# Explaining the Atomistic versus Ecological Fallacies in SES-Health Gradients

*By* WILLIAM PARKER AND JOHANNES SPINNEWIJN*

A long literature in economics and public health has examined the relationship between socioeconomic status (SES) and health, often relying on area-based measures when individual-level data are unavailable. Using comprehensive Dutch administrative data, we build on work estimating income-health gradients (e.g., Chetty et al. (2016)) and study how the level of aggregation systematically shapes estimates by inducing so-called *ecological* or *atomistic* biases. Individual and area income proxy for different channels (e.g., personal resources vs. local surroundings), but are obviously correlated. Aggregating to the area level reduces meaningful variation and attenuates non-linearities. We find that area-level estimates exaggerate the income gradient relative to individual-level estimates and are less robust to including individual and area-level controls respectively. Moreover, interaction analyses reveal that area exposure is highly unequal: the health of low-income households are far more sensitive to neighbourhood conditions than high-income households. Together, these findings shed light on the sources of SES–health gradients and show how data aggregation affects both interpretation and empirical estimates.

## I. The Ecological and Atomistic Fallacies

In research on health inequality the unit of interest is not always the unit of observation. Researchers often analyse information at higher levels of aggregation than would be ideal, substituting area-based measures for unavailable individual-level data. For example, there are over 1,500 papers on PubMed and 250 on EconLit from the past decade that examine health inequality using SES observed at the area level only. This approach raises concerns about the ecological fallacy, defined by Robinson (1950) as the error of inferring individual-level relationships from area-level correlations. The converse error — the atomistic or individualistic fallacy — involves ignoring contextual influences by focusing exclusively on individual characteristics. Although less frequently discussed in health disparities research, this fallacy is potentially equally problematic, as it can obscure neighbourhood or other local determinants of health (Subramanian et al., 2009). Distinguishing between ecological and individual-level relationships is not merely methodological but may reflect substantively different causal pathways. Individual income proxies for socioeconomic status and available household resources specifically, while neighbourhood income reflects broader contextual and compositional characteristics—such as local amenities, social environment, and shared exposure to area-level conditions—that shape observed health gradients (Diez Roux and Mair, 2010).

We can conceptualise the ecological fallacy and the atomistic fallacy using an omitted variable bias framework. Let us assume a data-generating process where individual and area-level factors *both* play a role in health and that these factors are well proxied by income, measured at the individual or area-level respectively. To fix ideas, we observe an individual $i$ living in a geographical area $j$, with income level $Y_{i,j}$ and measured health $H_{i,j}$. Both the ecological model and the atomistic model are biased relative to a 'unified' model, as they fail to separate individual effects from area-level effects.

* Parker: London School of Economics (w.d.parker@lse.ac.uk); Spinnewijn: London School of Economics and CEPR (j.spinnewijn@lse.ac.uk). We thank Amy Finkelstein, Jon Kolstad, Michael Stepner and participants of the 2026 ASSA meeting for valuable comments and suggestions, the Dutch Central Bureau of Statistics for data support, and Manuel De Faria and Rasmus Duret for excellent research assistance. We thankfully acknowledge financial support from UC Berkeley Haas Center for Healthcare Marketplace Innovation and the European Research Council through grant 101200402.

We consider the following simple linear additive model:

$$H_{ij} = \beta_{\text{indiv}} Y_{ij} + \beta_{\text{area}} \overline{Y}_j + \nu_{ij} \qquad \text{(Unified)}$$
$$\overline{H}_j = \beta_{\text{eco}} \overline{Y}_j + \nu_j \qquad \text{(Ecological)}$$
$$H_{ij} = \beta_{\text{atom}} Y_{ij} + \nu_{ij} \qquad \text{(Atomistic)}$$

