Be careful with ecological associations

Roumeliotis, S.; Abd ElHafeez, S.; Jager, K.J.; Dekker, F.W.; Stel, V.S.; Pitino, A.; ... ; Tripepi, G.

Received: 1 February 2021 | Accepted: 6 February 2021
DOI: 10.1111/nep.13861

REVIEW ARTICLE

NEPHROLOGY WILEY

Be careful with ecological associations

Stefanos Roumeliotis[1] | Samar Abd ElHafeez[2] | Kitty J. Jager[3] | Friedo W. Dekker[4] | Vianda S. Stel[3] | Annalisa Pitino[5] | Carmine Zoccali[6] | Giovanni Tripepi[6]

[1]Division of Nephrology and Hypertension, 1st Department of Internal Medicine, AHEPA Hospital, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

[2]Epidemiology Department, High Institute of Public Health-Alexandria University, Alexandria, Egypt

[3]Department of Medical Informatics, Academic Medical Center, Amsterdam Public Health Research Institute|, University of Amsterdam, Amsterdam, The Netherlands

[4]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[5]Institute of Clinical Physiology (IFC-CNR), Rome, Italy

[6]Institute of Clinical Physiology (IFC-CNR), Clinical Epidemiology and Physiopathology of Renal Diseases and Hypertension of Reggio Calabria, Reggio Calabria, Italy

Correspondence
Giovanni Tripepi, Clinical Epidemiology of Renal Diseases and Hypertension, Grande Ospedale Metropolitano BMM, Via Vallone Petrara, Reggio Calabria, Italy.
Email: gtripepi@ifc.cnr.it

Abstract

Ecological studies are observational studies commonly used in public health research. The main characteristic of this study design is that the statistical analysis is based on pooled (i.e., aggregated) rather than on individual data. Thus, patient-level information such as age, gender, income and disease condition are not considered as individual characteristics but as mean values or frequencies, calculated at country or community level. Ecological studies can be used to compare the aggregated prevalence and incidence data of a given condition across different geographical areas, to assess time-related trends of the frequency of a pre-defined disease/condition, to identify factors explaining changes in health indicators over time in specific populations, to discriminate genetic from environmental causes of geographical variation in disease, or to investigate the relationship between a population-level exposure and a specific disease or condition. The major pitfall in ecological studies is the ecological fallacy, a bias which occurs when conclusions about individuals are erroneously deduced from results about the group to which those individuals belong. In this paper, by using a series of examples, we provide a general explanation of the ecological studies and provide some useful elements to recognize or suspect ecological fallacy in this type of studies.

KEYWORDS
confounding, ecological fallacy, ecological studies

1 | INTRODUCTION

Ecological studies belong to the family of observational studies and are commonly used in public health research.[1] The main characteristic of an ecological study is that the statistical analysis is based on aggregated (i.e., on population/community-level variables) rather than on individual data.[2] There are three main types of ecological studies: geographical, longitudinal and migration. A typical aim of a geographical ecological study is to compare geographical areas in terms of aggregated population/community health indicators and/or to investigate the association between demographic and socioeconomic data and given health or social condition. A longitudinal ecological study aims to assess time-related trends of the frequency of a pre-defined disease/condition as well as to identify factors aggregated at population level explaining these changes over time in specific populations. A migration ecological study considers migrant populations as units of analysis and can be used to discriminate genetic from environmental causes of geographical variation in disease.

Ecological studies have the advantage to be cheap and provide a rather fast answer to a research question, and can be used to assess/generate various hypotheses. For example, to compare the aggregated prevalence and incidence data of a given condition across different areas (e.g., the burden of poverty), to investigate the relationship between a population-level exposure and a specific disease or

Stefanos Roumeliotis and Samar Abd ElHafeez contributed to the study.

