

**University College London**  
**Institute of Health Informatics**  
**MSc Module in Public Health Data Science ASSESSMENT**  
**Development of a Composite Health Index for England**

YSBL5

## **1. BACKGROUND**

### **a. A statement of the context and setting**

As Metge *et al* (1) writes, “a composite index is a mathematical combination of several indicators or measures in order to form a single number”. This single number could be used to track an entire set of indicators, and would allow for identifying differences between localities and regions.

In the last issue of the Annual Report of the Chief Medical Officer (2), Professor Sally Davies recognizes the importance of public health not only as a financial cost, but as an important asset that needs to be tracked in order to be improved, so it can contribute widely to both economy and happiness of the nation. To be able to track this, she proposes the creation of a Composite Health Index comprising subgroups of health outcome measures, modifiable risk factors and the social determinants of health, that could be disaggregated by composition, be followed alongside the overall index, and mirror the multi-faceted determinants of the population’s health. The government should track this index, she argues, to know where to focus its efforts in improving the health of England.

## **2. AIMS**

### **a. A description of the aim for the index you have developed**

The index formulated aims to provide a simplified tool to track the most important health issues for the general population in England, detailed in Table 1, with the ability of being disaggregated into different subgroups of indicators across different localities. This is important to identify where public policy would be most effective.

## **3. METHODS:**

### **a. Description of data items included in the index**

The data items included are the five most influential in each subgroup and were selected according to a Principal Component Analysis (PCA), with the condition to track the general population, pertain to Local Authorities, present the most recent data, and have less than 15% of missing data. When necessary, a discretionary selection after PCA was needed for choosing indicators that don’t encompass others.

The subgroup Health Outcome Measures contained the indicator most represented in the Dimension-1, “92488 - Mortality rate from causes considered preventable”, but since it covers a broad range of preventable diseases, it was preferable to use other indicators that are narrower in their tracking. Because of this, it was preferable to keep the other indicators that track preventable causes of death.

In Social Determinants of Health Indicators “90630 - Children in low income families (all dependent children under 20)” and “10101 - Children in low income families (under 16s)” are close in their representation in Dimension-1, so the indicator 10101 was replaced by the indicator “11201 - Violent crime (including sexual violence) - hospital admissions for violence”.

For modifiable risk factors, the indicator “93014 - Percentage of physically active adults - current method” was not used since its counterpart, indicator “93015 - Percentage of physically inactive adults - current method” precedes it and both share similar representation, which gave way for indicator “20601 - Reception: Prevalence of overweight (including obesity)” that covers another important risk factor. Indicator “93078 - Average number of portions of fruit consumed daily (adults)” was removed from the selection since it would be included in indicator “93077 - Proportion of the population meeting the recommended '5-a-day' on a 'usual day' (adults)” because fruits are an essential part of the ‘5-a-day’ programme (3, 4).

Table 1: Indicators included in the Composite Health Index.

Subgroup	Indicator ID	Indicator Name
<b>Health Outcome Measures</b>	40401	Under 75 mortality rate from all cardiovascular diseases
	40701	Under 75 mortality rate from respiratory disease
	40402	Under 75 mortality rate from cardiovascular diseases considered preventable
	40502	Under 75 mortality rate from cancer considered preventable
	40702	Under 75 mortality rate from respiratory disease considered preventable
<b>Social Determinants of Health</b>	91872	Deprivation score (IMD 2015)
	90630	Children in low income families (all dependent children under 20)
	11201	Violent crime (including sexual violence) - hospital admissions for violence
	90356	Fuel poverty
	11202	Violent crime (including sexual violence) - violence offences per 1,000 population
<b>Modifiable Risk Factors</b>	20602	Year 6: Prevalence of overweight (including obesity) (10-11 yrs)
	93077	Proportion of the population meeting the recommended '5-a-day' on a 'usual day' (adults)
	93015	Percentage of physically inactive adults - current method
	20601	Reception: Prevalence of overweight (including obesity)
	92443	Smoking Prevalence in adults (18+) - current smokers (APS)

## b. Description of the methods for creation of the index

The creation of the index was undertaken using a data driven approach, where each individual indicator was included in the first step as an item belonging to the Public Health Outcomes Framework from Public Health England (ProfileID = 19) using the FingertipsR package Version 0.2.2 from the CRAN repository for the R programming language (5). As first classification, 168 indicators were extracted from the PHE database using the argument for Local Authorities (AreaTypeID = 101). This data went through filters that selected the latest update for each indicator, values for District and Unitary Authorities, tracked the general population (Sex = "Persons").

