SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Application of census-based ethnic classification, self and external classification of ethnicity in surveys to study of Roma health inequalities

by Ferenc Vincze

Supervisor: Prof. Dr. János Sándor

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF HEALTH SCIENCES

DEBRECEN, 2021

Application of census-based ethnic classification, self and external classification of ethnicity in surveys to study of Roma health inequalitiesTitle of the thesis

By Ferenc Vincze, Nutritionist

Supervisor: Prof. Dr. János Sándor, MD, PhD, Habil.
Doctoral School of Health Sciences, University of Debrecen

Head of the **Examination Committee**: Prof. György Paragh MD, PhD, DSC
Members of the Examination Committee: Dr. Edit Paulik, PhD
Dr. László Ulicska, PhD

The Examination is held online and starts at 10:00 on 5th July, 2021.

Head of the **Defense Committee**: Prof. György Paragh MD, PhD, DSC
Reviewers: Dr. Karolina Kósa, PhD
Dr. Zsuzsanna Kiss Orsós, PhD
Members of the Defense Committee: Dr. Edit Paulik, PhD
Dr. László Ulicska, PhD

The PhD Defense is held online and starts at 11:00 on 5th July, 2021.

Publicity is provided online. If you wish to participate, please indicate it in a message sent to the e-mail address vincze.ferenc@med.unideb.hu by 2 p.m. on 2 July, 2021 latest. After the deadline it is no longer possible to connect to the PhD defense due to technical reasons.

Introduction

Health of the Roma population

The Roma population is the largest ethnic minority group in Europe. The common opinion is that the health status of this population is poor compared to the non-Roma population. Recently, this rather anecdotal belief has been convincingly supported by a growing number of quantitative investigations. The current urgency for intervention is hardly debatable, but effective interventions are hindered by our limited knowledge regarding details of the mechanisms that have led to this substantial population-level health deficit in the Roma population.

It has been thoroughly demonstrated that Roma people are overrepresented in marginalized communities, live in relatively unhealthy environments and have an unhealthier lifestyle than the non-Roma population. A large proportion of the literature on Roma health is devoted to communicable disease. Many papers report outbreaks or cases and not the epidemiology of disease, although some have drawn attention to low immunization coverage.

Cardiometabolic disorders (hypertension, heart disease, elevated lipid levels, type 2 diabetes and metabolic syndrome) and other noncommunicable diseases are the primary sources of health issues in the Roma population. Modifiable risk factors, such as high blood pressure, high cholesterol, high blood glucose, overweight and obesity, low fruit and vegetable intake, physical inactivity and tobacco and alcohol use, are responsible for unfavorable health outcomes. Studies have shown unfavorable birth outcomes in the Roma population. In addition, the Roma people's access to care is reduced, leading to high levels of unmet health needs.

The poor health status of Roma is associated with increased all-cause premature mortality. This may reflect either the lower quality of healthcare or the worse health status caused by unhealthier lifestyles or disadvantageous socioeconomic status.

Despite substantial uncertainties, the European Union considers this problem a high priority, necessitating additional systematic research on the role of the Roma ethnicity on health determinants, indicating that the scientific base must be strengthened to establish an adequate Roma-centric policy. The use of regular surveys to evaluate the health status of the Roma population by inserting the Roma ethnicity into the variables examined during data collection seems to be a technically simple and promising method to quantify some of these issues. Roma-specific survey results could be very informative. However, this approach is hindered by legal constraints (right to personal data protection of survey participants).

Challenges in Roma health monitoring

It seems that higher-quality publications on the health impact of Roma ethnicity use narrower case definitions or more specific settings. This approach results in knowledge that relies primarily on subgroup-specific findings, observations that cannot be extrapolated to the whole Roma population and limited knowledge on the public health importance of the Roma ethnicity as a risk factor for poor health. Furthermore, the Roma people are clearly not a homogeneous social stratum. Our knowledge on the role of this heterogeneity in determining the health status of this population is currently quite limited. Moreover, we do not know whether it is justified

at all to use this dichotomization (distinguishing Roma and non-Roma persons) in epidemiological investigations of the risk associated with Roma ethnicity.

Parallel application of self and external classification of ethnicity in Roma health surveys

Although a variability of methods have been applied in studies to evaluate Roma health status compared to non-Roma populations, the mainstream European approach for identifying Roma persons in censuses, surveys and clinical studies is self-reporting. Reports from Bulgaria, Spain, Slovakia, Serbia, Belgium and England apply this approach. Furthermore, this research attitude is also reflected in Hungarian publications on the health status and genetic susceptibility of the Roma people. Due to the fluidity of self-reported identity, which is influenced by societal attitudes, self-reported Roma classification is considered an approximation that underestimates the proportion of Roma persons and leads to biased results. These limitations manifest in multicountry studies.

The change in the proportion of the Roma population varies from one time point to the next. In health surveys, subjects who do not identify themselves as Roma are still often classified as such by experts or interviewers, hence the conclusion drawn by most researchers that self-identification underestimates the size of the Roma population.

Because many people are considered Roma by members of the general population based on external traits, it seems to be a useful approach for interviewers to classify survey participants to prevent biases caused by refused admission of Roma ethnicity in self-reporting. According to recent publications, nonself-reported ethnicity can be based on lifestyle, surname, and residence but not on explicitly defined racial characteristics. Health care staff members, officers, survey interviewers (with or without the support of a Roma community leader), and parents of children can perform this classification. Furthermore, the combination of self-reporting and a decisive external classification has also been practiced. The latter approach clearly shows that external classification is considered more reliable than self-reporting in certain cases. This heterogeneity highlights the lack of a standard methodology, as well as the limited comparative value of results from studies with an external Roma classification. Moreover, it is known that observer-reported identification is influenced by the observer's ethnicity and sex.

