Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-23-533

Reviewer A

Overall, the manuscript provides an elegant analysis of how geospatial data can inform epidemiological studies and how prior analyses based on Environmental exposure and SES risk at cancer diagnosis underestimate the potential risk factors and exposure in mesothelioma.
Reply: Thank you for your positive feedback.

I had some comments/questions for the authors:
Comment 1: Since the study reports that only 43.7% of patients had residential information available for up to 30 years, were there any specific characteristics for the patients with missing long term residential information?
Reply 1: As shown in Table 1, we found that in general patient characteristics (i.e., age at diagnosis, sex, race, Hispanic ethnicity, caner stage, and tobacco product use) for the subset of patients with up to 30-year residential history to be very similar to the overall population, as well as the subgroups with shorter than 30-year residential histories. Table 1 also showed that the 30-year group (n=444) differed from the others in terms of measures related to residential history, including duration lived at a census tract, and time-weighted-average REcan and REses. The main reason that the length of the residential history decreases with increasing look-back time window is likely that electronic recording of various sources used to compile address histories were not available (or became popular) before the 1980s.
In the revised manuscript, we added the following sentences to highlight these differences (page 13 lines 259-264), "The distribution of patient-level characteristics across subgroups defined by the retrospective observation time window remained largely similar to those found in the entire study sample (Table 1), with some notable differences. With longer residential history available spanning from 5-year to 30-year, the average moving distance increased from 12.5 to 19.4 kilometers, while the proportion of non-movers decreased from 13.8% to 3.4%. In addition, those with up to 30-year residential history had lower REcan and REses.".

Comment 2: Since the study is based in NY state with such a large multicultural population, Could you quantify any effect of our of state migrations or country of origin in this analysis that may impact results.
Reply 2: Thank you for your comment. While NY state is a large multicultural populous state, which could affect the residential mobility of our study population, we do not have sufficient data to assess the effect of state migration pattern on the residential history of our study population. We did, however, added some new results and discussions to describe the residential mobility of the study population. In general, we found that the majority of the tracts lived by our study population were within the NY state, and the moves across census tracts were largely within-state moves. The new

sentences are:
On page 12 lines 243-244, "We found that 3,793 (82.4%) addresses were in NYS, which spanned 2,140 census tracts. The remaining addresses were in other 41 states, encompassing 705 census tracts.".
On page 12 lines 247-249, "Approximately 14% of the patients only lived in one address (i.e., non-movers), and the median distance moved among the entire study population was 12.1 (IQR 1.7 – 193.9) kilometers (Table 1).".
On page 17 lines 362-365, "As NYS is a large multicultural populous state, it remains to be confirmed that if the same trend would hold true for mesothelioma patients diagnosed in a different period in NYS or in other states, since migration patterns may differ over time and/or among people residing in another state.".

Comment 3: Can you comment on the limitations of the Yost index percentile ranking at census tract level to quantify SES exposure.
Reply 3: We added the following sentences in the limitation section (page 18 lines 376-378), "In addition, Yost index was a composite measure based on seven individual census variables, and may not be a comprehensive measure of neighborhood SES (51, 53).".

Comment 4: I found Figures 1 and 3 hard to interpret and could be revised or added as supplementary appendix information.
Reply 4: We revised figure captions and added further explanations in the notes of the Figure 1 and Figure 3. The new legends of the two figures are:

Figure 1. Distributions of relative exposure to air toxic (REcan, upper panel) and relative exposure for socioeconomic status (REses, lower panel) from seven different estimates based on patient residential history information, and the p-values from comparisons (Wilcoxon signed-rank tests) of REcan and REses based on the address at cancer diagnosis to those from the remaining six alternative measures.

Note: The box plot showed the descriptive statistics of the REcan and REses values (i.e., median (the line inside the box), 25^{th} and 75^{th} percentiles (the width of the box, which was also the inter quartile range (IQR)), and 1.5*IQR from 25^{th} and 75^{th} percentiles (the width of the whiskers)). We used Wilcoxon signed-rank tests to compare REcan and REses at the cancer diagnosis address with six alternative estimates. For example, REcan from the first address lived was compared with REcan measured at cancer diagnosis address and the p-value was $4.6*10^{-9}$ (i.e. p-value< 0.0001); and significant difference was also found between time-weighted average REses using all addresses lived and REses based only on the address lived at cancer diagnosis (p-value = 0.00012).

