



An Ecological Evaluation of Vinyl Chloride Exposure and Liver Cancer Incidence and Mortality in Texas

Kevin M. Towle^{*1}, Stacey M. Benson², Natalie S. Egnot² and Gary M. Marsh²

¹Cardno ChemRisk, San Francisco, CA, USA; ²Cardno ChemRisk, Pittsburgh, PA, USA

Abstract

The goal of this analysis was to evaluate the association between county-level ambient vinyl chloride (VC) and county-level liver cancer incidence and mortality rates in Texas. Modeled county-level ambient VC data were obtained from the National Air Toxics Assessment. Age-adjusted county-level liver cancer incidence rates were abstracted from the Texas Cancer Registry and age-standardized county-level liver cancer mortality rates were obtained from the peer-reviewed literature. Multivariable imputation was utilized to impute incidence rates in counties with suppressed liver cancer incidence rates. Negative binomial and Poisson regression models were utilized to evaluate the association between county-level ambient VC and county-level liver cancer incidence and mortality rates, respectively, adjusted for county-level heavy drinking prevalence, hepatitis mortality rates, median income, and race (percent Hispanic). County-level ambient VC was not associated with county-level liver cancer incidence or mortality rates. Specifically, when compared to the lowest tertile of ambient VC, the middle (relative risk [RR]: 1.06, 95% confidence interval [CI]: 0.95–1.19) and highest (RR: 1.03, 95% CI: 0.90–1.17) tertiles of ambient VC were not associated with liver cancer incidence. Similarly, county-level ambient VC in the middle (RR: 0.95, 95% CI: 0.85–1.05) and highest (RR: 0.93, 95% CI: 0.82–1.05) tertiles were not associated with liver cancer mortality. This analysis suggests that county-level ambient VC is not associated with liver cancer incidence or mortality in Texas. Our study provides novel results regarding liver cancer risk from low-level non-occupational exposure to ambient VC.

Citation of this article: Towle KM, Benson SM, Egnot NS, Marsh GM. An ecological evaluation of vinyl chloride exposure and liver cancer incidence and mortality in Texas. *J Clin Transl Hepatol* 2021;000(000):000–000. doi: 10.14218/JCTH.2020.00073.

Keywords: Vinyl chloride; Liver cancer; Epidemiology; Incidence.

Abbreviations: ASL, angiosarcoma of the liver; ASPEN, Assessment System for Population Exposure Nationwide; ATSDR, Agency for Toxic Substances and Disease Registry; BRFSS, Behavioral Risk Factor Surveillance System; CDC, Centers for Disease Control and Prevention; EPA, Environmental Protection Agency; HCC, hepatocellular carcinoma; IARC, International Agency on Cancer Research; NAACCR, North American Association of Central Cancer Registries; NATA, National Air Toxics Assessment; NCHS, National Center for Health Statistics; NCI, National Cancer Institute; NPCR, National Program of Central Cancer Registries; PEL, Permissible Exposure Limit; PVC, polyvinyl chloride; RR, relative risk; SEER, Surveillance and Epidemiology End Results; VC, vinyl chloride.

Received: 26 July 2020; Revised: 05 October 2020; Accepted: 27 November 2020

***Correspondence to:** Gary Marsh, Cardno ChemRisk, 20 Stanwix Street Suite 505, Pittsburgh, PA 15222, USA. Tel: +1-412-694-7051, Fax: +1-415-896-2444, E-mail: gary.marsh@cardno.com

Introduction

Vinyl chloride (VC), also called vinyl chloride monomer, is a compound used in the manufacture of the polymer polyvinyl chloride (PVC). Due to its versatility and technical properties, PVC is used in various industrial and consumer plastic and vinyl products, including pipes, building supplies, automotive parts, clothing, and medical devices.¹ The PVC global market is projected to increase approximately 3% per year until 2021.²

In 2012, the International Agency on Cancer Research (IARC) classified VC as a group 1 carcinogen based on evidence from animal and occupational epidemiology studies.³ Specifically, IARC concluded that there was sufficient evidence in humans that VC causes angiosarcoma of the liver (ASL) and hepatocellular carcinoma (HCC) due to findings from two large, multicenter cohort studies (one in the USA and one in Europe) of workers at VC and PVC production or processing plants.^{3–5} IARC also discussed a meta-analysis on occupational exposure to VC that concluded that VC workers may experience an increased risk of HCC.^{3,6} The meta-analysis authors noted that findings may have been influenced by the under diagnosis of true ASL and that there was likely variability in the probability and level of exposure among production workers, as direct measurements of VC exposure were not reported in the underlying studies.⁶

The 2012 IARC evaluation was specific to workers occupationally exposed to VC who were potentially simultaneously exposed to other industrial chemicals and agents. It is important to note that IARC did not perform an evaluation of the association between non-occupational VC exposure and cancer risk. The Agency for Toxic Substances and Disease Registry (ATSDR) reported that inhalation of air containing VC is the most likely non-occupational exposure route for the general population.⁷ According to the Environmental Protection Agency (EPA), VC is present in ambient air due to discharge of exhaust gases from VC manufacturing or processing facilities or evaporation from areas where chemical wastes are stored.⁸ Ambient VC concentrations are typically around 1 µg/m³ or less, with concentrations near emission sources ranging from trace levels to over 2,600 µg/m³.^{7,9} For comparison, the Occupational Safety and Health Administration's Permissible Exposure Limit (PEL) is 1 ppm (2,500 µg/m³).¹⁰ While there is information on low-dose exposure among occupational cohorts, there is limited evidence available on the potential risk of liver cancer among individuals non-occupationally exposed to ambient VC.¹¹

