RESEARCH ARTICLE

REVISED A survey of quality of life indicators in the Romanian Roma population following the 'Decade of Roma Inclusion' [version 3; peer review: 2 approved, 1 approved with reservations]

Rebecca Powell Doherty[1,2], Pyrros A. Telionis[1,3], Daniel Müller-Demary[4], Alexandra Hosszu[4], Ana Duminica[4], Andrea Bertke[2], Bryan Lewis[1], Stephen Eubank[1]

[1]Biocomplexity Institute of Virginia Tech, Blacksburg, VA, USA
[2]Population Health Sciences, Virginia Tech, Blacksburg, VA, USA
[3]Department of Geography, Virginia Tech, Blacksburg, VA, USA
[4]Agentia Impreuna, Bucharest, Romania

v3 **First published:** 15 Sep 2017, **6**:1692
https://doi.org/10.12688/f1000research.12546.1
Second version: 08 Mar 2018, **6**:1692
https://doi.org/10.12688/f1000research.12546.2
Latest published: 13 Dec 2018, **6**:1692
https://doi.org/10.12688/f1000research.12546.3

Abstract

Background: This study explores how the Roma in Romania, the EU's most concentrated population, are faring in terms of a number of quality of life indicators, including poverty levels, healthcare, education, water, sanitation, and hygiene. It further explores the role of synthetic populations and modelling in identifying at-risk populations and delivering targeted aid.
Methods: 135 surveys were conducted across five geographically diverse Romanian communities. Household participants were selected through a comprehensive random walk method. Analyses were conducted on all data using Pandas for Python. Combining land scan data, time-use survey analyses, interview data, and ArcGIS, the resulting synthetic population was analysed via classification and regression tree (CART) analysis to identify hot-spots of need, both ethnically and geographically.
Results: These data indicate that the Roma in Romania face significant disparities in education, with Roma students less likely to progress beyond 8 th grade. In addition, the Roma population remains significantly disadvantaged with regard to safe and secure housing, poverty, and healthcare status, particularly in connection to diarrheal disease. In contrast, however, both Roma and non-Roma in rural areas face difficulties regarding full-time employment, sanitation, and water, sanitation, and hygiene infrastructure. In addition, the use of a

Open Peer Review

Reviewer Status ? ✓ ✓

	Invited Reviewers		
	1	**2**	**3**
version 3 (revision) 13 Dec 2018		✓ report	✓ report
version 2 (revision) 08 Mar 2018	? report	? report	✗ report
version 1 15 Sep 2017	? report	? report	✗ report

1. **Ciprian Marius Ceobanu**, Alexandru Ioan Cuza University, Iași, Romania
2. **Martin McKee**, London School of Hygiene and Tropical Medicine, London, UK
3. **Pilar Carrasco-Garrido**, King Juan Carlos University, Madrid, Spain

synthetic population can generate information about 'hot spots' of need, based on geography, ethnicity, and type of aid required. **Conclusions**: These data demonstrate the challenges that remain to the Roma population in Romania, and also point to the myriad of ways in which all rural Romanians, regardless of ethnicity, are encountering hardship. This study highlights an approach that combines traditional survey data with more wide-reaching geographically based data and CART analysis to determine 'hot spot' areas of need in a given population. With the appropriate inputs, this tool can be extrapolated to any population in any country.

Keywords
Roma, Romania, rural populations, water quality, healthcare, development, global health, decade of Roma inclusion

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Rebecca Powell Doherty (rpdoh15@vt.edu)

Author roles: Powell Doherty R: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Telionis PA**: Formal Analysis, Methodology, Validation, Writing – Review & Editing; **Müller-Demary D**: Data Curation, Methodology, Resources, Writing – Review & Editing; **Hosszu A**: Conceptualization, Data Curation, Investigation, Writing – Review & Editing; **Duminica A**: Data Curation; **Bertke A**: Conceptualization, Project Administration, Supervision, Writing – Review & Editing; **Lewis B**: Formal Analysis, Funding Acquisition, Methodology, Software, Supervision, Visualization, Writing – Review & Editing; **Eubank S**: Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work has been partially supported by the National Institutes of Health and National Institute of General Medical Sciences - Models of Infectious Disease Agent Study Grant 5U01GM070694-13, the Defense Threat Reduction Agency - Comprehensive National Incident Management System Contract HDTRA1-11-D-0016-0001, and the Virginia-Maryland College of Veterinary Medicine.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Powell Doherty R, Telionis PA, Müller-Demary D *et al.* **A survey of quality of life indicators in the Romanian Roma population following the 'Decade of Roma Inclusion' [version 3; peer review: 2 approved, 1 approved with reservations]** F1000Research 2018, **6**:1692 https://doi.org/10.12688/f1000research.12546.3

First published: 15 Sep 2017, **6**:1692 https://doi.org/10.12688/f1000research.12546.1

REVISED **Amendments from Version 2**

We have updated our manuscript to include new data analysis and the inclusion of a synthetic population model that demonstrates the utility of augmenting more traditional survey methods. This has resulted in a further five figures, including a visual depiction of our additional methodology, as well as geographical depictions of our 'hot-spot' analysis. We have also edited our manuscript for clarity in numerous areas, particularly taking into account the responses from our reviewers. We stress that our new data and the use of the synthetic population is a proof of concept only, and further iterations with larger survey sample sizes will be necessary to truly validate the method.

Pyrros A. Telionis has been added as a new author to the version 3-this change results from the addition of the new analysis and figures, to which Mr Telionis contributed substantially. As Bryan Lewis and Stephen Eubank are no longer affiliated with Virginia Tech, Rebecca Powell Doherty is the sole corresponding author for version 3. These changes have been agreed with all authors.

See referee reports

Introduction

In the years that followed independence and the democratic election of 1990, the southeastern European country of Romania received significant aid from the International Monetary Fund (IMF), World Bank (WB), European Bank for Reconstruction and Development (EBRD), European Investment Bank (EIB), the US Agency for International Development (USAID), and other donors[1]. This influx of investment enabled Romania to make great strides in multiple areas of development and meet a number of the goals set forth in the United Nations Millennium Development Goals (UN MDGs)[2]. In particular, the issues of severe poverty and hunger have significantly improved for ethnic Romanians and affluent minorities, with severe poverty (as defined by the United Nations) decreasing from 10 per cent to 4.1 per cent as of 2006[2]. In addition, maternal mortality has fallen by half to 17 deaths/100,000 births, infant mortality has decreased 25 per cent, and Romania has seen a significant decrease in adolescent pregnancy, concomitant with a significant increase in the use of modern contraceptives[2]. In the 1990's and early 2000's, vaccination rates, particularly for measles, improved to around 98 per cent, up from less than 70 per cent at the time of independence; HIV/AIDS cases have decreased and life expectancy for those living with HIV has increased dramatically; and there has been a significant decrease in domestic violence[2].

For the Roma, the second most numerous minority in the country (after Hungarians), however, such progress was not extended. Despite enjoying a reprieve from targeted discrimination during the Soviet era, Romanian independence brought on a renewal of oppressive policies and behaviours against the Roma. The Roma are Europe's most marginalised group[3], a minority population numbering between 10–12 million individuals across the continent and the UK[4]. Emerging from slavery in the late 19th century, they have historically faced discrimination in employment, education, and access to healthcare[5]. Numerous studies indicate Roma have a significantly reduced lifespan compared to non-Roma and suffer greater rates of communicable and waterborne diseases[6–8]. In multiple countries, they are less likely to have access to basic services, including a municipal water supply, waste water treatment, or trash disposal[9],. Romania boasts the largest concentration of Roma in the European Union (EU), at approximately 1.85 million individuals, representing 9.3 per cent of the overall population of 19.8 million, though official census numbers vary[4].

The addition of eastern European countries (including Bulgaria, Romania, and Hungary) to the EU in the mid-2000s has renewed interest in the well-being of this population, as indicated by the EU's targeted attempt to improve the circumstances of the Roma through the recently concluded *Decade of Roma Inclusion* (DRI), a ten year long initiative by twelve European countries to improve the socio-economic status and social standing of the Roma minority across the continent[10]. Numerous studies have explored the success of the DRI, both during its implementation and since its conclusion, and outcomes vary, depending on the sector and goal in question[8,11–13].

For such assessments, international aid agencies and non-governmental organizations often employ assessment surveys and interviews to determine the type and level of need in a particular area or for a disadvantaged population[14]. However, while such methods are useful for specific communities 'of interest' and can provide statistical support for straightforward claims or goals, they are of little use in identifying new areas and populations in need or addressing multi-faceted and complex issues. This proof-of-concept study explores the possibility of using a synthetic representation of Romania (down to the individual level) to predict currently unrecognized areas of need based on key variables from an assessment survey and a classification and regression tree (CART) analysis. The synthetic population was generated from the fusion of land-scan data, geographic census data and ethnicity statistics, time-use surveys, and our own needs assessment survey data. Our representation captures details about households and their quality of life, and is able to capture heterogeneities across geographic space. This approach augments the strength of the survey, in particular allowing the identification of potential areas of need without requiring additional resources to conduct a needs assessment in those regions. Furthermore, the synthetic population becomes an ideal foundation for dynamic simulations and can be used to identify sub-populations at greatest risk for infection during disease outbreaks.

Methods

Regional survey

We developed our survey by combining questions adapted from a validated WASH survey previously used for multiple use service strategy research (personal communication to *authors*) and the WHO core questions on drinking-water and sanitation[15] with questions related to demographics, socio-economic status, and healthcare access and history, we conducted 135 surveys each consisting of 56 total questions across five geographically diverse communities throughout Romania. The survey questions

were modified with the assistance of our NGO partner to appropriately reflect cultural characteristics in Romania. Communities were chosen from a list of those that had previously participated with *Agentia Impreuna* in education and anti-discrimination capacity-building programs for communities with prominent Roma populations. In addition, in an attempt to address geographical bias in improve the generalizability of our findings, communities were identified for their geographical diversity. Participating communities included central urban households, suburban communities, and very rural, mountainous regions. Communities were further distinct in the level of integration observed between the Roma population and the non-Roma, being fully integrated in some areas and completely separate in others. Household participants were selected through a comprehensive random walk method, with survey teams accompanied by both Roma and non-Roma community leaders. Survey teams varied the time of day they moved through any given community to ensure access to the full population, and interviews were conducted in areas throughout the community, with participants identified at their homes, as well as in shops and cafes. Identifying information for the participants was used only to ensure there was no duplication of household information. Any household with an individual over the age of 18 present and willing to participate, regardless of ethnicity, was included until the desired 30 surveys per community were achieved or there were no further willing participants. Interviews were conducted by trained volunteers who either spoke the national language (Romanian) or were accompanied by a certified translator. The team interviewed only one member of each household, who provided information about all members of the household. The specifics of participating communities are purposefully withheld to comply with the approval constraints of our ethics board.

