Investigating life expectancy predictors in England using the ONS Health Index

Rachel Mumford, MPhil Population Health Sciences
Supervised by Dr Louise Lafortune

UNIVERSITY OF CAMBRIDGE

1. Background

Life expectancy (LE) is an overarching indicator of health representing how long you can expect to live on average. **Healthy life expectancy (HLE)** refers to how long you can expect to live **in good health**.

Improvements in LE in England are slowing it remains lower than in other developed countries such as Japan[1]. There is increasing **inequality** in LE in England[2].

Many factors have been shown to be important predictors of LE, but these are context dependent. Limited research has been conducted on LE in England.

LE can be used to compare within and between countries. This uses spatially aggregated data. **Aggregating data over different spatial areas can lead to a statistical bias** known as the Modifiable Areal Unit Problem (MAUP)[3].

2. Aim and research questions

The aim was to **investigate the predictors for LE and HLE** in England and explore the potential of the **Office for National Statistics (ONS) Health Index** for this type of research. The project also investigated **whether** the **predictors** of LE **change when looking at different levels of granularity** for the same variables. The three research questions were:

1. Which individual variables in the ONS Health Index best **predict life expectancy** in England?
2. Which individual variables in the ONS Health Index best **predict healthy life expectancy** in England?
3. Do **the predictors of life expectancy change** when looking at lower-tier local authority (LTLA) data compared to upper-tier (UTLA) data?

3. Methods

Data

All three datasets (LE, HLE and Health Index) were open-source from the ONS using spatially aggregated data. LE and Health Index data were available at **LTLA (n=307)** and **UTLA (n=150)** level**,** HLE only at UTLA.

The **ONS Health Index** is a unique data source with **56 individual indicators** (e.g. healthy eating, pupil attainment) from multiple sources. These fall under under three domains: **healthy people, healthy lives, healthy places**. It is constructed so that higher values are always 'healthier' with variables benchmarked to the 2015 England average which scores 100[4].

Data from 2019 were used as there were fewer missing variables than in 2020-2021.

Analysis

All analysis was undertaken in R version 4.2.3[5]**. Collinearity testing** was undertaken using the variance inflation factor (VIF) function, and 19 variables (UTLA) and 11 variables (LTLA) removed prior to analysis.

Three multiple linear regression models were fit:

1. **Life expectancy** by the Health Index explanatory variables **at UTLA level**.
2. **Life expectancy** by the Health Index explanatory variables **at LTLA level**.
3. **Healthy life expectancy** by the Health Index explanatory variables **at UTLA level.**

Both-direction stepwise selection was used for **model simplification**.

4. Results

Significant predictors

Multiple variables were found to be significant predictors ($p<0.05$) of LE and HLE. These can be summarised in the following categories:

- Birth, childhood and young people
- Healthy behaviours and associated conditions
- Disability and frailty
- Mental health (Life expectancy only)

The **effect sizes were small** for all significant variables although some were slightly higher, such as for teenage pregnancy. Some variables (e.g. happiness, air pollution) showed **unexpected negative coefficients**. For air pollution, this did not occur when London datapoints were excluded suggesting an urban influence.

Differences at different spatial scales

Some but not all significant variables differed between the UTLA and LTLA models for LE. The LTLA model explained more of the variation. This suggests possible MAUP although also differences may also be due to the UTLA sample size.

5. Discussion

These results suggest **predictors of LE are complex,** with many factors contributing. Several factors identified here relate to **childhood** and the environment in which children are born and raised, suggesting this is a critical phase for shaping health. Other important factors include those relating to **diet and behaviours** like smoking, which is in line with previous studies[6]. However, a measure **of social and economic deprivation**, such as the index for multiple deprivation (IMD), was not included in the model and this may be the main driver of these relationships (fig 1) given variables such as teenage pregnancy are known to also be associated with deprivation[7].

High air pollution appeared to predict lower LE. This is probably due to **confounding variables relating to urban living** like wealth or access to services. Model overfitting also led to some other spurious negative coefficients e.g. happiness.