The organization of Roma-specific data collection in different settings to compare Roma to non-Roma disparities is hindered by the uncertain nature of the abovementioned classification methods. In reality, the basic choice between self- or external-reporting cannot be based on evidence because differences between results from interviewer-reporting compared to self-reporting-based surveys are not known in detail. We could not identify any publication in peer-reviewed international journals on health interview surveys with a twofold Roma ethnicity assessment (parallel application of self-reporting and interviewer reporting) that directly compared the results produced by different Roma definitions.

Representativeness of Roma health surveys

Numerous problems connected to representativeness can be identified because higher-quality publications on the health impact of Roma ethnicity use narrower case definitions or more specific settings. This approach results in knowledge that relies primarily on subgroup-specific

findings, observations that cannot be extrapolated to the whole Roma population and limited knowledge on the public health importance of Roma ethnicity as a risk factor.

Taking into account that surveys are inevitably subject to sampling bias due to the nonexistence of a Roma sampling frame, a census that provides data representative of the population could be a useful part of monitoring.

Self-declared ethnicity-related data based on national censuses have been available for decades in many countries. To avoid the well-known under-registration of Roma ethnicity, multiple questions used to be applied in census questionnaires. Governments ensure the regularity and feasibility of censuses required for monitoring. If a census collects data on health status and ethnicity along with the most important sociodemographic risk factors for health impairments, then it can establish a basis for comparative investigations of Roma and non-Roma populations regularly and in a feasible manner.

Objectives

The census collects data on health status and ethnicity along with the most important sociodemographic risk factors for health impairments; therefore, it can establish a basis for regular comparative investigations of Roma and non-Roma investigations in a feasible manner.

Our aim was to compare the occurrence of chronic diseases and activity limitations among the Roma and non-Roma populations.

Herein, we aimed to describe the risk and health impact of Roma ethnicity independent of sociodemographic factors using census-derived data.

We studied the added value of census-based data in monitoring the health gap between Roma and non-Roma populations.

The census-based investigation could not avoid self-reporting-related ethnicity misclassification and the remarkable underestimation of the Roma population size. There is a need to further develop the currently used methodology. Therefore, we aimed to analyze the added value of the development of the currently used ethnic classification methodology in population surveys.

The objective of our investigation was to assess the health status differences between Roma and non-Roma adults using both self-reporting and interviewer-reporting ethnicity for all participants.

We described the differences between the results achieved by self-reporting and interviewer-reporting classifications.

In this comparative study, we aimed to explore the rationale of utilizing the external Roma ethnicity classification.

Materials and methods

Prevalence of chronic diseases and activity-limiting disability in the Roma population

Setting

This study investigated the database of the last Hungarian national census conducted in 2011. The cross-sectional survey covered the entire Hungarian population. Data collection was performed from 1 October 2011 to 31 October 2011. The study population included all Hungarian citizens living in the country or staying temporarily abroad for a period of less than 12 months, as well as all foreign citizens and stateless persons living in Hungary for a period of more than 3 months.

Access to the database was provided by the Hungarian Central Statistical Office (HCSO). This study was approved and supervised by the institutional board of the HCSO, which is responsible for both the utilization of the census database and the protection of human rights in handling sensitive personal data (KSH/ADKI/1156/2014).

Explanatory variables and outcome measures

The first part of the census questionnaire focused on housing conditions (type of walls and presence of utilities, such as bathroom, flush toilet, electricity, water and hot water supply, and heating system). Because the response to these questions was compulsory, apart from homeless people and those who were living in institutions, all subjects' living conditions were characterized by these parameters. All factors were used as proxy indicators of deprivation in the analysis.

The second part of the census questionnaire was focused on personal characteristics. The responses were compulsory in this part of the questionnaire as well, apart from voluntary questions related to ethnicity, the presence of chronic disease, and activity limitations. Subjects were classified according to covariate variables, such as sex, age, marital status, highest level of education, and employment status. Subjects were classified according to covariate variables, such as sex, age, marital status, highest level of education, and employment status. The following age groups were used: 0–5, 6–17, 18–34, 35–59, 60–64, and 65+ years. The highest level of education was categorized as less than primary, primary, vocational, high school, and tertiary school. Individuals less than 6 years old or attending an educational institution were classified as not having completed their education. Employment status was described as working, unemployed, retired, receiving social benefits, dependent, or student. Single, coupled, divorced, and widowed were distinguished with respect to marital status.

Four questions referred to self-declared ethnicity. These questions included primary self-declared ethnicity ("Which ethnicity do you feel you belong to?"), secondary self-declared ethnicity ("Do you think you belong to another ethnicity too?"), first language ("What is your first language?"), and secondary language ("In what language do you usually speak with family members or friends?"). Those who responded with "Roma" to any of these questions were regarded as Roma people. All others were considered to be non-Roma people.

The primary outcome variable was the self-declared occurrence of at least one chronic disease. The type of disease was not requested. The existence of activity limitations among subjects

with chronic disease was investigated as a secondary outcome. Activity limitations were subgrouped according to the affected function: self-sufficiency, everyday life, learning-working, family life, transport, communication and community life.

Statistical analysis

First, the crude prevalence of having at least one chronic disease and activity limitations among those with chronic disease were calculated for the Roma and non-Roma subpopulations. The association with Roma ethnicity was evaluated using the $\chi 2$-test.

Next, internal indirect standardization was used to control for the confounding effects of age, sex, and level of education. The age- and sex-standardized prevalence ratios (SPRs) and age-, sex-, and education-standardized prevalence ratios (SPR_e) were calculated for both the Roma and non-Roma populations. Next, we assessed the risk associated with Roma ethnicity by the ratio of Roma-specific and non-Roma-specific standardized prevalence ratios ($RR = SPR_{Roma}/SPR_{non\text{-}Roma}$; $RRe = SPR_{e,Roma}/SPR_{e,non\text{-}Roma}$). The 95% confidence intervals (95% CI) of the calculated measures were used in the statistical evaluation.

