Regional health inequalities explained: empirical evidence from Czechia

ALEŠ BĚLOHRADSKÝ[1,2], DANIELA GLOCKER[3]

[1] Charles University, Faculty of Science, Department of Social Geography and Regional Development, Prague, Czechia; e-mail: ales.belohradsky@natur.cuni.cz
[2] Ministry of Finance of the Czech Republic, Prague, Czechia
[3] Organisation for Economic Co-operation and Development, Paris, France; e-mail: daniela.glocker@oecd.org

ABSTRACT   Health outcomes in Central and Eastern Europe have been generally improving over the last two decades. However, in Czechia, similar to other countries, the outcomes vary significantly across the regional dimension. In 2016, life expectancy at birth ranged from between 75.5 to 80 years – across 77 districts. This article empirically analyses the determinants of these differences using a wide range of explanatory variables that are available on a district level. Applying factor analysis, our findings indicate that the socio-economic situation within a district is the main driver – explaining a difference of up to four years in life expectancy at birth. Further, controlling for gender specific variables, the results suggest that for men alcohol consumption significantly reduces life expectancy. Both factors capturing the socio-economic situation and alcohol consumption exceed the impact of health care accessibility or environmental causes on life expectancy. The results remain robust when using more disaggregated information on 206 administrative districts – per geographic scale.

KEY WORDS   geography of health – regional inequalities – life expectancy – factor analysis – Czechia

1. Introduction

Health is a key driver not only for personal well-being but also for economic growth. The health status of the population is linked to economic performance through workforce participation and productivity. Health outcomes in turn are influenced by a set of determinants, ranging from access to and quality of care to living conditions and lifestyle choices of the population. As such, health status inequality is closely linked to socio-economic conditions – an important aspect of regional disparities (CSDH 2008, WHO 2013). While welfare states were often thought to tackle health inequalities, it is increasingly recognised that not only are health inequalities not reduced but also that they may even become more pronounced in certain areas (Mackenbach 2012). Thus, there is a need to explain causes of this development and to find a way to address them.

Health outcomes in Czechia, measured by life expectancy (both at birth and in older ages), healthy life years at birth, infant mortality, or amenable mortality, have improved considerably over the last two decades, slowly converging to OECD averages (OECD 2018). But within the country, health outcomes differ across regions. As such, life expectancy at birth across 77 districts ranges from 75.5 to 80 years. Moreover, this inequality has increased since the nineties, when a slightly smaller range of about four years has been reported (Blažek, Dzúrová 2000).

This article analyses the sources of regional variations in health outcomes. It provides a first empirical assessment of structural differences across Czech regions that determine health inequalities. Namely, the aim is to explain the variation in life expectancy across administrative districts in Czechia and to specify the main causes, which must be reflected in order to develop targeted policy solutions. For this purpose, an exploratory factor analysis with no a priori assumption on the number and character of factors is applied. The main hypothesis is that besides pure geographic factors, there is a list of other factors ranging from socio-economic to environmental conditions, which influence the region's average health outcomes. More specific hypotheses will be formulated in the next section based on the literature review.

A dataset bringing together spatially disaggregated data at the regional level is constructed, allowing for a systematic analysis of regional variations in health outcomes. Regional specific variables are regressed on health outcomes using a factor analysis approach. This paper adds to the literature by analysing a comprehensive list of explanatory variables on a disaggregated regional level. Cross-country studies usually explain the variance by different institutional, legal and regulatory settings. These framework conditions, however, are mostly the same for all regions within a country and thus cannot explain the inter-regional differences. In this regional analysis, new determinants are explored in their impact on health outcomes. A complex factor analysis is applied to identify the most influential implicit

drivers. Further, the results are compared across different territorial divisions, as the choice of scale can affect the results.

The article firstly discusses available literature analysing health inequalities and its results with a special attention to Central Europe and Czechia in particular. The next section provides a descriptive overview of health outcomes across Czech regions. The Section 4 describes the empirical strategy, i.e. the method of factor analysis, and the data used. Results of our estimations and robustness checks follow in the Section 5. The policy implications of the results conclude the study.

2. Literature review

Studies explaining regional variations in health outcomes tend to focus on cross-country comparisons. A long-run perspective on life expectancy evolution in Europe is provided by Leon (2011) showing the gap which has emerged between former communist bloc and the Western countries. Central Eastern European (CEE hereafter) countries including Czechia still have a gap in absolute terms vis-à-vis Western countries. Vågerö (2010) also highlights that health inequalities within country are widening more in post-communist countries than in the West. Richardson et al. (2014) find that the regional variation in Europe did not change much between 1991 and 2008 with the exception of males in Eastern Europe. Interestingly, females in Eastern Europe do not seem to be affected by economic aspects to such a degree. Dzúrová (2000) explains the emergence of regional disparities in Czechia using mortality rates by different abilities to adapt to new conditions in the capitalist regime.

Within a country, however, framework conditions such as the institutional set up, regulations and laws apply to all regions. Based on individual-level data, a second branch of literature focuses on determinants of self-rated health. A detailed literature review on the effect of income inequality is for example provided by Pickett and Wilkinson (2015) who strongly confirm the dependency of health on income. Many of these studies, however, use subjective self-assessed socio-economic status instead of any objective measures. They show that poor self-rating in one aspect is likely to be accompanied by poor self-rating in the other (Präg, Mills, Wittek 2016). Also a meta-analysis by Cundiff and Matthews (2017) shows that the link between health and subjective socio-economic status is stronger for self-rated health than for physical health outcomes. Assessing self-rated health in a spatial framework, Thomson et al. (2017) show that regional health levels are lowest in continental Europe (especially CEE) and highest in Scandinavia, Ireland and other isolated regions, especially around big cities. Within a country, Pantyley (2017) describes an urban-rural pattern in the case of South-Eastern Poland. He

shows that self-rated health is correlated mostly to age, life quality, social position and financial situation, i.e. the determinants which all support the huge gap between urban and peripheral regions.