VC and PVC plants tend to be clustered primarily along the Texas and Louisiana Gulf coast, which may result in potentially elevated ambient VC concentrations in these areas due to plant emissions. In 2017, Cicalese *et al.*¹² published an ecological study on the association between air pollution (arsenic, benzene, 1,3-butadiene, and VC) and liver cancer incidence in Texas, concluding that their findings suggest that VC is a "significant contributor" to the incidence

of liver cancer in Texas. As noted in our recent Letter to the Editor, this analysis had severe methodological limitations related to insufficient latency period, missing data due to data suppression, lack of adjustment for confounders, and inappropriate model selection.¹³ Due to these limitations and the data gap concerning the potential health effects of non-occupational ambient VC exposure, the objective of the current study was to perform a more rigorous and comprehensive ecological evaluation of the association between exposure to county-level ambient VC and both liver cancer incidence and mortality in Texas. Our analysis has the additional benefits of allowing for adjustment for important liver cancer risk factors, imputation of incidence rates for suppressed counties, consideration of both liver cancer incidence and mortality, and utilization of more conventional Poisson and spatial regression modeling.

Methods

Exposure data

Modeled county-level VC data were obtained from the 1996 National Air Toxics Assessment (NATA). NATA is the EPA's review of air pollutants in the USA, where emissions data from stationary sources (*e.g.*, large waste incinerators and factories), area and other sources (*e.g.*, small manufacturers, wildfires, dry cleaners) and mobile outdoor sources (*e.g.*, vehicles) are used as inputs in air quality models to estimate the county-level and census tract-level concentrations of various chemicals.¹⁴ The EPA uses the Assessment System for Population Exposure Nationwide (ASPEN) simulation model to combine emissions data with meteorological data to estimate concentrations of air pollutants, accounting for rate/location of release, wind speeds and direction, breakdown and transformation, and deposition.¹⁵ The 1996 assessment was selected to allow for sufficient disease latency, as this was the earliest dataset available. Specifically, we obtained the county-level median estimated annual average ambient VC ($\mu\text{g}/\text{m}^3$) for every county in Texas. Additionally, we performed a sensitivity analysis using the county-level 95th percentile ambient VC instead of the county-level median ambient VC.

NATA provides data on both ambient VC (concentration of VC in the open air) and exposure VC (concentration of VC that a person may breathe over time, accounting for time that a person spends indoors and outdoors).¹⁶ Due to the study objectives, we selected ambient VC for the final model; however, we also performed a sensitivity analysis using county-level median exposure VC.

Liver cancer data

Incidence: Liver cancer incidence rates were obtained from the Texas Cancer Registry (TCR). The TCR collects data through passive and active surveillance and meets the high quality data standards put forth by the Centers for Disease Control and Prevention (CDC) and the North American Association of Central Cancer Registries.¹⁷ Specifically, county-level liver cancer incidence rates (per 100,000 population), age-adjusted to the 2000 USA Standard Population, were queried for each county in Texas from 2006 to 2015.

Mortality: Liver cancer mortality rates were abstracted from the Mokdad *et al.*¹⁸ (2017) analysis of cancer mortality patterns across USA counties. Briefly, Mokdad *et al.*¹⁸ used death records from the National Center for Health Statistics (NCHS), population counts from the Census Bureau, the NCHS, and the Human Mortality Database, and validated

small area estimation models to estimate county-level mortality rates for various cancer endpoints. Age-standardized 2010 liver cancer mortality rates (per 100,000 population) were calculated using the USA Census population.¹⁸

Latency period: Our liver cancer incidence (2006–2015) analysis allowed for a 10–19 year latency period (ambient VC data from NATA 1996). Similarly, our liver cancer mortality (2010) analysis provided a 14 year latency period.

Potential confounding variables

Known risk factors of liver cancer were considered as potential confounding variables. County-level sex (percent male), race (percent Hispanic), and income (median household income in the past 12 months, inflation-adjusted dollars) were abstracted from the 2010 American Community Survey (5-year averages).^{19–21} Data from 2010 were selected from among the available years (2005–2017) to correspond to the midpoint of the outcome data. County-level age-standardized total cigarette smoking prevalence (non-daily and daily smoking; 2010) and age-standardized prevalence of heavy drinking (consumption, on average, of more than one drink per day for women and two drinks per day for men in the past 30 days; 2010) were reported in the peer-reviewed literature based on calculations using validated small area estimation methods and data from the Behavioral Risk Factor Surveillance System (BRFSS).^{22,23} County-level age-adjusted obesity (body mass index ≥ 30) prevalence data were collected from the CDC.²⁴ Lastly, age-standardized hepatitis mortality rates (per 100,000 population) for 2010 were obtained from the USA's Infectious Disease Mortality Rates dataset from the Institute for Health Metrics and Evaluation.²⁵

Incidence rate imputation for suppressed counties

The TCR suppresses incidence rate and cancer count data when there are fewer than 16 reported cancer-specific cases in a county, based on the threshold used by the National Program of Central Cancer Registries (NPCR) at the 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). Additionally, the TCR does not provide incidence rate, population size, or cancer count data for counties with a risk population of <1,000 persons. Of the 254 counties in Texas, 102 had fewer than 16 cases of liver cancer reported from 2006–2015 and 8 had risk populations <1,000 persons. Overall, 110 counties (43%) had missing liver cancer incidence data due to suppression or an insufficient at-risk population. Therefore, we used a step-wise selection ($p=0.10$) negative binomial regression model (accounting for over-dispersion in the reported incidence data) to perform multivariable imputation to estimate liver cancer incidence rates for counties with suppressed rate data in the TCR. Multiple imputation is a technique for handling missing values, where the distribution of the observed data is used to estimate multiple datasets (to reflect uncertainty around the true value) that are then pooled for statistical analysis.²⁶ Variables selected for the multivariable imputation model were county-level race (percent Hispanic), smoking prevalence, heavy drinking prevalence, hepatitis mortality rate, and median income. Three counties had imputed liver cancer incidence rates that were negative, which were set to zero for purposes of the statistical analyses. Imputed incidence rates for the remaining counties were rounded to the nearest tenth to match the data format of rates reported by the TCR.