Ethical statement

Surveys (Supplementary material 1 and Supplementary material 2) and procedures were approved by the Virginia Tech Institutional Review Board (IRB) prior to study implementation (VT IRB #16-475), and all interviews and analysis were carried out according to IRB protocol.

IRB protocol and participant protections

Informed consent was obtained from all individual participants included in this study. A brief explanation of the survey questions and the intended use of the data was provided to each participant, and the individual's agreement to participate in the survey interview was considered consent, as indicated by the IRB protocol. Further, interviewers ensured each participant understood that he or she could refuse to answer any question and could withdraw their consent at any time. Survey participation was anonymous, and no identifying information was retained. In addition, the IRB stipulated that location data for the participating villages remain unavailable, due to the vulnerable population and minority status of some study participants. All demographic information was self-reported, and those who were considered part of the Roma sample self-identified as either Roma or Rudar (a sub-set of Roma people who do not speak Romani), in response to a question that explicitly asked for their ethnicity (Dataset 1).

Synthetic Population Generation. In order to generate a synthetic population for Romania that would allow us to explore variables of interest based upon geographic location and ethnicity, we fused data sets from multiple sources (Figure 1). To establish our base population, we populated the land-scan data from the Global Population Project[16] with data from the U.S. Census Bureau International Database[17], which predicts global populations based on past census data and growth projections, along with time-use survey data from Russia (chosen as a substitute for specific similarities)[18], as there are no available time-use data from Romania[18]. We then used ArcGIS[19] to join this population to shape files that defined administrative regions of Romania, at the *judet* (county), city, town, and commune level[20]. Exporting our population, now defined geographically, to Python/Pandas[21], we merged it by geographic region to ethnicity data, counts of individuals reporting to be from various ethnic groups, obtained from the Romanian National Institute of Statistics[22]. We were then able to assign each household in the population an ethnicity (Roma or non-Roma), and identify regions of the country with concentrated Roma populations. Finally, we applied a CART analysis (Figure 2), based upon our pilot survey data, to the synthetic population, and exported data related to our variables of interest (insecure housing, education level, water quality, diarrheal rates, parameters of poverty, and urban *versus* rural communities) to ArcGIS for visualization.

Classification and Regresssion Tree (CART) Analysis. Using the synthetic population described, we used a classification and regression tree (CART) analysis to identify how ethnicity, household size, and age structure of the household predicted the responses to seven of the most significant quality of life indicators The resulting tree grouped many of the surveys into similar pools based on these three predictor values (Figure 2). Acknowledging the small sample size of our pilot survey (n=135), we aggregated categories containing only a single household into four larger groupings. Using independent variables previously identified, the univariate classification tree produced five overall categories based on this analysis. Each household in the population was then assigned an individual survey response from its corresponding pool based on the ethnicity, household size, and age structure of the household.

Primary data analysis

All data analyses were conducted via Pandas with Python (version 2.7.11 & 0.18.0) notebook and the software package Epipy[21,22] (Dataset 2–Dataset 3). Descriptive statistics were broken down by community, ethnicity, gender, age, household size, education level, marital status, employment, literacy, and geographical description (urban *versus* rural). WASH parameters were defined using the UN descriptions as provided in the DRI progress report through 2013, as well as the addition of a 'safe water score', which included the option of a private, protected well water source in addition to tap water in the home[10]. The overall WASH score for each participating household is an aggregate of the following UN parameters: indoor toilet (improved sanitation), indoor bathroom (improved sanitation II), piped water to tap (improved water source), and insecure housing (a 0–3 score reflecting the status of the floor, walls, and

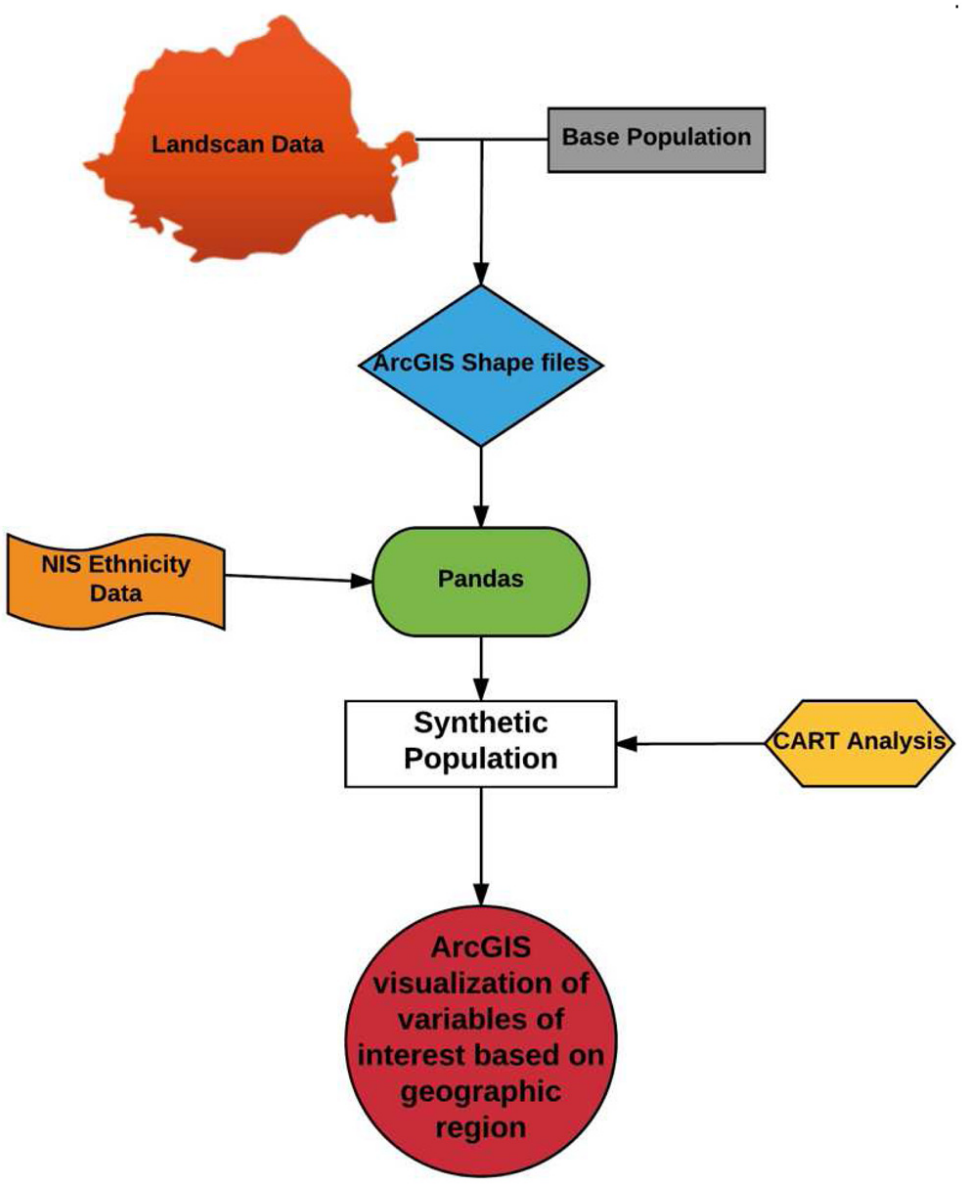

Figure 1. Work Flow to Generate Synthetic Population and ArcGIS mapping.

roof of a dwelling). The overall 'WASH Safe' score exchanged the improved water source parameter for the aforementioned safe water score. In addition, time to primary drinking water sources has been converted to a numerical scale, based on 15 minute intervals, up to one hour (0–4 scale). Distance to primary drinking water is indicated both by a percentage of those in each ethnic group who travel a kilometre or more and the average distance travelled by each group. Similar to the WASH score, the healthcare score is an aggregate of self-reported immunization, reported incidence of diarrheal event, access to primary care physician (PCP), and reported medical insurance status. Finally, the poverty score is an aggregate of available electricity in dwelling, available gas source in dwelling, and the UN indicator of severe poverty (surviving on 2USD/person/day or less). Univariate analyses compared the Roma sample to the non-Roma sample for each variable (using non-Roma as the reference population), as well as urban areas to rural ones (with urban areas as the reference population) for some parameters. Odds Ratios (ORs) with 95 per cent confidence intervals are reported, as are t-test results (95 per cent confidence interval) with accompanying p-value where appropriate.

Secondary data analysis and multivariate models

Multivariate linear regression analyses were conducted by using combinations of the four aggregate scores, as explained in primary analysis, and by including parameters that demonstrated significance in univariate modelling (Dataset 2–Dataset 3).

