Available online at www.sciencedirect.com
ScienceDirect
Procedia Environmental Sciences 36 (2016) 4 – 11
Procedia Environmental Sciences

International Conference on Geographies of Health and Living in Cities: Making Cities Healthy for All, Healthy Cities 2016

What's Wrong with Understanding Variation Using a Single-Geographic Scale? A Multilevel Geographic Assessment of Life Expectancy in the United States

Rockli Kim[a], S V Subramanian[a, b, *]

[a]Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA
[b] Harvard Center for Population and Development Studies, 9 Bow Street, Cambridge, MA 02138, USA

Abstract

There has been limited effort to consider multiple areal units or scales in understanding spatial and geographic processes. Treating observed differences in the results by choice of geographic unit of analysis simply as a nuisance is conceptually problematic and can be empirically misleading. We consider the existing research on geographic variations in life expectancy in the United States to demonstrate that prior county-level studies have overestimated the importance of the county level by omitting states. Future investigations should critically assess the relative importance of multiple geographic, spatial, and non-geographic contexts, including an assessment of what units/scales have been omitted.

Peer-review under responsibility of the organizing committee of Healthy Cities 2016

Keywords: Geographical variances; Multilevel modeling; Scales; Unit of analysis; Modifiable Area Unit Problem; Life expectancy; United States.

* Corresponding author. Tel.: +1-617-432-6299 ; fax: +1-617-432-3123.
E-mail address: svsubram@hsph.harvard.edu

Peer-review under responsibility of the organizing committee of Healthy Cities 2016
doi:10.1016/j.proenv.2016.09.002

1. Introduction

The sensitivity of geographic patterns to the choice of areal units is well known, and is commonly captured within the influential framework of "Modifiable Areal Unit Problem" (MAUP)[1], which highlights the fact that areal units are usually arbitrarily determined and, therefore, "modifiable", in the sense that they can be aggregated to form units of different sizes or spatial arrangements leading to different results[1]. The general idea that patterns and relationships observed at one analytical unit (whether individual or geographic/ecological) is well recognized[2,3]. The fundamental premise in these frameworks is that there intrinsically exists one ideal unit of analysis and inference; be it individual or one particular geographic scale[4,5]. With the advent of multilevel modeling[3,6-8], while there has been substantial efforts to simultaneously consider, especially the scales or units of individual and certain geographic aggregation[4,5,9], efforts to consider multiple geographic units/scales has been limited[10,11].

Meanwhile, the idea of considering geographic aggregation at multiple units was explicitly recognized by Harold Moellering and Waldo Tobler in their classic paper published in 1972 entitled, "Geographical Variances"[12]. In their paper, Moellering and Tobler went on to propose a methodological framework to simultaneously model variation at multiple geographic levels[12], outlining what can be considered as a precursor to the current multilevel models. Building on the foundational, but unfortunately less remembered, contribution of Moellering and Tobler, we present the thesis that treating observed differences in the results by choice of unit of analysis simply as a nuisance is conceptually problematic and can be empirically misleading (at worst) and in many instances provide an impoverished interpretation of the undertaken inquiry.

In order to exemplify our thesis, we consider the existing research on geographic variations in life expectancy in the United States (US). Extensive evidence shows increasing geographic disparity in premature mortality trends, indicating that not all areas have equally benefited from the economic and medical improvements. While the all-cause death rates in the US have reduced by 42.9% between 1969 and 2013[13], this national trend alone is inadequate to capture specific states and counties that are performing significantly differently. Substantial variation in life expectancy has been reported across the states[14, 15]. Many more studies have examined life expectancy at the county-level, which is the smallest unit for which mortality data are routinely available in the US, and have reported that between-county inequality has been steadily increasing in recent decades[16-18].

A distinct feature of existing assessments of geography of life expectancy in the US is an exclusive reliance on a single level - either at the state or county - as the unit of analysis. By focusing on a single geographic scale, prior studies have implicitly and/or explicitly treated their unit of interest as the primary driver of variability in life expectancy. For instance, in the county-level analyses, an implicit assumption is that the lowest level at which data is available equates with the appropriate unit of analysis. However, the relative importance of one unit can be truly examined only when multiple scales that are thought to influence the outcome are simultaneously considered[9,10]. Legislations, policies and programs that provide health care, economic assistance and social services are administered and implemented at both the county and state levels. Hence, the significant variation in mortality at the county level, as identified in previous county-level studies, may substantially attenuate once the county-state membership is explicitly modeled.

2. Methods

We used the publicly available county-level life expectancy estimates for 1961–1999 compiled by Ezzati and colleagues for the empirical exemplification[16]. The analytic data contains repeated cross-sections of 122,850 life expectancy estimates across 39 years at level-1, nested within 3,150 counties at level-2, nested within 51 states at level-3. We specified and estimated the following models. First, we ignored states and assumed repeated measurements of life expectancy to be nested only within counties (Model 1). We then estimated models ignoring counties and specified repeated measures to be nested within only states (Model 2). Lastly, we estimated a three-level model accounting for the entire hierarchical nesting structure of repeated measures in counties in states (Model 3). In order to visualize the geography of life expectancy by counties and states, we mapped the residuals estimated from each of the models. Technical details and interpretations of each of the models are provided in Appendix A.