Exploratory spatial data analyses

Exploratory spatial data analyses were performed to qualitatively evaluate the association between county-level ambient VC, liver cancer incidence and mortality rates, and potential confounders to inform statistical analyses and modeling. Additionally, the NATA 1996 dataset provided exposure VC data at both the county- and census tract-level. Therefore, we explored exposure VC at the census tract-level to assess the distribution of exposure VC within counties.

Statistical analyses

County-level ambient VC data were log (base 10) transformed (due to positive skewness), and then both ambient VC and covariate data were categorized into tertiles (low, medium, and high) (Table 1). Ambient VC and covariate data were modeled as categorized tertiles, with the lowest tertile serving as the reference group. A sensitivity analysis was performed using scored tertiles (1, 2, 3), instead of categorized tertiles (2 vs. 1 and 3 vs. 1).

For liver cancer incidence analyses, negative binomial regression models were utilized due to over-dispersed incidence data. For liver cancer mortality analyses, Poisson regression models were used, as the mortality data were not over-dispersed. For both models, unadjusted univariate analyses were performed for ambient VC and each covariate variable. Statistically significant predictors of liver cancer (county-level heavy drinking prevalence, hepatitis mortality rates, median income, and race [percent Hispanic]) were included in the adjusted multivariable liver cancer incidence and mortality models. These covariates were also statistically significantly associated with ambient VC using univariate linear regression. County-level smoking prevalence, obesity prevalence, and sex (percent male) were not statistically significant predictors of liver cancer incidence or mortality and were excluded from the final models; however, we performed a sensitivity analysis with all of these variables included. The calculated risk estimates for all models were incidence/mortality rate ratios (relative risk [RR]; exponentiated beta-coefficients from the models) with 95% confidence intervals (CIs). All analyses were performed using STATA 14.2.

Sensitivity analyses

In addition to tests previously described, various sensitivity analyses were performed to further assess the consistency of findings. Different alternative imputation methods were evaluated, including assuming that each suppressed county had 1) 8 cases of liver cancer (median of range of 0–15 cases), 2) 15 cases of liver cancer (maximum number of cases that would still be suppressed), and 3) a liver cancer incidence rate similar to the Texas average (9.5 cases per 100,000 population, as reported by the TCR). For the assumed 8 and 15 cases models, we were unable to calculate incidence rates for the counties with a risk population of <1,000 persons ($n=8$), as specific population sizes were not reported by the TCR. Therefore, we utilized the imputed incidence rates from the multivariable imputation method for these counties.

To examine spatial correlation, a geospatial regression was performed using GeoDa 1.12.1.131. It is likely that area-level factors influence exposure observations, where individuals living near each other experience exposures that are more similar than exposures experienced by individuals living further away. Spatial regression can account for this

Table 1. Descriptive results for county-level VC and covariates

Variable	Tertile 1			Tertile 2			Tertile 3		
	n ^a	Mean	Range	n ^a	Mean	Range	n ^a	Mean	Range
Median ambient VC, µg/m ³	85	7.87×10 ⁻⁶	4.12×10 ⁻⁶ – 1.78×10 ⁻⁵	85	4.45×10 ⁻⁵	1.86×10 ⁻⁵ – 8.99×10 ⁻⁵	84	2.19×10 ⁻³	9.10×10 ⁻⁵ – 1.02×10 ⁻¹
Median ambient VC – Log (base 10) transformed, µg/m ³	85	-5.24	-7.39 – -4.75	85	-4.40	-4.73 – -4.05	84	-3.41	-4.04 – -0.99
Heavy drinking prevalence, %	91	4.76	2.7–5.7	88	6.24	5.8–6.7	75	7.65	6.8–16.5
Hepatitis mortality rate, per 100,000	98	0.26	0.15–0.32	72	0.36	0.33–0.39	84	0.53	0.40–2.2
Median household income, \$	85	33,679	19,959–38,702	85	42,018	38,832–45,929	84	54,632	46,128–83,889
Race, percent Hispanic	85	10.7	0–17.0	85	24.3	17.1–35.5	84	58.1	35.8–98.3
Sex, percent male	85	48.4	44.9–49.1	85	49.8	49.1–50.6	84	54.2	50.7–68.4
Smoking prevalence, %	87	18.9	14.0–21.1	85	22.5	21.2–23.8	82	25.9	23.9–29.6
Obesity prevalence, %	91	28.0	23.3–29.1	79	29.7	29.2–30.3	84	32.0	30.4–36.8

^aIdentical values were assigned to the same tertile, which may result in an unequal distribution of the number of counties across the tertiles.