Chetty et al. (2016) provided a detailed analysis of the influence of personal income on life expectancy in the United States, showing that income is closely associated with longevity. Further, the authors find that differences between income groups have been increasing over time. At the same time, Chetty et al. point out that there is a significant regional variance in this association, which can be explained by health behaviours (namely smoking) more than accessibility of health care or environmental factors.

Besides the socio-economic factors, we can trace another stream of literature, which analyses differentials in health outcomes by environmental factors. Samoli et al. (2016) describe the influence of various sources of air pollution on selected diagnoses in London, showing that the changes in the air particle composition in time have direct effect on morbidity. Another environmental threat, the groundwater contamination, is addressed for Slovakia by Rapant and Krčmová (2007), who show that connecting sites with arsenic-contaminated water are associated with bigger health risk.

Based on the presented literature, it is possible to outline the potential sources of variations in regional health outcomes, which at the same time frame the hypotheses of this article: life expectancy would be higher in regions (1) with better general socio-economic conditions (Picket, Wilkinson 2015; Chetty et al. 2016), and (2) with lower environmental burden (Samoli et al. 2016). For the sake of completeness, it is also expected that life expectancy is positively affected by (3) better accessibility to health care (OECD 2018) and (4) micro-economic behavioural factors such as adverse lifestyle habits (Lee et al. 2012).

3. Regional variations in health outcomes in Czechia

The health system in Czechia features universal coverage with a broad health care basket and equal access to care regardless of social and economic background. Nevertheless, health outcomes vary significantly across its 77 districts, with life expectancy at birth differing by more than 4 years for both men and women. The larger cities such as Prague and Brno have the highest life expectancy, while districts in the Ústecký region in the north have the lowest one (Figure 1 and 2).

As the institutional set up, regulations and laws are the same for all districts, the variations in health expectancies are likely to reflect structural differences. Correlations of life expectancy and district aggregates show distinct patterns for socio-economic characteristics of the districts. Figure 3 shows the strong positive relationship of life expectancy at birth with education and election participation,
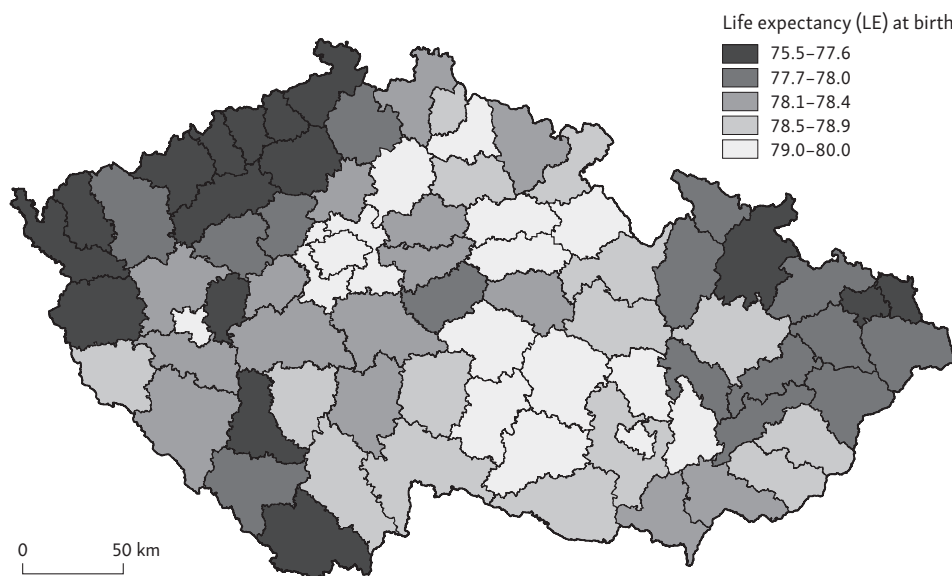
Fig. 1 – Life expectancy at birth across 77 statistical districts. Average for men and women over the years 2012–2016. Source: Own elaboration based on data from Czech Statistical Office.

whereas the relationship with medical infrastructure or environment seems less clear. Unemployment exhibits less strong negative relationship.

4. Estimating the determinants of regional health outcomes

This section applies factor analysis on spatially disaggregated data for Czechia to identify the drivers of regional variations in health outcomes. This approach allows to provide a first assessment of underlying processes that affect regional variations in health outcomes, which are key to develop targeted policy solutions.

4.1. Empirical strategy

Health outcomes are determined by a highly inter-correlated system of demographic factors, socio-economic situation, lifestyle, environmental and living conditions as well as access to and quality of health care (WHO 2012). To disentangle different factors that determine health outcomes, a factor analysis is applied. Compared to simple regression analysis, this procedure has the advantage of avoiding multicollinearity issues which hinder assessing the effect of individual