Hot Spot Generation. Using the Spatial Autocorrelation (Global Moran's I) tool in ArcGIS, which measures spatial autocorrelation based on feature locations and feature values, we analyzed each variable of interest to determine whether the pattern expressed

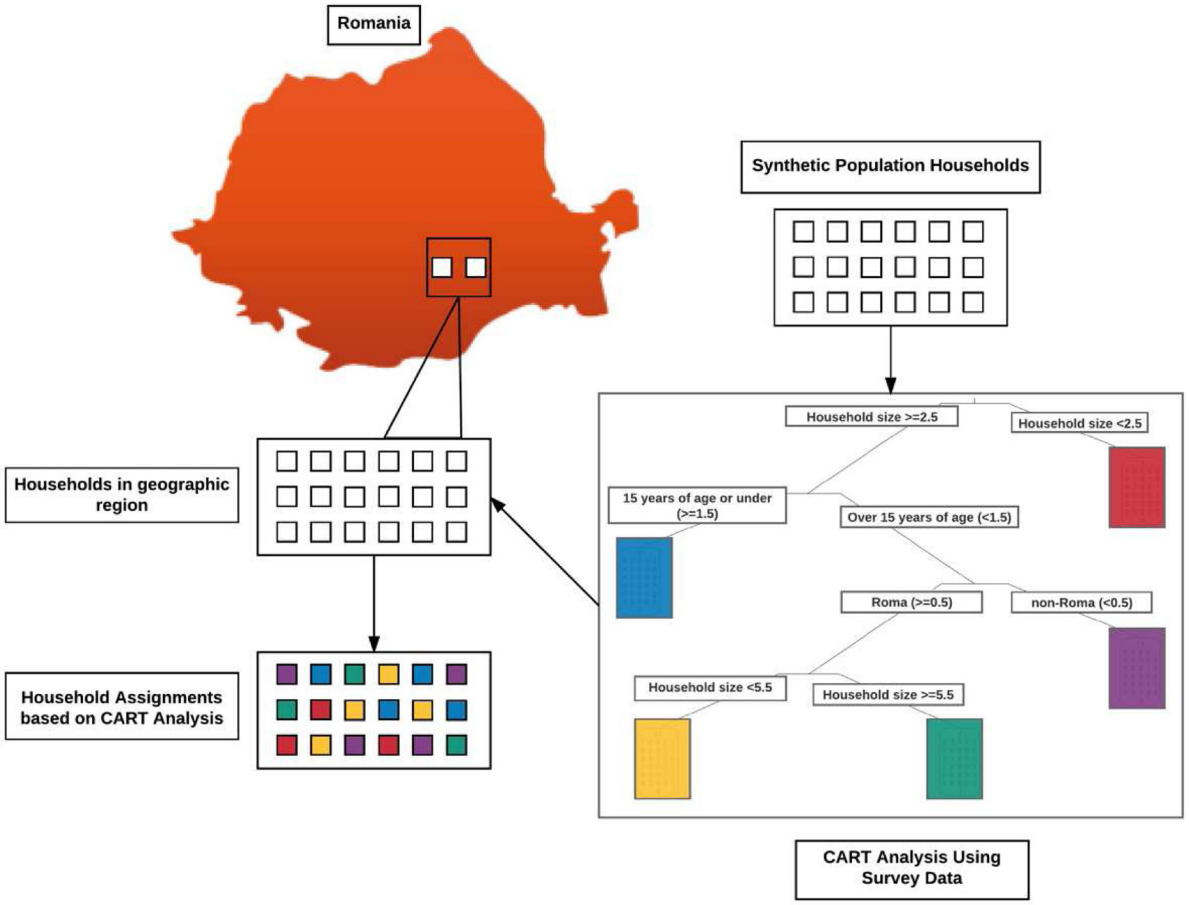

Figure 2. Work-flow for CART Analysis and categorical variable assignment in synthetic population.

in our population was random. Significant autocorrelation (non-random pattern or clustering) was determined by z-score and accompanying p-value (≤ 0.05). Significance or lack thereof suggests whether the independent variables upon which our synthetic population was built (household size and ethnicity) are appropriate indicators for our specific quality of life (QoL) parameters. For variables demonstrated to be significantly spatially auto-correlated, we progressed to Incremental Spatial Autocorrelation with a fixed distance measure to identify areas of intense need or 'hot spot' clustering.

Results

Population demographics

Analyses of demographic data and breakdown by percentage indicate our sample population is, overall, predominantly Roma (72.6 per cent *vs.* 27.4 per cent non-Roma), split evenly by sex (50.4 per cent Female, 49.6 per cent Male), and average approximately 47 years of age (Table 1). Three of the five sample communities are rural (more than 25km from a city centre), one is suburban (between 10–25km from a city centre), and one is urban (less than 10km from a city centre). There is no significant difference between Roma and non-Roma in the sample population on the basis of marital status, age, or sex. However, our data indicate notable disparities in level of education (secondary school completion for Roma *vs.* high school completion for non-Roma), household size (5.3 individuals for Roma *vs.* 4.2 individuals for non-Roma), and literacy rate (61 per cent literate Roma *vs.* 97.4 per cent literate non-Roma) (Table 1). Little difference is noted in full-time employment rates between the groups (26.6 per cent Roma *vs.* 32.4 per cent non-Roma), though some difference is observable between rural and urban communities (Table 1).

WASH, healthcare, poverty parameters

Using parameters utilized by the DRI in the 2011 progress report, univariate analysis indicates little difference between Roma and non-Roma with regard to specific WASH variables. The non-Roma are slightly more likely to have an indoor toilet (21.6 per cent non-Roma *vs* 17.3 per cent Roma) and bathroom (21.6 per cent non-Roma *vs* 20.4 per cent Roma), but the Roma are more likely than non-Roma to have tap (indoor or outdoor) water (20.4 per cent Roma *vs* 8.1 per cent non-Roma), whether piped in from a personal well or a municipal water source (Table 2). However, when considering all safe water options (including a protected well *without* a tap to the home or garden), non-Roma report greater accessibility (59.5 per cent non-Roma *vs* 50 per cent Roma). In addition, Roma are significantly more at risk to inhabit insecure housing, regardless of

Table 1. Study population demographics broken down by community. Romania, 2016. *M=male, F=female, FT=full-time, UE=unemployed, DL=day labour.*

	Population N (%)	*Sex % M(F)*	*Age of Respondents (Mean in years)*	*House-hold Size N (Mean no. persons)*	*Education Level of Respondents (Mean Grade Completed)*	*Marital Status % Partnership (Single)*	*Employment Status % FT (UE or DL)*	*Literacy of Families Overall (%)*	*Geographical Location*
Community 1									
Roma	*28 (96.6)*	*28.5 (71.4)*	*49.3*	*4.7*	*Secondary School (8th grade)*	*71.4 (28.6)*	*10.7 (89.3)*	*56*	--
Non-Roma	*1 (3.4)*	*100 (0)*	*52*	*5*	*Some University / College*	*100 (0)*	*0 (100)*	*80*	--
Overall	29	31 (69)	49.4	4.7	Secondary School (8th grade)	72.4 (27.6)	7.7 (89.7)	57.5	*Rural*
Community 2									
Roma	*24 (80)*	*75 (25)*	*48.1*	*5.9*	*Secondary School (8th grade)*	*87.5 (12.5)*	*16.7 (83.3)*	*58.9*	--
Non-Roma	*6 (20)*	*50 (50)*	*50.1*	*4.8*	*Some University / College*	*83.3 (16.7)*	*66.7 (33.3)*	*100*	--
Overall	30	70 (30)	45.8	5.7	Required School (10th grade)	86.7 (13.3)	*26.7 (73.3)*	66.7	*Rural*
Community 3									
Roma	*18 (60)*	*44.4 (55.6)*	*42.6*	*5.3*	*Secondary School (8th grade)*	*94.4 (5.6)*	*27.8 (72.2)*	*58.3*	--
Non-Roma	*12 (40)*	*58.3 (41.7)*	*50.1*	*4.2*	*High School + Vocational School*	*91.7 (8.3)*	*25 (75)*	*100*	--
Overall	30	50 (50)	45.6	4.9	Required School (10th grade)	93.3 (6.7	26.7(73.3)	74.4	*Suburban*

F1000Research 2018, 6:1692 Last updated: 26 JUL 2021

	Population N (%)	Sex % M(F)	Age of Respondents (Mean in years)	House-hold Size N (Mean no. persons)	Education Level of Respondents (Mean Grade Completed)	Marital Status % Partnership (Single)	Employment Status % FT (UE or DL)	Literacy of Families Overall (%)	Geographical Location
Community 4									
Roma	*13 (43.3)*	*46.2 (53.8)*	*42.7*	*6.1*	*Secondary School (8th grade)*	*84.6 (15.4)*	*23.1 (76.9)*	*48*	--
Non-Roma	*17 (56.7)*	*29.4 (70.6)*	*60.4*	*2.9*	*Required School (10th grade)*	*58.8 (41.2)*	*23.5 (76.5)*	*95.1*	--
Overall	30	36.7 (63.3)	52.7	4.3	Secondary School (8th grade)	70 (30	23.3 (76.7)	69.2	*Rural*
Community 5									
Roma	*15 (93.7)*	*66.7 (33.3)*	*35.4*	*4.7*	*High School (12th grade)*	*73.3 (26.7)*	*73.3 (26.7)*	*93.2*	--
Non-Roma	*1 (6.3)*	*100 (0)*	*40*	*4*	*Required School (10th grade)*	*100 (0)*	*100 (0)*	*100*	--
Overall	16	68.8 (31.2)	35.7	4.7	High School (12th grade)	75 (25)	75 (25)	93.6	*Urban*
Overall									
Roma	*98 (72.6)*	*51 (49)*	*44.8*	*5.3*	*Secondary School (8th grade)*	*75.5 (24.5)*	*26.6 (73.4)*	*61*	--
Non-Roma	*37 (27.4)*	*46 (54)*	*52.4*	*4.2*	*High School (12th grade)*	*75.6 (24.3)*	*32.4 (67.6)*	*97.4*	--
Total	135	49.6 (50.4)	46.9	4.8	Required School (10th grade)	80 (20)	28.1 (71.9)	72.3	--

F1000Research 2018, 6:1692 Last updated: 26 JUL 2021

Table 2. Univariate analyses. Romania, 2016. *Reference population for all variables is non-Roma. * indicates significance at 95% CI level. ** indicates significance at 90% CI level.*

		Roma	Non-Roma	t-statistic	p-value	Odds Ratio	95% CI
WASH	Improved Sanitation (Indoor Toilet, % yes)	17.3	21.6	0.567	0.57	1.31	0.51, 3.37
	Improved Sanitation II (Indoor Bathroom, % yes)	20.4	21.6	0.154	0.878	1.08	0.43, 2.71
	Improved Water Source (Piped water to tap, % yes)	20.4	8.1	-1.701	0.091**	0.34	0.1, 1.24
	Insecure Housing (% yes)	27.6	5.4	2.858	0.005*	6.65	1.5, 29.6
	Time to Primary Drinking Water Source (Mean, 0–4 scale, 15min intervals)	1.12	1.0	0.769	0.443	1.12	0.37, 3.43
	Distance to Primary Drinking Water Source (%, 1km or more)	12.2	10.8	0.124	0.901	1.15	0.35, 3.82
	Safe Water Source (tap or well, % yes)	50	59.5	0.978	0.329	1.47	0.8, 1.91
Healthcare	Moderate/Severe Diarrhea in Last Year (% yes)	58.1	40.5	-1.84	0.07**	2.04	0.94, 4.4
	Reports Immunization of any kind (% yes)	87.8	97.1	0.678	0.499	1.58	0.42, 5.96
	Medically Insured (% yes)	81.6	89.1	1.057	0.292	1.86	0.58, 5.9
	Access to PCP (% yes)	98	97	-0.231	0.818	0.75	0.07, 8.53
Poverty	Electricity in Home or Dwelling (% no)	13.2	2.7	1.804	0.07**	5.51	0.69, 43.68
	Piped or Tank Gas in Home or Dwelling (% no)	32.7	18.9	1.57	0.12	2.47	0.82, 5.24
	Spends more than $2/person/day (% no)	55.1	43.2	1.23	0.22	1.61	0.75, 3.45

geographical region, than non-Roma (27.6 per cent Roma *vs* 5.4 per cent non-Roma) (Table 2). Interestingly, while the Roma population have greater access to tap water (indoor or outdoor), they are less likely to use it as their primary drinking water source, demonstrated by the increased time and distance Roma are likely to travel to secure safe drinking water (12.2km Roma *vs.* 10.8km non-Roma; Table 2). Of interest, however, is the increased time all individuals in suburban and urban areas must travel to secure drinking water compared to their rural counterparts (16–30 minutes (1.2 on 0–3 scale) urban *vs.* 0–15 minutes (1.0 on 0–3 scale) rural) (Table 3).