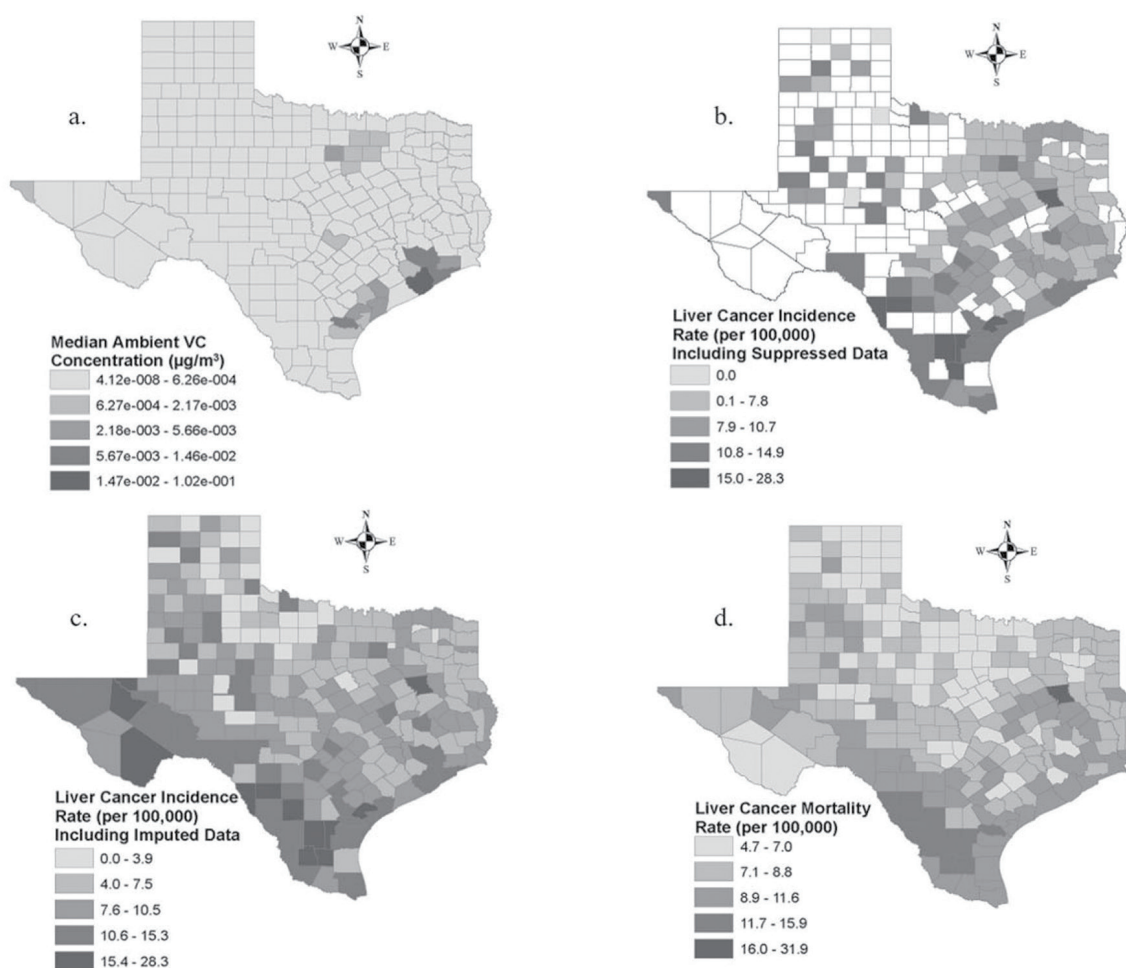


Fig. 1. Spatial distribution of county-level data. (A) Median ambient VC ($\mu\text{g}/\text{m}^3$). (B) Liver cancer incidence rate (per 100,000), including suppressed data. (C) Liver cancer incidence rate (per 100,000), including imputed data. (D) Liver cancer mortality (per 100,000).

spatial correlation, which is not possible with a negative binomial or Poisson regression models. Weights were created using rook contiguity (neighbor defined as a county sharing a common border), and scored tertiles were used for both exposure and covariate data.^{27,28} A spatial lag regression model was selected based on diagnostics from the ordinary least-squares regression. A spatial error regression model was also ran as an additional sensitivity analysis.

Results

VC exposure

Modeled county-level median ambient VC was reported for all 254 counties in Texas, ranging from $4.12 \times 10^{-8} \mu\text{g}/\text{m}^3$ (McMullen County) to $0.0102 \mu\text{g}/\text{m}^3$ (Brazoria County). The median ambient VC in Texas was $5.85 \times 10^{-4} \mu\text{g}/\text{m}^3$. Counties with imputed liver cancer incidence rates (suppressed counties) had a lower average county-level median ambient VC in comparison to counties with reported liver cancer incidence rates in the TCR (not suppressed counties) (Supplementary Table 1). The spatial distribution of median ambient VC by Texas county is shown in Fig. 1.

Exposure VC data were available at both the county-level

and census-tract level in the NATA 1996 dataset, and exploratory spatial data analyses revealed that within the same county, exposure VC varied across the census tract-level for some counties. For example, the largest difference was in Brazoria County, where the county-level median exposure VC was $0.073 \mu\text{g}/\text{m}^3$, while the census tract-level exposure VC within Brazoria County ranged from 0– $0.74 \mu\text{g}/\text{m}^3$. Out of the 254 counties, 27 (11%) had an order of magnitude or larger difference between the maximum census tract-level VC exposure and the median county-level exposure VC (e.g., similar to Brazoria County) (Supplementary Table 2). Only 6% of counties had an order of magnitude or larger difference between the census-tract level 95th percentile exposure VC and the county-level 95th percentile exposure VC values (Supplementary Table 2).