These are very stylised, in practice we would expect the true data generating process to be a dynamic, reflexive process between health, income and other inputs, however they serve to highlight the biases in question. It is straightforward to show that the ecological and atomistic specifications fail to separately identify $\beta_{\text{indiv}}$ and $\beta_{\text{area}}$. First, taking area means of the unified specification, we see the ecological coefficient simply sums the individual and area components, $\beta_{\text{eco}} = \beta_{\text{indiv}} + \beta_{\text{area}}$. By only using between-area variation, this model adds up the individual and area-level effects.[1] Second, using the familiar omitted variable bias results, the coefficient from the atomistic specification is $\beta_{\text{atom}} = \beta_{\text{indiv}} + \rho \cdot \beta_{\text{area}}$, where $\rho$ equals the between-area share of the total variance in incomes $Var(\overline{Y}_j)/Var(Y_{ij})$. The atomistic bias is stronger if $\rho$ is large, i.e. more of the income variation is between-areas (which the atomistic specification ignores), or if the contextual effect $\beta_{\text{area}}$ is large relative to the individual effect. Assuming that $\beta_{\text{indiv}}$ and $\beta_{\text{area}}$ have the same sign, both atomistic and ecological biases will result in an overestimate, compared to the unified specification.[2]

## II. Context and Data

Methodological research has previously shown that estimates of $\beta_{\text{eco}}$ may exaggerate, attenuate, or reverse the sign of $\beta_{\text{atom}}$, depending on within-area heterogeneity, population density, and the spatial scale of aggregation (e.g., Piantadosi, Byar and Green, 1988; Krieger et al., 2002; Davis, Mahar and Strumpf, 2023). Other work has used multi-level modelling to analyse the respective role of individual-level and area-level factors and potential heterogeneity by race, class, sex and time (e.g., Subramanian et al., 2009; Diez Roux and Mair, 2010; Decoster, Minten and Spinnewijn, 2021). We take a straightforward empirical approach, motivated by the power of administrative data to precisely document health gradients (e.g., Chetty et al., 2016). By focusing on a single regressor, income, which is measurable at the area and the individual level, we can directly compare the relative strength of these biases.

The Dutch data are well-suited to studying these issues, because we observe all relevant information at the individual level, alongside detailed information on place of residence, within a nested geographic hierarchy. Here we focus on both the municipality level (Dutch: *gemeenten*), which are comparable in population size to US counties, while neighbourhoods (Dutch: *buurten*) are approximately one-quarter the size of US census tracts in terms of population.

Our primary measure of health is a mortality-weighted chronic disease index (CDI) developed by Danesh et al. (2024), which quantifies the underlying mortality risk at old age arising from diagnosed chronic conditions. The CDI provides a continuous measure of health status for all individuals at all ages given their profile of chronic conditions and does not require long follow-up periods or age-specific calculations. This property improves statistical precision even in small geographic samples and enables more granular analysis of health disparities. It is also observable at prime working age, when income is most relevant as a measure of SES. Our primary measure of socioeconomic status is household disposable income—total income after taxes and transfers — which reflects both market earnings and redistribution and thus captures households' available material resources. Using

[1] This ecological estimator is identical to that for a hybrid specification, where the dependent variable is observed at the individual level, $H_{ij}$, but the regressor used is the area-level mean $\overline{Y}_j$. This type of specification is often used in a mediation analysis when certain mediators of interest are not available at the individual level.

[2] We can also analyse the 'classic' ecological fallacy question within this framework. If the atomistic $\beta_{\text{atom}}$ is the coefficient of interest, but we are constrained to estimate the ecological specification $\beta_{\text{eco}}$, then we have the following expression for the bias: $\beta_{\text{eco}}/\beta_{\text{atom}} - 1 = (1-\rho)/(\beta_{\text{indiv}}/\beta_{\text{area}} + \rho)$. The bias shrinks as the between-region share of income variation increases, or if the individual effect is strong relative to the area effect.

individual-level administrative data on income and health, we can construct area-based measures of income and health aggregated at the neighbourhood and municipal levels and quantify the magnitude of biases arising from both the ecological and individualistic fallacies.