Afterwards, the indicators were filtered according to the amount of missing values they contained, using a threshold of 15%. This highlighted five indicators, of which it was deemed preferable to keep indicators "92432 - Deaths from drug misuse", "91367 - Treatment completion for TB" and "90791 - HIV late diagnosis" since these track important aspects of healthcare. The missing values remaining in the data were replaced with the median value of each indicator.

A Shapiro-Wilk test was performed on each indicator to assess its normality. For those indicators that didn't present a normal distribution, a logarithmic rescaling was performed. Some indicators presented negative or zero values, which produced NaN values that were replaced with the median of its corresponding indicator. A polarity reversal was performed for those indicators where a higher value meant a better index. Then, a scaling was done on the whole data so different indicators that are measured in different scales can be compared.

PCA was performed with the `prcomp` function and the "factoextra" package from the CRAN repository (6) for identifying which indicators are the most influential. From the dimensions obtained, Dimension-1 explained 80% of the variance, so further analysis was focused solely on this dimension (Figure 1).

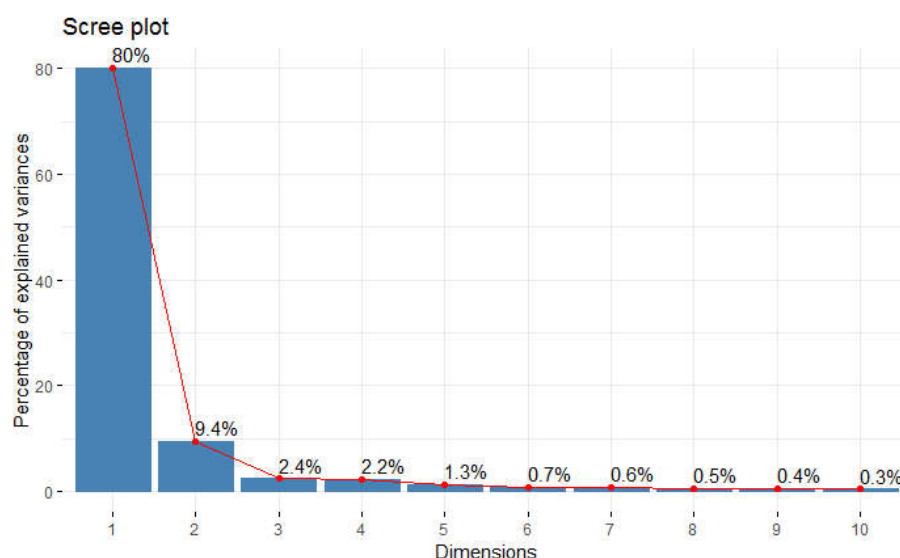


Figure 1: Dimension-1 of the PCA explains 80% of the variance.

Indicators were ranked according to their Quality of Representation in Dimension-1 (Figure 2) and from this ranking five indicators from each subgroup were selected as

explained above. For interpretation, a score was calculated with these indicators (Table 1), with the equation

$$\text{Health Score} = -1 \times (\text{Mean (Z-score)}) + 10$$

with higher values representing better health.

