Regional socio-economic factors influencing diabetes incidence: the case of Romania

Elena DRUICĂ*, Zizi GOSCHIN**, Cristian BĂICUȘ***

Abstract

We examine the relation between average net wage, urbanization rate, women density, life expectancy, medical infrastructure and medical human resources, and the incidence of total, insulin, and non – insulin diabetes among Romanians. We fitted three panel regression models with interaction terms using official data comprising of 41 Romanian counties analysed between 2007 and 2014. After controlling for age groups, we found that the share of women in the overall population moderates the influence of salary level on diabetes incidence for the total and non – insulin groups, while for the insulin – dependent category, urbanization rate was positively associated with the number of newly recorded patients. Health infrastructure was relevant only for the total, and the insulin – dependent categories. Our results are in line with the sizeable disparities in diabetes that exist within other European countries and are useful for regional decision-makers planning adequate healthcare services and target proper risk groups.

Keywords: diabetes, regional studies, Romania, healthcare, panel data models

Introduction

The diabetes burden across the world captured the academic interest in the field of health long ago. Usually, the concern is driven by the costs associated with the disease (Colagiuri *et al.* 2003; ADA, 2003; Dall *et al.*, 2010), or the complications deriving from this condition (Williams *et al.*, 2002; Jonsson, 2002; Gregg *et al.*, 2016), and eventually targets the idea of proper interventions designed by policy makers and health care authorities meant to prevent the disease (Tol *et al.*, 2013). Patient level studies as well as regional or nation – wide studies have been conducted in response to the steep increase in the incidence and prevalence of the

---

*Elena DRUICĂ is professor at the University of Bucharest, Romania; email: elena.druica@faa.unibuc.ro.
** Zizi GOSCHIN is professor at the University of Economic Studies Bucharest, Romania; email: zizi.goschin@csie.ase.ro.
*** Cristian BĂICUȘ is professor at the Carol Davila University of Pharmacy and Medicine, Bucharest, Romania; email: cbaicus@gmail.com.

disease, as well as a consequence of the significant complications, high costs and mortality deriving from it. Current trends are alarming in Europe as well, e.g. the prediction of 70% increase in young population cases in 2020 compared to 2005 (Patterson *et al.,* 2009). The first national study on diabetes and pre-diabetes disease prevalence revealed that in 2014, 68807 Romanians (representing 11.6% of the population aged 20 to 70) suffered from diabetes, the figures further increasing to 73740 individuals in 2015. The European Health Interview Survey reports that one in twenty Romanians live with diabetes, a figure that, although is still below the European average of 6.9%, is on the rise.

The issue can be discussed in relation to aging population, or a health system that underperforms, but additional circumstances like increased life expectancy, less healthier lifestyle leading to stress and obesity (Mărăcine and Ianole, 2010), to name only a few, may also play an important role in the overall distribution of diabetes. Given the importance of regional studies in planning local, specific approaches to prevention and healthcare (Tamayo *et al.,* 2013), and considering the increasing burden that diabetes places on the public health budget, this study puts on the spotlight several regional factors that can contribute to a better understanding of diabetes distribution in Romania. We explored several variables capturing the economic well-being, demography, as well as health infrastructure and medical human resources, likely to have an impact on the incidence of both insulin dependent and insulin independent diabetes across the country.

The importance of the study is twofold: on the one side, it is one of the very few regional studies that discuss diabetes incidence in Romania. Bringing clarifications regarding significant factors that may explain the distribution of the condition may help identifying specific risk groups and provide ground for tailored interventions. On another side, according to the International Diabetes Federation, the number of the EU citizens affected by diabetes was 33 million in 2010 and it is projected to increase to 38 million until 2030. Adding Romania on the list of the European countries that conduct specific investigations on diabetes is in line with the European Union initiative to support research in this area in an attempt to reverse the trend.

The paper is organized as follows: next section will briefly present the main literature that deals with factors responsible for diabetes incidence and prevalence by type. Section 2 introduces the data and the method while the empirical results section presents the main findings and discusses the implications. In the concluding section we summarize the results.

1. Literature review and variables selection

In this section we first clarify the concepts and set several working definitions. Secondly, we review the literature concerning factors that are associated with diabetes incidence, select the main predictors in our study, and suggest the preliminary functional form of the models. We searched Medline for articles published between 1966 and January 2017 with no language restriction, using the keywords “diabetes” and “socio - economic”. As the socio-economic factors documented in the literature became available, we continued our search using each of them as an independent keyword.

Diabetes mellitus is a group of chronic metabolic diseases characterized by high blood sugar levels (American Diabetes Association, 2009). There are two main types of diabetes: type 1 diabetes mellitus is an autoimmune disease resulting in pancreatic cell destruction, with the failure of pancreas to produce enough insulin; and type 2 diabetes mellitus, in which insulin resistance and deficiency co-exist (Cerasi, 1991). The first type of diabetes shows up early in life (juvenile-onset), and is labeled insulin-dependent, because its treatment consists of insulin. Type II diabetes, occurring in adulthood, is known as non-insulin-dependent because it is based on other medications (Reinehr, 2013). Nevertheless, insulin is often used as a last resort. This type of diabetes is generated by both genetic and environmental factors, the most important risk factor being obesity (Murea *et al.*, 2012).

Diabetes mellitus leads to pathologic changes involving blood vessels, nerves, skin, and eyes. Consequently, chronic complications are serious: end-stage chronic kidney disease, blindness, neuropathy, amputations of the lower extremities, myocardial infarction, and stroke (Cade, 2008; Deshpande *et al.*, 2008; Long *et al.*, 2011).

There is a considerable amount of research in what concerns the socio – economic and other related factors having an impact on the emergence and prevalence of diabetes. After conducting a cross – sectional study on a sample of African Americans and White non-Hispanics, Robbins et al. found that socio-economic status displays an important association with “type II” diabetes prevalence among women, while a similar association is less significant among men. The same authors found that poverty plays a far more important role in explaining the occurrence of the disease than other variables like education or occupational status. Another important result provided by this study is that the association between diabetes prevalence and the mentioned variables are in many cases independent of other risk factors (Robbins *et al.*, 2001).