Figure 3. Temporal changes of REcan and REses, shown as yearly changes and their 95% confidence intervals, within each of the five different look-back observation

windows prior to patient's cancer diagnosis.
Note: We examined temporal changes of REcan (i.e., relative exposure for non-asbestos air toxics) and REses (i.e., relative exposure for socioeconomic status), respectively, within five different look-back observation windows prior to patient's cancer diagnosis. The beta coefficient and their 95% confidence intervals from the generalized equation estimating equations (GEE) models were shown in the Y-axis. X-axis shows the observation windows of 5-year (n=974), 10-year (n=952), 15-year (n=913), 20-year (n=839), and 30-year (n=444) prior to cancer diagnosis, respectively. The models used time in years as the explanatory variable, which was treated as a continuous variable, while REcan and REses were the response variables, respectively. The models also adjusted for patient-level characteristics, including age at cancer diagnosis, sex, race, Hispanic ethnicity, cancer stage, and tobacco use status.

Reviewer B

Comment: Interesting paper for the analysis methodology used, however the pathology to which the methodology was applied is probably not the most suitable for this type of analysis and basically, to date, other certain etiological factors other than asbestos due to pollution are not recognized in the literature. In this case the analysis is conducted on non-asbestos air toxic as well as on SES. The paper does not analyze the numerous international experiences on the registration of mesotheliomas which, among other things, acquire information on the entire residential history of the case also with historical certificate of residence. It would be appropriate to discuss the literature on the subject, for example just to cite the most recent literature:

Kitamura Y, Zha L, Liu R, Shima M, Nakaya T, Kurumatani N, Kumagai S, Goji J, Sobue T. Association of mesothelioma deaths with neighborhood asbestos exposure due to a large-scale asbestos-cement plant. Cancer Sci. 2023 Apr 3. doi: 10.1111/cas.15802.

Gaitens JM, Culligan M, Friedberg JS, Glass E, Reback M, Scilla KA, Sachdeva A, Atalla A, McDiarmid MA. Laying the Foundation for a Mesothelioma Patient Registry: Development of Data Collection Tools. Int J Environ Res Public Health. 2023 Mar 11;20(6):4950. doi: 10.3390/ijerph20064950.

Magnani C, Mensi C, Binazzi A, Marsili D, Grosso F, Ramos-Bonilla JP, Ferrante D, Migliore E, Mirabelli D, Terracini B, Consonni D, Degiovanni D, Lia M, Cely-García MF, Giraldo M, Lysaniuk B, Comba P, Marinaccio A. The Italian Experience in the Development of Mesothelioma Registries: A Pathway for Other Countries to Address the Negative Legacy of Asbestos. Int J Environ Res Public Health. 2023 Jan 4;20(2):936. doi: 10.3390/ijerph20020936.

Lysaniuk B, Cely-García MF, Mazzeo A, Marsili D, Pasetto R, Comba P, Ramos-

Bonilla JP. Where are the landfilled zones? Use of historical geographic information and local spatial knowledge to determine the location of underground asbestos contamination in Sibaté (Colombia). Environ Res. 2020 Dec;191:110182. doi: 10.1016/j.envres.2020.110182.

Petrof O, Neyens T, Nuyts V, Nackaerts K, Nemery B, Faes C. On the impact of residential history in the spatial analysis of diseases with a long latency period: A study of mesothelioma in Belgium. Stat Med. 2020 Nov 20;39(26):3840-3866. doi: 10.1002/sim.8697.

Marinaccio A, Corfiati M, Binazzi A, Di Marzio D, Bonafede M, Verardo M, Migliore E, Gennaro V, Mensi C, Schallemberg G, Mazzoleni G, Fedeli U, Negro C, Romanelli A, Chellini E, Grappasonni I, Pascucci C, Madeo G, Romeo E, Trafficante L, Carrozza F, Angelillo IF, Cavone D, Cauzillo G, Tallarigo F, Tumino R, Melis M; ReNaM Working Group. The epidemiological surveillance of malignant mesothelioma in Italy (1993-2015): methods, findings, and research perspectives. Epidemiol Prev. 2020 Jan-Feb;44(1):23-30. English. doi: 10.19191/EP20.1.P023.014.

Vimercati L, Cavone D, Lovreglio P, De Maria L, Caputi A, Ferri GM, Serio G. Environmental asbestos exposure and mesothelioma cases in Bari, Apulia region, southern Italy: a national interest site for land reclamation. Environ Sci Pollut Res Int. 2018 Jun;25(16):15692-15701. doi: 10.1007/s11356-018-1618-x.

Vimercati L, Cavone D, Delfino MC, Caputi A, De Maria L, Sponselli S, Corrado V, Ferri GM, Serio G. Asbestos Air Pollution: Description of a Mesothelioma Cluster Due to Residential Exposure from an Asbestos Cement Factory. Int J Environ Res Public Health. 2020 Apr 12;17(8):2636. doi: 10.3390/ijerph17082636.