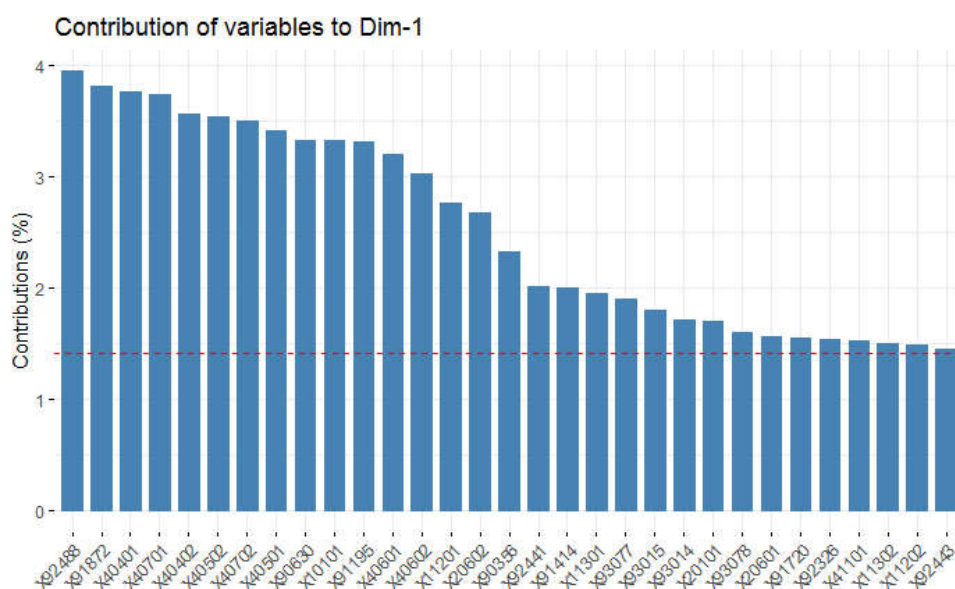


Figure 2: Contribution of each indicator to Dimension-1, up to indicator “92443 - Smoking Prevalence in adults (18+) - current smokers (APS)”. Red line indicates the average contribution of all indicators.

## 4. RESULTS

### a. Presentation of the overall index

The Composite Health Index comprising the five most relevant health indicators from the three subgroups allows to rank Local Authorities according to their health score. Descriptive statistics are detailed in Table 2, and relevant rankings are presented in Table 3 and 4 for the overall index in Local Authorities. Table 5 presents a ranking comprising regions in England. Figure 3 shows the distribution of scores across England.

Table 2: Descriptive statistics of the Composite Health Index across Local Authorities in England.

Mean	Standard deviation	Median	Maximum	Minimum
10.00	±0.791	10.11	11.44	7.49

Table 3: The 20 Local Authorities with the least health issues in England.

<b>Rank</b>	<b>Local Authority</b>	<b>Health Score</b>
1st	Waverley	11.44
2nd	Chiltern	11.42
3rd	Wokingham	11.40
4th	Vale of White Horse	11.33
5th	Hart	11.29
6th	Elmbridge	11.28
7th	South Cambridgeshire	11.27
8th	South Oxfordshire	11.25
9th	Richmond upon Thames	11.24
10th	Horsham	11.19
11th	St Albans	11.16
12th	East Devon	11.15
13th	Guildford	11.15
14th	Epsom and Ewell	11.14
15th	Rushcliffe	11.11
16th	Mole Valley	11.10
17th	East Hertfordshire	11.08
18th	Fareham	11.07
19th	Wealden	11.05
20th	Harborough	11.05

Table 4: The 20 Local Authorities with the most health issues in England.

<b>Rank</b>	<b>Local Authority</b>	<b>Health Score</b>
326th	Blackpool	7.49
325th	Manchester	7.55
324th	Kingston upon Hull	7.87
323rd	Middlesbrough	7.89
322nd	Liverpool	8.04
321st	Knowsley	8.09
320th	Nottingham	8.36
319th	Hartlepool	8.40
318th	Salford	8.41
317th	Sandwell	8.44
316th	Rochdale	8.45
315th	Blackburn with Darwen	8.49
314th	Burnley	8.53
313th	Wolverhampton	8.56
312th	Hyndburn	8.59
311th	Barking and Dagenham	8.62
310th	Sunderland	8.65
309th	Bradford	8.70
308th	Oldham	8.71
307th	St. Helens	8.73

Table 5: Regions in England ranked according to their Composite Health Index Score.

Rank	Area	Health Score
1st	South East region	11.12
2nd	South West region	11.06
3rd	East of England region	10.86
4th	London region	10.2
5th	East Midlands region	10.05
6th	West Midlands region	9.63
7th	Yorkshire and the Humber region	9.26
8th	North West region	9.01
9th	North East region	8.8