A similar strong and direct association between low economic status and diabetes incidence has been found in a cross – sectional study conducted by Larrañaga et al. involving over 65000 Spanish patients (Larrañaga *et al.*, 2005). This study stresses the idea that in assessing the risk profile of a patient, it is important to account for socio – economic variables, as low income settings are more likely to be characterized by higher prevalence of the disease. On the other side, a study conducted by Moradi *et al.* in Kurdistan found that diabetes prevalence shifted from low-income people to people with higher socio-economic status (Moradi *et al.*, 2016). The relation with other chronic diseases listed as complications or comorbidities of diabetes have also been discussed in relation to the socio-economic status. For example, education, income and occupation are strongly associated with cardiovascular diseases (Winkleby, 1992), and of these three determinants, Leigh (1993) found that education is by far a stronger predictor than the other two. The results obtained by Metcalf *et al.* (2008) support the previous findings.

While looking at similar determinants and finding similar associations to the ones found in the rest of the studies, Nordahl (2014) emphasizes the idea that not only the predictors as such play an important role in explaining the incidence of chronic diseases, but also the possible interactions among them. Tol *et al.* (2013) point toward the role that some variables play in moderating the relationship between income and diabetes incidence and finds that women in low income settings as well as life expectancy in the same low income context result in higher risk of diabetes. Lee *et al.* (2013) also found that socio-economic status is a good predictor of Type II diabetes, and that there are interactions with gender.

Other authors reported as well that low economic status, along with gender, lack of education, retirement, unemployed, and urban residence are good predictors of an increased diabetes prevalence (Peykari *et al.*, 2015), allowing for another potential predictor in our regional study, that is, urbanization rate. Last, but not least, demographic variables like age and gender structure of the population (Black 2002; Hosler and Melink, 2003; Rivera *et al.*, 2015), as well as the role of the environment (Leeder *et al.*, 2004), proved to be important in diabetes incidence by type. Yet, some other studies advocate the idea that higher risks for type II diabetes do not depend on socio-economic status, but rather on genetic inheritance and other environmental factors that need to be investigated by case (Gaillard, 1997).

In selecting the predictors of the regional incidence of total, insulin dependent and insulin independent diabetes, we begin with the average salary as a measure of wealth. There are sizeable disparities in diabetes between and within European countries, the income level being positively associated with its prevalence (Tamayo *et al.*, 2013; Patterson *et al.*, 2001), therefore the average regional wage needs to be tested in our model. Another important variable that plays a major role in the evolution of chronic diseases is education (Smith *et al.*, 2015). Since the share of Romanian population by education level lacks from the official statistical data by county, we used instead life expectancy at birth, a variable that captures the outcome of a far more complex number of factors concerning lifestyle. Not only life expectancy directly reflects the health status of the population, regardless the level of healthcare expenditures (Lubitz *et al.*, 2003), but it is also closely related to income distribution (Wilkinson, 1992) and education (van Baal *et al.*, 2016). Better education has been proven to raise life expectancy by up to 30 percent (Meara *et al.*, 2008).

Given that previous research found considerable disparities in diabetes prevalence between women and men (e.g. Wannamethee *et al.*, 2012), the share of female in total regional population will be included in the model and we expect it to be significant. Environment-related predictors used in previous studies to explain diabetes incidence are unavailable, in particular those related to CO2 emissions or other variables standing for pollution issues, but we consider that they can be largely captured by the urbanization rate. We used two broad population age groups (0 – 14 and 15 – 79), recommended by the International Diabetes Federation studies based on the age groups where each type of diabetes mainly occurs, as control variables.

Based on the previously discussed literature review, we set four research hypotheses as follows:

H1: *After controlling for age groups, the share of women in the overall population moderates salaries influence on diabetes incidence*

H2: *After controlling for age groups, life expectancy moderates salaries influence on diabetes incidence*

H3: *After controlling for age groups, higher degrees of urbanization are negatively associated with diabetes incidence*

H4: *A better healthcare infrastructure is negatively associated with diabetes incidence*

Each hypothesis will be tested by category of diabetes, resulting in 12 hypotheses in total. Since two of the hypotheses involve moderation effects, we will work with models with interaction that will be presented in Section 2.3.

2. Data and method

2.1. Data

Our data covers a time span of 8 years, between 2007 and 2014, and the 41 Romanian counties, resulting in a balanced panel with 328 observations in total. Two Romanian providers of official statistical data, namely The National Institute of Statistic and The National Institute of Public Health, were the source of the information we used. The R software, version 3.4.3, and several dedicated packages were used in analyzing our data. The “gplots” package (Warnes *et al*., 2016) displayed the heterogeneity by county (see Annex 1). The “tseries” package[1] provided the tests for stationarity. The “plm” package (Croissant and Millo, 2008) is a dedicated package that helps fit panel models, while the “lmtest” package (Zeileis and Hothorn, 2008) was the essential tool for conducting diagnostic tests.

Our dependent variables convey absolute diabetes incidence, in three categories: total number of new cases, new cases of insulin-dependent, and new cases of non-insulin diabetes. Based on the studies presented in the previous section, we derived a list of variables that could be used to explain the variability in our predicted variables. Both the independent and dependent variables are presented and described in Annex 2. We double checked the stationarity in R and EViews9SV and

[1] Trapletti, A. and Hornik, K. (2017), tseries: Time Series Analysis and Computational Finance. R package version 0, pp.10-40.

found that all the variables included in our analysis proved to be stationary in log. We applied the Levin, Lin and Chu test and the Breitung test in the category of common root tests, and we applied the lm, Pesaran and Shin test, the Fisher- ADF test, theFisher – PP test and the Hadri test in the category of individual root tests.

Particular attention has been paid to the health–related variables. The National Institute of Statistics provides a large variety of data: number of medical personnel (family doctors, specialists, nurses), number of doctors by specialization, number of hospitals and medical offices where the diagnosis can be confirmed, etc., but not all these indicators are accessible at county level. The main problem with the available variables is that they are highly correlated (Annex 3). One way to overcome this drawback is to construct a composite index (CI) that groups our variables, capturing all the individual information.

Using statistical indices in research is a common practice. These indices are usually derived from two or more dimensions of the phenomenon, and the final composite index is obtained as a weighted combination of them (Saisana and Tarantola, 2002). Building a composite indicator is considered a delicate task, prone to pitfalls for two reasons: 1) relevant indicators that should be part of the final index are sometimes unavailable and proxies need to be used instead, and 2) the manner in which the components of the final index are combined together is often controversial (Mazziotta and Pareto, 2013). Despite these problems, composite indices are used on a large scale and their effectiveness was confirmed by a long practice (OECD, 2008).