condition (e.g., the association between gross domestic product cycles and suicide trends across time in various countries or the link between air pollution concentration and lung cancer in different cities). On the other hand, ecological studies have several limitations. First, when performing an ecological study to investigate the relationship between the frequency of a risk factor and the burden of a given disease, investigators do not know whether all individuals affected by the disease of interest were really exposed to the risk factor being investigated, as it occurs in studies enrolling individuals as units of analysis. Moreover, ecological studies are prone to confounding,[3] a problem that is difficult to address in this type of studies.[4] Although an ecological study may reflect a causal (etiological) effect between a candidate risk factor and a disease,[5] the problem of ecological fallacy is frequently raised as a bias limiting the use of ecological associations as proofs of etiological links.[6] This bias occurs when conclusions about individuals are erroneously deduced from results about the group to which those individuals belong. From this perspective, we can conclude that a given association emerged in an ecological study is affected by ecological fallacy only by performing an individual-level analysis. In an ecological study performed in the late 19th century,[7] a direct association was reported between the proportion of Protestants in a series of Prussian communities and the frequency of suicides. The hypothesis that being Protestant is a risk factor for suicide is an example of ecological fallacy. In fact, a more careful evaluation of available data revealed that in the late 19th century, most of the suicides within the Prussian communities were committed by Catholics who, when in the minority because of the high percentage of Protestants in a given community, are isolated from a social point of view and for this reason at higher risk of suicide. The misleading conclusion from the ecologic study about the link between the proportion of Protestants and the frequency of suicide mentioned above is an example of ecologic fallacy.[6] Although the 'ecological fallacy' is a bias of primary importance, ecological studies have the advantage of accounting for 'individualistic fallacy' which occurs when individuals are assumed to be unaffected by the environment in which they live.[8] Here, we provide six examples: *Example 1* gives a hypothetical example of ecological fallacy; *Example 2* is an ecological study aimed at identifying the magnitude and the reasons of disparities to access cardiac procedures (coronary artery bypass, coronary angioplasty, and so on) after an acute myocardial infarction (MI) in Ontario, Canada; *Example 3* underscores the importance of ecological studies to discovery of previously unsuspected risk factors for a given disease (e.g., salt intake as a risk factor for hypertension); *Examples 4–6* underline the caution to be adopted when interpreting the results of an ecological study because of the bias due to ecological fallacy.

SUMMARY AT A GLANCE

This article describes the application of ecological studies to provide an answer to specific epidemiological questions. This paper focuses on the problem of ecological fallacy, a bias which occurs when conclusions about a given relationship in individuals are deduced from inferences about the group to which those individuals belong.

2 | EXAMPLE 1

We consider three different populations in which systolic blood pressure and serum creatinine were simultaneously assessed in a series of individuals. Within each population, at the individual-level analysis, systolic blood pressure and serum creatinine were directly and significantly interrelated (see Figure 1, left panel). Then, we decide to investigate the problem of the systolic blood pressure-creatinine link by an