**Aggregated Composite Health Index in England**

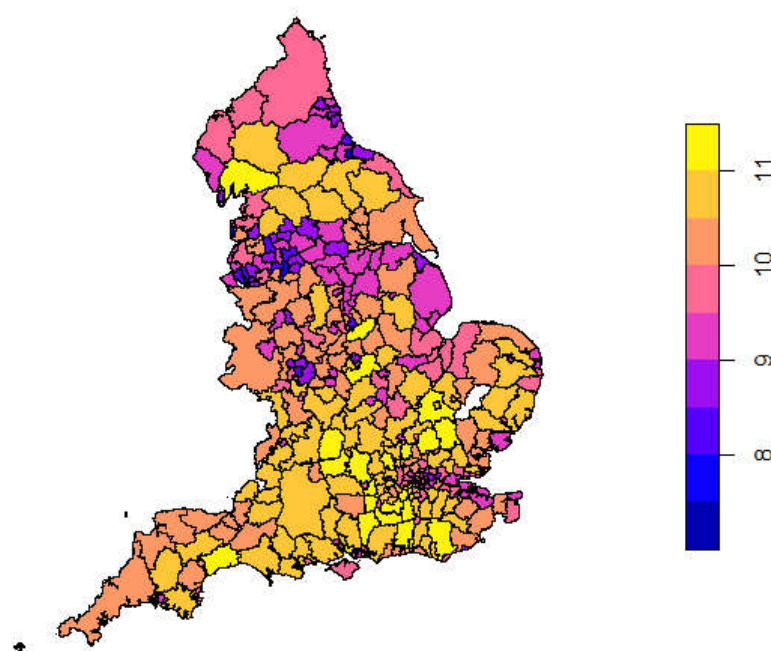


Figure 3: Distribution of Composite Health Index scores across England.

## b. Presentation of disaggregated index

The disaggregated Composite Health Index for Health Outcome Measures is presented. Descriptive statistics are detailed in Table 6, and relevant rankings are presented in Table 7 and 8 for the disaggregated index in Local Authorities. Figure 4 shows the distribution of scores across England.

Table 6: Descriptive statistics for the Health Outcome Measures subgroup.

Mean	Standard deviation	Median	Maximum	Minimum
10.00	±0.941	10.18	11.43	6.23

Table 7: The 20 Local Authorities with the best Health Outcome Measures readings in England.

<b>Rank</b>	<b>Area</b>	<b>Health Score</b>
1st	South Oxfordshire	11.43
2nd	Kensington and Chelsea	11.39
3rd	Hart	11.39
4th	Waverley	11.38
5th	Horsham	11.35
6th	Chiltern	11.34
7th	Vale of White Horse	11.34
8th	South Cambridgeshire	11.29
9th	East Dorset	11.29
10th	South Lakeland	11.29
11th	Mid Suffolk	11.27
12th	Cotswold	11.25
13th	Surrey Heath	11.25
14th	Babergh	11.24
15th	Suffolk Coastal	11.23
16th	Winchester	11.22
17th	South Norfolk	11.21
18th	Wokingham	11.20
19th	East Hertfordshire	11.19
20th	Guildford	11.19

Table 8: The 20 Local Authorities with the worst Health Outcome Measures readings in England.

<b>Rank</b>	<b>Area</b>	<b>Health Score</b>
326th	Manchester	6.23
325th	Blackpool	6.85
324th	Middlesbrough	7.24
323rd	Kingston upon Hull	7.39
322nd	Blackburn with Darwen	7.47
321st	Liverpool	7.52
320th	Knowsley	7.57
319th	Burnley	7.73
318th	Nottingham	7.76
317th	Salford	7.76
316th	Rochdale	7.82
315th	Corby	8.24
314th	Hastings	8.25
313th	Hyndburn	8.27
312th	Tameside	8.28
311th	Oldham	8.34
310th	Hartlepool	8.37
309th	Preston	8.40
308th	Lincoln	8.42
307th	St. Helens	8.43