The multidimensional nature of the healthcare infrastructure that we aim to gauge recommends the use of a CI. Besides the usual role of CI as a synthesis of the incorporated indicators, an additional reason is, in our case, the significant degree of correlation (between 0.454 and 0.853) among the initial variables. Given this multicollinearity issue, the only way to include the information contained in all variables in the regression model is by aggregating them. While high correlation among the variables included in composite indicators might be considered a limit (Salzman, 2003), it is generally accepted that it becomes a problem only when the number of variables is large (OECD, 2008), which is not our case. Due to limited data availability, we had to build the composite index based on only five indicators. In this process, we followed the standard guidelines set forth by the literature and especially the OECD methodology for building composite indicators (OECD, 2008). The first step is to analyze the relevance and suitability of the indicators, given that the strength of a CI emerges from the variables it incorporates. We did not reject any variable at this stage, aiming to retain as much information as possible.

A multivariate analysis was further conducted in order to reveal the structure of the data and to help select the appropriate weights for variable aggregation. To this end, we applied a principal components analysis (PCA), largely acknowledged as the most common method for identifying sub-groups and guiding the choice of weights (OECD, 2008). The solution indicated by PCA contains two principal components, accounting together for almost 84% of the total variation. The first component brings together four variables (public sector physicians per 1000 inhabitants, physician assistants and nurses per 1000 inhabitants, number of hospitals per 100,000 inhabitants and hospital beds per 1000 inhabitants), while the second one includes only one (private practice physicians per 1000 inhabitants). The weights recommended by PCA for the first component are quasi-equal, while the weighting scheme for aggregating our two components into the final CI is strongly unbalanced, with the first component getting 86%. We could not rely entirely on these results because there are some problems related to the application of PCA on our database. Firstly, the relatively low volume of our sample impedes the validity of Principal Component Analysis. A popular "rule of thumb" for PCA requires minimum 10 cases per variable, resulting 50 necessary cases, instead of the 41 available in our database. Secondly, PCA demands that the variables are evenly distributed among the components, while in our case all but one variable are included in the first component.

Consequently, we decided to adopt a slightly different grouping and weighting strategy. The literature states that weighting should be based not only on data properties, as revealed by PCA, but also on the theoretical framework (OECD, 2008). Weighting based on expert's opinions instead of statistical guidelines finds support both in literature (Munda, 2007) and practice (EC, 2004), and has been applied in building health care composite indices as well[2]; WHO, 2000). In line with such studies, and relying on theoretical considerations, especially on the recommendations of the medical experts from the National Institute of Public Health (our data provider), we grouped the variables according to their nature. We highlighted the two distinct dimensions of the data, namely Human Capital (including private and public sector physicians, physician assistants and nurses), and Material Infrastructure (referring to hospitals and hospital beds). This allowed us to give more strength (i.e. bigger weight) to human factor against material endowments. The resulting weights, 70% and 30% respectively, have been allotted considering the key importance of medical staff in making a quick and accurate diagnosis and in administrating the appropriate treatment, and are also grounded in previous theoretical research, stating that the weights should be more or less proportional with the components variances (Mazziotta and Pareto, 2013). Nevertheless, we applied an equal weighting scheme within each of these two groups, as suggested by our PCA analysis, and also accepted it as good practice when the indicators within a group are similar and the number of variables is low (OECD, 2008).

In order to allow aggregations, all variables have been previously transformed using the min-max method. This is a popular normalization procedure that rescales each variable relative to its extreme values (minimum and maximum), producing

[2] King's Fund - The sick list 2000, the NHS from best to worst and WHO - Overall Health System attainment.

values that range between 0 (worst performance) and 1 (best performance). The final Composite health infrastructure index, computed as a weighted average of the Human Capital and the Material Infrastructure dimensions, is sensitive to all the choices made during its construction, especially to the selected weighting scheme, which is the most controversial issue (Mazziotta and Pareto, 2013; OECD, 2008; Sharpe, 2004). Consequently, the sensitivity analysis, the final step of the procedure, was conducted by testing different combinations of weights. The resulting CI as well as its two components treated as independent variables have been included in different specifications of the regression model in order to validate the best result. The Composite health infrastructure index built as previously presented provided positive and highly significant coefficients in all our models, thus confirming that it was a good choice for our data.

2.2. The method

Our data account for two dimensions simultaneously: the cross – sectional one, in this case, the Romanian counties, and the time dimension, as we repeatedly measured our variables for the same statistical unit over eight years, 2007 to 2014. Therefore, we need to use a dedicated method to deal with both dimensions, and this falls in the area of panel data econometrics. Lately, the use of panel models has been on the rise due to some advantages it has over other methods. There are at least three reasons for preferring this type of analysis: data availability, a greater capacity for complex modeling compared to single cross-section or time series data, and a challenging methodology that claimed for improved econometric theory and dedicated software (Hsiao, 2007). In our case, two specific extra reasons make the panel regression model appropriate: its ability to control for the impact of potentially omitted variables, and its ability to capture heterogeneity across the statistical units under analysis.

There are several approaches in working with panel data, three of which will be used in this paper. The simplest is the pooled OLS model that treats all data as cross–sectional. As a consequence, a pooled model will apply the same coefficients to all the counties and will fail to account for the fact that the same county occurs repeatedly in our data set, at different moments in time. In terms of econometric theory, a pooled model is expected to have a high explanatory power, but usually suffers from misspecification and autocorrelation in errors.

Shifting to the category of models able to capture individual specificities, we will work under the assumption that there is something specific to each county and this might influence either diabetes incidence, or its predictors. As long as we accept the idea that this influence exists, we need to control for it. The standard way in which fixed and random effects models are doing this is by removing time – invariant specific characteristics, and then assessing the mere effect of the predictor on our dependent variable.

According to econometric literature, the main difference between fixed and random effects models is whether the specificities are correlated or not with the regressors. Unlike the pooled model, where all coefficients of the regression model were held constant, the fixed and random effects models will allow for each statistical unit under analysis to have its own intercept. A fixed effects model indicates that the individual specific factors captured by the intercepts are correlated with the regressors. In the case of random effects models, the differences across intercepts are due to random influences, independent from the regressors and can therefore be included in the error term. The decision on which model is the best fit is usually made based on two tests. The Lagrange Multiplier test for poolability helps deciding between the OLS pooled model and potential individual specificities. If the pooled model is rejected, the Hausman test comes into play and makes the final decision between the fixed and random effects model. A detailed presentation of the basic econometrics of panel data can be found in the classical reference in the field (Baltagi, 2001).