Life expectancy (LE) at birth

75.0–76.3
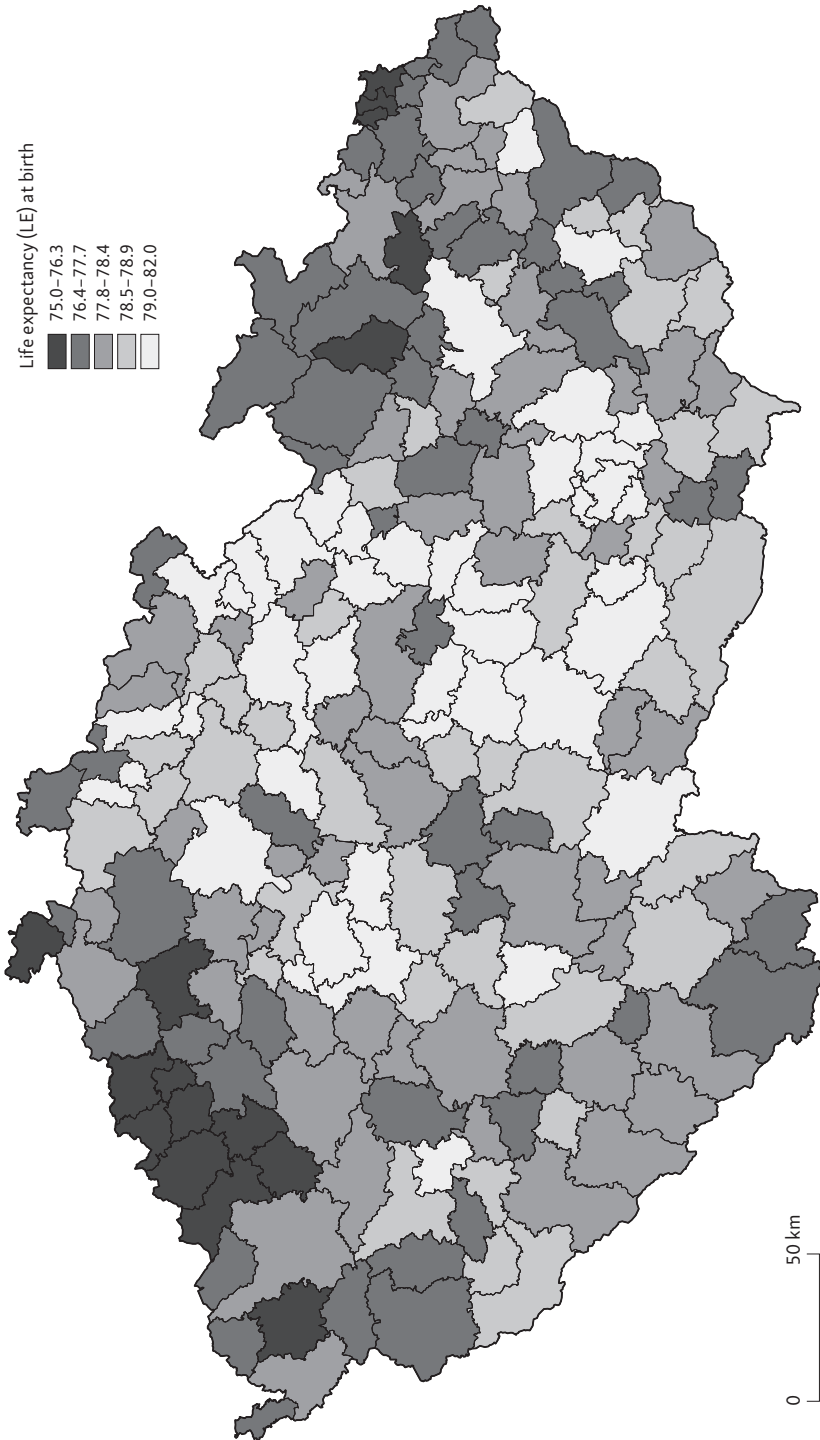76.4–77.7
77.8–78.4
78.5–78.9
79.0–82.0

0   50 km

Fig. 2 – Life expectancy at birth across 206 administrative districts. Average for men and women over the years 2012–2016. Source: Own elaboration based on data from Czech Statistical Office.

A. Outpatient doctors
(per 1,000 inhabitants, 2014)

B. Hospital beds
(per 1,000 inhabitants, 2014)

C. Tertiary education
(share of population aged 15+, 2011)

D. Unemployment
(2016)

E. Air pollution (particulate matters < 2.5 μm)
(yearly average μg/m³, 2012–2016)

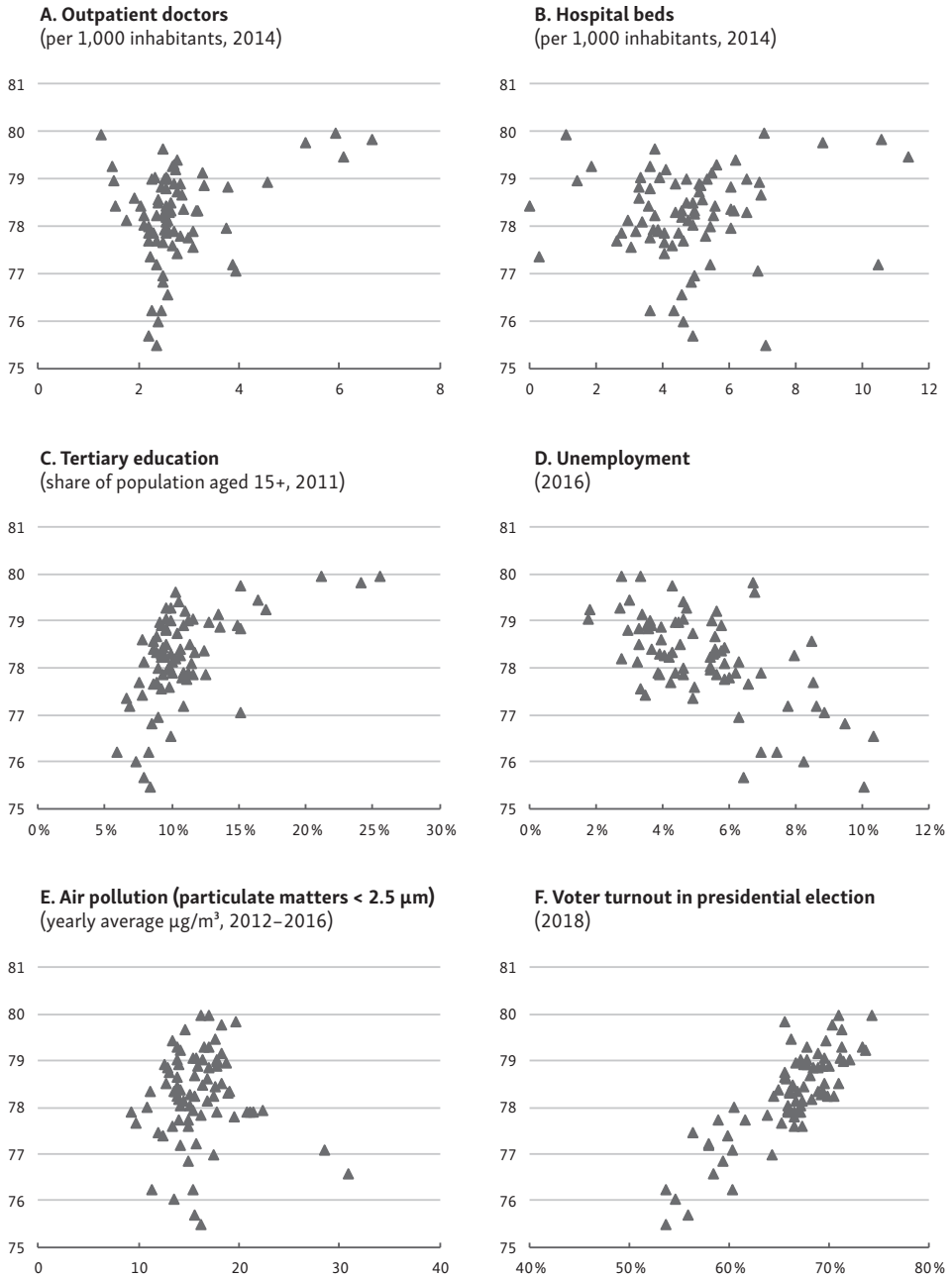**F. Voter turnout in presidential election**
(2018)