### Health Outcome Measures in England

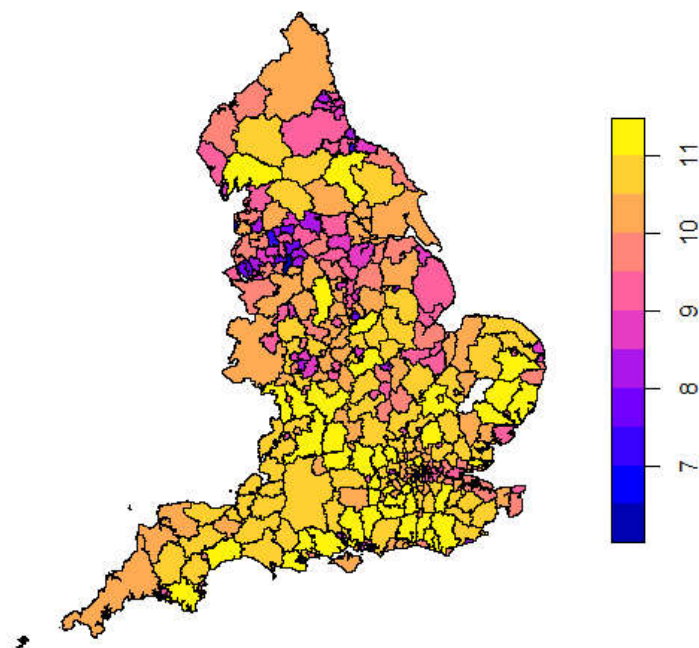


Figure 4: Distribution of Composite Health Index scores for Health Outcome Measures across England.

The disaggregated Composite Health Index for the Social Determinants of Health subgroup is presented next. Descriptive statistics are detailed in Table 9, and relevant rankings are presented in Table 10 and 11 for the overall index in Local Authorities. Figure 5 shows the distribution of Social Determinants of Health scores across England.

Table 9: Descriptive statistics for the Social Determinants of Health subgroup.

Mean	Standard deviation	Median	Maximum	Minimum
10.00	$\pm 0.823$	10.16	11.52	7.18

**Social Determinants of Health in England**

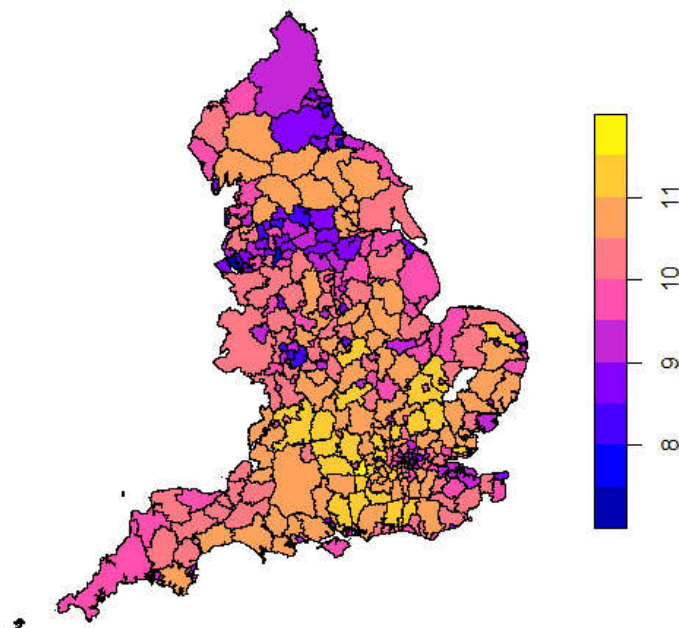


Figure 4: Distribution of Composite Health Index scores for Social Determinants of Health across England.



Table 10: The 20 Local Authorities with the best Social Determinants of Health readings in England.

<b>Rank</b>	<b>Area</b>	<b>Health Score</b>
1st	Wokingham	11.52
2nd	West Oxfordshire	11.39
3rd	South Oxfordshire	11.33
4th	Vale of White Horse	11.33
5th	Fareham	11.24
6th	Chiltern	11.24
7th	West Berkshire	11.22
8th	Cotswold	11.19
9th	Mid Sussex	11.18
10th	East Hampshire	11.18
11th	Uttlesford	11.16
12th	Waverley	11.15
13th	East Hertfordshire	11.15
14th	South Cambridgeshire	11.15
15th	Elmbridge	11.13
16th	Horsham	11.13
17th	Hart	11.10
18th	South Northamptonshire	11.08
19th	Bracknell Forest	11.08
20th	St Albans	11.08

Table 11: The 20 Local Authorities with the worst Social Determinants of Health readings in England.