Reply: Thank you for your comment and the references. We agree with the established clear etiology of mesothelioma due to asbestos exposure, whether it is because of direct occupational exposure (e.g., among works of industries involving the use of asbestos), para-occupational exposure (e.g., family members of workers who exposed to asbestos through work clothes at home), or environmental exposure (e.g., living at close proximity to industries that used asbestos, or natural environmental asbestos sources). The purpose of the current study, however, was not to identify non-asbestos exposure as a potential cause of mesothelioma, nor attempt to identify latency period of mesothelioma, as we only have mesothelioma cases and lack of a comparison group. Instead, we aimed to demonstrate differences in exposure assessment between using a single residential address information (i.e. address at cancer diagnosis) and residential history information. We showed the utility of incorporating residential history information to improve exposure assessment. As we stated in the methods, we used NATA data to estimate cancer exposure in this study, which is non-asbestos exposure. We added the following sentences to clarify and also cited the papers suggested by the reviewer.

On page 5 lines 75-81, “Workplace asbestos exposure is the dominant contributor to mesothelioma, though para-occupational (or take-home) exposure and environmental exposure to asbestos (e.g., residing in close proximity to asbestos sources) also play a critical role of mesothelioma development (5-9). Studies from mesothelioma registries in many countries have been instrumental in providing rich information, including specifically residential histories, to understand the impact of asbestos exposure at scale, and to investigate known and unknown asbestos exposure sources (8, 10-12).”.

On page 19 lies 403-407, “While the current study was a case-only design and specific to mesothelioma which has a known etiology, our methods can be applied to future studies with comparison groups to further study mesothelioma patients, as have been shown in other studies based on population-level mesothelioma registries where residential history information is available, as well as other cancer types (7-9, 11, 12, 31-33).”.

Some new references were also added on page 6 lines 101-103, “Some researchers were able to take advantage of the comprehensive national-level cancer registry and population health databases to obtain partial or complete residential history information (7, 8, 12, 31-33), …”.

Reviewer C

Comment 1: Be precise on some statements, like “SES also contribute to the poor prognosis and survival” but SES itself is neutral, won't result in poor. There is also a “Reses”, which I believe should be “REses”.
Reply: Thank you for pointing this out. We revised the sentence to () “…, as well as lower socioeconomic status (SES), also contribute to the poor prognosis and survival (2, 13-16),” on page 5 line 83, and corrected Reses to REses in the manuscript.

Comment 2: What is the support that you choose 6-month as the duration if there is only a single point address, and you mentioned the median time lived in an address was 2.2 years, why don’t you use this median value?
Reply: Thank you for the suggestions. We conducted the analyses using the median time lived in an address (i.e., using 2.2 years instead of 6 month). We clarified this method on page 8 lines 155-157, “If only a single point address was available as the last address, we assumed a duration of 2.2 years, which was the median length of residency at an address among the study population.”. The results from the new analyses are more or less the same as the old ones, with changes occurring in the decimal places. We have used these updated results in the revised manuscript.

Comment 3: There is lacking the description of the tracts you have mapped to, like # of tracts, in/out NY state area. May consider including a map to present the tracts, probably

with the corresponding REcan and Reses in years.
Reply: Thank you for raising this point. For privacy concerns, we are not able to include a map, as mesothelioma is a rare cancer and the study population is small. However, we added additional information about residential mobility of the study population as shown below.
On page 12 lines 243-244, "We found that 3,793 (82.4%) addresses were in NYS, which spanned 2,140 census tracts. The remaining addresses were in other 41 states, encompassing 705 census tracts.".
On page 12 lines 247-249, "Approximately 14% of the patients only lived in one address (i.e., non-movers), and the median distance moved among the entire study population was 12.1 (IQR 1.7 – 193.9) kilometers (Table 1).".

Comment 4: Mesothelioma could have long latency from the first exposure to diagnosis, it is reasonable that the residence at diagnosis does not contribute a lot, the authors have proven that by looking back years on residential history and SES. Those methods would help to track the inducement and survival conditions for mesothelioma patients. And it could be further developed to adapt studies on other cancers that may be related to residence and SES.
Reply: Thank you for your comment. We added a sentence in the discussion (page 19 lines 399-403), "In addition, the availability of longitudinal exposure patterns based on patient residential history may have important implications, such as tracking the inducement and survival conditions for mesothelioma patients, investigating susceptible time window and latency period, in studying disease development of mesothelioma and different cancer types." to further highlight the implication of using residential history to study the impact of residence and SES on other cancers.