Then, we used multiple logistic regression modeling to investigate the influence of Roma ethnicity, independent of housing conditions (wall quality, public utilities for bathrooms, flush toilets, electricity, water and hot water supply, and heating system) and personal characteristics (sex, age, level of education, employment, and marital status). Associations were quantified by odds ratios (ORs) and corresponding 95% CIs.

Finally, point estimates and 95% CIs of the relative risks estimated by standardization or logistic modeling were used to describe the impact of Roma ethnicity on chronic disease and the occurrence of activity limitation. Attributable risk fractions among Roma (AR_{Roma}) and the whole population ($AR_{Population}$) were computed with corresponding 95% CIs. Statistical computation was performed using STATA 14.0 (Stata Corporation, College Station, TX) software.

The application of self- and external classification of ethnicity in a population-based health interview survey

Setting

The survey was implemented in 2015 and covered 20 of 175 Hungarian districts. The list of 3500 persons above 18 years was prepared by randomization from the population registry of the entire population (N = 965,680) with a residential place in the studied districts. Data collection was performed if the subjects signed informed consent forms. Being aware of the highly sensitive nature of our investigation from an ethical point of view and preventing ethical restrictions in utilizing results from our investigation, all ethical regulations were strictly respected throughout the entire study. Because the data collection had been implemented in different areas of Hungary, neither an institutional nor a regional ethical committee were suitable to evaluate our research plan. Therefore, the detailed study protocol and the applied questionnaire were reviewed and approved by the highest-level committee of the Medical Research Council by the Ethical Committee of the Hungarian National Scientific Council on Health (15563-2/2015/EKU 0111/15). The questionnaires were anonymized before entering the

data into an electronic database. Records without personal identifiers were archived and processed according to the ordinance of the ethical approval.

Explanatory variables and outcome measures

The Hungarian adaptation of the European Health Interview Survey was used for data collection. The questionnaire contains information about general health status, diseases, accidents, functionality, lifestyle, social capital, access to health care, access to preventive services, adherence to drug consumption and oral health. A total of 52 indicators were investigated. Trained interviewers completed the questionnaires. All health indicators were dichotomized before analysis.

Each respondent's ethnicity was identified by themselves and by the interviewers. Questions to assess self-reported ethnicity in the previous Hungarian Census 2011 were added to the survey questionnaire. To counter the low response rate for the ethnicity item, two questions were included. These questions asked about the ethnicity to which the participant belongs ("Which ethnicity do you feel you belong to?") and about the other ethnicity to which he or she also belongs ("Do you belong to another ethnicity?"). The primary and secondary self-reported ethnicities were registered in this way. The self-reported Roma category included all interviewees who reported Roma ethnicity either primarily or secondarily.

The interviewers' Roma classification was part of the questionnaire, which was completed by the interviewers without asking the participants and without informing them about the registered data. The interview was completed at the home of participants. The interviewers' observations on the visible characteristics and living conditions of the interviewees during the interview formed the bases of categorization. There were no other specific rules for the interviewers' Roma identification. Similar to the governmental protocol in the United States, identification of ethnicity can be performed by an observer, despite the acknowledged limitations of external classifications and the practical impossibility of constructing instructions for the observers' classifications.

Statistical analysis

The sociodemographic determinants of the willingness to declare Roma ethnicity were analyzed by multivariate logistic regression. The frequencies of sex, age, education, marital status, employment, and number of persons in the household among participants who self-reported as Roma were compared to those of persons assessed as Roma by only interviewer-based assessment.

The associations between Roma ethnicity and health status indicators were investigated in multivariate logistic regression models applying self-reported and interviewer-reported Roma classifications separately. These models were controlled for age, sex, education, and employment status. Associations were evaluated by the adjusted odds ratios (ORs) and their corresponding 95% confidence intervals (95% CIs). Results from the two approaches were compared using 95% confidence intervals by indicators to determine the differences between the two Roma definitions in evaluating the differences between Roma and non-Roma individuals. Statistical analysis was performed using SPSS 18 (SPSS package for Windows, Release 18; SPSS, Chicago, IL, USA).

Results

Chronic diseases caused health loss among the Roma population

There were 9,937,628 subjects who participated in the Hungarian census in 2011. The restriction of this population to Hungarian citizens living in Hungary resulted in a database of 9,794,318 persons. Because the response rate for ethnicity and chronic disease-related questions was 78.36%, the database that could be used to analyze the determinants of chronic disease contained 7,674,607 records. In the case of logistic regression modeling, 166,366 records were excluded because housing condition data were not available. This exclusion reduced the response rate to 76.66%. Based on self-declarations, 3.83% (294,189 persons) were classified as Roma individuals, and 21.19% (1,626,447 persons) had at least one chronic disease in the study population.

The response rate for the question about activity limitation among those with chronic disease was 87.09%. This resulted in a database of 1,416,424 persons with chronic disease who reported activity limitations (in the multivariate modeling, 40,763 whose housing condition data were not available were also excluded, resulting in an 84.58% response rate for that analysis). There were 38,800 (2.74%) Roma persons in this database.

Sex representations among the Roma and non-Roma people were significantly different because of the higher proportion of males in the Roma population. The Roma age structure deviated significantly from the non-Roma age composition. The Roma were overrepresented in younger age groups and underrepresented in older age groups. The level of education was much lower among Roma than among non-Roma. The percentage of individuals who had primary education or less than primary education was 53.31% in the Roma population. All other personal and housing-related deprivation indicators applied in the multiple logistic regression models indicated the poorer status of the Roma population in comparison to the non-Roma population.

Prevalence of chronic diseases in the Roma population

The crude prevalence of having at least one chronic disease was significantly lower in the Roma population (14.75%) than in the non-Roma population (21.45%). However, significant risk elevation was observed in the Roma population after adjusting for age and sex (RR = 1.41) and after adjusting for age, sex, and education (RR_e = 1.11). We found that Roma ethnicity was a significant risk factor for having at least one chronic disease in the multivariate logistic model when controlled for deprivation indices (OR = 1.17).

