Spatial and Spatio-temporal Epidemiology 7 (2013) 11–24

Contents lists available at SciVerse ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Original Research

How to choose geographical units in ecological studies: Proposal and application to campylobacteriosis ☆

Julie Arsenault [a,f,*], Pascal Michel [b,f], Olaf Berke [c], André Ravel [a,f], Pierre Gosselin [d,e,f]

[a] Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec, Canada J2S 7C6
[b] Laboratoire de lutte contre les zoonoses d'origine alimentaire, Agence de la santé publique du Canada, Faculté de médecine vétérinaire, 3200 rue Sicotte, C.P. 5000, Saint-Hyacinthe, Québec, Canada J2S 7C6
[c] Department of Population Medicine, University of Guelph, Guelph, Ontario, Canada N1G 4C3
[d] Institut national de santé publique du Québec (INSPQ), Beauport, 945 Avenue Wolfe, Québec, Québec, Canada G1V 5B3
[e] Centre hospitalier universitaire de Québec (CHUQ), 2705 boulevard Laurier, Sainte-Foy, Québec, Canada G1V 4G2
[f] Groupe de recherche en épidémiologie des zoonoses et santé publique, Université de Montréal, 3200 Sicotte, Saint-Hyacinthe, Québec, Canada J2S 7C6

ARTICLE INFO

Article history:
Received 11 November 2010
Revised 20 February 2013
Accepted 17 April 2013
Available online 26 April 2013

Keywords:
Campylobacteriosis
Ecological study
Geographical unit
Modifiable areal unit problem (MAUP)

ABSTRACT

In spatial epidemiology, the choice of an appropriate geographical unit of analysis is a key decision that will influence most aspects of the study. In this study, we proposed and applied a set of measurable criteria applicable for orienting the choice of geographical unit. Nine criteria were selected, covering many aspects such as biological relevance, communicability of results, ease of data access, distribution of exposure variables, cases and population, and shape of unit. These criteria were then applied to compare various geographical units derived from administrative, health services, and natural frameworks that could be used for the study of the spatial distribution of campylobacteriosis in the province of Quebec, Canada. In this study, municipality was the geographical unit that performed the best according to our assessment and given the specific objectives and time period of the study. Future research areas for optimizing the choice of geographical unit are discussed.

1. Introduction

In epidemiology, the study of the spatial distribution of diseases has become more popular in the last decade following new methodological developments and ease of access to geographical information systems. These studies are useful for evaluating hypotheses linking disease occurrence to environmental determinants, but also for identifying regions with unexpectedly high or low incidence. In practice, such investigations are often planned as ecological studies and careful attention needs to be given to their design to minimize the effects of biases, including the well-described ecological bias and the related concept of the modifiable areal unit problem (MAUP) (Waller and Gotway, 2004). MAUP occurs when conclusions of a study conducted for a given dataset are influenced by the way data are aggregated, either in terms of scale and/or boundary delineation. Thus, one of the most crucial elements to consider during study design is the choice of a geographical unit for analysis (Osypuk and Galea, 2007). For the purposes of this study, we defined a *geographical framework* as a set of boundaries delineating an administrative (i.e. census) or natural organization of the territory (i.e. watershed). These frameworks usually include different subsets at various scales. The areas defined by the boundary of a geographical framework at a defined scale were termed a *geographical unit*.

* Corresponding author at: Faculté de médecine vétérinaire, Université de Montréal, 3200 rue Sicotte, Saint-Hyacinthe, Québec, Canada J2S 7C6. Tel.: +1 450 773 8521x86040; fax: +1 450 778 8129.

E-mail addresses: julie.arsenault@umontreal.ca (J. Arsenault), pascal.michel@phac-aspc.gc.ca (P. Michel), oberke@uoguelph.ca (O. Berke), andre.ravel@umontreal.ca (A. Ravel), pierre.gosselin@inspq.qc.ca (P. Gosselin).

