Response to the Comments on Cicalese et al.: “An Ecological Study of the Association between Air Pollution and Hepatocellular Carcinoma Incidence in Texas”

Loren Raun† Ali Shirafkan∥ Laura Campos† Daria Zorzi∥ Mauro Montalbano† Colin Rhoads∥ Valia Gazis∥ Katherine Ensor∥ Cristiana Rastellini†∥∥ Luca Cicalese†

†Department of Surgery, University of Texas Medical Branch, Galveston, TX, USA; ∥Department of Neurosciences and Cell Biology, University of Texas Medical Branch, Galveston, TX, USA; †Department of Statistics, Rice University, Houston, TX, USA

Dear Editor

In response to the comments [1, 2] regarding the article by Cicalese et al., “An Ecological Study of the Association between Air Pollution and Hepatocellular Carcinoma Incidence in Texas” [3], the authors acknowledge minor editorial errors but stand behind the methodology and conclusions of the paper. There appears to be a misunderstanding concerning what these data and results suggest; we aim to clarify these points and respond to critiques that the authors knowingly published results based on limited assumptions.

First of all, no cause-effect relationship is stated or implied by the authors. When done properly, statistical control for lurking and confounding variables aims to focus simply on the correlation between 2 random variables. After controlling for as many variables as possible, significant relationships discovered through correlation alone do not imply causation; this requires further exploration. By carefully excluding other common causes of hepatocellular carcinoma (HCC) in Texas, the authors explore a previously unstudied (but toxicologically well-known) cause of liver cancer – airborne toxics, such as 1,3-butadiene, arsenic, benzene, and vinyl chloride. Two kinds of regression were performed on the data for distinct reasons.

Logistic regression tests the association of a positive HCC rate with risk factors (including vinyl chloride), while gamma regression tests the association of the magnitude of positive HCC rates with these factors. It should also be noted that the logistic regression included all data from the Texas Cancer Registry and National Air Toxics Assessment (NATA), while the gamma regression only included the subset of these data with a non-zero HCC incidence rate. The authors acknowledge some inconsistencies between the results of the 2 regressions, though they generally measure different relationships.

Below, each claim of the commenter is numbered, with the authors’ responses following each claim. Quoted material is reproduced from either the comments submitted to the editor [1, 2] (numbered) or the published paper [3] (bulleted).

To the many claims by Krock [1] of the Vinyl Institute, the authors provide the following responses:

1. “The latency period for angiosarcoma of the liver and HCC is much longer than the 7- to 13-year duration used in the study (36–48 years)” [1].
   - The cited latency periods account for a latency period between exposure and death from HCC, not between exposure and cancer incidence.
- Defining cancer latency due to chemical exposure is admittedly challenging because it is impossible to pinpoint the exact moment exposure occurred. Accordingly, the authors can find no published estimates of latency periods for HCC incidence due to prolonged environmental exposure to vinyl chloride, though documented cases of both angiosarcoma and HCC incidence in vinyl chloride workers exist within the stated range of 7–13 years [4]. In a 2014 paper, a measure of years from onset to diagnosis for all liver cancers (the closest the authors can find to a latency-to-diagnosis measure) was 10.8 years [5].
- Numerous risk assessment models consider the first exposure to carcinogenic chemicals to induce a linear cancer dose-response where the risk of developing cancer occurs at first exposure.
- The authors neglected to match HCC rates with NATA Texas counties reporting VCM [vinyl chloride monomer] emissions from point sources” [1].
- No attempts to reconcile NATA estimates with point sources were performed because this is addressed in the modeling methodology of the NATA. With differing methodologies between each NATA, point source emissions have been handled differently, but more recent models are considered to be more accurate.
- “No explanation is provided why [the] hypothesis would not apply to Washington, DC, listed as having the highest HCC rate” [1].
- While this analysis could be performed, Texas was chosen due to its variability in sampling locations and large sample size. Texas is tied for the third highest rate (behind Washington, DC, and Louisiana and tied with Mississippi), though 95% confidence intervals of the top 5 incidence rates by state/district overlap, implying an inconclusive significant difference between these rates.
- “Because the high rate of HCC observed in Texas is not explained by other risk factors, and given that Texas has the second largest petrochemical industry and agricultural industry in the nation, we hypothesize that chronic exposure to selected air toxins may contribute to the increased risk. Texas has a high density of emissions in some counties and a low density in others, resulting in varying levels of hazardous air pollutant concentrations by county. We believe the variation provides a good basis to conduct an ecologic investigation of the association between the rate of HCC with HCV and known carcinogenic pollutants such as vinyl chloride, arsenic, benzene, and 1,3-butadiene” [3].
- “The authors state that a statistically significant relationship exists between many air toxins and HCC incidence. With differing methodologies between each NATA, point source emissions have been handled differently, but more recent models are considered to be more accurate.”
- “The authors note that the highest NATA value for VCM was 0.01 μg/m³, and the screening threshold for VCM is 1.2 μg/m³,” and further comments regarding lifetime risk [1].
- Comparison to risk thresholds or risk levels is welcomed for the discussion. The authors demonstrate a potential relationship between VCM and HCC at concentrations lower than published levels.
- “In terms of vinyl chloride, the screening level is 1.2 μg/m³, while the highest average concentration in the years evaluated was well below this level at 0.0106 μg/m³ ... If future research confirms an association between increased risk of HCC and vinyl chloride exposure, the screening level may not be protective” [3].
- “The authors provide limited qualifying statements regarding “the first analysis” and some inconsistencies in their data but did not take those restrictions into account when drawing conclusions” [1].
- The authors do not attribute a cause-effect relationship to any emissions and HCC but simply demonstrate that a statistically significant relationship exists between many air toxins and HCC incidence. Conclusions are appropriately qualified and remain within the realm of statistical relationship, not causation. “While not confirmatory, our study indicates a need for further investigation” [3].
- “The authors [do not] account for the most likely cause of increased occurrence – hepatitis C or B virus infection” [1].
- “The known etiologies of HCC are comprised of hepatitis C (HCV) and B (HBV) virus infection, nonspecific cirrhosis, and alcohol-induced liver disease” [3].
- “The HCV prevalence and the number of prisons in each county were also included in the analysis to control for confounding from these variables” [3].
- “HBV prevalence data [per county] are currently not available and should be included in future research if possible” [3]. However, since 1991, Texas has experienced a steady decline in the incidence of acute hepatitis B. Currently, only acute hepatitis B is reportable in Texas. The reported incidence for acute hepatitis B in 2015 was 159 cases (Texas Department of Health). Moreover, an antiviral therapy is available. Steadily declining HBV infection in Texas in the last decades (as also nationally) does not explain the increasing rate of HCC observed.
- “No statistically significant association was found between the presence of a positive rate of HCC within a county and the prevalence of HCV in a county regardless of air pollution assessment year. HCV was retained in the final model because it is a known risk factor” [3].
To the claims by Gennissen et al. [2] of the European Council of Vinyl Manufacturers, the authors provide the following responses:

1. **The median latency period for HCC differs from the latency period used in this study.**
   - See above.

2. **Irreproducibility of methods.**
   - All data sources are disclosed and publically available.
   - See Data section of the Methods section as well as supplementary material.
   - “The analysis was conducted using data obtained from publicly available sources of HCC incidence, air pollutant concentrations for vinyl chloride, arsenic, benzene, and 1,3-butadiene, HCV prevalence, obesity prevalence, and the number of prisons in a county” [3].
   - All variables are defined for each dataset.
   - Excluded data are given with rationale.
   - Statistical models are defined by their names.
     i. Logistic regression: "Nonparametric generalized additive model" [3].
     ii. Gamma regression: "Nonparametric generalized additive gamma regression model" [3].

3. **In the logistic regression method, a distinction/division is made between counties with a positive HCC rate and the other counties. The authors implicate that the distinguishing feature between the 2 groups is liver cancer. The methodologic confounder concerning this assumption is based on the fact that all zero and low rates (i.e., less than 16 cases) are suppressed. By doing this, the number of the population at risk has an essential distinguishing effect on whether a county is placed in the positive HCC rate group or the other group” [2].
   - The Texas Cancer Registry suppresses rates less than 16 cases for statistical stability, in adherence to the same data reporting norms of the National Program of Central Cancer Registries (NPCR) at the Centers for Disease Control and Prevention (CDC), the North American Association of Central Cancer Registries (NAACCR), and the Surveillance and Epidemiology End Results (SEER) Program at the National Cancer Institute (NCI).
   - Section Logistic Regression: “The first analysis examined the relationship between a positive HCC rate (i.e., the presence of a rate; zero and low rates were suppressed) for each of the 254 counties in the state according to the risk factors” [3].
   - Section Gamma Regression for Counties with Reported HCC Rates: “Of the counties in Texas, the HCC rates were suppressed in 108 counties for the 2000–2013 period and in 137 counties for the 2006–2013 subset. Therefore, the second analysis examined the 146 and 117 counties where a HCC rate was reported for the respective time period” [3].

We thank Mr. Krock and Dr. Boeckx, the Vinyl Institute and the European Council of Vinyl Manufacturers, and the editors of Liver Cancer for providing us an opportunity to respond to comments on “An Ecological Study of the Association between Air Pollution and Hepatocellular Carcinoma Incidence in Texas” [3]. This paper provides novel insights into the correlation between many different airborne toxins and HCC incidence in Texas and begs for further investigation.

### References

1. Krock R: Vinyl Institute comments on Cicalese et al.: "An ecological study of the association between air pollution and hepatocellular carcinoma incidence in Texas". Liver Cancer 2018, DOI: 10.1159/000486432.
2. Gennissen A, Boeckx M, Sevenster A: European Council of Vinyl Manufacturers comments on Cicalese et al: "An ecological study of the association between air pollution and hepatocellular carcinoma incidence in Texas". Liver Cancer 2018, DOI: 10.1159/000486528.
3. Cicalese L, Raun L, Shirafkan A, et al: An ecological study of the association between air pollution and hepatocellular carcinoma incidence in Texas. Liver Cancer 2017; 6:287–296.
4. Evans DMD, Williams WJ, Kung IT: Angiosarcoma and hepatocellular carcinoma in vinyl chloride workers. Histopathology 1983; 7:377–388.
5. Nadler DL, Zurbenko IG: Estimating cancer latency times using a Weibull model. Adv Epidemiol 2014, DOI: 10.1155/2014/746769.