Each of the age- and sex-adjusted; age-, sex-, and education-adjusted; and multiply adjusted estimations revealed a significant impact of Roma ethnicity on chronic disease occurrence, both within the Roma population and in the whole population.

Prevalence of activity-limiting disability in the Roma population

Each unadjusted frequency of activity limitation outcome was significantly more frequent in the Roma population than in the non-Roma population. Age- and sex-standardized measures exhibited a similar pattern. The self-sufficiency limitation was more frequent in the Roma population (RR = 1.75). The activity limitation in everyday life showed a significantly higher

frequency in the Roma population (RR = 1.64). The Roma ethnicity was also associated with an increased risk of limitation in the Roma population in the fields of learning-working (RR = 1.59), family life (RR = 1.60), transport (RR = 1.54), communication (RR = 1.44), and community life (RR = 1.14). Each age- and sex-adjusted risk ratio and attributable fraction was highly significant.

According to the SPR_e for activity limitations, there was no evidence of differences between the Roma and non-Roma populations with respect to self-sufficiency or community life-related activities. Significant risk elevation was observed in the fields of everyday life activities (RR_e = 1.18, AR_{Roma} = 15.25; $AR_{Population}$ = 0.53), learning-working (RR_e = 1.19, AR_{Roma} = 16.97; $AR_{Population}$ = 0.98), family life (RR_e = 1.22, AR_{Roma} = 18.03; $AR_{Population}$ = 0.89), and transport (RR_e = 1.16, AR_{Roma} = 13.79; $AR_{Population}$ = 0.39) in Roma with chronic disease. Activity limitation in communication was significantly less frequent in Roma people (RR_e = 0.86, AR_{Roma} = −16.28; $AR_{Population}$ = −0.64).

In the multiple logistic regression analyses, Roma ethnicity was associated with a higher risk of activity limitation in the fields of everyday life activities (OR = 1.20, AR_{Roma} = 16.67; $AR_{Population}$ = 0.58), learning-working (OR = 1.24, AR_{Roma} = 19.35; $AR_{Population}$ = 1.19), family life (OR = 1.22, AR_{Roma} = 18.03; $AR_{Population}$ = 0.89), and transport (OR = 1.03, AR_{Roma} = 2.91; $AR_{Population}$ = 0.08). There was no significant difference in the activity limitation between the Roma and the non-Roma populations in the fields of self-sufficiency, communication, or community life.

The application of self- and external classification of ethnicity in a population-based health interview survey

The response rate of the survey was 69.2%, with 2421 participants. Because the number of Roma adults was small among the population older than 65 years, the statistical evaluation was restricted to the age range of 18–64 years (nine self-reported and nine interviewer-reported Roma among 572 subjects older than 65). Ultimately, the investigation focused on 1849 subjects.

There were 124 self-reported Roma subjects, whereas 179 people were categorized as Roma ethnicity by interviewers, of whom 61 individuals were identified only by the observers.

Socio-economic status of the respondents

There was no difference between Roma and non-Roma samples with respect to sex and marital status composition. The Roma age distribution was shifted towards the younger age groups. Economic activity and the level of education were significantly higher in non-Roma individuals. Roma households were larger than non-Roma households. The differences between the Roma and non-Roma populations were similar, independent of whether ethnicity was assessed via self-report or interviewer report.

According to the multivariate logistic regression analysis, employed Roma individuals were less willing to declare (OR = 3.49) their Roma ethnicity than economically inactive Roma individuals. Similar underreporting of Roma ethnicity was observed in younger age groups with borderline significance (OR = 5.06).

Descriptive health status indicators for Roma

According to the crude descriptive measures, the general health status of the Roma is inferior to that of the non-Roma. There was no difference between Roma and non-Roma individuals with respect to accident frequency or adherence in drug consumption.

Apart from the equal crude prevalence of cardiometabolic disorders, chronic disorders showed a higher occurrence among the Roma. A higher prevalence was observed for cardiometabolic diseases, the general chronic disease occurrence in the Roma does not deviate significantly from that of the non-Roma. Geographical access to health care is similar among Roma and non-Roma individuals, while access in terms of time was worse for Roma individuals than for non-Roma individuals. Lifestyle indicators are disadvantageous among the Roma population, but two indicators (prevalence of obesity and hearing loss) show no association with Roma ethnicity.

Almost all functional statuses and oral health-related indicators were worse in the Roma population. Indicators related to social capital were similar in Roma and non-Roma populations. The only exception was that the Roma likely face more difficulty when they need help from neighbors. Ethnic differences in the use of preventive services vary depending on the service.

There are five indicators (difficult to see clearly; not easy to receive help from the neighbors if he/she would need it; cholesterol level was measured in the last year; blood glucose level was measured in the last year; and pulled out teeth because of dental caries or loose teeth) for which the conclusions regarding differences between Roma and non-Roma individuals were not the same when assessing ethnicity by self-reporting vs. interviewer-reporting. Each of the observed differences suggested that Roma status is worse if the interviewer-reporting approach is applied and is equal to non-Roma status if the method of self-reported ethnicity is applied.

Roma ethnicity as a health determinant independent of socioeconomic status

Using logistic regression to investigate differences between the characteristics of the two Roma definitions compared to the non-Roma population, it was found that for 31 indicators, there were no remarkable differences, whereas there were significant differences for 14 variables based on both Roma definitions. Differences between interviewer-reported and self-reported Roma ethnicity-based ORs were observed for seven indicators. However, the deviations of odds ratios from self-reporting and interviewer-reporting analyses were the same for these seven indicators, and the corresponding confidence intervals exhibited a wide overlap.