Fig. 3 – Life expectancy at birth (2012–2016) by various factors.
Source: Own elaboration based on data from Czech Statistical Office.

explanatory variables. Rather, it is assumed that the explanatory variables capture common underlying factors, which influence health outcomes (see e.g., Sharma 1996, for a detailed description of factor analysis methodology). Put more formally, factor analysis finds a few common variables q that reconstruct linearly the p original variables:

$$x_{ij} = f_{i1} \lambda_{1j} + f_{i2} \lambda_{2j} + \cdots + f_{iq} \lambda_{qj} + e_{ij},$$
$$\text{with } i = 1, ..., n; j = 1, ..., p; k = 1, ..., q,$$

where x_{ij} is the value of the i-th observation on the j-th variable, f_{ik} is the i-th observation on the k-th common factor, λ_{kj} is the set of linear coefficients called the factor loadings, and residual e_{ij} is known as the j-th variable's unique factor. The standardised vector $\tilde{x}$ (mean 0, variance 1) of observed variables can then be described by the following system of regressions:

$$\tilde{x} = F\Lambda' + e,$$

where F is a $n \times q$ matrix of factors, Λ represents the $p \times q$ factor loading matrix and e is a vector $n \times p$ of errors with diagonal covariance equal to the $p \times p$ uniqueness matrix Ψ. The common factors F and the error vector e are assumed to be uncorrelated. The estimation of factor loadings uses the eigenvalue structure of correlation matrices, similar to the principal component analysis. Under the factor model, the correlation matrix of $\tilde{x}$, denoted as Σ, is decomposed as:

$$\Sigma = \Lambda\Phi\Lambda' + \Psi$$

As such, the factor loading solution is not unique and there are several factoring and rotation methods. In the following analysis, the principal factors and Varimax rotation is used. It is therefore assumed that the common factors are uncorrelated such that $\Phi = I$. The orthogonal rotation gives a new loading matrix with $\Lambda^* = \Lambda C$ and $\Sigma = \Lambda^*\Lambda^{*'}$ by identifying the transformation matrix C. The varimax rotation determining the matrix C maximises the variance of squared loadings. Further, the number of factors is chosen based on the standard Kaiser's criterion, i.e. the factors with an eigenvalue greater than 1. The $n \times q$ matrix of estimated factor scores $\hat{F}$ for all observations n can then be expressed as:

$$\hat{F} = \tilde{x}\Sigma^{-1}\Lambda^*$$

In a final step, the factor scores are regressed on the dependent variable (y), with v representing a normally distributed residual term:

$$y = \hat{F}\gamma + v$$

In the following analysis, the dependent variable will reflect regional health outcomes, namely life expectancy, whereas the observable regional indicators are used to capture underlying factors that determine health inequalities.

4.2. Data

The dataset brings together spatially disaggregated information for the 77 districts of Czechia (and also for 206 administrative districts, if available). Indicators include lifestyle and environmental factors, socio-economic characteristics and medical infrastructure. Bringing these indicators together provides a basis to assess the underlying processes that determine regional differences in health outcomes.

To identify variations in health outcomes across regions, indicators that proxy the average health status of the population have to be identified. In the following analysis, life expectancy at birth (from the Czech Statistical Office, hereafter CZSO) is considered as the main measure for health outcomes. One of the limitations of this measure is that it disregards the share of life spent in bad health, which can also differ by location. In this respect, healthy life expectancy would be a more appropriate indicator, however it is generally not available on the required regional level. Other health outcome measures used in the literature could be considered, such as infant or amenable mortality. Infant mortality in Czechia is generally low (OECD 2018). Thus, on regional level, numbers are not sufficient to conduct an empirical assessment. On the contrary, regional observations for life expectancy at birth are sufficient and well accessible. Moreover, disaggregation by gender allows for additional robustness checks. A further measure that will be used to proxy the health status of the population is life expectancy at the age of 65, which in addition to life expectancy at birth is one of the most widely available and used summary measures (Joumard, André, Nicq 2010).

Health outcomes are usually linked to a variety of determinants, capturing socio-economic, lifestyle and environmental conditions, as well as medical infrastructure and quality. To capture potential determinants, following indicators were used:

Socio-economic situation: The health status of the population is often found to be closely linked to socio-economic characteristics (Pickett, Wilkinson 2015). For example, higher income is closely linked to higher educational level and a lower risk of unemployment. Furthermore, individuals with a better socio-economic status tend to face fewer constraints regarding (financial) access to health care, have higher health literacy and better access to relevant health information. These characteristics are highly interrelated, though their effect on health can be examined separately, such as for the education (Van Baal et al. 2016, Sasson 2016). Moreover, all these variables are likely to be correlated with the urbanisation of the respective region (Pantyley 2017). To capture these dependencies, following indicators are included (if available, values for 2016 are used for all indicators):
- Average unemployment rate in the district and the share of population with tertiary education (both CZSO).

