Peer Review Report

Review Report on Associations between gender gaps in life expectancy, air pollution, and urbanization: A global assessment with Bayesian spatiotemporal modeling

Original Article, Int J Public Health

Reviewer: Mauricio Tec
Submitted on: 06 Nov 2022
Article DOI: 10.3389/ijph.2023.1605345

EVALUATION

Q 1 **Please summarize the main findings of the study.**

The paper investigates spatiotemporal trends in the gender gap in life expectancy (GGLE) and gender-specific life expectancy (LEm and LEf). The statistical methodology and analysis are almost identical to a recently published study using the same dataset but looked only at life expectancy without considering the gender gap [1]. The main findings are (a) that most countries have mixed GGLE trends, being generally lower in Africa and South-East Asia; (b) the statistical model suggests that lower air quality is associated with a higher gender gap.

Q 2 **Please highlight the limitations and strengths.**

Strengths
1. The topic is relevant.
2. Accommodating spatial effects in the statistical method is important.
3. Bayesian inference allows for uncertainty quantification.

Limitations
1. Not carefully written. There are multiple broken sentences.
2. Repeated misuse of the word effects suggests a causal relationship when the method only supports the discovery of statistical associations.
3. The authors particularly interpret the regression coefficients to imply a causal relation and suggest strong policy implications.
4. Some plots are redundant.
5. Contrary to their flexible spatial model, their analysis of temporal trends could be more extensive. In particular, it gives most countries an extremely poor fit to the U-shape in GGLE.
No code or data for reproducibility.

Q 3 **Please provide your detailed review report to the authors. The editors prefer to receive your review structured in major and minor comments. Please consider in your review the methods (statistical methods valid and correctly applied (e.g. sample size, choice of test), is the study replicable based on the method description?), results, data interpretation and references. If there are any objective errors, or if the conclusions are not supported, you should detail your concerns.**

Major comments:
---
1. (line 436) "It demonstrates that improving air quality may close the GGLE difference."
This causal statement is too strong. Your analysis is based on statistical correlation and does not use any formal causal inference method.

2. Given the methodological similarities with Wang et al. (2022), more comparisons and discussion would be expected. In particular, it should be cited as a previous work using the same statistical model in a similar context

3. The linear trend model is not appropriate for characterizing spatiotemporal variations. For example, the authors can see Knorr-held (2000) [2] for a model that allows for flexible temporal variations and spatial/temporal variations using autoregressive Gaussian models (as used in this paper).

4. I really don't see the value of having different maps for b0 + b1i and b1i alone (Fig 2b and 2c) since one is just an offset from the other. Related, can you report the statistical significance of the trends b0 + b1i for the GGLE? It would seem as if the vast majority will not be significant, owing to the poor fit of linear time trend given the most common U-shape. An effort to identify the significant ones (perhaps in a table or figure) would be useful.

5. One would expect the regression coefficients of GGLE to be simply the subtraction of the coefficients of the LEm and LEf models. At least, this would be the case under unconfoundedness (with a well-specified outcome model). However, we see in Table 1 that these identity doesn't hold for PM2.5, even when using spatial random effects. Here, both LE models have near-zero coefficients. Why is then the coefficient of GGLE larger and more significant?
I intuit that this scenario could happen when the confounders for the gap are not the same as for life expectancy. Is this the case here? This type of reasoning could, in part, justify why we need a separate analysis of GGLE and not simply look at LEm and LEf and compare the coefficients.

Minor comments:
---
1. How is the graph for the BYM2 model constructed? Do islands (e.g. Australia, Japan) not have any neighbors to adjust for spatial correlations? The BYM2 prior needs adjustment for the number of connected components in the graph [3]. How did you adjust for it?

2. (line 405) How is spatial confounding different from an omitted confounder (with a strong spatial autocorrelation)?

3. (line 172) This sentence makes it sound like the original BYM is not a spatial model, while the BYM2 is simply a re-parametrization.

4. Can you report the statistical significance of the trends b0 + b1i for the GGLE? It would seem as if the vast majority will not be significant, owing to the poor fit of linear time trend given the most common U-shape. An effort to identify the significant ones (perhaps in a table or figure) would be useful.

5. Why is vi given a random independent prior? The U-shape in GGLE indicates that it is strongly auto-correlated. An autoregressive prior would be more appropriate.

6. What is the probability distribution of the prior of phi? Only an inequality was mentioned.

Some typos/broken English:

(line 23) show an ... has a
(line 46) a reversal of this .... that the
(line 56) net of whatever
(line 131) "Both" shouldn't be capitalized
(line 217) Something is off with the term having lnPM_it and lnUP_it in eq (4), perhaps a missing indicator I(j=region(l))?
(line 441) should b.

References
---
[1] Wang, S., Ren, Z., Liu, X. and Yin, Q., 2022. Spatiotemporal trends in life expectancy and impacts of economic growth and air pollution in 134 countries: A Bayesian modeling study. Social Science & Medicine, 293, p.114660.
[2] Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in medicine, 19(17-18), pp.2555-2567.
[3] Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S.J., Gelman, A. and DiMaggio, C., 2019. Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. Spatial and spatio-temporal epidemiology, 31, p.100301.

PLEASE COMMENT

Q 4 Is the title appropriate, concise, attractive?

Yes

Q 5 Are the keywords appropriate?

Yes

Q 6 Is the English language of sufficient quality?

Needs improvement.

Q 7 Is the quality of the figures and tables satisfactory?

Yes.

Q 8 Does the reference list cover the relevant literature adequately and in an unbiased manner?)

I am not an expert in the literature on life expectancy. But I do suggest a better comparison and discussion with Wang (2022).

QUALITY ASSESSMENT

Q 9 Originality

Q 10 Rigor

Q 11 Significance to the field

Q 12 Interest to a general audience

Q 13 Quality of the writing

Q 14 Overall scientific quality of the study

REVISION LEVEL

Q 15 Please make a recommendation based on your comments:

Major revisions.