In addition to physical infrastructure, we analysed the differences between Roma and non-Roma with regard to key factors contributing to overall health status. Roma are more than twice as likely to report at least one household member suffering from moderate to severe diarrhoea (lasting more than 3 days) than non-Roma (58.1 per cent Roma *vs* 40.5 per cent non-Roma; OR 2.04) (Table 2). In addition, while there is little difference in access to a primary care physician between the groups, Roma are approximately 1.5 times less likely to report having received an immunization of any kind (87.8 per cent Roma *vs* 97.1 per cent non-Roma; OR 1.58) and fewer Roma possess medical insurance (81.6 per cent Roma *vs* 89.1 per cent non-Roma; OR 1.86) than non-Roma (Table 2).

Finally, we used the UN definition of extreme poverty (2USD/person/day or less) in addition to two other variables as an overall indicator of impoverished conditions (Table 2). Roma report a slightly greater, though not significant, incidence of lacking working electricity in their homes or dwellings (13.2 per cent Roma *vs* 2.7 per cent non-Roma), as well as lacking piped gas and/or the ability to purchase gas tanks (32.7 per cent Roma *vs.* 18.9 per cent non-Roma, p=0.12) (Table 2). Moreover, Roma report greater incidences of severe poverty (2USD/day/person or less) than non-Roma (55.1% per cent *vs.* 43.2 per cent) (Table 2), although overall, those in rural areas are significantly more susceptible to extreme poverty than those in suburban or urban communities (61.8 per cent rural *vs.* 32.6 per cent urban) (Table 3).

Multivariate analyses

Following univariate analysis, we used general multivariate linear regression analysis for four distinct models, combining categories that indicated a specific score (WASH, WASH Safe, poverty, healthcare) or approached a level of significance

in the univariate analysis (Table 4). These analyses further demonstrate the significant ($\alpha = 0.05$) disparity between Roma and non-Roma.

A multivariate combination of demographic variables further highlights the difference in education level and household size between Roma and non-Roma. Roma households are significantly larger than non-Roma households, but whether this is a correlation with birth rate or the presence of multiple generations in a single dwelling is beyond the scope of this study. Furthermore, Roma individuals are far less likely to complete required education (10^{th} grade) than non-Roma individuals (MOD1; Table 4). In our univariate analysis, we broke down the score categories to their individual components and identified significant factors to further explore. Multivariate analysis of these parameters points to insecure housing as having the strongest correlation with being Roma, followed by access to tap water (improved water source), and less significantly, the occurrence of moderate or severe diarrhoea (MOD2; Table 4).

Finally, we analysed our four score categories, using two different approaches. We first analysed the WASH score, as defined by the DRI, together with the healthcare and poverty scores (MOD3; Table 4). Healthcare and poverty equally significantly correlate

Table 3. Geographical univariate analysis. Romania, 2016. *Reference population for all variables is urban. * indicates significance at 95% CI level.*

	Rural	*Urban*	*t-statistic*	*p-value*	*Odds Ratio*	*95% CI*
Time to Primary Drinking Water Source (Mean, 0–4 scale, 15min intervals)	*1.0*	*1.2*	*1.306*	*0.194*	*0.53*	*0.19, 1.49*
Spends more than $2/person/day (% no)	*61.8*	*32.6*	*3.323*	*0.001**	*3.3*	*1.58, 7.08*

Table 4. Multivariate analysis modelling. Romania, 2016. *All models use non-Roma as reference. * indicates significance at 95% CI level. ** indicates significance at 90% CI level.*

MOD1	**Regression coefficient**	**p-value**	**95% Confidence Interval**
Property Documents	0.0854	0.279	0.069, 0.240
Education Level	0.2613*	0.001	0.100, 0.422
Household Size	0.2362*	0.002	0.083, 0.389
Employment Status	0.0505	0.559	-0.119, 0.220
MOD2	**Regression coefficient**	**p-value**	**95% Confidence Interval**
Improved Water Source	-0.1914*	0.05	-0.383, -0.0000465
Moderate/Severe Diarrhea	0.1302**	0.08	-0.016, 0.276
Electricity in Dwelling	0.1802	0.139	-0.058, 0.419
Insecure Housing	0.2860*	0.001	0.111, 0.461
MOD3	**Regression coefficient**	**p-value**	**95% Confidence Interval**
WASH Score	-0.4104*	0.017	-0.747, -0.074
Healthcare Score	0.3407**	0.066	-0.022, 0.704
Poverty Score	0.3391*	0.013	0.070, 0.608
MOD4	**Regression coefficient**	**p-value**	**95% Confidence Interval**
WASH Safe Score	-0.250	0.203	-0.521, 0.111
Healthcare Score	0.3277**	0.083	-0.042, 0.698
Poverty Score	0.3305*	0.02	0.052, 0.609

with being Roma. The WASH score, however, is negatively correlated to the Roma, indicating that Roma individuals actually have an advantage over non-Roma individuals. To further investigate this question, we ran an additional analysis with healthcare and poverty, but substituting our WASH Safe score (MOD4; Table 4). The significant difference observed in healthcare and poverty remains, but when protected well water is included alongside tap water in the definition of improved or safe water sources, the disparity associated with WASH is eliminated.

Dataset 1. Coded survey data

http://dx.doi.org/10.5256/f1000research.12546.d177233

Romania, 2016. Excel file of compiled responses to survey questions. Coded and de-identified. Numerical code corresponds to responses as indicated on the study surveys (Supplementary material 1 and Supplementary material 2).

Dataset 2. Python Notebook data analysis and statistics

http://dx.doi.org/10.5256/f1000research.12546.d177234

Romania, 2016. Python Notebook analysis of survey data.

Dataset 3. Python Notebook data analysis and statistics

http://dx.doi.org/10.5256/f1000research.12546.d177235

Romania, 2016. Python Notebook analysis of survey data, exported as a PDF file.

CART Analysis and *Hot Spot* Generation

Following the CART-based assignment of categories to the synthetic population, we used ArcGIS to determine, at the *judet* (county) level, which regions of Romania are most in need of development and/or government aid based on seven key parameters. We reduced each parameter to a binary distinction during the generation of the population in order to simplify the visualization process, and all parameters are presented on a continuous scale using standard deviation from the mean.

Poverty Parameters. First, we visualized the availability of electricity to households throughout the country (Figure 3.3A). Analysis of survey responses indicated the presence or absence of electricity in a household was a significant distinction between Roma and non-Roma families. Our visualization (darker regions indicate areas of greater risk and/or need) demonstrates that households most likely to lack electricity are clustered in the middle of the country where Brasov, Sibiu, and Mures counties meet, and extend into the North-West corner into Bihor, Salaj, and Satu-Mare counties. Additional areas at risk are observed along the southern border in Dolj county, as well as in select areas near the capitol, Bucharest.

We then visualized the level of severe poverty, defined as the inability to spend more than U.S. $2 per person in a household per day (Figure 3.3B). Households with the greatest number of individuals at risk or currently experiencing severe poverty, shown by dark regions on the map, are in the middle portion of the country where Brasov, Sibiu, and Mures counties meet. Additional regions of risk include Caras-Severin county in the far West and various smaller pockets along the Eastern border counties.

Healthcare and Infrastructure. Analysis of insecure housing, rates of diarrheal disease, and the lack of access to an 'improved water source' (as defined by the WHO,[15]) led us to explore these issues in our population as parameters indicative of health status and deficiencies in essential infrastructure (Figure 3.3C–E). In general, these three parameters mimic the same patterns demonstrated by our poverty parameters, with a concentration of areas of need in the Central region and North-West corner of the country. Specifically, areas with prominent levels of insecure housing are where Brasov, Sibiu, and Mures counties meet in the middle portion of the country, along with regions in the North-West counties of Bihor, Salaj, Cluj, and Bistrita-Nasaud (Figure 3.3C). Additional smaller pockets in the South-West include areas of Caras-Severin, Mehedinti, and Dolj counties.

The lack of access to an 'improved water source', in contrast to insecure housing, is not a widespread issue, nor is it concentrated in one particular region, with only small pockets of affected areas spread throughout the country (Figure 3.3D). The areas with the highest rates of diarrheal disease cluster in Sibiu and Mures counties (excepting Brasov), as well as in Bihor, Salaj, Bistrita-Nasaud, and Satu-Mare. Additionally, clusters of high diarrheal disease rates are observed in the more southern county of Arges. Isolated clusters are also identified along the South-Western and North-Eastern border counties (Figure 3.3E).

Education and Geographical Classification. To provide a geographical classification to frame our other variables of interest, we visualized all of Romania to identify urban *versus* rural areas. The most urban areas are lightest (such as the capital of Bucharest), while the most rural areas are darker shades and predominantly align with the Carpathian mountain range and the boundary of the country with the Black Sea (Figure 3.3F).

We next visualized areas of the country in which portions of the population have not completed beyond an eighth-grade education or are at risk for not doing so (Figure 3.3G). The regions most in need of aid in this area are again Mures and Sibiu counties in the central portion of the country, Bihor and Satu-Mare counties in the North-West, and Arges, Giurgiu, and Dambovita counties in the Southern portion of the country. In addition, we observe some at risk areas in and around the city limits of the capital, Bucharest.