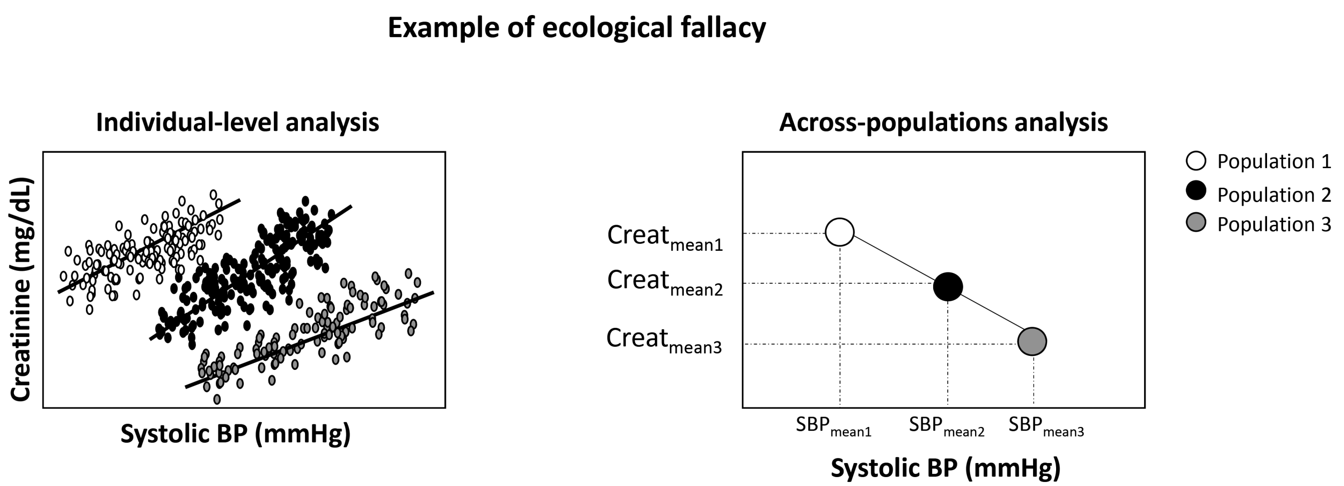

FIGURE 1 Example of ecological fallacy (see Example 1 for details). SBP_{mean1} and $Creat_{mean1}$ = averages of systolic blood pressure and creatinine, respectively, in individuals belonging to the Population 1; SBP_{mean2} and $Creat_{mean2}$ = averages of systolic blood pressure and creatinine, respectively, in individuals belonging to the Population 2; SBP_{mean3} and $Creat_{mean3}$ = averages of systolic blood pressure and creatinine, respectively, in individuals belonging to the Population 3

ecological approach, using the same data, that is, by an across-populations analysis (see Figure 1, right panel). To do this, we calculate the mean values of serum creatinine and the mean values of systolic blood pressure separately in the three populations and look for the correlation among these average values. Given the fact that mean systolic blood pressure is lower in population 1 than in populations 2 and 3 whereas the average creatinine displays an opposite pattern (it is higher in population 1 than in populations 2 and 3), in an ecological analysis we find an inverse relationship between creatinine and systolic blood pressure. In this case, the type of population acts as a confounder in the link between systolic BP and serum creatinine. Thus, the results of the across-populations analysis do not match with those of individual-level analysis, resulting in an ecological fallacy.[6] If the conclusion about the direction of the systolic blood pressure-creatinine link had been drawn from the ecological analysis, we would have erroneously concluded that this link was an inverse one. In general, it is difficult to adjust for confounders in ecological studies, mainly because of the relatively low number of units of analysis (in the across-populations analysis of Example 1, only three groups are available and for this reason, data adjustment for confounding is practically impossible).

3 | EXAMPLE 2

Alter et al. performed an ecological study[9] to assess the relationship between the socioeconomic status of patients, their access to specific cardiac procedures (coronary artery bypass, coronary angioplasty, and so on), and short-term (12 months) mortality rate after an acute MI in Ontario, Canada. Although the universal healthcare system in Canada aims to avoid discrimination, patients living in higher-income areas have the highest rates of cardiac procedures and the lowest 12-month mortality rates. By using an ecological approach to provide an answer to their research question, Alter et al did not really study the effect of socio-economic status of patients but rather the effect of living in a neighbourhood with a certain socio-economic status. The same authors also performed a follow-up study,[10] by using individual-level clinical and socioeconomic status in 3407 patients who were hospitalized for acute MI in 53 large-volume hospitals in Canada from December 1999 to February 2003, and found that the difference in mortality rate after acute MI was largely explained by differences in baseline cardiovascular risk factors among patients rather than the disparity in the process of care. This finding is of paramount importance from a public health perspective because it suggests that universal healthcare by itself does not eradicate health disparities in Canada and underlines that the management of cardiovascular risk factors and the promotion of healthy behaviour are absolute public health priorities in the country, particularly in the poorest strata of the Canadian population. The peculiarity of this study is that the investigators provided an answer to a public health research question (*Does income affect the access to specific cardiac procedures in Canada?*) in two steps: (1) by using an ecological study and (2) by performing an individual-level analysis, this latter providing a plausible explanation of the inequalities emerged in the ecological study.

4 | EXAMPLE 3

Several epidemiological and experimental studies have consistently demonstrated a significant relationship between salt intake and blood pressure. The first evidence about this relationship emerged in ecological studies. In 1960, an across-population link between salt intake and blood pressure was firstly described by Dahl[11] who found a linear relationship between the mean sodium intake and the burden of hypertension in five different populations. Dahl also reported that the burden of hypertension was relatively low in populations with salt intake below 4–5 g salt/day and hypothesized that salt intake increased the chance of high blood pressure in humans. In 1979, another ecological study by Froment et al.,[12] using data derived from 28 populations, reported that an increase of 100 mmol/day of sodium intake associated with a 10 mmHg average increase in blood pressure. Although these ecological studies were prone to several bias, confounding, and other methodological problems, they have had the merit to pave the way in 1988 to an international observational study (the INTERSALT study[13]) carried out on 10 074 individuals in 32 countries worldwide to investigate the salt-blood pressure relationship. In this large study, each patient underwent both blood pressure and sodium urine measurements. In an individual-level analysis (n = 10 074), patients with 24 h urinary sodium excretion higher than 100 mmol displayed an average increase of 6 mmHg of systolic blood pressure as compared to patients with 24 h urinary sodium below this level. Of note, the effect derived by this individual-level analysis was almost identical to that observed in an across-population analysis of the same study using countries as units of analysis (n = 52). The ecological analysis showed that a 100 mmol increase of the sample median of 24 h sodium excretion associated with 5–7 mmHg increase of the sample median of systolic blood pressure. Thus, individual- and population-level analyses provided on average similar results. However, although the effect size of salt intake on blood pressure was similar on average between across-populations analysis (n = 52 populations) and individual-level analysis (n = 10 074), the estimation of such an effect obtained using individual data is more precise (i.e., with a narrower 95% confidence interval) than that obtained by using aggregated data.