Liver cancer incidence

Liver cancer incidence rates (per 100,000 population) ranged from 0 (Cottle, Sterling, Lipscomb, and Sherman Counties) to 28.3 (Brooks County) in the dataset from the TCR (not suppressed counties), and ranged from 0 (Armstrong, Franklin, and Throckmorton Counties) to 18.6 (Jim Hogg County) in the imputed dataset (for suppressed counties). The average liver cancer incidence rate in the data-

Table 2. Liver cancer incidence and mortality rate ratios

Variable ^a	Category, tertile	Liver cancer incidence		Liver cancer mortality	
		Unadjusted RR (95% CI)	Adjusted RR (95% CI) ^b	Unadjusted RR (95% CI)	Adjusted RR (95% CI) ^b
Median ambient VC	Low (ref)	NA	NA	NA	NA
	Medium	1.01 (0.87–1.17)	1.06 (0.95–1.19)	0.94 (0.84–1.04)	0.95 (0.85–1.05)
	High	1.11 (0.95–1.28)	1.03 (0.90–1.17)	0.97 (0.87–1.07)	0.93 (0.82–1.05)
Heavy drinking	Low (ref)	NA	NA	NA	NA
	Medium	1.39 (1.21–1.59)*	1.38 (1.23–1.56)*	1.08 (0.97–1.20)	1.07 (0.95–1.20)
	High	1.65 (1.43–1.89)*	1.66 (1.46–1.88)*	1.21 (1.09–1.35)*	1.22 (1.08–1.37)*
Hepatitis	Low (ref)	NA	NA	NA	NA
	Medium	1.27 (1.10–1.47)*	1.14 (1.01–1.28)*	1.12 (1.00–1.25)*	1.08 (0.97–1.22)
	High	1.38 (1.21–1.59)*	1.16 (1.03–1.30)*	1.26 (1.14–1.40)*	1.19 (1.06–1.33)*
Race, percent Hispanic	Low (ref)	NA	NA	NA	NA
	Medium	1.21 (1.06–1.39)*	1.23 (1.09–1.39)*	1.00 (0.89–1.11)	1.00 (0.89–1.11)
	High	1.69 (1.48–1.93)*	1.68 (1.50–1.88)*	1.26 (1.14–1.40)*	1.22 (1.10–1.36)*
Income	Low (ref)	NA	NA	NA	NA
	Medium	0.90 (0.78–1.04)	0.87 (0.78–0.97)*	0.94 (0.85–1.04)	0.94 (0.84–1.04)
	High	0.75 (0.65–0.87)*	0.76 (0.67–0.85)*	0.80 (0.72–0.89)*	0.83 (0.74–0.93)*

^aVC data were log transformed, and then both VC and covariate data were divided into tertiles (low, medium, and high). ^bAdjusted for heavy drinking prevalence, hepatitis mortality rates, income, and race (percent Hispanic). **p*-value <0.05. Abbreviations: CI, confidence interval; RR, rate ratio; VC, vinyl chloride.

set from the TCR was 9.0, while the average liver cancer incidence rate in the imputed dataset was 7.3. The spatial distribution of liver cancer incidence rates by Texas county is shown in Fig. 1 (for both suppressed and imputed counties). Based on visual inspection, there was no evidence of a spatial association between county-level ambient VC and liver cancer incidence rates (Fig. 1).

County-level ambient VC was not associated with county-level liver cancer incidence in either the univariate or multivariable negative binomial regression models. Specifically, in the unadjusted model, the middle and highest tertiles of county-level ambient VC were associated with liver cancer incidence RRs of 1.01 (95% CI: 0.87–1.17) and 1.11 (95% CI: 0.95–1.28), respectively, when compared to the lowest tertile of ambient VC. The absence of association between county-level ambient VC tertile and liver cancer incidence persisted (RR: 1.06, 95% CI: 0.95–1.19 and RR: 1.03, 95% CI: 0.90–1.17 for the middle and highest tertiles, respectively) upon adjustment for county-level heavy drinking prevalence, hepatitis mortality rate, race (percent Hispanic) and median income. Increasing tertiles of county-level heavy drinking prevalence, hepatitis mortality rate, and race (percent Hispanic) were associated with a statistically significant increased rate of liver cancer incidence, while increasing tertiles of county-level median income were associated with a statistically significant decreased rate of liver cancer incidence (Table 2).

Liver cancer mortality

Liver cancer mortality rates (per 100,000 population) ranged from 4.72 (Hartley County) to 31.9 (Anderson County). The average liver cancer mortality rate was 8.1. The spatial distribution of liver cancer mortality rates by Texas county is shown in Fig. 1. Based on visual inspection, there was no evidence of a spatial association between county-level am-

bient VC and liver cancer mortality rates (Fig. 1).

County-level ambient VC was not associated with county-level liver cancer mortality in either the univariate or multivariable Poisson regression models. In the unadjusted model, the middle and highest tertiles of county-level ambient VC had liver cancer mortality RRs of 0.94 (95% CI: 0.84–1.04) and 0.97 (95% CI: 0.87–1.07), respectively, when compared to the lowest tertile of county-level ambient VC. In the adjusted model, the middle (RR: 0.95, 95% CI: 0.85–1.05) and highest (RR: 0.93, 95% CI: 0.82–1.05) tertiles of ambient VC were not associated with liver cancer mortality, when compared to counties with the lowest tertile of ambient VC (Table 2). The highest tertile of county-level heavy drinking prevalence, hepatitis mortality rate, and race (percent Hispanic) were associated with a statistically significant increased rate of liver cancer mortality, when compared to counties with the lowest tertile. In contrast, counties with the highest tertile of county-level median income were associated with a statistically significant decreased rate of liver cancer mortality (Table 2).

