$Income, na.rm=TRUE),
                     main="Density Plot of Income (Zoomed-In)",
                     xlab="Income (hundreds, pounds)",
                     ylab="Density",
                     col="salmon",
                     lwd=2,
                     xlim=c(min(data$Income), 1500))

source("clean.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
source("models.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
source("vis.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
```