2.3. The models: In our analysis, we worked with the following three model specifications:

(Model 1) *log(Total new cases of diabetes) ~ log(Average Net Wage) +log(Composite health infrastructure index) + log(Population 0-14) + log(Population 15-79) + log(Density women)+ log(Urbanization rate) + log(Life expectancy)+ log(Average Net Wage)* log(Density women) + log(Average Net Wage)* log(Life expectancy)*

(Model 2) *log(New cases of insulin diabetes) ~ log(Average Net Wage) + log(Composite health infrastructure index) + log(Population 0-14) + log(Population 15-79) + log(Density women) + log(Urbanization rate) + log(Life expectancy) + log(Average Net Wage)* log(Density women) + log(Average Net Wage)* log(Life expectancy)*

(Model 3) *log(New cases of noninsulin diabetes) ~ log(Average Net Wage) + log(Composite health infrastructure index) + log(Population 0-14) + log(Population 15-79) + log(Density women) + log(Urbanization rate) + log(Life expectancy) + log(Average Net Wage)* log(Density women) + log(Average Net Wage)* log(Life expectancy)*

We conducted a preliminary testing of the functional form of the models using Ramsey’s regression specification error test available in the “lmtest” package in R and found that the functional form is appropriate.

3. Empirical results

For each of the functional forms specified in Section 2.3, we run the pooled, fixed and random effects model. According to the econometric theory, only one of these versions is appropriate in each case. The Lagrange Multiplier (LM) test for poolability, a test that helps the choice between the pooled OLS and a model that accounts for counties specificities, was conducted in three different settings, considering individual effects, time effects and twoways, or both.

After conducting the LM tests, we found that for the total new cases of diabetes as well as for insulin independent diabetes, the pooled OLS model should be rejected. In the next stage, the Hausman test applied to the same models indicates that, in choosing between random and fixed effects models, the right model for total diabetes as well as for non-insulin diabetes is fixed effects. However, for the total incidence category, the recommendation was to include both individual and time effects, while in the case of non-insulin diabetes, only individual effects were relevant. Based on the same LM test, for the insulin – dependent diabetes, the right choice is the pooled, or OLS model. Table 1 presents the estimated coefficients in each case, after using the Arellano method to correct for heteroscedasticity and serial correlation. In all three cases, we applied the Durbin Watson specification test for panel data in R and found that the d- statistic is not statistically significant.

The first interesting result is that for the first and third model, concerning the total number of cases and the non – insulin categories, a model that accounts for the existence of a certain heterogeneity at county level is the right choice, while for the insulin diabetes, the OLS models assuming that all the coefficients are constant provide better fits for the data. Put in different words, once the OLS model is preferred over the fixed effects, it means that all counties share similarities in insulin diabetes incidence, without particular specificities. However, putting all diabetes categories together, the incidence shows specificities across counties. A possible explanation of this result may regard the structure of the total number of cases, which includes several categories of diabetes that are specific to certain regions. Previous research shows, for example, that malnutrition diabetes displays the highest regional inequality (Druica and Goschin, 2016) among all reported categories, a fact that may induce heterogeneity at county level in the total incidence. The same reference source indicates that, for the insulin dependent category, the cases are rather evenly distributed, which may support the idea that the constant coefficients in the OLS model are appropriate to describe the regional distribution of the disease.

Another important result regards the differences in the explanatory power of the three models. While the overall model explains nearly 70% of the variations in diabetes incidence, the second model concerning the insulin dependent cases can only explain 26% of the variation. For the third model, the explanatory power is higher, reaching slightly over 50%.

Table 1. The estimated coefficients for each model, corrected for heteroscedasticity and serial correlation (Arellano method). All variables are in log (Standard errors in parentheses)

Variables	Total (fixed effects twoways)	Insulin dependent (OLS)	Insulin independent (fixed individual effects)
Intercept	dropped	539.304 (331.904)	dropped
Average net wage (log)	-77.302* (36.109)	-144.605 (96.520)	-12.457 (58.651)
Composite health index (log)	0.261*** (0.052)	0.621*** (0.118)	0.106 (0.086)
Population age group 0 – 14 (log)	0.874*** (0.051)	0.833*** (0.148)	0.906*** (0.089)
Population age group 15–79 (log)	0.306*** (0.066)	0.524*** (0.144)	0.196* (0.085)
Women density (log)	178.973*** (55.563)	159.261 (142.666)	190.031** (64.858)
Urbanization rate (log)	-0.072 (0.114)	-0.976*** (0.261)	0.129 (0.198)
Life expectancy (log)	-40.907. (23.262)	-102.206 (86.433)	15.706 (46.015)
Average net wage (log) * Women density (log)	-48.205** (16.073)	-46.027 (40.321)	-49.726** (18.685)
Average net wage (log) * Life expectancy (log)	10.440 (7.072)	26.187 (24.873)	-4.890 (13.394)
R^2	68,58%	26,28%	51,63%
Adjusted R^2	62,09%	24,19%	43,10%
F statistic	65.721 (p-value: < 2.22e-16)	12.596 (p-value: < 2.22e-16)	32.964 (p-value: < 2.22e-16)

** - significant at 0.05 level; ** - significant at 0.01 level; *** - significant at 0.001 level*

Source: Authors' calculations

The overall model

We found that all the selected variables were significant at 5%, except for life expectancy, which is marginally significant, and urbanization rate that proved to be statistically insignificant. The direct relation between the composite health index and the total number of new cases recorded points most likely toward an improved diagnosis capacity.

The average net wage holds a negative sign, which is in line with previous research that states that higher income is associated with lower incidence of chronic diseases, including diabetes. Women density holds a positive sign; a similar result can be found in other studies proving that diabetes incidence is higher among women. The interaction term showing the moderating effect of women share in the overall population on the association between salary and total diabetes incidence expressed in absolute value is statistically significant and holds a negative sign. The result shows that other factors held constant, higher salaries result in lower diabetes incidence in those areas with higher women density.

Life expectancy is not statistically significant either as a single predictor or in interaction with the average net wage, and so is urbanization rate. Of the four hypotheses set at the beginning of our study, in the case of total diabetes incidence, we fail to reject H1 at 5% level, while H2, H3 and H4 must be rejected.

The insulin dependent category

On the list of the variables selected in the model that were not statistically significant, we found the average net salary, women density and life expectancy. Previous research shows that the insulin – dependent category is less likely to be associated with gender (Dahlquist *et al*., 2011), a result that is also suggested by our research. In terms of healthcare infrastructure, we found that the composite health care infrastructure index is significant and has a positive impact on the dependent variable. We expect that a better health infrastructure would provide the ground for detecting and handling a higher number of cases, and would keep them into the records, especially with this category of diabetes where detection and care need to be available in the early stage of a patient's life. The fact that the urbanization rate is negatively correlated with the explained variable is in line with the idea that urbanization rate is a proxy for improved living standards, increased income, broader healthcare access and utilization. None of the interaction terms were statistically significant, so of the four hypotheses that we set at the beginning of our study, only H3 fails to be rejected for the insulin – dependent category.