Sensitivity analyses

County-level ambient VC was consistently not associated with county-level liver cancer incidence or mortality rates across various sensitivity analyses (Supplementary Table 3). For example, both county-level 95th percentile ambient VC and median exposure VC were not associated with liver cancer incidence or mortality rates. Additionally, no association was observed between county-level median ambient VC and liver cancer incidence or mortality rates when the final multivariable model also included the covariates of county-level smoking prevalence, obesity prevalence, and sex (percent male). Similarly, no association was observed when each suppressed county was assigned the Texas average liver cancer incidence rate. When each suppressed county was

assigned 8 or 15 cases of liver cancer, the highest tertile of county-level ambient VC was associated with a statistically significant decreased rate of liver cancer incidence, when compared to the lowest tertile of ambient VC. Further, county-level ambient VC was not associated with county-level liver cancer incidence in both the spatial lag and spatial error regression models, while county-level ambient VC was associated with a statistically significant decreased rate of liver cancer mortality in both the spatial lag and spatial error regression models (Supplementary Table 3).

Discussion

This ecological analysis provides no evidence that Texas county-level ambient VC is associated with either county-level liver cancer incidence or mortality rates in Texas. This study has the strength of evaluating ambient VC levels that the general population could experience. There is evidence that high cumulative exposure to VC in occupational settings (*i.e.* $\geq 2,271$ ppm-years or $>2,500$ ppm-years) is associated with an increased risk of HCC.^{11,29} However, the potential health effects of lower exposure to ambient levels (*e.g.*, emissions from manufacturing and processing facilities) have not been well studied. In our analysis, low-level county-level ambient VC (maximum concentration of $0.0102 \mu\text{g}/\text{m}^3$ [approximately 4 ppt]) was not associated with county-level liver cancer incidence or mortality rates in Texas. Our findings contrast with the conclusions of the Cicalese *et al.*¹² (2017) analysis, which reported that VC is a significant contributor to the incidence of liver cancer in Texas (Cicalese and colleagues did not specify if they used ambient or exposure VC concentrations).

This analysis considered multiple well-known risk factors of liver cancer as potential confounders, including sex, race, heavy drinking prevalence, hepatitis mortality rates, smoking prevalence, income, and obesity. Our analysis suggests that county-level heavy drinking prevalence, hepatitis mortality rates, median income, and race (percent Hispanic) were all significant predictors of liver cancer incidence and mortality. These results are in agreement with other studies that have reported that heavy drinking and hepatitis virus infections are associated with an increased risk of liver cancer.^{30–32} It has also been reported that income is inversely associated with liver cancer incidence, likely due to differences in risk factors across income levels.³³ Additionally, the American Cancer Society reports that race is a risk factor of liver cancer, with higher rates of liver cancer occurring among Asian Americans, Pacific Islanders, and Hispanics/Latinos compared to other races and ethnicities among the USA's population.³⁴ The consideration of potential confounding factors is an important strength of our study that builds upon the only other published study on this topic. While the Cicalese *et al.*¹² (2017) study did adjust for county-level hepatitis C infection prevalence, the number of prisons in each county (as a proxy of hepatitis infection), and obesity prevalence, they did not consider several other important risk factors, suggesting that their results may be susceptible to bias.

A limitation of the underlying data in our analysis was the high prevalence of missing liver cancer incidence rate data (43% of counties). Rather than simply excluding the missing data, this analysis has the strength of using various methodologies to impute liver cancer incidence rates for the suppressed counties. Counties with suppressed liver cancer incidence data had a lower average population size than counties with reported incidence rates (Supplementary Table 1). This is logical, as the TCR suppresses counties with less than 16 cases of cancer and counties with an insufficient at-risk population size. Additionally, counties with imputed incidence rates (suppressed counties) had a lower

average and lower bound median ambient VC than counties with reported incidence rates (not suppressed counties), by an order of two magnitudes (Supplementary Table 1). Therefore, this study increased the county sample size and representativeness by including a wider range of exposure across populations in Texas. Additionally, exclusion of the suppressed counties may result in bias if the excluded counties (due to low liver cancer counts) were also counties with low ambient VC. In this scenario, exclusion of the suppressed counties could result in a biased overestimate of the association.

We used NATA data from 1996, TCR data from 2006–2015, and mortality data from 2010, which allows for a minimum of 10 years between exposure and disease diagnosis. Studies that do not allow for sufficient induction and latency periods for liver cancer result in biased risk estimates, as the short latency time periods are potentially biologically irrelevant, as any impact on disease would not yet have been clinically diagnosed or detected. Liver cancer latency (not specific to a certain exposure) has been estimated to be 10.8 years, defining latency as the time from cancer initiation to diagnosis.³⁵ It has also been reported that the median latency for angiosarcoma of the liver and HCC deaths among workers at VC or PVC plants are 36 and 48 years, respectively.¹¹ Another study reported that the latency for malignant hepatoma (predominantly ASL, but HCC and cholangiocellular carcinoma were also noted) among PVC production workers ranged from 12 to 34 years.³⁶ It should be noted that these latency estimates are specific to disease mortality rather than disease diagnosis. We included the earliest NATA data available but it is possible that the latency periods in this study may not be sufficient for some disease diagnosis or mortality.