Cumulative Risk. Upon visualization of each individual QoL variable, we generated a map to indicate cumulative need across all variables (Figure 3.3H). Unsurprisingly, areas of greatest cumulative need mimic the patterns identified in individual variables and are concentrated in the central portion of the country in Brasov, Sibiu, and Mures counties. Additional regions include areas of Arges and Dambovita counties, along with isolated clusters throughout the country.

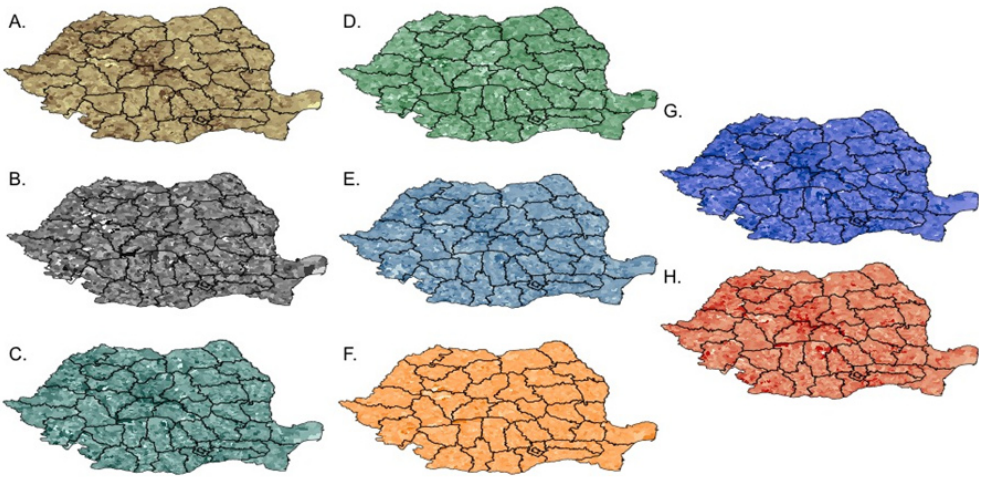

Figure 3. Visualization of quality of life parameters. ***A****) Lack of Electricity,* ***B****) Severe poverty,* ***C****) Insecure Housing,* ***D****) Lack of 'improved water source,* ***E****) High incidence of diarrheal disease,* ***F****) Urban versus rural distribution,* ***G)*** *Prevalence of lack of education and* ***H****) Cumulative risk.* Following the assignment of categories to the synthetic population, we used ArcGIS to determine, at county level, what regions are most in need of development and/or government aid based on key parameters. We reduced each parameter to a binary distinction during the generation of the population, so as to simplify the visualization process, and all parameters are presented on a continuous scale using standard deviation from the mean.

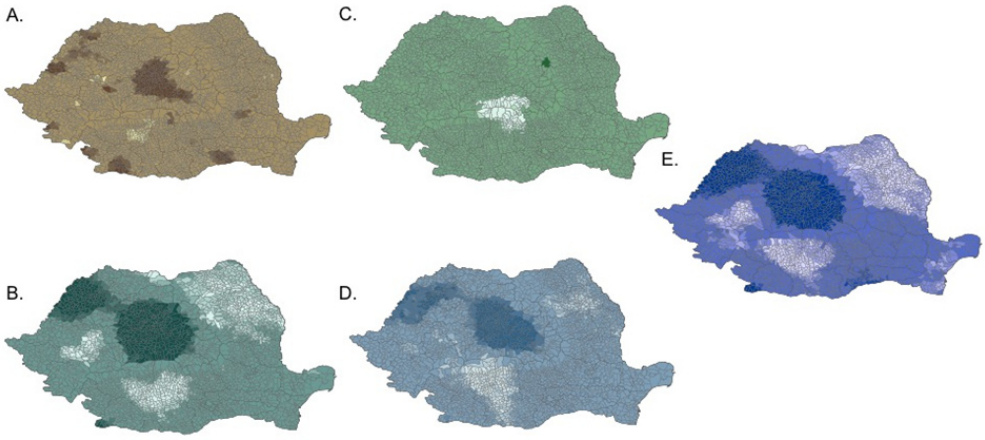

Figure 4. Hot spot analysis for quality of life parameters. ***A****) Lack of Electricity,* ***B****) Insecure Housing,* ***C****) Lack of 'improved water source',* ***D****) High incidence of diarrheal disease, and* ***E****) Lack of education beyond 8th grade.* Using spatial autocorrelation (Global Moran's I), each variable was analyzed to determine whether the pattern expressed in the population was random. Significant autocorrelation (non-random pattern or clustering) was determined by z-score and accompanying p-value (p=0.05). Significance or lack thereof suggests whether the independent variables upon which our synthetic population was built (household size and ethnicity) are appropriate indicators for our specific QoL parameters. For variables demonstrated to be significantly spatially auto-correlated, we progressed to Incremental Spatial Autocorrelation with a fixed distance measure to identify areas of intense need (dark shading) or 'hot spot' clustering.

Parameter Correlation and Hot Spot Analysis. We used the spatial autocorrelation Global Moran's I together with the Incremental Spatial Autocorrelation test to both validate our model and determine whether the pattern of clustering for each variable was significant. This analysis provides additional information beyond initial ArcGIS visualization (Figure 3.3), as it allows analysis down to the commune level and highlights the distinct patterns exhibited by the various parameters. Variables that did not demonstrate significance using Global Moran's, including geographical classification and severe poverty, were not carried through to hot spot visualization.

Analysis of the lack of electricity variable demonstrated that this variable is significantly geographically auto-correlated (z-score = 24.802, p-value<0.0001) and aligns with prior visualization, showing intense hotspots of need concentrated predominantly in the central portion of the country (dark areas on the map). In addition, hot spots are observed just south of

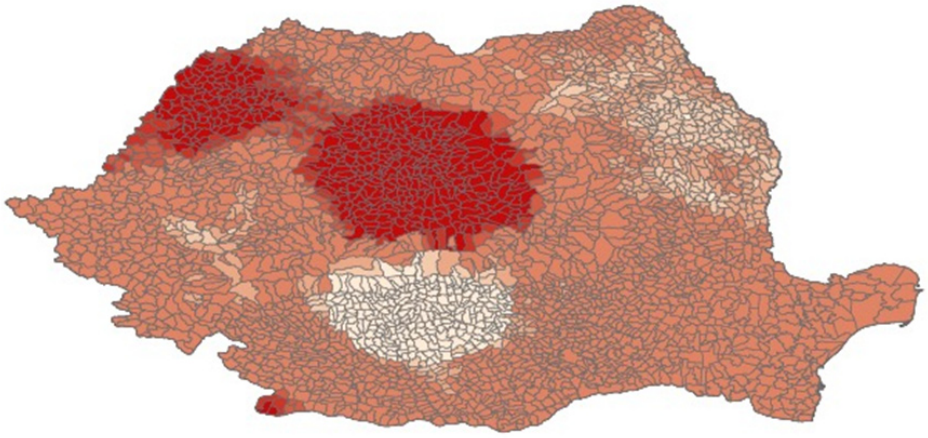

Figure 5. Hot Spot Analysis for Cumulative Risk. Hot spot analysis visualization of cumulative need across the country. Including all significantly auto-correlated parameters, with the addition of the measure of severe poverty, need of *any kind* is significantly auto-correlated in Romania (z-score = 11.5, p-value < 0.0001) and most apparent in the central portion of the country and extending to the North-West corner (dark red). These areas correlate with locations of Roma communities.

Bucharest, and in select communes in the South-West and North-West portions of the country (Figure 3.4A).

Likewise, analysis of the variable indicating areas with a strong prevalence of insecure housing was also significantly clustered by geographical region (z-score = 15.46, p-value<0.0001). It too aligns with previous visualization, as well as with *some* areas that are in need of access to electricity (Figure 3.4B). However, comparing the two variables, there are also communes that exhibit a need for better housing that are, paradoxically, not deficient in access to electricity, particularly in the North-West corner of the country.

Analysis of the 'improved water source' metric demonstrated that, while significant (z-score = 2.179, p-value = 0.029), the pattern of clustering is not as strong as in other variables, highlighting only one small hot spot throughout the country (Figure 3.4C).

Autocorrelation analysis of rates of diarrheal disease indicated significant geographical clustering (z-score = 8.548, p-value < 0.0001) and also appeared to be most concentrated in the central and North-West judets (Figure 3.4D). Analyzing more deeply, we observe numerous communes that appear in both the electricity variable and the housing variable. Alternatively, communes in Bihor and Satu-Mare counties in the North-West corner demonstrate particularly high rates of diarrheal disease and insecure housing, but not a significant lack of electricity.

The education variable (Figure 3.4E), indicating communes and/or judets with a significant number of individuals at risk for or already failing to progress beyond eighth grade, demonstrates a clustering pattern most similar to the insecure housing variable (Figure 3.4B). Significantly auto-correlated (z-score = 18.499, p-value < 0.0001), hot spots are most intense in the central and North-West counties. Much like the prevalence of insecure housing and, to a lesser extent, the lack of electricity, hot spots also appear in the Southern regions of the country, outside Bucharest and throughout Dolj county. The clustering pattern observed in these three parameters is distinctly different from that which appears in diarrheal disease and water quality analysis.

Finally, we utilized our analysis to visualize hot spots of cumulative need across the country (Figure 3.5). Including all significantly auto-correlated parameters, with the addition of the measure of severe poverty, need of *any kind* is significantly auto-correlated in Romania (z-score = 11.5, p-value < 0.0001) and, not surprisingly, most apparent in the central portion of the country and extending to the North-West corner.

Discussion

A number of studies have examined the various factors *the Decade of Roma Inclusion* (DRI) sought to address in Roma communities across the EU, both during the implementation of the project and since its conclusion in 2015[5,10,12,24,25]. Unfortunately, while some improvements did occur, a number of studies indicate the DRI did not achieve its stated goals in the areas of education, housing, employment, and health status of Roma in participating countries[26,27]. Our study supports these conclusions, particularly with regard to education, healthcare, and poverty. However, disparities that other studies have highlighted in multiple countries with regard to employment and sanitation do not necessarily occur in Romania[25,28–30]. Rather, both the Roma and non-Roma in rural Romania face similar challenges regarding access to full-time employment and water, which are exacerbated by a lack of municipal sanitation services in over 800 Romanian communities[31]. The lack of significant difference between Roma and non-Roma in our sample in relation to indoor toilets and bathrooms does not indicate that either ethnic group has an advantage, but rather all those who reside in rural communities face a disadvantage, regardless of ethnicity.