The insulin – independent category

The third is a fixed effects model that shows that, after controlling for age group, none of the explanatory variables are statistically significant except for women density, which holds a positive sign, as well as its interplay with the average net wage, confirming previous studies stating that the type II diabetes is associated with gender. The result also confirms previous findings that show that the relationship between the level of income and diabetes incidence is moderated by the share of women in the overall population. For the insulin – independent category, the composite health infrastructure index is not statistically significant. We therefore fail to reject H1, but H2, H3 and H4 are rejected.

Conclusions, discussions and limitations of the study

Setting official data comprising of 41 administrative units analysed between 2007 and 2014 as a departure point, our study examined the relation between average net wage, urbanization rate, women density, medical infrastructure and medical human resources, all these controlled for the age structure of the Romanian population, with the absolute incidence of insulin, non – insulin and total cases of diabetes among Romanians.

We organized our research around four hypotheses. The first research hypothesis states that after controlling for age groups, the share of women in the overall population moderate salaries influence on diabetes incidence. This hypothesis cannot be rejected either for the overall model, or for the insulin – independent case. In terms of implications, this result helps identifying a specific target group of women with low income, which being at higher risks, needs more care both in terms of preventive actions.

The second research hypothesis states that after controlling for age groups, life expectancy moderates salaries influence on diabetes incidence. No group confirms this hypothesis and moreover, life expectancy is not statistically significant except for the overall model, but even in this case it is only marginally significant, at 10% level. This lack of significance may be due to the fact that life expectancy does not vary too much across counties, or that its interaction with the average salary does not result in groups with similar characteristics as those discussed by Tol *et al.* (2013).

The third research hypothesis states that after controlling for age groups, higher degrees of urbanization are negatively associated with diabetes incidence. Our study revealed that only the insulin – dependent category is associated with urbanization rate, and that the association is negative, as expected. Therefore, we fail to reject H3 for this category, while it is rejected for the overall model as well as for the insulin – independent group. The result is an indication that urbanization in Romania is still a ground for an improved quality of life, and not necessarily a source of stress and psychological pressure like it has already become in many highly developed countries.

The fourth hypothesis states that a better healthcare infrastructure is negatively associated with diabetes incidence. This hypothesis has to be rejected in all three cases. In fact, the variable displayed a statistically significant positive association with the overall incidence, as well as for the insulin – dependent category. We appreciate that the importance of this result is twofold. First, of the variables included in the model, little can be done in terms of directly targeted interventions. A higher concern for medical resources however, both material and human, is a field where intervention can be implemented, and our results provide guidance. Our computations revealed large differences in the county values of the Composite Healthcare Index (Figure 3), and for its two components - Human Capital Index and Material Infrastructure Index (Figures 1 and 2, respectively), over 2007-2014. The maps indicate a West-East divide, with Western regions displaying a higher healthcare endowment that mirrors their better economic development compared to Eastern regions.

Secondly, the positive sign in the Composite Healthcare Index may be an indication that the available Romanian healthcare infrastructure is not efficient in preventing and curing the disease, but only in identifying it and keep it into records. A negative sign would have invited to further investigations regarding the mechanisms that can be created in terms of reducing the incidence. For now, all we can hope for is that the percent of undiagnosed patients with diabetes, a problem across Europe, will decrease with the improvement of healthcare infrastructure.

An important limitation of this study lies in the fact that our model is static, only capturing the relation between variables at the time they were being examined. Previous researches document the lag between the exposure to risk factors and the disease as such, a fact that is impossible to be accounted for in our model. Even with this important limitation, our study was able to confirm the existence of several associations that have been documented before in relation to insulin-dependent and independent diabetes, and provides a solid ground for extending this type of research.

References

American Diabetes Association (2003), Economic Costs of Diabetes in the U.S. in 2002, *Diabetes Care*, 26(3), pp. 917-932.

American Diabetes Association (2009), Diagnosis and Classification of Diabetes Mellitus, *Diabetes Care*, 32(1), pp. 62–67.

Baltagi, B.H. (2001), *Econometric Analysis of Panel Data*, 3rd edition. Hoboken, NJ: John Wiley & Sons.

Black, S. (2002), Diabetes, Diversity and Disparity: What do we do with the evidence? *American Journal of Public Health*, 92(4), pp. 543 – 548.

Cade, W.T. (2008), Diabetes-Related Microvascular and Macrovascular Diseases in the Physical Therapy Setting, *Physical Therapy*, 88(11), pp. 1322–1335.

Cerassi, E. (1991), Insulin resistance, insulin deficiency, and non-insulin-dependent diabetes mellitus, *Diabetes Res Clin Pract*, 14(2), pp.37-45.

Colagiuri, S., Colagiuri, R., Conway, B., Grainger, D. and Davey, P. (2003), *DiabCo$t Australia: Assessing the Burden of Type 2 Diabetes in Australia.* Canberra, Australian Capital Territory, Australia: Diabetes Australia.

Croissant, Y. and Millo, G. (2008), Panel Data Econometrics in R: The plm Package, *Journal* *of Statistical Software*, 27(2), pp. 1-43.

Dahlquist, G.G., Nystrom, L. and Patterson, C.C(2011), Swedish Childhood Diabetes Study Group; Diabetes Incidence in Sweden Study Group. Incidence of type 1 diabetes in Sweden among individuals aged 0-34 years, 1983–2007: an analysis of time trends, *Diabetes Care*, 38(8), pp. 1754–1759.

Dall, T.M., Zhang, Y., Chen, Y.J., Quick, W.W., Yang, W.G. and Fogli, J. (2010), The economic burden of diabetes, *Health Aff (Millwood)*, 29(2), pp. 297–303.

Deshpande, A.D., Harris-Hayes, M. and Schootman, M. (2008), Epidemiology of Diabetes and Diabetes-Related Complications, *Physical Therapy*, 88(11), pp. 1254–1264.

Druica, E. and Goschin, Z. (2016), Does economic status matter for the regional variation of malnutrition-related diabetes in Romania? Temporal clustering and spatial analyses, *Journal of Applied Quantitative Methods*, 11(4).

European Commission (2004), *Composite Indicator on e-business readiness*, Brussels, DG JRC.

Gaillard, T.R., Schuster, D.P., Bossetti, B.M., Green, P.A., and Osei, K. (1997), Do sociodemographics and economic status predict risks for type II diabetes in African Americans? *Diabetes Educ.*, 23, pp. 294–300.