In the self-report-based Roma analysis, a body mass index (BMI) above the normal value was associated with a lower risk of respiratory system disorders (OR: 0.64), whereas respiratory system disorders were associated with a higher risk (OR: 1.88) among the Roma population. However, the use of glasses or contact lenses (OR: 0.47) and blood glucose measurement in the last year (OR: 0.65) were less frequent in the Roma based on interviewer-reporting analysis. Furthermore, obstructive pain hindering physical activity in the last 4 weeks (OR: 2.23), bleeding gums (OR: 1.87) and lost teeth (OR: 1.85) were more frequent in the Roma in the interviewer-reporting analysis.

The positive correlation between the point estimates for ORs using the two approaches was strong ($r = 0.840$, $p < 0.001$), with three outliers (risk of road traffic accidents, not taking medicine for respiratory diseases, and tooth cavities without dental filling).

Discussion

Prevalence of chronic diseases and activity-limiting disability in the Roma population

The crude prevalence of chronic diseases in the Roma population was remarkably lower than in non-Roma populations. The opposite difference was observed for each studied activity limitation, which can be attributed, theoretically, to the slower disease progression, resulting in fewer complications or to faster prognosis, resulting in early death. In light of the published results from many settings on the worse disease prognosis for Roma populations and some data on the elevated mortality risk among Roma populations, the latter explanation seems to be more likely.

The analyses controlled for the younger age structure of the Roma population, with age and sex adjustment, demonstrating that Roma people exhibit an elevated risk both for chronic disease occurrence and for each studied activity limitation. These results suggest that both disease development and prognosis in those with disease is accelerated in the Roma population.

After standardization by level of education, the risk of chronic disease was significantly mitigated but remained significant. This indicates that a considerable part of the age- and sex-adjusted excess risk was partly attributable to the poor education status of the Roma people.

The elevated age-, sex-, and education-corrected risk of activity limitations proved to be significant in the field of everyday life, learning-working, family life, and transport. In the field of communication, this risk was reduced in Roma people.

The more extensive adjustment by multiple regression modeling confirmed the findings achieved by age, sex, and education adjustment. The only qualitative difference was that the activity limitation in communication was not significant in the regression modeling. Two quantitative modifications were also observed. The risk estimation was significantly higher for chronic disease occurrence and was significantly decreased for activity limitations in transport in the regression models.

The impact of Roma ethnicity independent of sociodemographic factors

On the basis of the multiple regression modeling, the impact of Roma ethnicity on chronic disease occurrence was 14.53%, showing that the Roma ethnicity is a distinct and important risk factor. On the other hand, the Roma ethnicity did not prove to be a risk factor among Roma people with chronic disease with regard to each of the studied activity limitations. In different fields, the impact varied between -4.17 and +19.35%. The estimated population-level impact of 0.39% for chronic diseases and between -0.12 and +1.19% for activity limitations is modest at first glance.

Taking into consideration that the Roma population is considerably underestimated by the census, because the actual size of the Roma population may be 870,000, as estimated by a Hungarian study focused on the estimation of the real population size using external Roma identifiers based on active contributions of the local governments, the corrected attributable risk fraction for chronic disease risk is 1.16% (95% CI: 1.10–1.22).

The primary strengths of the census derived analysis

The primary strength of this census-based investigation is that it included the whole country, avoiding selection bias via preparation of a sampling frame. Furthermore, the census database comprised nearly 10 million respondents, ensuring high statistical power for the analyses.

The achieved response rates (78.36% for chronic disease, 87.09% for activity limitation among persons with chronic disease) were relatively high compared to those observed in the Hungarian implementations of the European Health Interview Surveys (72% in 2009 and 62% 2014), resulting in reduced selection bias in the census-based evaluation compared to the surveys.

However, this investigation could not avoid self-report-derived ethnicity misclassification and the remarkable underestimation of the Roma population size. Although this validity issue has to be acknowledged, the results from parallel applications of self-report and interviewer classifications in health surveys suggest that this validity issue has minor importance in the investigation of population-level health status differences between Roma and non-Roma people. This is a likely consequence of the fact that the self-declared Roma subjects tended to live in more segregated circumstances, have worse health status, and be less educated than those who did not declare Roma identity.

Differences in health status indicators for the Roma population due to the parallel application of self and external classification of ethnicity

Our observation confirmed the common belief that observer reports are more effective in identifying Roma adults than the self-reporting approach. In the case of Roma adults, the intention not to admit one's Roma ethnicity is stronger than the misclassification by an observer who assesses the Roma ethnicity by obtaining information during the interview. In fact, application of the observer-reported Roma classification resulted in 1.44-fold more identified Roma individuals (N = 179) than the application of only the self-identification (N = 124) approach.

According to the evaluation of the sociodemographic differences between only observer-identified and self-identified Roma adults, working Roma adults are more willing to reject admission to Roma ethnicity. Since one of the most important social characteristics of the Roma is their exclusion from the labor market, this profile suggests that Roma individuals who can break out of this marginalized social position through employment may have a secretive attitude regarding their ethnicity. It seems that this subgroup can be reached by the application of observer reports classifying Roma individuals in health data collection.

The crude descriptive analysis showed significant differences between Roma and non-Roma groups for 35 of the 52 indicators investigated. There was only one indicator shown to reflect better conditions in the Roma BMI above normal value ($\geq$25 kg/m^2). According to the majority of studied indicators (30 in self-reporting and 35 in observer-reporting analyses), the health status of the Roma was disadvantageous compared to that of the non-Roma. Each difference between self-reporting and observer-reporting results showed the Roma health status was more disadvantageous in the case of observer reporting. The added value of observer reporting in Roma health studies can be presented by these 5 of 52 investigated indicators. This increased efficacy of the observer-reporting approach in demonstrating health status differences between

the Roma and non-Roma ethnicities also confirms the reduced reliability of self-reporting of the Roma ethnicity.