<b>Rank</b>	<b>Area</b>	<b>Health Score</b>
326th	Blackpool	7.18
325th	Liverpool	7.40
324th	Manchester	7.68
323rd	Middlesbrough	7.71
322nd	Kingston upon Hull	8.09
321st	Knowsley	8.10
320th	Burnley	8.23
319th	Birmingham	8.27
318th	Wolverhampton	8.35
317th	Bradford	8.41
316th	Sunderland	8.44
315th	Sandwell	8.45
314th	Hartlepool	8.46
313th	Blackburn with Darwen	8.49
312th	Stoke-on-Trent	8.51
311th	Nottingham	8.51
310th	Halton	8.54
309th	Newcastle upon Tyne	8.54
308th	Rochdale	8.60
307th	Leicester	8.60

The disaggregated Composite Health Index for the Modifiable Risk Factors subgroup is presented next. Descriptive statistics are detailed in Table 12, and relevant rankings are presented in Table 13 and 14 for the overall index in Local Authorities. Figure 6 shows the distribution of Social Determinants of Health scores across England.

Table 12: Descriptive statistics for the Modifiable Risk Factors subgroup.

Mean	Standard deviation	Median	Maximum	Minimum
10.00	±0.755	9.99	11.77	8.14

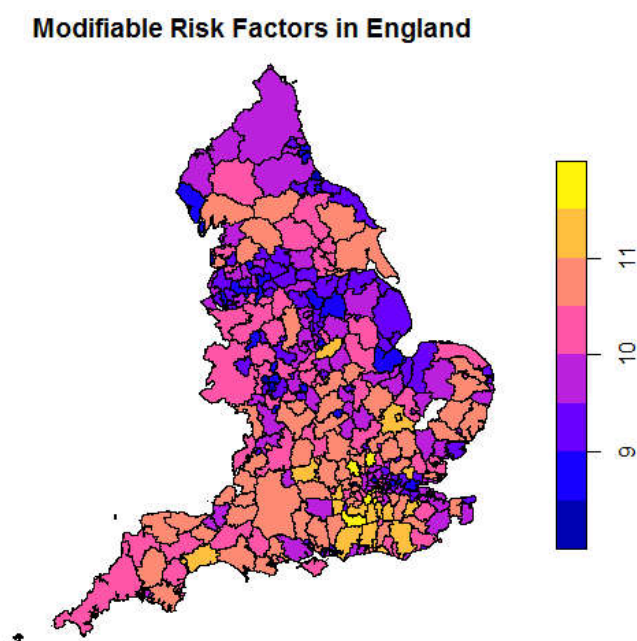


Figure 5: Distribution of Composite Health Index scores for Modifiable Risk Factors across England.

Table 13: The 20 Local Authorities with the best Modifiable Risk Factors readings in England.

Rank	Area	Health Score
1st	Waverley	11.77
2nd	Richmond upon Thames	11.75
3rd	Chiltern	11.66
4th	Elmbridge	11.56
5th	Epsom and Ewell	11.51
6th	St Albans	11.51
7th	East Devon	11.5
8th	Guildford	11.49
9th	Wokingham	11.49
10th	Rushcliffe	11.4
11th	Hart	11.39
12th	South Cambridgeshire	11.37
13th	Christchurch	11.36
14th	Vale of White Horse	11.33
15th	Mole Valley	11.24
16th	Tonbridge and Malling	11.23
17th	Cambridge	11.22
18th	Chichester	11.2
19th	Brighton and Hove	11.16
20th	Brentwood	11.09

Table 14: The 20 Local Authorities with the worst Modifiable Risk Factors readings in England.

Rank	Area	Health Score
326th	Kingston upon Hull	8.14
325th	Barking and Dagenham	8.27
324th	Hartlepool	8.37
323rd	Sandwell	8.38
322nd	Blackpool	8.42
321st	Wolverhampton	8.5
320th	Brent	8.54
319th	Knowsley	8.6
318th	Barrow-in-Furness	8.62
317th	Newham	8.67
316th	Slough	8.68
315th	Sunderland	8.69
314th	Manchester	8.72
313th	Middlesbrough	8.72
312th	Hyndburn	8.73
311th	Cannock Chase	8.75
310th	South Holland	8.77
309th	Nottingham	8.81
308th	Salford	8.87
307th	Walsall	8.87

## 5. DISCUSSION

### a. Interpretation of results, including their strengths and weaknesses

The creation of a Composite Health Index that encompasses key health indicators is important for simplifying the healthcare landscape and provide a high level interpretation of its development through time and across different geographical areas. Nevertheless, there needs to be a level of discernment into which indicators are more relevant than others, since including all of them becomes impractical at the time when policy makers need to make decisions regarding policy due to the excessive amount of generalisation obtained (7). A balance between specificity and generalisation of the information being tracked needs to be achieved.