Gregg, E.W., Sattar, N. and Ali, M.K. (2016), The changing face of diabetes complications, *Lancet Diabetes Endocrinol*, 4(6), pp. 537–547.

Hosler, A. and Melnik, T. (2003), Prevalence of Diagnosed Diabetes and Related Risk Factors: Japanese Adults in Westchester County, New York, *American Journal of Public* Health, 93(8), pp: 1279 – 1281.

Hsiao, C. (2007), Panel data analysis—advantages and challenges, *Test*, 16(1), pp. 1 – 22.

Jonsson, B. (2002), Revealing the cost of type II diabetes in Europe, *Diabetologia*, 45, pp. S5–S12.

Larrañaga, I., Arteagoitia, J.M., Rodriguez, J.L., Gonzalez, F., Esnaola, S. and Piniés, J.A. (2005), Socio-economic inequalities in the prevalence of Type 2 diabetes, cardiovascular risk factors and chronic diabetic complications in the Basque Country, Spain, *Diabet Med.*, 22(8), pp. 1047-1053.

Lee, D.S., Kim, Y.J. and Han, H.R. (2013), Sex differences in the association between socio-economic status and type 2 diabetes: data from the 2005 Korean National Health and Nutritional Examination Survey (KNHANES), *Public Health*, 127(6), pp. 554-560.

Leeder, S. Raymond, S. Greenberg, H., Liu, H. and Esson, K. (2004), *Race Against Time: The Challenge of Cardiovascular Disease in Developing Economies*, New York, NY: Columbia University.

Leigh, J.P. (1993), Multidisciplinary findings on socioeconomic status and health, *American Journal of Public Health*, 83(2), pp. 289–290.

Long, A.N. and Dagogo-Jack, S. (2011), The Comorbidities of Diabetes and Hypertension: Mechanisms and Approach to Target Organ Protection, *Journal of Clinical Hypertension (Greenwich, Conn.)*, 13(4), pp. 244–251.

Lubitz, J., Cai, L., Kramarow, E. and Lentzner, H. (2003), Health, Life Expectancy, and

Health Care Spending among the Elderly, *The New England Journal of Medicine*, 349(11), pp. 1048-1055.

Mărăcine, V. and Ianole, R. (2010), How Knowledge Determines Demand Dynamics-A New Perspective for Evaluation. Study Case for Healthcare Services, *Journal of Economic Computation and Economic Cybernetics Studies and Research*, (4), pp. 5-22.

Mazziotta, M. and Pareto, A. (2013), Methods for constructing composite indices: one for all or all for one?, *Rivista Italiana di Economia Demografia e Statistica*, LXVII(2), pp. 67 – 80.

Meara, E. R., Seth Richards, and David M. Cutler (2008), The Gap Gets Bigger: Changes in Mortality And Life Expectancy, By Education, 1981–2000, *Health Affairs*, 27(2), pp. 350-360.

Metcalf, P.A., Scragg, R.R., Schaaf, D., Dyall, L., Black, P.N., and Jackson, R.T. (2008), Comparison of different markers of socioeconomic status with cardiovascular disease and diabetes risk factors in the Diabetes, Heart and Health Survey, *The New Zealand Medical Journal.*, 121(1269), pp. 45-56.

Moradi, G., Majdzadeh, R., Mohammad, K., Malekafzali, H., Jafari, S., and Holakouie-Naieni, K. (2016), Is the status of diabetes socioeconomic inequality changing in Kurdistan province, west of Iran? A comparison of two surveys, *Medical journal of the Islamic Republic of Iran*, 30,375.

Munda, G. (2007), *Social multi-criteria evaluation*, Springer-Verlag, Heidelberg, New York, Economics Series.

Murea, M., Ma, L. and Freedman, B.I. (2012), Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications, *The Review of Diabetic Studies: RDS*, 9(1), pp. 6–22.

Nordahl, H. (2014), Social inequality in chronic disease outcomes, *Danish Medical Journal*, 61(11), B4943.

OECD (2008), *Handbook on Constructing Composite Indicators.Methodology and user guide*, Paris: OECD Publications.

Patterson, C.C., Dahlquist, G., Soltesz, G., and Green, A. (2001), Europe and diabetes is childhood-onset type I diabetes a wealth-related disease? An ecological analysis of European incidence rates, *Diabetologia*, 44(Suppl. 3), pp. B9–16.

Patterson, C.C., Dahlquist, G., Soltesz, G., Green, A. and EURODIAB ACE Study Group (2009), Incidence trends for childhood type 1diabetes in Europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study, *The Lancet*, 373(9680), pp. 2027-2033.

Peykari, N., Djalalinia, S., Qorbani, M., Sobhani, S., Farzadfar, F., and Larijani, B. (2015), Socioeconomic inequalities and diabetes: A systematic review from Iran, *Journal of Diabetes and Metabolic Disorders*, 14(8).

Reinehr, T. (2013), Type 2 diabetes mellitus in children and adolescents, *World Journal of Diabetes*, 4(6), pp. 270–281.

Rivera, L.A., Lebenbaum, M, and Rosella, L.C. (2015), The influence of socioeconomic status on future risk for developing Type 2 diabetes in the Canadian population between 2011 and 2022: differential associations by sex, *International Journal of Equity Health*, 14(101).

Robbins, J., Vaccarino, V., Zhang, H., and Kasl, S. (2001), Socioeconomic Status and Type 2 Diabetes in African American and Non-Hispanic White Women and Men: Evidence from the Third National Health and Nutrition Examination Survey, *American Journal of Public Health*, 91(1), pp. 76 – 83.

Saisana, M. and Tarantola, S. (2002), *State-of-the-art report on current methodologies and practices for composite indicator development. Report EUR 20408 EN*, European Commission–Joint Research Centre, Ispra.

Salzman, J. (2003), *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being*, Ottawa: Center for the Study of Living Standards.

Sharpe, A. (2004), *Literature Review of Frameworks for Macro-indicators*, Centre for the Study of Living Standards, Ottawa, CAN.

Smith, W.C., Anderson, E., Salinas, D., Horvatek, R., and Bakerad P.D. (2015), A meta-analysis of education effects on chronic disease: The causal dynamics of the Population Education Transition Curve, *Social Science & Medicine, Medicine*, 127, pp. 29-40.

Tamayo, T., Rosenbauer, J., Wild, S.H., Spijkerman, A.M., Baan, C., Forouhi, N.G., Herder, C., and Rathmann, W. (2014), Diabetes in Europe: an update, *Diabetes Research and Clinical Practice*, 103, pp. 206-217.