5 | EXAMPLE 4

In an ecological study, Ralph et al.,[14] tested the hypothesis that population-wide use of diuretics might be associated with the incidence of end-stage renal disease (ESRD). The authors used aggregated data trends of both ESRD incidence and use of antihypertensive drugs in the United States. Renal failure data were obtained from the United States Renal Data Service, and drug information obtained from IMS Health (Fairfield, CT). They found a direct and significant relationship between the annual changes in diuretic distribution with concomitant annual changes in the incidence of ESRD (r = 0.75, p = 0.03) and drew the conclusion that diuretic therapy could be a risk factor of ESRD. They concluded that their results *'imply that diuretics appear to permit, induce or possibly accelerate renal* *disease in a small but significant proportion of diuretic-treated patients*'. Such a statement clearly suggests an etiological role for ESRD of the diuretics use at individual level, a pathogenetic hypothesis which demands to be formally tested by using individual-level rather than population-level data. In fact, the ecological study is not well suited for hypothesis testing because of unmeasured and uncontrolled confounding and from this perspective it is prone to ecological fallacy. Thus, the hypothesis generated in the Ralph's study requires further individual-level studies to confirm or deny a potential role of diuretics in the risk of ESRD in the community.

6 | EXAMPLE 5

A study by Pickett et al.,[15] further illustrates the concept of ecological fallacy. These authors performed a cross-sectional, ecological study to assess the association between child wellbeing and three macroeconomic measures, including material living status (assessed by average income), social status (assessed by income inequality) and social exclusion (assessed by children living in relative poverty) in rich, developed countries. In this cross-national comparison, data from 23 of the top 50 richest countries in the world were included. Child wellbeing was assessed by the overall Unicef index of wellbeing. Low scores of the index indicated worse outcomes. Income inequality was calculated as the ratio of the annual, total household income that was received by the richest 20% of the population divided by that received by the poorest 20%. Child relative poverty, another outcome of the study, was assessed as the proportion of children between 0 and 17 years old living on less than half the national median annual household income. By using Pearson's correlations, the authors found that the overall Unicef index of wellbeing was strongly, linearly and inversely correlated with the income inequality and the proportion of children living in relative poverty ($r = -0.64$, $p = 0.001$ and $r = -0.67$ and $p = 0.001$), respectively. The authors concluded that in rich, developed countries, child wellbeing improvements might depend more on decrease of inequality income and the percentage of children living in relative poverty, rather than on other measures targeting further economic growth. However, the authors cannot be sure that within a certain country the children that lived in relative poverty had also low levels of wellbeing. Thus, also in this case a problem of ecological fallacy can be suspected when the authors infer that children with lower wellbeing status were more likely to live in households with relative poverty. Therefore, an individual-level analysis is required to confirm the study hypothesis.

7 | EXAMPLE 6

Here, we focus on a study by Donnan et al.,[16] who investigated the association between antibiotic exposure and resistance to these drugs in a repeated cross-sectional study in 1995 and 1996. Data on antibiotic resistance were collected based on culture and sensitivity tests of midstream urine collection. Samples with and without trimethoprim resistant Gram-negative bacteria were identified and then linked to patient records. The crude Spearman rank correlation between practice prescribing of trimethoprim and resistance was very low (rho = −0.039) and not significant and this remained after adjustment for some potential confounders. The researchers concluded that the general practices with the highest rates of trimethoprim prescribing have similar proportions of patients with trimethoprim resistance as those practices with the lowest rates of trimethoprim prescribing. To apply these findings at the level of individuals, we need to take care of the ecological fallacy. To this scope, the authors linked patient records indicating prescription of trimethoprim and the urine test results. In this analysis, they found that trimethoprim resistance was significantly associated with age, sex and past exposure to trimethoprim or other antibiotics. The association with trimethoprim resistance was strongest for people who had recently been exposed to trimethoprim. For this reason, the researchers concluded that the analysis of data at the practice level obscured important associations between antibiotic prescribing and resistance at the individual level. Thus, care must be taken when extrapolating data either to individuals within the area level of measurement, or to a higher population level due to the liability to potential bias as ecological fallacy. Moreover, causation cannot be determined.