Since the Roma ethnicity covaried positively with deprivation, the indicators for Roma-to-non-Roma differences, without adjustment for sociodemographic status, are obviously not informative regarding the role of Roma ethnicity in influencing risk. The indicators corrected by sociodemographic factors confirmed the results from univariate analyses, such that the Roma health status was shown to be inferior to that of the non-Roma health status (with the exception of BMI above 25 kg/m^2). However, the number of adjusted indicators with statistically significant Roma-to-non-Roma differences was remarkably reduced compared to unadjusted indicators (self-reporting: 16 out of 30; observer-reporting: 19 out of 35).

The primary strengths of the parallel application of self- and external classification of Roma ethnicity

Indicators with statistically significant differences between Roma and non-Roma individuals that could be interpreted differently by self-reporting and observer reporting do not unequivocally support the higher sensitivity of observer reporting. Observer reporting showed higher efficacy for five of seven indicators, while self-reporting proved to be more effective for two of seven indicators. Our results suggest that the increased efficacy of observer reporting in Roma identification and in crude descriptive evaluation is not accompanied by higher efficacy in the evaluation of sociodemographically adjusted indicators.

The present study was a population-based investigation with the sample selected at random. The size of the non-Roma population was considerably large, ensuring relatively precise reference values for Roma-specific risk evaluation. The quality of the collected data was ensured by the application of questions from the European Health Interview Survey, which was also tested in a Hungarian national survey. The health-determining role of ethnicity could be studied with control for deprivation because Roma-specific risks were adjusted for a number of sociodemographic factors. The primary strength of this study is the parallel use of self-reporting and interview-reporting identification, allowing a direct comparison of the two methods.

Limitations of the studies

The census-based investigation could not avoid self-report-derived ethnicity misclassification and the remarkable underestimation of the Roma population size. There was also a misclassification with respect to the studied outcomes due to self-declaration. Because health status was not described in detail, the importance of this potential bias may be modest. On the other hand, the low specificity of the outcome classification restricts the utility of census-derived data in elaborating and monitoring interventions.

The most important limitations of this population-based study are the low response rate and the weak statistical power due to the relatively small number of Roma subjects in the studied sample. This small number of Roma subjects likely resulted in a type II error, which is responsible for the lack of any observable differences between the Roma risks computed by the two approaches, whereas many Roma-to-non-Roma differences were detected by both methods.

We could not investigate the added value of interviewer-reported ethnicity assessment, as an extra question in the survey added to questions on self-reported ethnicity. Odds ratios for interaction could not be computed by logistic regression models with terms for interaction between self-reported and interviewer-reported Roma ethnicity in the case of many indicators due to the small number of Roma participants in our survey. Therefore, the direct measure for the added value of interviewer-reporting ethnicity assessment as an additional question could not be computed using our database.

Conclusions

Taking into account that the census can estimate the health impact of the Roma ethnicity in the Roma population and within the entire population, it is sufficient to use the census-collected data to obtain reliable sociodemographic factors to adjust the estimations of Roma ethnicity-related risks. However, the self-report-derived misclassification, which needs further investigations, seems not to be strong enough to prevent the application of census-derived findings in intervention planning and evaluation, the addition of the census to public health monitoring of the health of the Roma population seems to be justified.

However, the only available census-based evaluations do not meet the criteria of public health monitoring, primarily because the health status assessment is not detailed enough to provide decision-makers with an identifiable intervention target. Only the magnitude of Roma ethnicity-related health problems and the potential health gain that can be achieved by effective interventions in Roma can be identified by the census. The details of such interventions can be elucidated by population surveys and setting specific epidemiological investigations.

The census-derived data suggest a significant impact in the Roma population for chronic disease occurrence accompanied by a worsening prognosis influence on activity limitations in everyday life, learning-working, family life, and transport but not in self-sufficiency, communication, or community life. It is estimated that 1.16% of chronic diseases are attributed to the Roma ethnicity in the entire population. This measure is not directly comparable to the percentage of total loss of disability-adjusted life years (DALY). Since the consequences of chronic diseases are variable, leading to premature death and functional impairments, this $AR_{Population}$ is in the same range as the percentage of total DALY from road injuries (1.52%), breast cancer (1.30%), cardiomyopathy and myocarditis (1.24%), and alcohol use disorders (1.04%).

Although census data cannot determine in detail the targets for interventions, they can be used to estimate the general importance of Roma ethnicity-related problems and the urgency of intervention, which is not achievable using population-level surveys or setting specific epidemiological investigations, according to previous experience. A census-based impact assessment of Roma ethnicity on chronic disease occurrence of 1.16% was not formerly described by analytic investigations or household surveys. Since sociodemographic risk adjustment can be performed using census data, the census provides useful data, despite the uncertainties of self-declared ethnicity assessments. Altogether, the census, in addition to population-level surveys and subgroup-specific epidemiologic investigations, can significantly improve the health-monitoring system of the Roma population.

Although the young and employed Roma seem to be less willing to declare their Roma ethnicity than the older and unemployed Roma, there is no remarkable discrepancy in survey conclusions in the difference between Roma and non-Roma adults' health status whether we used ethnicity data based on self-reporting or interviewer reporting. Based on our observations adjusted by sociodemographic status, both approaches for ethnicity identification are equally applicable in surveys, and it seems that the hesitation to insert self-reported Roma ethnicity into the set of surveyed indicators due to the assumed uncertain nature of self-identification is not justified.

The health status differences between the Roma and non-Roma populations were much larger than those between self-reported and interviewer-reported Roma populations. Therefore, issues related to the value of self-reported Roma ethnicity data are not reasonable to prevent extending these surveys by Roma-specific data collection, although Roma identification based on the combination of self-reporting and interviewer-reporting approaches yields remarkably larger Roma subgroups in surveys.

Primary statements and results

The importance of multivariate analyses in the estimation of chronic illness and disability prevalence in the Roma population

The crude prevalence of chronic diseases in the Roma population was remarkably lower than among non-Roma populations. The opposite difference was observed for each studied activity limitation. The analyses controlled for age and sex demonstrated that Roma people have an elevated risk both for chronic disease occurrence and for each studied activity limitation. After extending the standardization by level of education, the risk of chronic disease was significantly mitigated but remained significant. The elevated age-, sex-, and education-corrected risk of activity limitations proved to be significant in the field of everyday life, learning-working, family life, and transport.