Notably, our findings indicate that, in some instances, the Roma appear to have a slight advantage over non-Roma (Table 4). Using the DRI definition of piped water to an indoor or outdoor tap, our analyses indicate Romanian and other non-Roma individuals lag behind the Roma in 'improved water sources'. However, when one accounts for the prevalence of private, protected wells (WASH Safe score), the disparity is minimized and no longer significant (Table 4). We postulate this distinction is indicative of how our survey collected this type of data, and future iterations will refine how we classify 'safe' and 'improved' water sources.

Of additional interest is the key indicator that those in suburban and urban areas, Roma and non-Roma alike, take longer to reach their chosen primary drinking water sources than do their rural counterparts. However, this statistic is potentially ambiguous. The urban community included in this study reported overwhelmingly that it had recently been subject to a contamination of the municipal water supply with coliform bacteria and, thus, the majority of residents therein reported the need to purchase water rather than use the taps available in their homes. It was not possible to collect data regarding the behaviour of these residents prior to the contamination event. Furthermore, the suburban community included here recently experienced the loss of a bridge, connecting the far side of the river to the village centre on the other side. Those individuals stranded on the far side of the bridge (predominantly Roma) reported numerous problems with their wells, requiring them to travel 5km or more to the nearest crossing to reach a shop or market until the bridge is restored. Therefore, this statistic is potentially a reflection of the walking or driving time that would otherwise be unnecessary.

Despite the evidence presented that Roma and non-Roma alike are subjected to ineffective sanitation and hygiene services throughout the country, one should note that the Roma population still reports a greater incidence of diarrheal disease *and* a reduced rate of immunization than the non-Roma population. There are potentially a number of reasons for this. Unlike in other countries[5,30], the Romanian Roma report fairly equivalent rates of medical insurance and access to primary care, but the type of treatment received when care is sought was beyond the scope of this study and may be a contributing factor. Indeed, Roma individuals have elsewhere reported poor health related to both their unhygienic circumstances and the care they receive[25,32,33]. In addition, as has already been noted, both literacy rates and overall levels of education are significantly decreased in the Romanian Roma population. This is in contrast to education rates in Roma populations of other countries, as the educational component of the DRI has been lauded as the most successful portion of the initiative, albeit only for primary school attendance[26,27]. Rates of disease and healthcare status overall are inversely associated with education[34], which may offer another possible explanation for the disparity in diarrheal disease rates. It is important to consider, however anecdotally, the Roma do report some knowledge of personal water treatment and safety (data not shown), through the use of salt or lime in personal wells and a commitment to boiling water before drinking or cooking if possible. However, the lack of infrastructure and services works against these individual and imperfect efforts. Furthermore, for those Roma who do have access to tap water (municipal or otherwise), many of them report using an alternative primary water source. While these same individuals indicate that they believe their tap water to be safe (data not shown), their daily activities are in direct contrast to this assertion.

While the population data are of interest, our primary focus is using that data to demonstrate the utility of our CART analysis and *hot spot* generation tool. Recognizing the limited nature of our population size and to corroborate the validity of our approach, we searched for areas in Romania with development issues that were previously identified using more traditional methods. In particular, the areas identified as having a high prevalence of individuals experiencing insecure housing, lack of electricity and diarrheal disease (Figures 3.3 and 3.4) align with areas known for informal settlements, populated predominantly by Roma families, found in the suburban and urban areas surrounding the North-Western city of Cluj and the far North-West town of Baia Mare[35]. These areas extend westward into Bihor and Salaj counties, as well as southward into Mures county, the sites predicted to be the most concentrated hot spots on our maps. Our methodology also identifies incorporated areas (villages, cities, etc.) that suffer from specific issues. For example, the village of Holbav and numerous others in Brasov county have been highlighted as areas with energy poor communities with little indication of infrastructure improvements on the horizon[36]. These villages are in the central region of Romania and fall in the most intense hot spot for lack of electricity, as predicted by our model. Furthermore, in a case study by Vincze, the city of *Calafat* in Dolj county was characterized following the demise of its manufacturing economy[37]. The study highlighted the particular problems facing the Roma community in that area, noting a lack of formal employment along with inadequate housing and precarious government services. This portion of Dolj county is highlighted as a hot spot for housing, education, and generalized need in our model. These areas coincide with those identified in our model as regions of intense need across multiple variables and also boast large concentrations of Roma. Thus, our model provides corroborating evidence to demonstrate how the Roma minority in Romania are consistently at risk in key quality of life indicators and frequently lack access to basic services. However, as indicated by our survey data, non-Roma within these areas are likely also at risk.

Interestingly, our model only produces a small hot spot in the Eastern portion of Romania as an area of need related to 'improved water source' access. At first glance, this suggests that the WASH infrastructure in the country is better than initially anticipated. However, while there are no true hot spots, there are also no 'cold' spots. These results indicate that limited access to clean, reliable water sources is a ubiquitous problem across the country and *not* confined to specific geographic regions aligning with the Roma minority.

This model and subsequent analyses serve as an example of the utility of synthetic populations and how their use in conjunction

with traditional surveys, time-use data, and census data can augment the conclusions generated from those kinds of data. Using a model such as ours, conclusions of greater complexity can be made. While it is possible to analyze survey data for information, that analysis is severely restricted to the area in which the survey was conducted and the questions that were asked. Furthermore, the analysis only achieves a summary view of the population. In contrast, merging survey data with population statistics and conducting analyses via the synthetic population allows one to identify geographical regions with similar characteristics and populations with key identifiable traits, and combine the two to extrapolate conclusions beyond the original survey regions. This approach requires fewer on-the-ground resources and allows conclusions to be visualized in an accessible fashion for use in project proposals and grant justifications.

Limitations and future directions

The primary limitation of this study is the use of a small sample (n = 135) of survey respondents to generate the categories necessary for CART analysis and random household assignment. Constraints of limited time, funding and personnel, which are often factors in community-based public health studies, inhibited our ability to interact with more than 30 households per community and restricted the study to five communities. Future iterations will seek to obtain a more robust survey sample size for integration into the synthetic population. While acknowledging this limitation, we do note our ability to validate the predictions of the existing model via identification of similar conclusions from more traditional methods, thereby suggesting that the methodology is sound. Thus, using this type of model, conclusions can be drawn and applied to a larger population and geographic area even with limited resources and sample size, providing a valid methodology to conduct similar studies to highlight hot spots of need. Similar methodologies could also be applied to geographic areas with restricted access due to geography or political unrest, which limits the ability to assess needs within these areas. Additionally, subsequent studies can use these and other data to generate detailed models that explore specific initiatives that could be implemented to address discrepancies in equality and access, and progress the literature around Roma health disparities beyond analysis and into intervention testing.

Conclusions

The model and approach demonstrated herein provides a useful tool to identify and predict both areas of need and the type of need required in a given region. Furthermore, this approach allows populations to be separated based on ethnicity and other characteristics, and to determine if subpopulations require different kinds of assistance compared to the majority. Therefore, we assert this approach can and should be utilized by non-profit organizations, NGOs, and government funding agencies to more appropriately focus valuable time and resources during project planning and development to ensure aid reaches those who are in greatest need.

Data availability

Dataset 1: Coded survey data. Romania, 2016. Excel file of compiled responses to survey questions. Coded and de-identified. Numerical code corresponds to responses as indicated on the study surveys (Supplementary material 1 and Supplementary material 2).

DOI, 10.5256/f1000research.12546.d177233[38]

Dataset 2: Python Notebook data analysis and statistics. Romania, 2016. Python Notebook analysis of survey data.

DOI, 10.5256/f1000research.12546.d177234[39]

Dataset 3: Python Notebook data analysis and statistics. Romania, 2016. Python Notebook analysis of survey data, exported as a PDF file.

DOI, 10.5256/f1000research.12546.d177235[40]

Grant information

This work has been partially supported by the National Institutes of Health and National Institute of General Medical Sciences - Models of Infectious Disease Agent Study Grant 5U01GM070694-13, the Defense Threat Reduction Agency - Comprehensive National Incident Management System Contract HDTRA1-11-D-0016-0001, and the Virginia-Maryland College of Veterinary Medicine.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We would like to thank Dr. Ralph P. Hall for his assistance developing the surveying instrument, and all the staff members at *Agentia Impreuna* for helping to make this project possible. We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their suggestions and comments. In addition, we are grateful to the numerous friends and family who contributed financially to support our time in the field, with particular thanks to George K. Agardi, Sr. and Dr. Susan Evans.

Supplementary material

Supplementary material 1: Quality of Life Survey, English. Romania 2016. Survey questions provided in English.

Click here to access the data.

Supplementary material 2: Quality of Life Survey, Romanian. Romania, 2016. Survey questions provided in Romanian.

Click here to access the data.