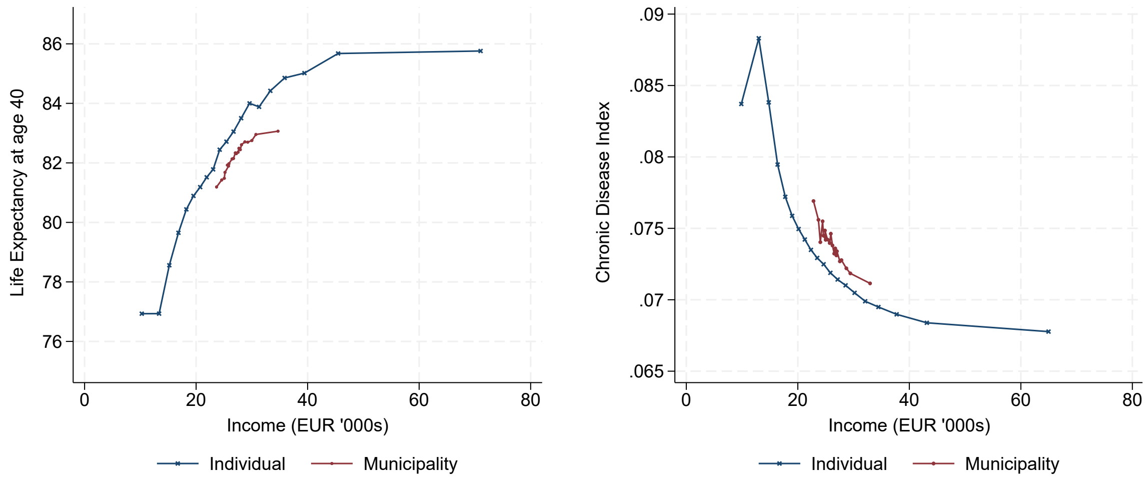


Figure 1. : Income–Health Gradients, Individual and Municipality Level

*Notes:* Household disposable income, expressed in 2013 Euros. Income ventile bins are constructed within age and sex. Life expectancy is computed as per Chetty et al. (2016) based on ages 40-80. Chronic disease index is taken from Danesh et al. (2024), based on ages 25-64. Area series are derived from municipality means.

As shown in Figure 1, there is a strong gradient between income and health, observed at the individual level, as well as the area level. The pattern in gradients for life expectancy and CDI is very similar, albeit reversed. The income distribution becomes markedly more compressed as we increase the geographic aggregation, seen as a horizontal compression of the quantiles. This is because 98% of the variation in income is within municipalities (88% within neighbourhoods).[3] By the same token, the vertical compression of the quantiles shows how the health distribution shrinks with geographic aggregation as over 99% of the variation in CDI is within municipalities (98% within neighbourhoods). We also note the strong nonlinearity in the mortality and CDI gradients, but the concavity of the income-health gradient is largely attenuated for the area-based measures, due to Jensen's Inequality.

## III. Results: Income Specifications and Health Gradients

Our main results compare income–health gradients across the three specifications: atomistic, ecological, and unified. Table 1 reports coefficient estimates from each specification, with age and sex effects removed. Given the concavity observed in Figure 1, we use logged CDI as the dependent variable and an alternative specification with income expressed in percentiles (see right panel of Table 1). The coefficients on $Y_{ij}$ and $\overline{Y}_j$ are directly comparable across specifications. For example, under the atomistic income-level specification, a EUR 10k increase in household income is associated with a 2.5% decline in mortality-weighted chronic disease.

Differences across the three specifications are substantial. First, comparing the two reduced-form specifications directly, the ecological estimate $\hat{\beta}_{\text{eco}}$ yields an income–health gradient that is 142 per-

[3] Although most income variation occurs within areas, there is meaningful spatial income sorting: the correlation between individual and area income is 0.167 at the municipality level (0.342 at the neighbourhood level).

cent larger than the atomistic estimate. That is, a EUR 10k increase in area-level income is associated with a 6.2% decline rather than a 2.5% decline in mortality-weighted chronic disease.[4] Second, within the unified specification, the area-level coefficient $\hat{\beta}_{\text{area}}$ remains larger than the individual-level coefficient $\hat{\beta}_{\text{indiv}}$, but the individual-level coefficient is still sizeable, suggesting that these coefficients reflect separate associations with health outcomes. Third, and relatedly, the reduction in the area-level estimate when going from the ecological to the unified regression is substantial, from $\hat{\beta}_{\text{eco}} = 0.62$ to $\hat{\beta}_{\text{area}} = 0.42$, and dominates the reduction in the individual-level estimate when going from the atomistic to the unified regression, from $\hat{\beta}_{\text{atom}} = 0.26$ to $\hat{\beta}_{\text{indiv}} = 0.21$. These patterns are consistent with the omitted-variable-bias expressions, given the observed correlation between individual and area income ($\rho = 0.122$).