As an ecological study, this analysis is limited to examining ambient VC, potential confounders, and liver cancer incidence/mortality rates at the county-level rather than the individual level. Ecological studies are inherently limited by specification bias, aggregation bias, and temporal ambiguity.^{37,38} For example, while our data suggests that county-level heavy drinking prevalence is associated with liver cancer incidence/mortality rates, we are unable to confirm that individuals with liver cancer are high consumers of alcohol. On the other hand, our use of a large number of relatively small, homogeneous counties as the units of analysis and the multiple covariates available for each of those counties helped to offset the general limitations inherent in ecological analyses.

Another limitation is potential exposure misclassification at the county-level due to the spatial distribution of exposure VC. This limitation of NATA data is noted by the EPA when it is suggested that "modeling results should not be used to draw conclusions about local exposure concentrations or risk".³⁹ Exploratory spatial analyses also revealed that there was variability in exposure VC at census tracts within the same county. County-level median concentrations may be a spatially diluted value that is not an adequate representation of exposures at specific locations, such as underestimating exposure at residences potentially near point sources or overestimating exposures at residences far away from point sources. While spatial variability in exposure VC was noted between the county-level and census tract-level, the maximum census tract-level and median county-level concentrations were not markedly different for the majority of counties. Additionally, we explored this potential dilution effect by evaluating the 95th percentile ambient VC for each county in Texas. Since the available data indicated that census tract-level and county-level 95th percentile values were similar for over 90% of the counties, we do not believe that potential spatial heterogeneity substantially impacted our conclusions. Therefore, a strength of this analysis is the examination of spatial heterogeneity through these various sensitivity analyses.

Conclusions

Overall, our findings suggest that county-level ambient VC is not associated with county-level liver cancer incidence or mortality in Texas. Strengths of this analysis include adjusting statistical analyses for potential confounding by known risk factors for liver cancer, allowing time for the development of liver cancer following VC exposure, and utilizing analytical imputation methods to address missing data. Our ecological study provides novel results regarding liver cancer risk from exposure to low-level ambient VC.

Funding

This work was funded by The Vinyl Institute, a USA trade organization that represents manufacturers of vinyl, vinyl chloride monomer, and vinyl additives and modifiers. The preparation of the manuscript and the conclusions drawn are the exclusive professional work product of the authors and may not necessarily be those of Cardno ChemRisk or The Vinyl Institute.

Conflict of interest

Authors KMT, SMB, NSE, and GMM are employed by Cardno ChemRisk, a consulting firm that provides scientific advice to the government, corporations, law firms, and various scientific/professional organizations. GMM is also professor emeritus of biostatistics and founding director of the Center for Occupational Biostatistics and Epidemiology at the University of Pittsburgh, Graduate School of Public Health.

Author contributions

Participated in the study design, data analysis, and manuscript writing (KMT, SMB, NSE, GMM).

References

- [1] ChemicalSafetyFacts.org. Chemical safety facts: Polyvinyl chloride - Uses, benefits and safety facts. Available from: <https://www.chemicalsafetyfacts.org/polyvinyl-chloride>.
- [2] Somheil T. Study: global PVC demand to grow 3.2% annually through 2021. Available from: <https://www.plasticstoday.com/study-global-pvc-demand-grow-32-annually-through-2021>.
- [3] World Health Organization, International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans. Available from: <https://monographs.iarc.fr/wp-content/uploads/2018/08/14-002.pdf>.
- [4] Mundt KA, Dell LD, Austin RP, Luippold RS, Noess R, Bigelow C. Historical cohort study of 10 109 men in the North American vinyl chloride industry, 1942-72: update of cancer mortality to 31 December 1995. *Occup Environ Med* 2000;57:774-781. doi:10.1136/oem.57.11.774.
- [5] Ward E, Boffetta P, Andersen A, Colin D, Comba P, Daddens JA, *et al*. Update of the follow-up of mortality and cancer incidence among European workers employed in the vinyl chloride industry. *Epidemiology* 2001;12:710-718. doi:10.1097/00001648-200111000-00021.
- [6] Boffetta P, Matisane L, Mundt KA, Dell LD. Meta-analysis of studies of occupational exposure to vinyl chloride in relation to cancer mortality. *Scand J Work Environ Health* 2003;29:220-229. doi:10.5271/sjweh.725.
- [7] Toxicological profile for vinyl chloride. Available from: <https://www.atsdr.cdc.gov/toxprofiles/TP.asp?id=282&tid=51>.
- [8] Vinyl chloride: Hazard summary. Available from: <https://19january2017snapshot.epa.gov/sites/production/files/2016-09/documents/vinyl-chloride.pdf>.
- [9] National Toxicology Program. Vinyl chloride. CAS No. 75-01-4. Report on Carcinogens, 11th Edition - Carcinogen Profiles. 12th ed. Research Triangle Park, NC; U.S. Dept. of Health and Human Services, Public Health Service, National Toxicology Program 2004:115-120.
- [10] Permissible exposure limits (PELs)-annotated tables. Available from: <https://www.osha.gov/annotated-pels>.
- [11] Mundt KA, Dell LD, Crawford L, Gallagher AE. Quantitative estimated ex-

posure to vinyl chloride and risk of angiosarcoma of the liver and hepatocellular cancer in the US industry-wide vinyl chloride cohort: mortality update through 2013. *Occup Environ Med* 2017;74:709-716. doi:10.1136/oemed-2016-104051.