References

1. European Investment Bank: **EIB loan signatures in Romania amounted to EUR 211 million in 2015**. 2015.
2. United Nations Development Programme: **Millennium Development Goals in Romania**. 2007.
3. European Union: **European parliament resolution of 31 January 2008 on a European strategy on the Roma**. 2008.
Reference Source
4. World Bank: **Roma**. 2015.
Reference Source
5. Cook B, Wayne GF, Valentine A, *et al.*: **Revisiting the evidence on health and health care disparities among the Roma: a systematic review 2003-2012.** *Int J Public Health.* 2013; **58**(6): 885–911.
PubMed Abstract | Publisher Full Text
6. Hajioff S, McKee M: **The health of the Roma people: a review of the published literature.** *J Epidemiol Community Health.* 2000; **54**(11): 864–9.
PubMed Abstract | Publisher Full Text | Free Full Text
7. Zeman CL, Depken DE, Senchina DS: **Roma health issues: a review of the literature and discussion.** *Ethn Health.* 2003; **8**(3): 223–49.
PubMed Abstract | Publisher Full Text
8. Parekh N, Rose T: **Health inequalities of the Roma in Europe: a literature review.** *Cent Eur J Public Health.* 2011; **19**(3): 139–42.
PubMed Abstract | Publisher Full Text
9. Gladdis K: **Life inside the Romanian gypsy ghetto that is so grim the town mayor sealed it off behind a wall**. *Daily Mail.* 2013.
Reference Source
10. Friedman E: **Decade of Roma Inclusion Progress Report 2005-2013**. 2014.
Reference Source
11. Flecha A: **Healthier lives for European minority groups: school and health care, lessons from the Roma.** *Int J Environ Res Public Health.* 2013; **10**(8): 3089–111.
PubMed Abstract | Publisher Full Text | Free Full Text
12. Fésüs G, Östlin P, McKee M, *et al.*: **Policies to improve the health and well-being of Roma people: the European experience.** *Health Policy.* 2012; **105**(1): 25–32.
PubMed Abstract | Publisher Full Text
13. Gould R: **Roma rights and Roma expulsions in France: Official discourse and EU responses.** *Critical Social Policy.* 2015; **35**(1): 24–44.
Publisher Full Text
14. Dale CR: **Development Planning: Concepts and Tools for Planners, Managers and Facilitators**. London: Zed Books; 2004.
Reference Source
15. World Health Organization: **Core Questions on Drinking-Water and Sanitation for Household Surveys**. 2006.
Reference Source
16. **Digital Raster Data-Romania [Internet]**. Oak Ridge National Laboratory. 2013.
Reference Source
17. **International Data Base-Romania [Internet]**. 2016.
Reference Source
18. Lester A: **Global Population NDSSL Tech Report**. 2016.
19. ESRI: **ArcGIS Desktop: Release 10**. Redlands, CA: Environmental Systems Research Institute; 2011.
20. Hijmans R: **GADM database of Global Administrative Areas**. 2015.
Reference Source
21. McKinney W, editor: **Data Structures for Statistical Computing in Python**. *9th Python in Science Conference*; SciPy. 2010.
Reference Source
22. NIS Romania: **Populația stabilă după etnie – județe, municipii, orașe, comune**. In: Romania INdS, editor. 2011.
23. Rivers C: **Modeling Emerging Infectious Diseases for Public Health Decision Support**. Blacksburg, VA: Virginia Polytechnic Institute and State University; 2015.
Reference Source
24. Curcic SM, Maj A, Shayna P: **Inclusion, Integration or Perpetual Exclusion? A Critical Examination of the Decade of Roma Inclusion, 2005-2015.** *Educ Res Eval J.* 2014; **13**(3): 257–67.
Publisher Full Text
25. Masseria C, Mladovsky P, Hernández-Quevedo C: **The socio-economic determinants of the health status of Roma in comparison with non-Roma in Bulgaria, Hungary and Romania.** *Eur J Public Health.* 2010; **20**(5): 549–54.
PubMed Abstract | Publisher Full Text
26. Bojadjieva A: **Roma Inclusion Index**. Budapest, Hungary: Decade of Roma Inclusion Secretariat Foundation; 2015.
Reference Source
27. Jovanovic Z: **Why Europe's "Roma Decade" Didn't Lead to Inclusion**. In: Foundations OS, editor. 2015.
Reference Source
28. Molnár A, Adány R, Adám B, *et al.*: **Health impact assessment and evaluation of a Roma housing project in Hungary.** *Health Place.* 2010; **16**(6): 1240–7.
PubMed Abstract | Publisher Full Text
29. Janevic T, Jankovic J, Bradley E: **Socioeconomic position, gender, and inequalities in self-rated health between Roma and non-Roma in Serbia.** *Int J Public Health.* 2012; **57**(1): 49–55.
PubMed Abstract | Publisher Full Text
30. Kühlbrandt C, Footman K, Rechel B, *et al.*: **An examination of Roma health insurance status in Central and Eastern Europe.** *Eur J Public Health.* 2014; **24**(5): 707–12.
PubMed Abstract | Publisher Full Text
31. Duminica G: **Raport Anual, 2015**. Bucharest, Romania: Agentia Împreună; 2015.
32. Casals M, Pila P, Langohr K, *et al.*: **Incidence of infectious diseases and survival among the Roma population: a longitudinal cohort study.** *Eur J Public Health.* 2012; **22**(2): 262–6.
PubMed Abstract | Publisher Full Text
33. Teira R, Suárez-Lozano I, Lozano F, *et al.*: **Characteristics and outcome of HIV infection in gypsies in the Spanish VACH Cohort.** *Enferm Infecc Microbiol Clin.* 2010; **28**(5): 266–72.
PubMed Abstract | Publisher Full Text
34. Regidor E, De Mateo S, Calle ME, *et al.*: **Educational level and mortality from infectious diseases.** *J Epidemiol Community Health.* 2002; **56**(9): 682–3.
PubMed Abstract | Publisher Full Text | Free Full Text
35. Small Steps PROJECT: **Romania-Cluj-Napoca Transylvania**. 2017.
Reference Source
36. Gaedtke F: **In Transylvania, energy poverty persists in the developed world**. 2016.
Reference Source
37. Vincze E: **Precarization of Working Class Roma through Spatial Deprivation, Labor Destitution and Racialization.** *Review of Sociology.* 2015; **25**(4): 58–85.
Reference Source
38. Powell Doherty R, Telionis PA, Müller-Demary D, *et al.*: **Dataset 1 in: A survey of quality of life indicators in the Romanian Roma population following the 'Decade of Roma Inclusion'.** *F1000Research.* 2018; **6**: 1692.
http://www.doi.org/10.5256/f1000research.12546.d177233
39. Powell Doherty R, Telionis PA, Müller-Demary D, *et al.*: **Dataset 2 in: A survey of quality of life indicators in the Romanian Roma population following the 'Decade of Roma Inclusion'.** *F1000Research.* 2018; **6**: 1692.
https://www.doi.org/10.5256/f1000research.12546.d177234
40. Powell Doherty R, Telionis PA, Müller-Demary D, *et al.*: **Dataset 3 in: A survey of quality of life indicators in the Romanian Roma population following the 'Decade of Roma Inclusion'.** *F1000Research.* 2018; **6**: 1692.
https://www.doi.org/10.5256/f1000research.12546.d177235

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 3

Reviewer Report 31 January 2019

https://doi.org/10.5256/f1000research.19000.r41911

✓ **Pilar Carrasco-Garrido**
Medical Microbiology and Immunology and Nursing Department, King Juan Carlos University, Madrid, Spain

This is a relevant manuscript from a public health standpoint. In this version after reviewing and incorporating the comments of the different reviewers, quality has improved.

Methodological analysis has been extended. The discussion of the results and their limitations are now properly defined, despite the limitation of the sample size.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 03 January 2019

https://doi.org/10.5256/f1000research.19000.r41912

✓ **Martin McKee**
European Centre on Health of Societies in Transition, London School of Hygiene and Tropical Medicine, London, UK

This is almost entirely a new paper, with very extensive additional analysis. This recognizes the previously noted limitations but uses a modelling approach to address them, taking the paper in a somewhat different direction. While it is important to continue to view the paper as a

methodological rather than empirical paper, seen in this light I feel it is much improved. I would prefer that you received comments from someone with greater expertise in the methods used than me, but as far as I can see it now addresses my previous concerns.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 19 April 2018

https://doi.org/10.5256/f1000research.15471.r31657

?

Ciprian Marius Ceobanu
Faculty of Psychology and Education Sciences, Department of Psychology, Alexandru Ioan Cuza University, Iași, Romania

Despite the small improvements there still remain important issues.

The comparison with Burkina-Faso is completely uninspired and demonstrate that the author(s) is not aware at all of the whole image related to Roma population.

Another important issue is related to the extremely small size of the sample. Only 135 persons (on average 27 from each region) is extremely unsignificant from a statistical point of view. Also, there is a large number of Roma population that came from Romania and currently live in different EU countries, on a long-term basis; are they included in the same population?

Again, the paper still remains a collection of results from official statistics without adding anything new or relevant.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 03 April 2018

https://doi.org/10.5256/f1000research.15471.r32678

Pilar Carrasco-Garrido
Medical Microbiology and Immunology and Nursing Department, King Juan Carlos University, Madrid, Spain

Although the objectives are adequate and the subject is relevant, the important limitations of the study are still there.

Possibly, the small sample size is a problem. This aspect is a methodological problem that can lead to conclusions in the study that are not adequate.

Sorry but, I do not recommend the acceptance of the article.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 22 March 2018

https://doi.org/10.5256/f1000research.15471.r31658

Martin McKee
European Centre on Health of Societies in Transition, London School of Hygiene and Tropical Medicine, London, UK

While the authors have added some clarity, I remain concerned about the small sample size and other limitations. I don't think they have enough data to draw any conclusions and it is not obvious to me that it has been rewritten as a proof of concept paper. For example, the abstract has, in the conclusions "This paper demonstrates..." whereas I really don't think it does (although in fact, what is concluded is quite well known). I still think it could be revised to be much less ambitious.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public health – and have written extensively on Roma health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Version 1

Reviewer Report 02 January 2018

https://doi.org/10.5256/f1000research.13585.r28652

Pilar Carrasco-Garrido
Medical Microbiology and Immunology and Nursing Department, King Juan Carlos University, Madrid, Spain

This is a relevant manuscript from a public health standpoint because one of the main contributions of the present work is to determine quality of life indicators in the Romanian Roma, but methodologically it has significant shortcomings:

- The comparison between the population of Burkina Faso and the population of Romania is not adequate. They are very different populations. This aspect is a methodological problem.
- Discussion of the results is limited. Some aspects are not adequately discussed.
- There is no information about the survey's non-response rate.
- Few references and some of them unrelated to the purpose of the study. It does not seem correct to incorporate a press article as a reference.

Sorry but, I do not recommend the indexing of the article.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
No

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 12 December 2017

https://doi.org/10.5256/f1000research.13585.r26652

?

Martin McKee
European Centre on Health of Societies in Transition, London School of Hygiene and Tropical Medicine, London, UK

The introduction is well written and provides a good overview of the situation facing the Roma population. There are a few more recent references that could be included, such as an evaluation of the Decade of Roma inclusion in Hungary but, in general, the authors have found most of the relevant information.

The fundamental challenge facing anyone doing research among the Roma population in this region is how to develop a sampling frame. There are numerous methodological problems, in particular varying degrees of assimilation (see, for example, work by K Kosa). Previous studies, such as that by the UNDP or in Hungary, have used Roma communities, identifiable by their socio economic and physical characteristics, while recognising that this is imperfect. However, this paper would benefit from a more detailed description of the communities from which the samples were drawn, in particular, how they relate to Romania as a whole. Given that, in many parts of Romania, Roma live in distinct settlements, separate from the Romanian population, even within individual villages, could the authors comment on any implications that their sampling strategy had for generalisability?

Given the high levels of distrust that many Roma, justifiably, have, some studies have sought to ensure involvement of Roma fieldworkers, or at least, involvement of community leaders. Can the

authors comment on what measures they took in this regard?