– Spatial segregation of minority populations that face disadvantages in living conditions and access to relevant infrastructure can be a main driver of regional variation. The share of inhabitants living in social exclusion – defined as locality where more than 20% of persons are living in inadequate conditions and are inhabiting a physically or symbolically delimited space – is considered. This information is based on estimates by Čada et al. (2015) and provided by the Ministry of Labour and Social Affairs.
– Voter turnout (for the last presidential election) is a common measure for civic engagement and governance (OECD 2011) and is used as proxy for overall social inclusion. Presidential election of 2018 is used in order to avoid a potential effect of different candidates and parties across regions (CZSO), although it creates a slight time inconsistency.
– Population density is included as a control variable capturing urbanisation levels (CZSO).
– From the Register of Economic Subjects operated by the Czech Statistical Office, information on the business structure is gathered, measured as the shares of firms by main specialisations. More specifically, the shares of firms specialised in mining sector (NACE B) as a measure for health-risky occupations and the share of firms in the knowledge sector (NACE J-O, R) are considered.

Environmental factors: Quality of life factors, such as access to clean water and air, but also social environmental factors, e.g. liveable communities, are crucial for health outcomes of the population. Environmental risk factors can both directly affect health, i.e. through environmental pollutants, as well as indirectly through a lower quality of life (WHO 2012; Samoli et al. 2016; Rapant, Krčmová, 2007).
– Air pollution is measured by the concentration of SO_2, and particulate matters with a diameter either between 10 and 2.5 μm or smaller than 2.5 μm. All data are gathered from maps provided by the Czech Hydro-meteorological Institute.
– Safety and deterioration of the social environment are captured by the number of crimes per population (CZSO).

Medical infrastructure: Access to quality health care is crucial for explaining health outcomes. As there is a lack of health care quality indicators in Czechia, this study uses quantitative indicators that try to proxy the potential access to primary and emergency care.
– The number of doctors (GPs and specialists) and number of hospital beds per population indicates the availability of medical infrastructure and the ease of access to care. CZSO provides these data only on the level of 77 districts. For further detail, we included also number gathered from the statistics of the Institute of Health Information and Statistics (ÚZIS). Similarly, the number of pharmacies per population shows a better access to medicaments.

– Average travel distance to the nearest hospital is calculated as an additional accessibility measure. It is a narrow distance calculated from the localization of hospitals as recorded by the Institute of Health Information and Statistics to every municipality and averaged for the district.

Lifestyle factors: Adverse lifestyle factors, such as the consumption of alcohol and tobacco, an unhealthy diet and insufficient physical exercise, are closely linked to certain types of (chronic) diseases adding to a higher premature mortality rate (Lee et al. 2012, Loring 2014, Yokota et al. 2018). To proxy individual risky behaviour across districts, following indicators are constructed:
– To proxy alcohol consumption, hospitalization rates for alcohol related disorders (diagnosis F10) and hospitalization and mortality rates for alcoholic liver diseases (diagnosis K70) are used (i.e. hospitalizations or deaths for the respective diagnosis per 1,000 inhabitants aged 15–64). As these statistics are relatively volatile at the regional level, we have used 6-years averages (2011–2016) as recorded by the Institute of Health Information and Statistics.
– Information on gambling as another type of risky behaviour is measured by the number of casinos and bars with slot machines per population (provided by the Ministry of Finance).

5. Empirical results

Applying the factor analysis to the dataset as described above, five relevant factors are identified (Table 1). The first factor is generally composed by density of doctors and medical facilities, therefore categorised as "health care accessibility". The second factor is mostly saturated by knowledge-intensive business specialisation, population density, the share of population with tertiary education and also number of crimes per population. As these characteristics are associated with the degree of urbanisation, this factor is categorised as "urbanisation". The third factor is categorised as "socio-economic deprivation" as it is negatively connected to voter turnout, and positively to unemployment, the share of population living in social exclusion and gambling. The fourth factor "alcoholic consumption" is comprised by a combination of hospitalisation and mortality rates for alcohol-related diseases, whereas the last factor "environmental pollution" is capturing the measures for air pollutants. The identified factors reflect the main categories described above and are robust for various specifications.

Regarding the assumptions of the factor analysis, both the sample size and correlations between the variables are appropriate. Kaiser-Meyer-Olkin test of sampling adequacy gives for all specifications the value between 0.7 and 0.8 and Bartlett's test of sphericity confirms the intercorrelation between the variables

Table 1 – Final rotated factor loadings, 77 districts

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
Hospital doctors per pop.	*0.90*	0.28	0.05	0.02	0.12	0.09
Hospital beds per pop.	*0.88*	0.18	0.15	0.00	0.04	0.12
Outpatient doctors per pop.	*0.82*	0.47	–0.05	0.00	0.01	0.06
Dentists per pop.	*0.70*	0.31	–0.19	0.18	–0.08	0.19
Pharmacies per pop.	0.32	–0.07	–0.07	0.07	0.06	0.62
Knowledge specialization	0.32	*0.82*	0.23	–0.02	0.10	0.10
Population density	0.34	*0.81*	0.00	0.05	0.07	0.15
Tertiary education	0.30	*0.76*	–0.45	–0.02	0.17	0.07
Crimes per pop.	0.30	*0.76*	0.39	0.04	0.03	0.12
Distance to nearest hospital	–0.21	–0.46	0.00	–0.12	–0.30	0.42
Voter turnout	0.01	–0.04	*–0.95*	–0.14	0.00	0.07
Social exclusion	0.04	0.13	*0.67*	0.13	0.04	0.38
Unemployment	0.12	–0.01	*0.60*	0.43	0.12	0.30
Gambling (casinos per pop.)	0.11	–0.01	*0.56*	0.11	0.05	0.42
Share of mining firms	–0.10	0.01	0.43	0.07	–0.24	0.58
Mortality K70 per pop.	0.01	0.04	0.19	*0.86*	0.13	0.18
Hospitalizations K70 per pop.	0.02	0.03	0.10	*0.84*	0.17	0.23
Hospitalizations F10 per pop.	0.07	–0.08	0.17	*0.74*	0.11	0.32
Air pollution – PM2.5	0.09	0.22	–0.04	0.42	*0.75*	0.19
Air pollution – SO2	0.04	0.20	0.57	0.40	*0.58*	0.08

Note: Numbers in italic gray highlight the greatest loadings for each factor. It shows which indicators are good measures of a given factor. A varimax rotation has been used.
Source: Own analysis. Based on data from the Czech Statistical Office, Institute of Health Information and Statistics (ÚZIS), Czech Hydrometeorological Institute, Ministry of Finance and Ministry of Labour and Social Affairs.

and thus the adequacy of the factor analysis. From the correlation matrix, the nearly unitary correlation is between air pollution measured by particular matters in the air of two different sizes. Therefore, we have employed only one of them, the particulate matter with a diameter smaller than 2.5 μm.