The strengths of the Index developed is that is built with a data driven approach, with no initial bias on which indicators are more important. The results of the PCA strongly suggests which indicators are more relevant, making them more suitable candidates to be included, but a level of discretion is necessary while choosing what data needs to be included to avoid over representation. This could be a weak point in the creation of the index, since interpretation is subject to human error (7). Another weak point that could be argued is that only the general population is tracked, which could mask differences in health between the sexes. A re-engineering of the index would not be difficult since all it is needed is to change the initial filters for sex in the code provided.

The results obtained suggest there is a strong influence of lifestyle decisions, given primarily by modifiable risk factors, and economic struggle given by the social determinants of health, both influencing heavily on the outcomes measures, and the areas where the two former indicators tend to be the worst, the latter follows suit. Although not

included in the list of indicators chosen to compose the index, the most influential indicator, 92488, suggests that prevention of disease would be the most impactful health policy that could be carried out, either via public education (with initiatives like the 5-a-day programme) or heavier taxes on items known to cause disease like tobacco and heavily processed foods.

**b. How the index could be used to improve the health of the public by conducting one of the following:**

**i. A brief stakeholder analysis**

Stakeholders like the Health Secretary and Policymakers, could use the information given by this composite index to direct public policy.

The British Medical Association alongside GPs would be in an expert position with coercion power, with the interest of improving the healthcare landscape and benefiting from a diminished work load on them. The NHS, Public Health England and academics, would also be in an expert position to act in those areas that need most attention, with the possibility of coercion power against policymakers and lobbyists.

The media would be flanked influenced by both the government and big companies in the tobacco, food and alcohol industry, which could use their advertisement money as bargain power against the media.

Patients affected by preventable diseases could be in a position of power against policy makers and politicians eager to win elections, and also could change their behaviour when they understand the impact of risk factors.

**ii. A brief media engagement and dissemination strategy**

It should be headed most effectively by a senior figure in an expertise position, be it a GP, researcher, or a member from an organization like the NHS or Public Health England.

The message should contain key findings like a call towards improving the government focus on prevention, use appropriate language and numbers showing what are the most pressing issues affecting health, with visuals and infographics focusing on balanced diets, and smoking prevention. Also there needs to be a call to improve the level of economic deprivation in which many areas go through, since this also affects heavily on health quality.

Channels considered should include policy reports, briefing papers, press releases, and public info campaign, to be able to reach as much as relevant receivers as possible, like policymakers, politicians, strong stakeholders and the wider public.

**6. REFERENCES**

(1) Metge C, Chateau D, Prior H, Soodeen R, De Coster C, Barré L. Composite Measures/Indices of Health and Health System Performance. Winnipeg, MB: Manitoba Centre for Health Policy, August 2009.

(2) Various Authors. Annual Report of the Chief Medical Officer, 2018 Health 2040 – Better Health Within Reach.

(3) National Health Service. Why 5 A Day??. <https://www.nhs.uk/live-well/eat-well/why-5-a-day/> Accessed 27-04-2019.

- (4) World Health Organization. Promoting fruit and vegetable consumption around the world. <https://www.who.int/dietphysicalactivity/fruit/en/> Accessed 27-04-2019.
- (5) Sebastian Fox, Julian Flowers, Simon Thelwall,, Duncan Gormansway, Carl Ganz, David Whiting, Crown Copyright 2018. Package 'fingertipsR'. <https://cran.r-project.org/web/packages/fingertipsR/index.html> Accessed 27-04-2019.
- (6) Alboukadel Kassambara, Fabian Mundt. Package 'factoextra'. <https://cran.r-project.org/web/packages/factoextra/index.html> Accessed 27-04-2019.
- (7) Michela Nardo, Michaela Saisana, Andrea Saltelli, Stefano Tarantola. Tools for Composite Indicators Building. 2005, European Commission, EUR 21682 EN, Institute for the Protection and Security of the Citizen, JRC Ispra, Italy.