The greatest problem in this paper is the very small sample size. Overall, less than 100 Roma respondents were included and only 37 non-Roma. Given the numerous problems involved in sampling in a study such as this, this is really far too few from which to draw any meaningful conclusion. This is noted in the limitations but I'm not really convinced that a study of this size can be regarded as much more than a pilot. I would suggest that it is described in this way, with many more caveats than there are at present.

Minor points:

I'm not sure that it is appropriate to use the words of Soviet rule for the countries of south-eastern Europe. Arguably, Romania was one of the most independent of the Soviet bloc states.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
No

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public Health – and have written extensively on Roma health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 06 November 2017

https://doi.org/10.5256/f1000research.13585.r26651

Ciprian Marius Ceobanu
Faculty of Psychology and Education Sciences, Department of Psychology, Alexandru Ioan Cuza University, Iași, Romania

The main issue is related to the size of the sample and its geographical distribution. It is hard to generalize over the entire Roma minority the conclusions of the study even if the statistical approach is appropriate.

The geographical distribution of the Roma population is pretty different over the Romanian national territory. The lack of indication of the geographical area of the subjects of the sample is a flaw. Another issue regards the random walk method for sampling which, in our opinion, is not representative for the entire Roma population. Maybe a multilevel sampling would be more appropriate than a simple random walk.

Despite the fact that the conclusions of the study are correct, these are pretty well known to all levels; also these are commonplaces that were specified in the documents of Decade of Roma Inclusion as directions for future action. There are a lot of significant reports that draw the same conclusions (see The World Bank documents for instance[1]). From this point of view, there is no original approach to the Roma problem in Romania.

The conclusion following the Decade of Roma Inclusion is that despite the efforts that there were made, there still remain lots of issues regarding the integration of the Roma population. Also,solving great structural problems of Romania will certainly improve the Roma population situation.

References

1. Human Development and Sustainable Development teams Europe and Central Asia: Diagnostics and Policy Advice on the Integration of Roma in Romania [Romanian]. *World Bank Group*. 2014.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Comments on this article

Version 1

Author Response 04 Mar 2018

Rebecca Powell Doherty, Virginia Tech, Blacksburg, USA

Please find below a detailed, point-by-point response to each query raised by the three published reviewers. We thank the reviewers for their time and effort. The changes to the manuscript, based on the below, can be seen in the revised document, version two.

Reviewer #1: Approve with Revisions

1. The main issue is related to the size of the sample and its geographical distribution. It is hard to generalize over the entire Roma minority the conclusions of the study even if the statistical approach is appropriate.

We agree with the reviewer that the size of the sample is relatively small. To address this issue, we have reframed the work as another reviewer recommended, as proof of concept. The data described here is utilized in a computer simulation designed to identify areas at high risk of certain health factors. In preliminary analysis, the simulation holds up when compared to on-the-ground knowledge and experience, as well as other published evidence. With regard to the geographical distribution, we purposely chose to visit Roma villages and settlements that were different from one another in an attempt to reduce the bias that inevitably occurs with this type of research, as it is logistically impossible for our small team to reach the entire Roma minority in Romania.

2. The geographical distribution of the Roma population is pretty different over the Romanian national territory. The lack of indication of the geographical area of the subjects of the sample is a flaw. Another issue regards the random walk method for sampling which, in our opinion, is not representative for the entire Roma population. Maybe a multilevel sampling would be more appropriate than a simple random walk.

As indicated above, our geographical distribution was actually varied. Of the five communities we visited, they ranged from extremely urban and close to the capitol to rural and in the far north of the country. We chose these villages and settlements specifically for their geographical diversity and, in particular, because they aligned with the non-Roma majority in their areas very differently. Some were extremely well-integrated and others entirely separate. We do not share the geographical area of the subjects as this level of anonymity was a requirement of our ethical approval to conduct the study, as the IRB

considered the population to be 'vulnerable'.

With regard to the random walk method, we respectfully disagree that this was not appropriate for sampling. Varying the time of day that we walked through these areas, and ensuring we went into shops as well as knocked on gates and doors allowed us to reach greatest number of people throughout a given area without the intimidation of a formal gathering, as well as ensuring the anonymity of our participants.

Our manuscript has been updated to reflect, where possible, some of these details.

3. Despite the fact that the conclusions of the study are correct, these are pretty well known to all levels; also these are commonplaces that were specified in the documents of Decade of Roma Inclusion as directions for future action. There are a lot of significant reports that draw the same conclusions (see The World Bank documents for instance[1]). From this point of view, there is no original approach to the Roma problem in Romania.

We respectfully disagree. Our approach has indeed taken elements from other studies, although the reports that the reviewer references are not properly peer-reviewed, and it is important that independent research verifies the information put forth by such entities as The World Bank and even the World Health Organization. Importantly, however, we believe our study to be significant in that our survey brings together questions of demographics, public health, and education, along with added insights into the level of trust (or lack thereof) that the Roma minority has in its government and fellow citizens. Further to this, there are few studies that focus exclusively on Romania, and we disagree with the notion that it is appropriate to compare Roma in Hungary, for example, with the Roma population in Romania, and indeed, our findings demonstrate clear differences in the Romanian Roma from what is reported elsewhere in the literature.

4. The conclusion following the Decade of Roma Inclusion is that despite the efforts that there were made, there still remain lots of issues regarding the integration of the Roma population. Also, solving great structural problems of Romania will certainly improve the Roma population situation.

We agree this is the case, and as mentioned above, these data will be used to pilot a simulation that allows us to pinpoint areas of high risk for structural problems, educational deficits, and numerous other public service categories to ensure the appropriate type of aid is delivered to the right place. This has implications for the Roma population within Romania of course, but also for those living elsewhere, and if other data sets are available, other minorities as well.

Reviewer #2: Approve with Revisions

1. The introduction is well written and provides a good overview of the situation facing the Roma population. There are a few more recent references that could be included, such as an evaluation of the Decade of Roma inclusion in Hungary but, in general, the authors have found most of the relevant information.

No specific response required.

2. The fundamental challenge facing anyone doing research among the Roma population in this region is how to develop a sampling frame. There are numerous methodological problems, in particular varying degrees of assimilation (see, for example, work by K Kosa). Previous studies, such as that by the UNDP or in Hungary, have used Roma communities, identifiable by their socio economic and physical characteristics, while recognising that this is imperfect. However, this paper would benefit from a more detailed description of the communities from which the samples were drawn, in particular, how they relate to Romania as a whole. Given that, in many parts of Romania, Roma live in distinct settlements, separate from the Romanian population, even within individual villages, could the authors comment on any implications that their sampling strategy had for generalisability?

We thank the reviewer for this thoughtful commentary, and we agree there are numerous problems with the available methodology, regardless of which is chosen. Our sample choice closely resembles that of what was used in Hungary, targeting specific Roma communities. We appreciate this is indeed imperfect, but we chose communities that our NGO partner had connections to and allowed us to identify community leaders (see response to next question) with whom we could work. These villages and settlements were purposefully geographically diverse, ranging from extremely urban and central to the capitol to rural and in the far north of the country with a large Hungarian population. Appreciating that our sample size, due to resources and logistics, is extremely small, we felt the geographic diversity to be a significant factor in our ability to generalise our findings to the whole of the Roma minority in Romania. We absolutely maintain, and indeed is part of the reason for this study, that our findings are specific to Romania and that there are too many variables to extrapolate to Roma living in other countries.

Our manuscript has been modified to include, where possible, some of this detail.

3. Given the high levels of distrust that many Roma, justifiably, have, some studies have sought to ensure involvement of Roma fieldworkers, or at least, involvement of community leaders. Can the authors comment on what measures they took in this regard?

As indicated in our methods section, all of our work was conducted in collaboration with a Roma-centric NGO based in Bucharest. The NGO assisted us in identifying appropriately diverse communities that may be receptive to speaking with us. In addition, as our NGO partner had extensive knowledge of the communities and a presence in them, we were able to identify community leaders who accompanied us as we moved through the areas. Importantly, the community representatives we worked with were both Roma and non-Roma. Despite this, we do note that it was more difficult to connect with individuals who did not identify as Roma, which is the reason for the small sample size in our data set.

4. The greatest problem in this paper is the very small sample size. Overall, less than 100 Roma respondents were included and only 37 non-Roma. Given the numerous problems involved in sampling in a study such as this, this is really far too few from which to draw any meaningful conclusion. This is noted in the limitations but I'm not really convinced that a study of this size can be regarded as much more than a pilot. I would suggest that it is described in this way, with many more caveats than there are at present.

We thank the reviewer for this suggestion, and we absolutely agree the sample size is quite small. As noted, this is due to a number of resource and logistic constraints, but we have modified our paper to

reflect this study be considered as a pilot or proof-of-concept study.

5. I'm not sure that it is appropriate to use the words of Soviet rule for the countries of south-eastern Europe. Arguably, Romania was one of the most independent of the Soviet bloc states.

We thank the reviewer for this insight and agree there is complexity in the discussion of how different countries functioned under communist leadership. We have, therefore, removed the reference to Soviet rule from the manuscript.

Reviewer #3: Not Approved

This is a relevant manuscript from a public health standpoint because one of the main contributions of the present work is to determine quality of life indicators in the Romanian Roma, but methodologically it has significant shortcomings:

1. The comparison between the population of Burkina Faso and the population of Romania is not adequate. They are very different populations. This aspect is a methodological problem.

We apologise that our paper was written in such a way as to confuse this reviewer. There is no comparison between Romania and Burkina Faso in our work. Our paper has been edited to ensure further clarity.

2. Discussion of the results is limited. Some aspects are not adequately discussed.

As the reviewer did not specifically indicate the ways in which our results are inadequately addressed, we are unable to make direct changes. However, we do feel our discussion of results is appropriately limited to the data we have and in keeping with the limitation of our sample size, as we mention in our limitations section.

3. There is no information about the survey's non-response rate.

As we were a small team and logistics was complicated, this is not data that we collected. We do note, however, that our surveys were conducted as interviews and as such there was not a 'non-response' rate, but rather individuals who simply did not want to talk to us and were not therefore included in our study.

4. Few references and some of them unrelated to the purpose of the study. It does not seem correct to incorporate a press article as a reference.

We respectfully disagree and feel that all of our references are appropriate and pertain to our work.

Competing Interests: I am the primary author of the study.