Using the identified factors to explain variations in life expectancy, the regression results (Table 2) show strong and unambiguous effects. The estimates are interpreted as deviations from the average life expectancy across all districts (which is captured by the regression constant). Thus, the estimation results show that urban regions are slightly better off. The most urbanised area is Prague, where life expectancy is approximately nine months higher than the average just because of urban amenities (see the note below the Table 2). Better work opportunities that positively affect quality of live may be one explanation. Health care accessibility, reflecting the density of doctors, also causes a difference in life expectancy of about 1.4 years between the district with the lowest and highest accessibility. The main driver of regional variation in health outcomes is, however, linked to socio-economic conditions. The most deprived districts are estimated to have a 2.6 years lower life expectancy than the average and 4.0 years lower than the best performing district.

Table 2 – Regression results after factor analysis for 77 former districts

	Min	Max	LE at birth	LE at 65	LE (fem) at birth	LE (m) at birth
Health care accessibility	–2.77 (Prague-West)	3.35 (Pilsen-City)	0.23 *** (0.05)	0.23 *** (0.04)	0.23 *** (0.06)	0.23 *** (0.06)
Urbanisation	–1.26 (Svitavy)	5.62 (Prague-City)	0.13 ** (0.05)	0.14 *** (0.04)	0.08 (0.06)	0.18 ** (0.06)
Socioeconomic deprivation	–1.77 (Prague-West)	3.11 (Most)	–0.82 *** (0.05)	–0.47 *** (0.04)	–0.80 *** (0.06)	–0.84 *** (0.06)
Alcoholic consumption	–1.33 (Č. Budějovice)	3.12 (Vsetín)	–0.20 *** (0.05)	0.06 (0.04)	0.04 (0.06)	–0.44 *** (0.06)
Environmental pollution	–1.83 (Jeseník)	3.00 (Karviná)	–0.06 (0.06)	–0.02 (0.05)	–0.06 (0.06)	–0.06 (0.06)
Constant (Average life expectancy)			78.23 *** (0.05)	17.71 *** (0.04)	81.23 *** (0.05)	75.38 *** (0.06)
Adjusted R sq.			0.81	0.68	0.75	0.79

Note: Statistical significance notation: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Standard errors in parentheses. Dependent variable is always life expectancy, but it differs in separate columns (for separate regression models): it is life expectancy at birth, at 65, and then at birth separately for females and males in the last column. Independent variables are the factors obtained from the factor analysis. Minimum and maximum values express the range for every factor (computed for the whole sample, i.e. for both sexes), while the mean value is always 0. It measures a deviation from average and as such it is comparable between different factors. For example, the life expectancy at birth in Prague City is estimated as 5.62 × 0.13 = 0.75 years higher than the average, which is about 9 months.
Source: Own analysis.

Environmental pollution also lowers health outcomes, though only marginally and the results are statistically insignificant. On the other side, alcoholic consumption shows a clear downgrading effect, particularly in the case of men.

There is also an interesting gender difference in the urbanisation effect. It seems that for females, the city amenities do not bring such an advantage, or they are offset by negative aspects connected to urbanisation. One has to take into account that these effects are separated from the socio-economic aspects, which certainly are correlated with the urban-rural pattern.

5.1. Robustness checks

To assess the need of the factor analysis model, first a simple regression was conducted. Testing for collinearity indeed indicated multicollinearity of several variables. As such, factor analysis as alternative to determine underlying processes is the preferred approach. Within the factor analysis, the results have been tested against different specifications. The main outcomes were consistent across alternative specifications. However, using the life expectancy at the age of 65 as an alternative measure for health outcomes, the relative importance of factors

Table 3 – Final rotated factor loadings, 206 administrative districts

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
Voter turnout	*−0.91*	0.03	−0.09	−0.04	0.12
Social exclusion	*0.79*	0.09	0.12	−0.06	0.30
Unemployment	*0.62*	−0.06	0.30	−0.08	0.43
Gambling (casinos per pop.)	*0.41*	0.08	0.01	0.19	0.60
Share of mining firms	0.25	0.06	−0.10	−0.24	0.79
Knowledge specialization	0.31	*0.82*	0.08	0.04	0.18
Tertiary education	−0.41	*0.81*	0.17	0.02	0.14
Population density	0.15	*0.68*	0.23	0.08	0.41
Dentists per pop.	−0.08	*0.43*	−0.03	0.29	0.55
Air pollution – PM2.5	0.00	0.15	*0.85*	−0.02	0.24
Air pollution – SO2	0.49	0.16	*0.76*	−0.04	0.15
Distance to nearest hospital	−0.09	−0.28	−0.29	0.17	0.71
Pediatricians per pop.	−0.06	−0.16	−0.19	0.01	0.73
Specialists per pop.	−0.02	0.36	−0.12	*0.57*	0.51
Pharmacies per pop.	0.03	0.00	−0.04	*0.49*	0.75
GPs per pop.	−0.08	−0.13	−0.19	0.20	0.72

Note: Data for number of doctors per population differ by their source from data for 77 districts presented in Table 1, thus these loadings are not directly comparable. A varimax rotation has been used.
Source: Own analysis. Based on data from the Czech Statistical Office, Institute of Health Information and Statistics (ÚZIS), Czech Hydrometeorological Institute, Ministry of Finance and Ministry of Labour and Social Affairs.

slightly changes in favour of health care accessibility and urbanisation. Socio-economic conditions remain the most important factor, but the difference to other factors is not as big as for the life expectancy at birth.