**Robustness and Heterogeneity.** The estimates and relative magnitudes are consistent when using income percentiles instead of income levels. The relationship between income and health becomes more linear, making the linear regression framework more suitable. The relative difference between the ecological and atomistic estimates become smaller and the relative difference between the area-level and individual-level estimates mostly disappears. Relatedly, the ecological bias has become even stronger in relative terms, while the atomistic bias has become smaller. Of course, the interpretation now differs as the estimated effects are associated with changing percentiles in the area-level and individual-level income distribution respectively.Partitioning by sex, we find that income gradients are somewhat steeper for females, despite a lower level effect. The balance between individual versus area gradients is relatively consistent between sexes however. Partitioning by age, we find that gradients for younger groups are driven relatively more by individual income than area income, compared to older cohorts, who arguably have been exposed for longer to area effects. This results in a 80% ecological bias for 25-35 year-olds, declining to 41% for those aged 55-65.

**Other socio-economic controls.** An alternative interpretation of the large ecological bias relative to the atomistic bias is that the area-level estimates are more sensitive to including individual-level controls than the individual-level estimates are to including area-level controls. For example, $Y_{ij}$ and $\overline{Y}_j$ could be incomplete measures of latent individual SES, which in turn links to health. Given that the ecological bias is greater than the atomistic bias, this would then suggest that $Y_{ij}$ is a more robust measure of latent SES than $\overline{Y}_j$. To test this, we add individual SES controls to the unified specification, notably wealth percentile, household composition, work status, and whether foreign born. This indeed reduces the area-level estimate substantially, from 4.2% to 2.0%. While it remains high, it raises the concern that the area-level estimate continues to pick up unobserved individual-level factors. Clearly, controlling for individual SES factors substantially reduces the individual-income effect too. To explore the robustness of the individual income channel to area-level factors one would like to consider including area fixed-effects, as in multi-level modelling: however, including $\overline{Y}_j$ is sufficient to capture all between-area variation correlated with $Y_{ij}$, so including area fixed-effects would not change the $\beta_{\text{indiv}}$ estimate.[5]

**Interactions.** While the regression specifications in Table 1 mean that each bias can be represented in a single statistic, they assume that the effects of $Y_{ij}$ and $\overline{Y}_j$ are linear and additive. To more flexibly model the relationship between individual income, area income, and health, we can instead use a specification with interactions between deciles of $Y_{ij}$ and $\overline{Y}_j$:

$$H_{ij} = \sum_{a=1}^{10} \sum_{d=1}^{10} \beta_{a,d} \, \mathbf{1}\big\{Q(\overline{Y}_j) = a, Q(Y_{ij}) = d\big\} + \nu_{ij}$$

where $Q(x)$ is a quantile function that maps $x$ to its decile rank. Figure 2a plots $\hat{\beta}_{1,d}$ in red and $\hat{\beta}_{10,d}$

[4]This is a gap larger than may be expected by inspecting the right-hand panel of Figure 1, but this is partly driven by the non-linear relationship and partly by age and gender effects, which absorb relatively more of the individual gradient than the area gradient.

[5]Relatedly, results are qualitatively similar at the municipality level instead of the neighbourhood level, but we find a higher individual-level estimate $\beta_{\text{indiv}}$ as the share of income variation across municipalities is much smaller than across neighbourhoods.