8 | CONCLUSIONS

Ecological studies should be considered as the first step, especially for public health objectives, to generate hypotheses at the population level. Depending on the research question, researchers should consider the possibility to replicate the results obtained by population-level analyses in individual-level studies. This type of studies plays an important role in public health research and are of interest both for policy makers and clinicians. A fundamental question when reading/interpreting an ecological study is whether the ecological study is the appropriate tool to provide the answer to the research question being addressed or it would have been more appropriate to apply another study design, for example, an individual-level observational study or a randomized controlled trial. The conclusions of an ecological study should be carefully evaluated in order to assess whether they are biologically plausible, whether alternative explanations exist to interpret the results and whether all potential confounders were taken into account in the data analysis. When reading an ecological study we should be always aware of the possibility of an ecological fallacy whereby potentially misleading causal inferences might be generated.

ORCID

Giovanni Tripepi https://orcid.org/0000-0002-3580-4602

REFERENCES

1. Morgenstern H. Ecologic studies. In: Rothman KJ, Greenland S, eds. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998:459-480.
2. Greenland S. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int J Epidemiol*. 2001;30:1343-1350.

3. Van Stralen KJ, Dekker FW, Zoccali C, Jager KJ. Confounding. *Nephron Clin Pract*. 2010;116(2):c143-c147.
4. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions—a primer for the clinician. *Clin Epidemiol*. 2017;9:185-193.
5. Szklo M, Nieto FJ. Epidemiology. *Beyond the Basics*. 2nd ed. Sudbury, MA: John and Bartelett Publishers; 2004.
6. Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol*. 1988;127:893-904.
7. Morgenstern H. Ecological studies in epidemiology: concepts, principles, and methods. *Ann Rev Public Health*. 1995;16:61-81.
8. Loney T, Nagelkerke NJ. The individualistic fallacy, ecological studies and instrumental variables: a causal interpretation. *Emerg Themes Epidemiol*. 2014;19(11):18.
9. Alter DA, Naylor CD, Austin P, Tu JV. Effects of socioeconomic status on access to invasive cardiac procedures and on mortality after acute myocardial infarction. *N Engl J Med*. 1999;341:1359-1367.
10. Alter DA, Chong A, Austin PC, et al. Socioeconomic status and mortality after acute myocardial infarction. *Ann Intern Med*. 2006;144:82-93.
11. Dahl LK: Possible role of salt intake in the development of essential hypertension, in Cottier P, Bock KD (eds): *Essential Hypertension—An International Symposium*. Berlin, Springer-Verlag, 1960, pg. 53-65.
12. Froment A, Milon H, Gravier C. Relation entre consommation sodee et hypertension arterielle. Contribution de l'epidemiologie geographique. *Rev Epidemiol Sante Publiaue*. 1979;27:437-454.
13. Intersalt. An international study of electrolyte excertion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research Group. *Br Med J*. 1988;297:319-328.
14. Hawkins RG, Houston MC. Is population-wide diuretic use directly associated with the incidence of end-stage renal disease in the United States? *Am J Hypertens*. 2005;18:744-749.
15. Pickett KE, Wilkinson RG. Child wellbeing and income inequality in rich societies: ecological cross sectional study. *BMJ*. 2007;335(7629):1080.
16. Donnan PT, Wei L, Steinke DT, et al. Presence of bacteriuria caused by trimethoprim resistant bacteria in patients prescribed antibiotics: multilevel model with practice and individual patient data. *BMJ*. 2004;328:1297.

How to cite this article: Roumeliotis S, Abd ElHafeez S, Jager KJ, et al. Be careful with ecological associations. *Nephrology*. 2021;26:501–505. https://doi.org/10.1111/nep.13861