1877-5845/$ - see front matter © 2013 The Authors. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.sste.2013.04.004

It is generally recommended that the choice of the unit for spatial analysis should be theory-driven, with the objective of testing hypotheses about specific chains of causation that might link disease occurrence location with potential risk factors (Macintyre et al., 2002; Gregorio et al., 2005). This choice is fundamental because biological and epidemiological mechanisms essential to the dynamic of a disease process at one geographical scale can be unimportant or nonexistent at another (Gotway and Young, 2002). Despite this recommendation, delineation of the geographical unit for studying spatial patterns of diseases has been perceived as a conundrum, for which an operational satisfactory solution still needs to be found (Macintyre et al., 2002; Gregorio et al., 2005; Gauvin et al., 2007). In this context, one of the challenges in choosing the appropriate geographical unit relates to a lack of adequate conceptualization and measurement of the effect of place on health (Macintyre et al., 2002). In addition, the choice of geographical unit is often limited by data availability, either because data on the precise geographical location of cases are not routinely collected in health-related databases or are not disclosed to researchers for privacy protection reasons, or because primary data collection is too expensive (Macintyre et al., 2002; Diez Roux, 2004a; Osypuk and Galea, 2007). Unit choice also represents a compromise between having a unit large enough to get reliable rates and not blurring meaningful local variation (Gregorio et al., 2005; Osypuk and Galea, 2007). For all of these reasons, it is recommended that the relevance of the geographical unit be evaluated prior to any analysis (Boscoe and Pickle, 2003; Diez Roux, 2004b; Osypuk and Galea, 2007; Riva et al., 2008). To our knowledge, there is however no guideline for this task available in the literature.

This paper is presented in two sections. In the first section, we propose a set of practical criteria as a guide for the choice of geographical unit of analysis for ecological studies of infectious diseases. In the second section, we present an application of these criteria in the study of the spatial distribution of campylobacteriosis in Quebec, Canada.

2. Part I – Proposal of criteria for ecological studies

Nine criteria were selected for evaluating and comparing geographical units in the context of ecological studies investigating spatial associations between infectious disease occurrence and environmental characteristics. These criteria were derived from a literature review and from discussions with experts in this field. They cover theoretical considerations (biological relevance), extrinsic considerations (communicability of results, data access), covariate distribution (intra-unit homogeneity), case and population distribution (% of areas with sufficient population size, completeness of geocoded events, variation in population size), and shape of area (variation in areal size, compactness).

2.1. Criterion 1: biological relevance

Biological relevance was defined as whether measured exposure variables accurately and comprehensively depict the hypotheses studied (Osypuk and Galea, 2007). Diverse scales representing different processes might be of interest when studying the spatial patterns of disease (Diez-Roux et al., 2001; Osypuk and Galea, 2007). This criterion was selected for reduction of measurement errors and thus improvement of study validity (Osypuk and Galea, 2007). The biological relevance criterion is more likely to be met when geographical units are purposively created for the problem under study. For example, the use of a geographical unit based on delineation of various landscapes would probably be the most biologically relevant for studying the influence of landscape characteristics on the risk of a particular disease.

2.2. Criterion 2: communicability of results

The communicability of results was defined as the degree of familiarity of the geographical unit for various end-users. Maps based on familiar frameworks do not need additional information to be understood and the information they convey is more easily grasped and recalled (Lewandowsky et al., 1993). The exchange and translation of information between researchers and public health authorities or local stakeholders is considered to be an important public health objective, allowing for efficient implementation of interventions (Lebel et al., 2007). The evaluation of this criterion is highly dependent on the targeted end-users. For instance, the use of watershed geographical units would be highly relevant for people working in watershed management, whereas municipality units are more appropriate for a general audience.

2.3. Criterion 3: data access

The availability of data was defined as the possibility of obtaining appropriate data in a timely manner, and is related to feasibility and validity issues. Data access includes issues related to the availability of existent databases versus the need for field sampling, the type of agreements required for data acquisition, the amount of time needed for data validation and processing prior to analyses, and the errors caused by transforming the data into the appropriate geographical unit. Data access is usually maximized by selecting available data that has already been collected for other purposes, such as census data.

2.4. Criterion 4: intra-unit homogeneity

Intra-unit homogeneity was defined as the level of homogeneity in exposure variables within the areas forming the geographical unit (Gauvin et al., 2007; Flowerdew et al., 2008; Grady and Enander, 2009). When aggregated data are used for the study of an underlying individual-based model, high intra-unit homogeneity reduces the impact of ecological bias from aggregated values as an approximation for individual level data (Salway, 2003; Riva et al., 2008). Furthermore, not all risk factors or determinants of health, such as population immunity or social environment, are reducible to individual level analogs (Reijneveld et al., 2000; Osypuk and Galea, 2007). Such