The estimation results may be biased due to omitted variables. Not all indicators that would influence health outcomes are available on a regional disaggregated level. For example, potential differences across districts in the quality of care are unobservable due to a lack of quality and performance indicators for the medical sector. Sorting of better skilled medical personnel into more attractive regions (i.e. with less crime and lower air pollution) could for example create differences in the quality of treatment. However, to identify potential sorting and quality differences and their effect on health outcomes requires better data that are currently not available. While such a bias might occur, it can be expected that the effect is captured to some degree through the factors describing social deprivation, urbanisation and access to health care.

As a further robustness check, the analysis was extended to the 206 administrative districts around municipalities with extended powers as the 77 former districts are often considered as too big to capture the most pronounced regional heterogeneity. However, some data are not available at the administrative district level, namely alcohol consumption and doctor density as recorded by the Czech Statistical Office. For the robustness check, the latter is proxied by numbers

Table 4 – Regression results, 206 administrative districts

	Min	Max	LE at birth		LE at 65		LE (fem) at birth		LE (m) at birth	
Health care accessibility	−1.80	3.58	0.12 (0.06)	.	0.15 (0.05)	**	0.21 (0.08)	**	0.01 (0.08)	
Urbanisation	−1.53	6.21	0.30 (0.05)	***	0.24 (0.04)	***	0.20 (0.06)	**	0.40 (0.06)	***
Socioeconomic deprivation	−1.60	4.18	−0.88 (0.05)	***	−0.49 (0.04)	***	−0.83 (0.06)	***	−0.92 (0.06)	***
Environmental pollution	−2.06	3.74	−0.17 (0.05)	**	−0.01 (0.04)		−0.02 (0.06)		−0.33 (0.06)	***
Constant (Average life expectancy)			78.14 (0.04)	***	17.67 (0.04)	***	81.19 (0.05)	***	75.24 (0.06)	***
Adjusted R sq.			0.66		0.47		0.52		0.61	

Note: Statistical significance notation: . $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Standard errors in parentheses. Source: Own analysis.

obtained from the National Register of Health-Care Providers. The factor analysis based on the 206 administrative districts results in four factors, as data on alcoholic consumption is missing.

Using the four factors, the regression shows consistent results to the main analysis. The overall explanatory power is lower, as we are missing data on alcoholic consumption and using less reliable measures on health care infrastructure. Therefore, results for the factor called "health care accessibility" (Factor 4 in Table 3) are not directly comparable with the results for 77 districts and the factor loadings provide a less clear interpretation. However, we can highlight that the socio-economic deprivation remains as the most important factor and it is even more pronounced than using bigger districts. On the other hand, urbanisation factor seems to be less relevant. The missing information on adverse behaviour (alcohol consumption) is likely to affect the results for environmental pollution – including the gender asymmetry (Table 4).

6. Conclusions and policy implications

The estimation results provide a first comprehensive assessment of determinants that affect regional variation in health outcomes in Czechia. The results are in line with previous studies suggesting a high dependency of health on income and other socio-economic characteristics (Pickett, Wilkinson 2015; Blažek, Dzúrová 2000). Further, the results in this study highlight that structural differences across Czech districts are linked to the location of vulnerable population groups, which require a holistic approach for policy solutions.

The analysis shows that the effect of socio-economic deprivation is strongest in explaining regional differences in life expectancy. This is in line with studies that suggest that social status, poor living conditions and risky behaviour related to health are main drivers of morbidity and inferior health status (Úřad vlády ČR 2017; Rodriguez, Derecho 2009). As the socio-economic deprivation factor captures, among others, the share of population living in inadequate conditions, unemployment, and the degree of civic engagement, policy actions to reduce inequalities in health outcomes could focus on improving liveability in communities, ensuring the availability of affordable quality housing and providing sufficient access to services. Moreover, it is possible to separate the effect of living in cities, which is also connected to the socio-economic aspect. From the estimation results, living in the city improves life expectancy, which can be caused by various urban amenities, but more importantly by job characteristics and by generally better education. Considering the gender differences in the results, men are taking more advantages from living in the city than women, working in significantly more favourable environment than in peripheries.

Although the estimates of socially deprived localities do not exclusively capture minority populations, in 3 out of 4 these localities the Roma population constitutes the majority (Čada 2015). As such, ethnic background and discrimination could play a role in observing health inequalities. Indeed, some studies estimate that the Roma face significant health inequalities with a significantly lower life expectancy (up to 15 years) and an infant mortality rate that is twice the national average (IOM 2016; Úřad vlády ČR 2017). While the Office of the Government supports programmes focused on reducing health inequalities in socially excluded areas with a high share of Roma population, there remains a lack of systematic support to train and increase the awareness of health and social assistants with respect to social, cultural and language barriers of minority groups (Úřad vlády ČR 2017). Targeted policies at minority groups should include preventive and educational programmes, in combination with health providers and social assistants being sensitised for minority population's risk factors and specific needs.

Moreover, substance abuse and behavioural risk factors, which have severe health consequences, were found to be predominant in socially deprived localities (Čada 2015). But also more generally, alcoholic consumption further explains some of the variation in life expectancy, especially in the case of men. Thus, access to preventive programmes can play a crucial role in reducing health inequalities. For example, a study provided by Sassi (2015) based on estimates for 2010 suggests that alcohol preventive programmes could significantly reduce deaths from chronic diseases and injuries: a programme of brief interventions by primary care physicians, aimed at people drinking too much, is estimated to lead to a total annual gain in the Czech population of 26,000 years of life in good health. Cost-benefit

estimations show that prevention has the potential to cut health expenditures for chronic diseases and injuries by up to USD PPPs 8 per person per year in Czechia (Sassi 2015). As alcohol consumption in Czechia tends to be higher in districts with worse environmental conditions, in addition to preventive programmes, the alleviation of environmental burden of the most polluted areas could be an important support for healthy lifestyle – although the direct effect is smaller than the one for alcoholic consumption in the presented results.