The differences between the LE UTLA and LTLA models suggest that **data granularity is important**, and that HLE data should be made available at more granular levels.

Limitations

Some of the unexpected results are likely due to the **model overfitting** due to small sample size and use of a high number of variables.

Some variables in the index are also **self-reported** and may be biased.

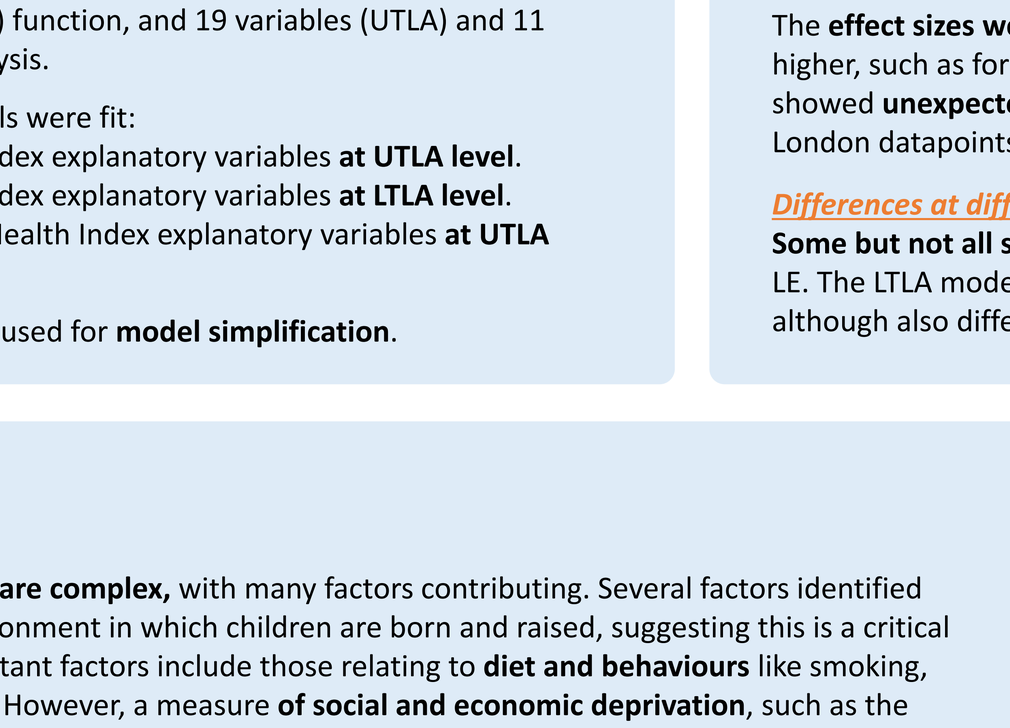

Figure 1. Conceptual framework displaying potential relationships between model variables and deprivation.

6. Conclusion

This study identifies factors which may be important in **predicting LE and HLE** at the area level.

Understanding the drivers of inequalities in these health outcomes is important as this can help **target interventions** and prevent the perpetuation of intergenerational poor health outcomes.

Further work will aim to incorporate a measure of **socioeconomic deprivation** to control for this as a possible confounding variable.

The **ONS Health Index** does provide an interesting tool for investigating public health questions, **but it has limitations** including with respect to some of the sources and sample size.

Publication of **granular data is essential** as MAUP may lead to incorrect conclusions and fewer datapoints restricts analysis.

7. References

1. Leon et al., (2019) Lancet Public Health, 4(11):e575-e82;
2. Marmot, M et al. (2016) Annals of Epi, 26(4):238-40;
3. Buzelli, M, (2020) Int. Encyclopedia of Human Geography, 169;
4. ONS, (2022) Health Index methods and developments, ONS website;
5. R Core Team ,(2021) R foundation;
6. Mackenbach, J et al. (2019) Lancet Public Health 4(10):e529-e37;
7. Cook, S et al. (2015) Obst, Gyn, and Reprod Med, 25(9):243-8.