Tol, A., Sharifirad, G., Shojaezadeh, D., Tavasoli, E., and Azadbakht, L. (2013), Socio-economic factors and diabetes consequences among patients with type 2 diabetes, *Journal of Education and Health Promotion*, pp. 2-12.

van Baal, P. Peters, F. Mackenbach, J. and Nusselder, W. (2016), Forecasting differences in life expectancy by education, *Population Studies,* 70(2), pp. 201-216.

Wannamethee, S.G., Papacosta, O., Lawlor, D.A., Whincup, P.H., Lowe, G.D., Ebrahim, S., and Sattar, N. (2012), Do women exhibit greater differences in established and novel risk factors between diabetes and non-diabetes than men? The British Regional Heart Study and British Women's Heart Health Study, *Diabetologia,* 55(1), pp 80–87.

Warnes, R.G., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2016), gplots: Various R, Programming Tools for Plotting Data. R package version 3.0.1.

Wilkinson, R.G. (1992), Income distribution and life expectancy, *BMJ: British Medical Journal*, 304(6820), pp.165-168.

Williams, R. Van Gaal, L. and Lucioni, C. (2002), Assessing the impact of complications on the costs of type II diabetes, *Diabetologia*, 45, pp. S13–S17.

Winkleby, M.A., Jatulis, D.E., Frank, E. and Fortmann, S.P. (1992), Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease, *American Journal of Public Health*, 82(6), pp. 816–820.

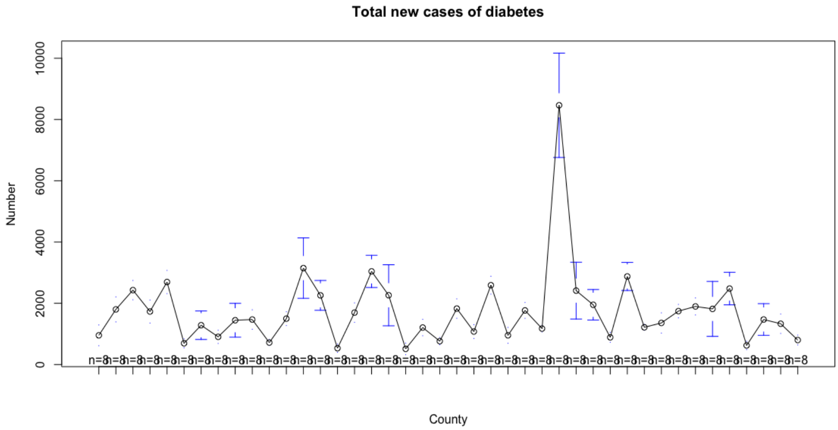
World Bank Database (2010), *Romania, GDP expressed in $US, constant at 2010 level* (retrieved from http://data.worldbank.org/indicator/NY.GDP.MKTP.KD?locations =RO).

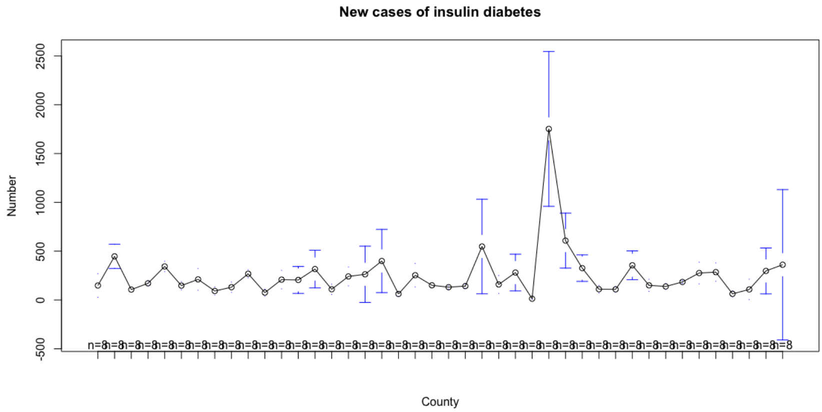
World Health Organization (2002), *Global assessment of the state-of-the-science of endocrine disruptors* (retrieved from http://www.who.int/ipcs/publications/ new_issues/endocrine_disruptors/en/).

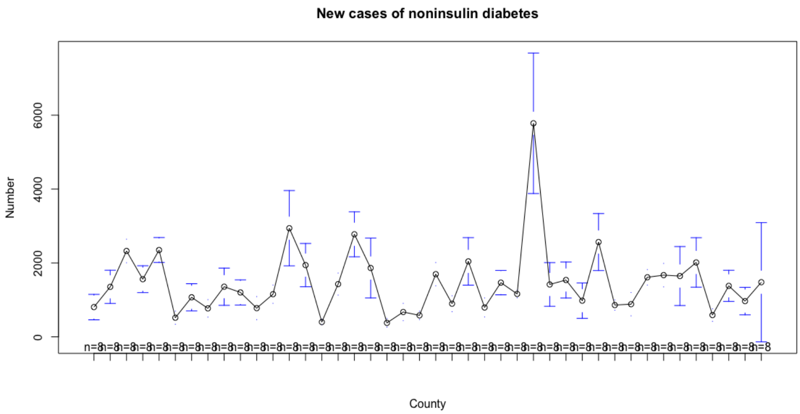
World Health Organization Reports (2016), *The 2016 Global report on Diabetes* (retrieved from http://apps.who.int/iris/bitstream/10665/204874/1/WHO_NMH_NVI_16.3_ eng.pdf).

Zeileis, A. and Hothorn, T. (2002), Diagnostic Checking in Regression Relationships, R News, 2(3), pp. 7-10.

Annex 1. Heterogeneity across counties, by category of diabetes







Annex 2. a. Preliminary list of explanatory variables.

Variable	Definition	Data source
Average net wage	The net nominal earnings (constant Romanian lei) is obtained by subtracting the taxes from the gross nominal earnings	National Institute of Statistics (NIS) and own computation to adjust for inflation.
Composite health index	The composite health infrastructure index combines two dimensions: - Human Capital: private and public sector physicians, physician assistants and nurses, all measured per 1000 inhabitants; - Material Infrastructure: hospitals per 100,000 inhabitants and hospital beds per 1000 inhabitants. All values are re-scaled by the min-max normalization procedure, in the range 0-1.	Own computations based on NIS data
Population 0_14	The number of persons aged 0-14 years.	NIS and own computations for population age group.
Population 15_79	The number of persons aged 15-79 years.	NIS and own computations for population age group.
Women density	Number of women relative to total population (%).	NIS and own computations
Urbanization rate	Urban relative to total population (%)	NIS and own computations
Life expectancy	Life expectancy at birth is the estimation of average number of years a newborn is likely to live.	NIS

Note: All variables are available in English at the Romanian Institute of Statistics, Tempo Online Database, at http://statistici.insse.ro/shop/?lang=en

Annex 2. b. List of explained variables.