Table 1—: Health-Income Gradients, by Specification

| | *Income levels (EUR '000s)* | | | | *Income percentiles* | | | |
|---|---|---|---|---|---|---|---|---|
| | Atom. | Ecol. | Unif. | Unif. + SES | Atom. | Ecol. | Unif. | Unif. + SES |
| $\beta_{\text{atom}}, \beta_{\text{indiv}}$ | -0.257 | | -0.206 | -0.060 | -0.160 | | -0.136 | -0.039 |
| | (-72.2) | | (-72.6) | (-44.0) | (-123.8) | | (-153.2) | (-62.4) |
| $\beta_{\text{eco}}, \beta_{\text{area}}$ | | -0.623 | -0.417 | -0.202 | | -0.324 | -0.188 | -0.094 |
| | | (-43.3) | (-32.3) | (-26.6) | | (-65.1) | (-39.3) | (-22.9) |
| Adj. $R^2$ | 0.009 | 0.007 | 0.012 | 0.058 | 0.016 | 0.008 | 0.018 | 0.058 |

| | *Females only* | | | *Males only* | | | *Ages 25–35* | | | *Ages 55–65* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Atom. | Ecol. | Unif. | Atom. | Ecol. | Unif. | Atom. | Ecol. | Unif. | Atom. | Ecol. | Unif. |
| $\beta_{\text{atom}}, \beta_{\text{indiv}}$ | -0.319 | | -0.252 | -0.198 | | -0.163 | -0.174 | | -0.157 | -0.300 | | -0.236 |
| | (-71.3) | | (-70.9) | (-67.1) | | (-65.3) | (-63.6) | | (-59.4) | (-53.7) | | (-50.4) |
| $\beta_{\text{eco}}, \beta_{\text{area}}$ | | -0.783 | -0.533 | | -0.460 | -0.295 | | -0.240 | -0.133 | | -0.875 | -0.619 |
| | | (-42.1) | (-31.6) | | (-42.5) | (-30.3) | | (-33.4) | (-20.1) | | (-35.9) | (-27.8) |
| Adj. $R^2$ | 0.011 | 0.008 | 0.015 | 0.008 | 0.005 | 0.010 | 0.004 | 0.002 | 0.005 | 0.012 | 0.010 | 0.016 |

*Notes:* This table reports coefficients (scaled $\times 100$) from regressions of log(CDI) on household disposable income using 2013 data of all Dutch residents aged between 25 and 65. Individuals $N_i = 8.43$m, Neighbourhood areas $N_j = 13.3$k. Age$\times$sex effects are absorbed. t-statistics in parentheses, clustered at the neighbourhood level. SES controls include household wealth percentile, household composition, work status, and whether foreign born. Results in the lower panel use income in levels (EUR '000s).

in blue for each $d \in (1, 2, ..., 10)$. The individual income gradient is roughly twice as steep for those living in a bottom decile income area, compared to those living in the top decile. Conversely, Figure 2b plots $\hat{\beta}_{a,1}$ in red and $\hat{\beta}_{a,10}$ in blue for each $a \in (1, 2, ..., 10)$. The area income gradient is relatively pronounced for those with bottom decile individual incomes, but very muted for those with top decile incomes.

This muted area–income gradient at the top of the individual income distribution complicates the interpretation of the area effect reported in Table 1. As high-income individuals show little response to neighbourhood income while low-income individuals show much stronger responses, the linear area coefficient $\hat{\beta}_{\text{area}}$ masks this heterogeneity and is not reflecting a common contextual effect. Health outcomes of lower-income individuals appear to be more sensitive to neighbourhood conditions, in line with earlier work (e.g., Chetty et al., 2016). Still, we cannot exclude that this apparent increase in area sensitivity for poorer individuals is due to measurement error: perhaps neighbourhood average income is a more informative measure of SES for low income households, than for high income ones; or conversely that individual income is a more informative measure of SES in low-income areas than in rich areas. Regardless of the interpretation, this interaction also reinforces the convexity of the health-income gradient.

**Summary & research directions.** Aggregation choices are not innocuous. We show that the estimated SES–health gradient varies by as much as 25–140% depending on whether SES is measured at the individual or neighbourhood level and on the comparison being made. Disentangling these tightly correlated channels requires research designs that align identification with the channel of interest. When the individual channel is central, panel data on health — such as repeated measures of the CDI — allow within-person changes in income and health to be isolated, (e.g., Danesh et al., 2024). When neighbourhood conditions are the object of interest, movers designs that exploit residential relocation, as in Finkelstein, Gentzkow and Williams (2021) and Parker (2026), offer a credible strategy for separating place effects from individual selection. More generally, progress in this literature will depend on designs that jointly model individual and contextual SES, rather than treating aggregation as a secondary modelling choice.