The risk of Roma ethnicity was independent of sociodemographic factors

Our results showed that the more extensive adjustment by multiple regression modeling significantly decreased the crude risk of chronic diseases and the activity limitations. Controlling the analyses for demographic factors (age, sex) is essential but not sufficient in analyzing ethnic data because a considerable part of the age- and sex-adjusted excess risk is attributable to the poor education status of the Roma people. Similar conclusions could not be drawn if we used more extensive adjustment by multiple regression modeling. In conclusion, these results support the theory that Roma ethnicity is an independent risk factor because only a part of the total risk can be explained by the age, sex, and education differences between Roma and non-Roma populations.

The impact of Roma ethnicity independent of sociodemographic factors using census-derived data

The corrected total attributable risk fraction for chronic disease risk was 1.16% (95% CI: 1.10–1.22). Although this measure is not directly comparable to the percentage of total loss of disability-adjusted life years (DALY), since the consequences of chronic diseases are variable, leading to premature death and functional impairments, this attributable risk is in the same range as the percentage of total DALY from road injuries (1.52%), breast cancer (1.30%), cardiomyopathy and myocarditis (1.24%), and alcohol use disorders (1.04%).

The added value of census-based data to monitoring the health gap between Roma and non-Roma populations

The census provides useful data, despite the lack of detailed information about the respondent's health status, and the low specificity of the outcome classification restricts the usefulness of census-derived data in elaborating and monitoring interventions. According to our results, the magnitude of Roma ethnicity-related health problems and the potential health gain that can be achieved through interventions in Roma populations can be evaluated by the census.

The self-reporting and interviewer-reporting ethnicity classification was used to assess the health status of the Roma population

Although observer reports are more effective in identifying Roma adults than the self-reporting approach, they cannot significantly improve the efficacy of data-driven health policy

formulation. Our results show that it is not worth undertaking the elaboration of methodology that can handle all the sensitive (historical, legal, and ethical) issues related to the external Roma classification.

Limitation of the self-reporting and interviewer-reporting ethnicity classification in Roma health surveys

Health status differences between the Roma and non-Roma populations were much larger than those between self-reported and interviewer-reported Roma populations. Therefore, issues related to the limitations of self-reported Roma ethnicity data do not seem to preclude extending these surveys through external Roma classification, although Roma identification based on the combination of self-reporting and interviewer-reporting approaches yielded remarkably larger Roma subgroups in the surveys.

Publication list

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/180/2021.PL
Subject: PhD Publication List

Candidate: Ferenc Vincze
Doctoral School: Doctoral School of Health Sciences

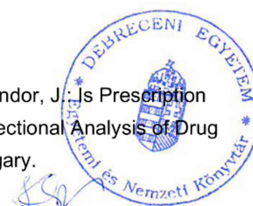
List of publications related to the dissertation

1. **Vincze, F.**, Földvári, A., Pálinkás, A., Sipos, V., Janka, E. A., Ádány, R., Sándor, J.: Prevalence of Chronic Diseases and Activity-Limiting Disability among Roma and Non-Roma People: A Cross-Sectional, Census-Based Investigation.
   Int. J. Environ. Res. Public Health. 16, 1-15, 2019.
   DOI: http://dx.doi.org/10.3390/ijerph16193620
   IF: 2.849

2. Janka, E. A., **Vincze, F.**, Ádány, R., Sándor, J.: Is the Definition of Roma an Important Matter? The Parallel Application of Self and External Classification of Ethnicity in a Population-Based Health Interview Survey.
   Int. J. Environ. Res. Public Health. 15, 353-374, 2018.
   DOI: http://dx.doi.org/10.3390/ijerph15020353
   IF: 2.468

List of other publications

3. Llanaj, E., **Vincze, F.**, Kósa, Z., Bárdos, H., Diószegi, J., Sándor, J., Ádány, R.: Deteriorated Dietary Patterns with Regards to Health and Environmental Sustainability among Hungarian Roma Are Not Differentiated from Those of the General Population.
   Nutrients. 13 (721), 1-15, 2021.
   DOI: http://dx.doi.org/10.3390/nu13030721
   IF: 4.546 (2019)


DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár

4. Begashaw, B., Harsha, N., Kőrösi, L., **Vincze, F.**, Czifra, Á., Ádány, R., Sándor, J.: Is Prescription Nonredemption a Source of Poor Health Among the Roma? Cross-Sectional Analysis of Drug Consumption Data From the National Health Insurance Fund of Hungary.
   Front. Pharmacol. 12, 1-9, 2021.
   DOI: http://dx.doi.org/10.3389/fphar.2021.616092
   IF: 4.225 (2019)

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

5. Nagy-Pénzes, G., **Vincze, F.**, Bíró, É.: Contributing Factors in Adolescents' Mental Well-Being: the Role of Socioeconomic Status, Social Support, and Health Behavior.
Sustainability. 12 (22), 1-17, 2020.
DOI: http://dx.doi.org/10.3390/su12229597
IF: 2.576 (2019)

6. Llanaj, E., **Vincze, F.**, Kósa, Z., Sándor, J., Diószegi, J., Ádány, R.: Dietary Profile and Nutritional Status of the Roma Population Living in Segregated Colonies in Northeast Hungary.
Nutrients. 12 (9), 1-21, 2020.
DOI: http://dx.doi.org/10.3390/nu12092836
IF: 4.546 (2019)

7. Nagy-Pénzes, G., **Vincze, F.**, Sándor, J., Bíró, É.: Does Better Health-Related Knowledge Predict Favorable Health Behavior in Adolescents?
Int. J. Environ. Res. Public Health. 17, 1-12, 2020.
DOI: http://dx.doi.org/10.3390/ijerph17051680
IF: 2.849 (2019)