Variable	Definition	Data source
Total number of cases	The number of new cases of patients diagnosed with diabetes recorded within a specific year.	The National Institute of Public Health
Number of insulin – dependent cases	The number of new cases of insulin – dependent patients recorded within a specific year.	The National Institute of Public Health
Number of insulin – independent cases	The number of new cases of insulin – independent patients recorded within a specific year.	The National Institute of Public Health

Note: The National Institute of Public Health has provided all variables, based on our written request.

Annex 3. Correlation matrices

3.1. Independent variables

Variable	Log(Average net wage)	Log(Composite infrastructure index)	Log(Population 0 – 14)	Log(Population 15 – 79)	Log (Women Density)	Log (Urbanization rate)	Log (Life expectancy)
Log(Average net wage)	1	0,29	0,20	0,05	0,4	0,23	0,21
Log(Composite infrastructure index)		1	0,35	0,28	0,40	0,65	0,37
Log(Population 0 – 14)			1	0,38	0,11	0,24	0,24
Log(Population 15 – 79)				1	0,25	0,26	0,28
Women Density					1	0,35	0,15
Urbanization rate						1	0,19
Life expectancy							1

Source: Authors' work

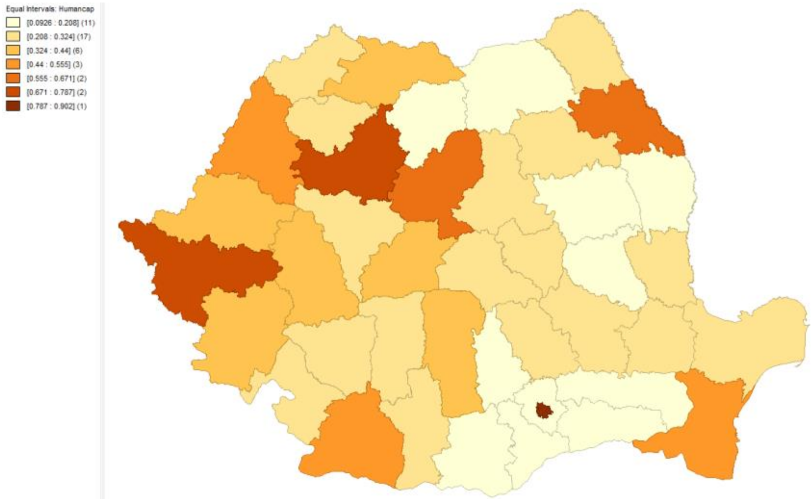
3.2. Healthcare related variables

	Nurses	Doctors – public	Doctors – private	Hospital beds	Hospitals - number
Nurses	1	.752**	.474**	.784**	.454**
Doctors – public		1	.525**	.853**	.733**
Doctors – private			1	.547**	.494**
Hospital beds				1	.760**
Hospitals - number					1

Source: Authors' work

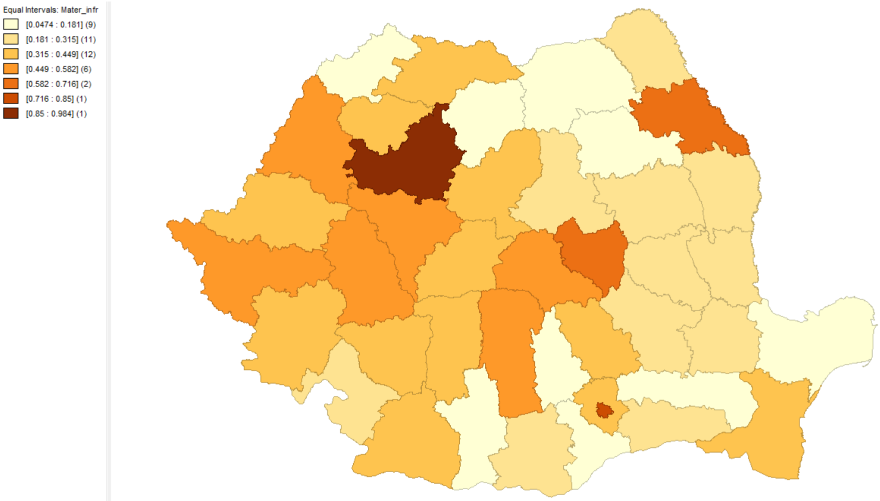
Annex 4. Maps

Figure 1. Health Care Human Capital Index by county – average values 2007-2014



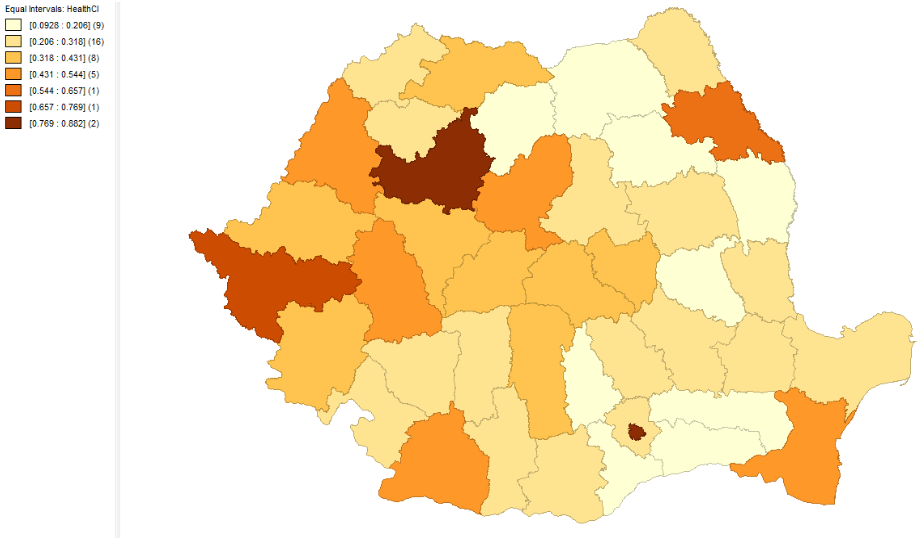
Source: own representation

Figure 2. Health Care Material Infrastructure Index by county – average values 2007-2014



Source: : own representation

Figure 3. Health Care Composite Index by county – average values 2007-2014



Source: : own representation