Access to medical care is crucial for improving the health outcomes of the population. While the density of doctors is high across Czechia, differences explain some variation in life expectancy across districts. As highlighted in OECD (2018), rural areas are facing some challenges to attract sufficient doctors to compensate the effect of an ageing society and doctors' workforce. Already, some first programmes to incentivise doctors setting up practices in underserved areas have been initiated. However, rather than solely focusing on improving the quantity, a focus should also be put on equal access to quality health care. Unfortunately, data capturing the quality of care are not yet available and remain subject for further research.

References

BLAŽEK, J., DZÚROVÁ, D. (2000): The decline of mortality in the Czech Republic during the transition: a counterfactual case study. In: Cornia, G.A., Paniccià, R. (eds.): The mortality crisis in transitional economies. Oxford University Press, Oxford, 303–327.

ČADA, K., ed. (2015): Analysis of socially excluded localities in the Czech Republic. Prague, European Social Fund in the Czech Republic.

CHETTY, R., STEPNER, M., ABRAHAM, S., LIN, S., SCUDERI, B., TURNER, N., BERGERON, A., CUTLER, D. (2016). The association between income and life expectancy in the United States, 2001–2014. JAMA, 315, 16, 1750–1766.

CSDH (2008): Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. World Health Organization, Geneva.

CUNDIFF, J., MATTHEWS, K. (2017): Is subjective social status a unique correlate of physical health? A meta-analysis. Health Psychology, 36, 12, 1109–1125.

DZÚROVÁ, D. (2000): Mortality differentials in the Czech Republic during the post-1989 socio–political transformation. Health & Place, 6, 4, 351–362.

IOM (2016): Implementation of the National Roma Integration Strategy and Other National Commitments in the Field of Health, Czech Republic, A multi-stakeholder perspective report on 2005–2014 developments. International Organisation of Migration, Brussels.

JOUMARD, I., ANDRÉ, C., NICQ, C. (2010): Health Care Systems: Efficiency and Institutions. OECD Economics Department Working Papers, No. 769, OECD Publishing, Paris.

LEE, I., SHIROMA, E.J., LOBELO, F., PUSKA, P., BLAIR, S.N., KATZMARZYK, P.T. (2012): Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. The Lancet, 380, 9838, 219–229.

LEON, D. (2011): Trends in European life expectancy: a salutary view. International Journal of Epidemiology, 40, 2, 271–277.

LORING, B. (2014): Alcohol and inequities: guidance for addressing inequities in alcohol-related harm. World Health Organization, Regional Office for Europe, Copenhagen.

MACKENBACH, J. (2012): The persistence of health inequalities in modern welfare states: The explanation of a paradox. Social Science & Medicine, 75, 4, 761–769.

OECD (2011): Compendium of OECD Well-Being Indicators. OECD Publishing, Paris.

OECD (2018): OECD Economic Surveys: Czech Republic 2018. OECD Publishing, Paris.

PANTYLEY, V. (2017): Health inequalities among rural and urban population of Eastern Poland in the context of sustainable development. Annals of Agricultural and Environmental Medicine, 24, 3, 477–483.

PICKETT, K., WILKINSON, R. (2015): Income inequality and health: A causal review. Social Science & Medicine, 128, 316–326.

PRÄG, P., MILLS, M., WITTEK, R. (2016): Subjective socioeconomic status and health in cross-national comparison. Social Science & Medicine, 149, 84–92.

RAPANT, S., KRČMOVÁ, K. (2007): Health risk assessment maps for arsenic groundwater content: application of national geochemical databases. Environmental Geochemistry and Health, 29, 2, 131–141.

RICHARDSON, E.A., PEARCE, J., MITCHELL, R., SHORTT, N.K., TUNSTALL, H. (2014): Have regional inequalities in life expectancy widened within the European Union between 1991 and 2008? The European Journal of Public Health, 24, 3, 357–363.

RODRIGUEZ, S.N., DERECHO, N.R., eds. (2009): Health and the Roma Community, analysis of the situation in Europe: Bulgaria, Czech Republic, Greece, Portugal, Romania, Slovakia, Spain. Fundación Secretariado Gitano, Madrid.

SAMOLI, E., ATKINSON, R.W., ANALITIS, A., FULLER, G.W., BEDDOWS, D., GREEN, D.C., MUDWAY, I.S., HARRISON, R.M., ANDERSON, H.R., KELLY, F.J. (2016): Differential health effects of short-term exposure to source-specific particles in London, UK. Environment international, 97, 246–253.

SASSI, F., ed. (2015): Tackling Harmful Alcohol Use: Economics and Public Health Policy. OECD Publishing, Paris.

SASSON, I. (2016). Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010. Demography, 53, 2, 269–293.

SHARMA, S. (1996): Applied Multivariate Techniques. John Wiley & Sons, New York.

THOMSON, K.H., RENNEBERG, A.-C., MCNAMARA C.L., AKHTER, N., REIBLING, N., BAMBRA, C. (2017): Regional inequalities in self-reported conditions and non-communicable diseases in European countries: Findings from the European Social Survey (2014) special module on the social determinants of health. European Journal of Public Health, 27, 1, 14–21.

Úřad vlády ČR (2017): Zpráva o stavu romské menšiny v České republice za rok 2016. Office of the Government of the Czech Republic, Prague.

VÅGERÖ, D. (2010): The East–West Health Divide in Europe: Growing and Shifting Eastwards. European Review, 18, 01, 23–34.

VAN BAAL, P., PETERS, F., MACKENBACH, J., NUSSELDER, W. (2016). Forecasting differences in life expectancy by education. Population studies, 70, 2, 201–216.

WHO (2012): Environmental health inequalities in Europe. World Health Organization, Regional Office for Europe, Copenhagen.

WHO (2013): Health 2020 – A European policy framework and strategy for the 21st century. World Health Organization, Regional Office for Europe, Copenhagen.

YOKOTA, R.T., NUSSELDER, W.J., ROBINE, J.M., TAFFOREAU, J., CHARAFEDDINE, R., GISLE, L., DEBOOSERE, P., VAN OYEN, H. (2018): Contribution of chronic conditions to smoking differences in life expectancy with and without disability in Belgium. European journal of public health, 28, 5, 859–863.