8. Kasabji, F., Alrajo, A., **Vincze, F.**, Kőrösi, L., Ádány, R., Sándor, J.: Self-Declared Roma Ethnicity and Health Insurance Expenditures: a Nationwide Cross-Sectional Investigation at the General Medical Practice Level in Hungary.
Int. J. Environ. Res. Public Health. 17, 1-17, 2020.
DOI: http://dx.doi.org/10.3390/ijerph17238998
IF: 2.849 (2019)

9. Pálinkás, A., Kovács, N., Sipos, V., **Vincze, F.**, Papp, M. C., Czifra, Á., Ádány, R., Sándor, J.: Az indikátoralapú teljesítményértékelésre épülő forráselosztás hatékonysága Magyarországon a felnőtteket ellátó háziorvosi praxisokban.
Orv. Hetil. 160 (39), 1542-1553, 2019.
DOI: http://dx.doi.org/10.1556/650.2019.31464
IF: 0.497

10. Nagy, A. C., Kovács, N., Pálinkás, A., Sipos, V., **Vincze, F.**, Szőllősi, G. J., Ádány, R., Czifra, Á., Sándor, J.: Improvement in quality of care for patients with type 2 diabetes in Hungary between 2008 and 2016: results from two population-based representative surveys.
Diabetes Ther. 10 (2), 757-763, 2019.
DOI: https://doi.org/10.1007/s13300-019-0582-x
IF: 3.179

11. Sándor, J., Pálinkás, A., **Vincze, F.**, Sipos, V., Kovács, N., Jenei, T., Falusi, Z., Pál, L., Kőrösi, L., Papp, M. C., Ádány, R.: Association between the General Practitioner Workforce Crisis and Premature Mortality in Hungary: cross-Sectional Evaluation of Health Insurance Data from 2006 to 2014.
Int. J. Environ. Res. Public Health. 15 (7), 1388-, 2018.
DOI: http://dx.doi.org/10.3390/ijerph15071388
IF: 2.468

DEBRECENI EGYETEM

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

12. Nagy, A. C., Kovács, N., Pálinkás, A., Sipos, V., **Vincze, F.**, Szőllősi, G. J., Csenteri, O. K., Ádány, R., Sándor, J.: Exploring quality of care and social inequalities related to type 2 diabetes in Hungary: nationwide representative survey.
    Prim. Care Diabetes. 12 (1), 1-13, 2018.
    DOI: http://dx.doi.org/10.1016/j.pcd.2017.12.004
    IF: 2.008

13. Sándor, J., Pálinkás, A., **Vincze, F.**, Kovács, N., Sipos, V., Kőrösi, L., Falusi, Z., Pál, L., Fürjes, G., Papp, M. C., Ádány, R.: Healthcare utilization and all-cause premature mortality in Hungarian segregated roma settlements: evaluation of specific indicators in a cross-sectional study.
    Int. J. Environ. Res. Public Health. 15 (9), 1-13, 2018.
    DOI: http://dx.doi.org/10.3390/ijerph15091835
    IF: 2.468

14. Sándor, J., Nagy, A. C., Jenei, T., Földvári, A., Szabó, E., Csenteri, O. K., **Vincze, F.**, Sipos, V., Kovács, N., Pálinkás, A., Papp, M. C., Fürjes, G., Ádány, R.: Influence of patient characteristics on preventive service delivery and general practitioners' preventive performance indicators: a study in patients with hypertension or diabetes mellitus from Hungary.
    Eur. J. Gen. Pract. 24 (1), 183-191, 2018.
    DOI: http://dx.doi.org/10.1080/13814788.2018.1491545
    IF: 1.617

15. Sipos, V., Pálinkás, A., Kovács, N., Csenteri, O. K., **Vincze, F.**, Szőllősi, G. J., Jenei, T., Papp, M. C., Ádány, R., Sándor, J.: Smoking cessation support for regular smokers in Hungarian primary care: a nationwide representative crosssectional study.
    BMJ Open. 8, 1-8, 2018.
    DOI: http://dx.doi.org/10.1136/bmjopen-2017-018932
    IF: 2.376

16. Sándor, J., Nagy, A. C., Földvári, A., Szabó, E., Csenteri, O. K., **Vincze, F.**, Sipos, V., Kovács, N., Pálinkás, A., Papp, M. C., Fürjes, G., Ádány, R.: Delivery of cardio-metabolic preventive services to Hungarian Roma of different socio-economic strata.
    Fam. Pr. 34 (1), 83-89, 2017.
    DOI: http://dx.doi.org/10.1093/fampra/cmw102
    IF: 1.675

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

17. Földvári, A., Kovács, N., Sipos, V., Merth, G., **Vincze, F.**, Szűcs, M., Sándor, J.: Estimation of incidence, prevalence, and age-at-diagnosis of myasthenia gravis among adults by hospital discharge records.
    Wien. Klin. Wochen. 127 (11-12), 459-464, 2015.
    DOI: http://dx.doi.org/10.1007/s00508-015-0796-5
    IF: 0.836

Total IF of journals (all publications): 44,032
Total IF of journals (publications related to the dissertation): 5,317

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

08 April, 2021

Acknowledgments

I would like to thank my supervisor, Prof. Dr. Sandor Janos, for allowing me to work at the Department of Public Health and Epidemiology, University of Debrecen and for always assisting me in discussing the research results with her professional guidance.

I would like to express my gratitude to Prof. Dr. Róza Ádány and Prof. Dr. Margit Balázs for all of their support during my PhD study.

I am thankful for my colleague Eszter Anna Janka for her always generous help and critique during my experiments.

I would like to express my gratitude to my family, my sister and my parents for their generous support, patience, and constant encouragement. Finally, I would like to thank each of my friends for their spiritual support and for always being with me.

I would like to express my sincere respect for my colleagues.