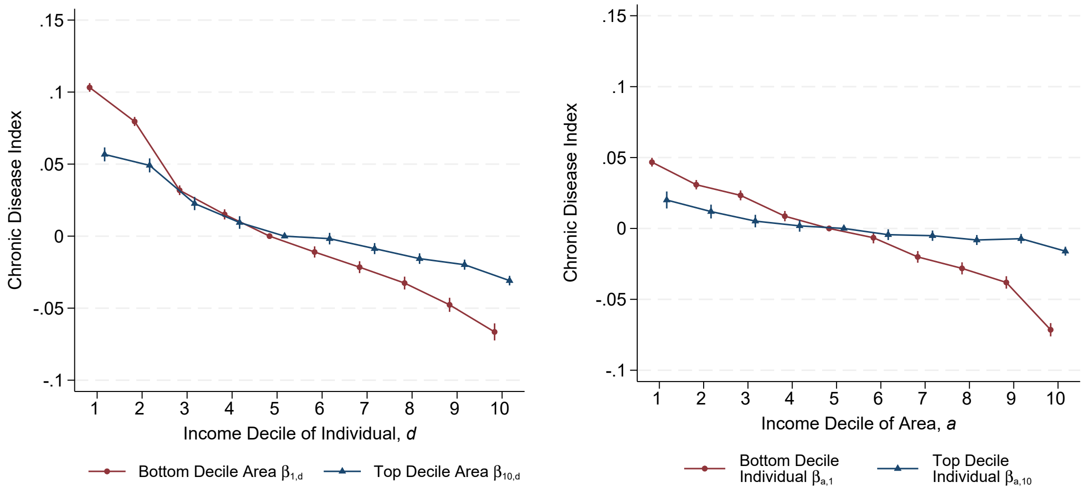


Figure 2. : Decile Interactions Between Neighbourhood and Individual Gradients

*Notes:* Income is household-level disposable income, deciles are constructed within age and sex. Chronic disease index is taken from Danesh et al. (2024), based on all Dutch residents ages 25-64, N=8.43m. Area denotes neighbourhood of residence (*NL: Buurt*).

## REFERENCES

**Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler.** 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014." *JAMA*, 315(16): 1750–1766.

**Danesh, Kaveh, Jonathan T. Kolstad, William D. Parker, and Johannes Spinnewijn.** 2024. "The Chronic Disease Index: Analyzing Health Inequalities Over the Lifecycle." National Bureau of Economic Research Working Paper 32577.

**Davis, Laura E, Alyson L Mahar, and Erin C Strumpf.** 2023. "Agreement between Individual and Neighborhood Income Measures in Patients with Colorectal Cancer in Canada." *JNCI: Journal of the National Cancer Institute*, 115(5): 514–522.

**Decoster, André, Thomas Minten, and Johannes Spinnewijn.** 2021. "The income gradient in mortality during the COVID-19 crisis: evidence from Belgium." *The Journal of Economic Inequality*, 19(3): 551–570.

**Diez Roux, Ana V, and Christina Mair.** 2010. "Neighborhoods and Health." *Annals of the New York Academy of Sciences*, 1186(1): 125–145.

**Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams.** 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review*, 111(8): 2697–2735.

**Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson.** 2002. "Geocoding and Monitoring of US Socioeconomic Inequalities in Mortality and Cancer Incidence." *American Journal of Epidemiology*, 156(5): 471–482.

**Parker, William.** 2026. "Zooming in on Health: Using High-Resolution Spatial Data to Understand the Geography of Health Inequality." PhD diss. London School of Economics, PhD thesis chapter.

**Piantadosi, Steven, David P. Byar, and Sylvan B. Green.** 1988. "The Ecological Fallacy." *American Journal of Epidemiology*, 127(5): 893–904.

**Robinson, W. S.** 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review*, 15(3): 351–357.

**Subramanian, S V, Kelvyn Jones, Afamia Kaddour, and Nancy Krieger.** 2009. "Revisiting Robinson: The Perils of Individualistic and Ecologic Fallacy." *International Journal of Epidemiology*, 38(2): 342–360.