- [12] Cicalese L, Raun L, Shirafkan A, Campos L, Zorzi D, Montalbano M, *et al*. An ecological study of the association between air pollution and hepatocellular carcinoma incidence in Texas. *Liver Cancer* 2017;6:287-296. doi:10.1159/000475776.
- [13] Marsh GM, Towle KM. Methodological limitations to the analysis by Cicalese *et al*. *Liver Cancer* 2018;7:299-300. doi:10.1159/000489022.
- [14] Technology transfer network 1996 national-scale air toxics assessment: Overview: The 4 steps. Available from: <https://archive.epa.gov/airtoxics/nata/web/html/4steps.html>.
- [15] Technology transfer network 1996 national-scale air toxics assessment: The ASPEN model. Available from: <https://archive.epa.gov/airtoxics/nata/web/html/asp.html>.
- [16] National air toxics assessment: NATA frequent questions. Available from: <https://www.epa.gov/national-air-toxics-assessment/nata-frequent-questions>.
- [17] Texas Health and Human Services. About the Texas cancer registry. Available from: <https://www.dshs.texas.gov/tcr/about.aspx>.
- [18] Mokdad AH, Dwyer-Lindgren L, Fitzmaurice C, Stubbs RW, Bertozzi-Villa A, Morozoff C, *et al*. Trends and patterns of disparities in cancer mortality among US counties, 1980-2014. *JAMA* 2017;317:388-406. doi:10.1001/jama.2016.20324.
- [19] United States Census Bureau. Age and Sex dataset S0101. 2009-2013 American Community Survey 5-year estimates. Available from: https://www.tiogacountyny.com/media/2527/vill-waverly-acs-2009_2013.pdf.
- [20] United States Census Bureau. American Community Survey: Median income in the past 12 months (in 2010 inflation-adjusted dollars) dataset S1903. Available from: https://faast.waterboards.ca.gov/attachments/proposal_24274/attachment_63395.pdf.
- [21] United States Census Bureau. American Community Survey 5-year data, ACS demographic and housing estimates (Table DP05). Available from: https://www.tiogacountyny.com/media/8917/village-of-spencer-acs-2015_2019-data-profiles.pdf.
- [22] Dwyer-Lindgren L, Mokdad AH, Srebotnjak T, Flaxman AD, Hansen GM, Murray CJ. Cigarette smoking prevalence in US counties: 1996-2012. *Popul Health Metr* 2014;12:5. doi:10.1186/1478-7954-12-5.
- [23] Dwyer-Lindgren L, Flaxman AD, Ng M, Hansen GM, Murray CJ, Mokdad AH. Drinking patterns in US counties from 2002 to 2012. *Am J Public Health* 2015;105:1120-1127. doi:10.2105/AJPH.2014.302313.
- [24] CDC. Overweight & obesity. Available from: <https://www.cdc.gov/obesity/data/index.html>.
- [25] IHME. United States infectious disease mortality rates by country 1980-2014. Available from: <http://ghdx.healthdata.org/us-data>.
- [26] Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc* 1996;91:473-489. doi:10.1080/01621459.1996.10476908.
- [27] Anselin L. Exploring spatial data with GeoDa: A workbook. Available from: https://www.researchgate.net/profile/Lubna_Hamzalouh/post/Is_it_possible_to_use_spatial_autocorrelation_on_UTM_data_to_evaluate_association_between_road_casualties_and_situation_of_road_structural_obstacles/attachment/59d62b8879197b8077989e6e/AS:34336560348480@1458876147212/download/Anselin++2005++Exploring+Spatial+Data+with+GeoDa+A+Workbook-annotated.pdf. Accessed March 6, 2005.
- [28] Anselin L, Syabri I, Kho Y. GeoDa: An introduction to spatial data analysis. *Geograph Anal* 2006;38:5-22. doi:10.1111/j.0016-7363.2005.00671.x.
- [29] Mastrangelo G, Fedeli U, Fadda E, Valentini F, Agnesi R, Magarotto G, *et al*. Increased risk of hepatocellular carcinoma and liver cirrhosis in vinyl chloride workers: synergistic effect of occupational exposure with alcohol intake. *Environ Health Perspect* 2004;112:1188-1192. doi:10.1289/ehp.6972.
- [30] El-Serag HB, Mason AC. Risk factors for the rising rates of primary liver cancer in the United States. *Arch Intern Med* 2000;160:3227-3230. doi:10.1001/archinte.160.21.3227.
- [31] Perz JF, Armstrong GL, Farrington LA, Hutin YJ, Bell BP. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* 2006;45:529-538. doi:10.1016/j.jhep.2006.05.013.
- [32] Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, *et al*. Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *Br J Cancer* 2015;112:580-593. doi:10.1038/bjc.2014.579.
- [33] Anyiwe K, Qiao Y, De P, Yoshida EM, Earle CC, Thien HH. Effect of socioeconomic status on hepatocellular carcinoma incidence and stage at diagnosis, a population-based cohort study. *Liver Int* 2016;36:902-910. doi:10.1111/liv.12982.
- [34] Liver cancer risk factors. Available from: <https://www.cancer.org/cancer/liver-cancer/causes-risks-prevention/risk-factors.html>.
- [35] Nadler DL, Zurbenko IG. Estimating cancer latency times using a Weibull model. *Adv Epidemiol* 2014;2014:746769. doi:10.1155/2014/746769.
- [36] Leilbach WK. A 25-year follow-up study of heavily exposed vinyl chloride workers in Germany. *Am J Ind Med* 1996;29:446-458.
- [37] Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health* 1995;16:61-81. doi:10.1146/annurev.pu.16.050195.000425.
- [38] Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982;72:1336-1344. doi:10.2105/ajph.72.12.1336.
- [39] EPA strongly cautions that these modeling results should not be used. Available from: <https://archive.epa.gov/airtoxics/nata/web/html/hapem-page-2new3.html>.