Nontuberculous mycobacterial disease and molybdenum in Colorado watersheds.

E.M. Lipner, Ph.D, M.P.H.[1,2], J. French, Ph.D[3], C.R. Bern, Ph.D[4],

K. Walton-Day, Ph.D [4], D. Knox, Ph.D[5], M. Strong, Ph.D[1],

D.R. Prevots, Ph.D, M.P.H.[6*] and J.L. Crooks, Ph.D, M.S.[1,2*]

*Senior Authors

[1]National Jewish Health, Denver, CO USA; [2]Colorado School of Public Health, Aurora, CO USA; [3]Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO USA; [4]Colorado Water Science Center, United States Geological Survey, Denver, CO USA; [5]University of Colorado-Boulder, Boulder, CO USA; [6]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD USA

Address for correspondence:

Ettie M. Lipner, Ph.D, M.P.H.; Center for Genes, Environment, and Health, National Jewish Health, 1400 Jackson Street. Denver, CO 80602. Tel: 303-398-1861, Email: LipnerE@njhealth.org

The following authors participated in the conception of the study: EML. In the collection of the data: EML, CRB, KWD. In the design of the study: EML, KWD, JLC, DRP, DK, CRB. In the analysis and interpretation: EML, KWD, JLC, DRP, JF, MS, CRB. Drafting the manuscript for important intellectual content: EML, KWD, JLC, DRP, JF, MS, CRB, DF.

SUPPORT:

EML was partially supported by the Cystic Fibrosis Foundation. KWD and CB were supported by the Toxic Substances Hydrology Program, USGS. DRP was supported by the Division of Intramural Research, NIAID. JF was supported by NSF awards 1463642 and 1915277. MS and JLC were supported by NSF award 1743597. DK was unfunded.

All authors report no conflict of interest.

Article Summary Line: Increasing molybdenum concentrations in the source water of Colorado watersheds is significantly associated with NTM disease risk.

Running title: NTM and molybdenum in Colorado watersheds.

Keywords: nontuberculous mycobacteria; watersheds; molybdenum; spatial; Poisson; source water

Total word count: 3498

Abstract word count: 146

ABSTRACT

Nontuberculous mycobacteria (NTM) are environmental bacteria which may cause chronic lung disease. Environmental factors that favor NTM growth likely increase the risk of NTM exposure within particular environments. We aimed to identify water quality characteristics associated with NTM disease across Colorado watersheds. We conducted a geospatial, ecological study, associating data from patients with NTM disease treated at National Jewish Health and water-quality data from the Water Quality Portal. Water-quality characteristics associated with disease were assessed using generalized linear models with Poisson-distributed discrete responses. We observed a highly robust association between molybdenum in the source water and disease risk. For every 1-unit increase in the log concentration of molybdenum in the source water, disease risk increased by 16.4%. The risk of NTM varies by watershed and is associated with watershed-specific water quality characteristics. These findings may contribute to mitigation strategies to decrease the overall risk of exposure.

INTRODUCTION

Nontuberculous mycobacteria (NTM) are environmental organisms and opportunistic pathogens responsible for an increasingly high burden of lung disease in North America, and indeed worldwide (1, 2). More than 190 NTM species have been identified to date (3); they have been isolated from a variety of environmental sources, primarily soil and water, which are the natural reservoirs of these organisms. Environmental conditions related to soil properties, natural water, and the characteristics of engineered water systems, including the biofilms that form in hospital and municipal water supplies, are believed to contribute to increased concentrations of NTM leading to greater potential for NTM exposure (4). While exposure to NTM is extremely common and the disease is rare, distinct geographic variability of disease has been demonstrated in both general and high-risk populations (5-11). Hawaii, Florida and California have consistently shown high disease prevalence (8, 10, 11). These geographic differences are not explained by host related factors, but rather are due to variation in regional environmental conditions. Specific soil and water related factors that favor NTM growth and persistence likely increase the risk of NTM exposure within particular environments. Previous epidemiologic studies (6, 8, 11-15) demonstrate that specific environmental factors may interact to create conditions favorable for increased concentrations of NTM organisms, thereby increasing the individual exposure risk in a given environment. However large gaps remain in our understanding of the geographic variability of NTM.

Identifying determinants of the regional ecology and environmental sources of NTM is of major public health importance. The rapidly aging U.S. population has greater risk for developing NTM disease. Explaining the increasing prevalence trends is critical,

as NTM patients undergo lengthy and complex treatment regimens, and are often re-infected despite initial cure. The lack of evidence-based guidance on environmental risk factors is a large unmet need for populations at risk for this disease.

In our previous study (14), we demonstrated an increased risk of NTM disease within specific watersheds in Colorado. To further explore these findings, we sought to explain why we observed higher disease risk in these areas. We used an ecological design with water quality data collected or hosted by the U.S. Geological Survey, U.S. Environmental Protection Agency and National Water Quality Monitoring Council and NTM data from patients who were resident in the State of Colorado and treated at National Jewish Health (NJH) a leading respiratory hospital in Denver. We aimed to identify whether water-quality characteristics across watersheds increase the risk of NTM disease in Colorado.

METHODS

Data Collection

Patient data were obtained from the NJH Electronic Medical Record database. Our study population comprised all patients with a diagnosis of NTM treated at NJH and who were resident in Colorado during the study period, from February 2008 through January 2018. Patient address, NTM species, and patient demographic information were extracted. Body site isolation data were not available, therefore NTM disease includes pulmonary and extra-pulmonary. This study was approved by the NJH Institutional Review Board (HS-3148).

NTM species

NTM species from patient isolates are listed in Supplementary Table 1.

Socio-demographic data:

Center for International Earth Science Information Network (CIESIN), Columbia University.

Gridded population density datasets for total population as well as age and racial/ethnic categories were obtained for Colorado during 2010 from the Socioeconomic Data and Applications Center (SEDAC) (16). This dataset contains the following racial/ethnic categories: White alone, Black or African American alone, Asian alone, American Indian and Alaska Native alone, Native Hawaiian and Other Pacific Islander alone, Some Other Race alone, and Two Or More Races. Since Colorado has a majority white population, for the analysis we created two categories: "White alone" versus all other race/ethnic groups, termed "Non-White".

Environmental exposure data:

Watershed boundaries were obtained from the Watershed Boundary Dataset (17), and have been previously described (14). In this analysis we used the Hydrologic Unit Code (HUC)-10 watershed level.

Water-Quality Data Compilation

We obtained water-quality data from the Water Quality Portal (WQP) (18), a water quality database sponsored by the U.S. Geological Survey (USGS), the U.S. Environmental Protection Agency (USEPA) and the National Water Quality Monitoring Council (NWQMC). We extracted surface-water data that included 62 water-quality characteristics from 7,174 unique sampling locations collected in Colorado from January 1, 2000 through December 31, 2018. Data. We examined the filtered water-sample

129 fractions (filtered indicates that water was passed through a 0.45 micrometer filter (19)).
130 (Supplementary table 2).

131 **Statistical Analysis**

132 Analysis of data was performed using the R packages, "rgdal" (20), "sp" (21), "arm"
133 (22). All water-sample sites were aggregated by watershed and the median value of each
134 water-quality characteristic was calculated for each watershed using ArcGIS 10.2 (Esri,
135 Boston, Mass). Water-quality characteristics were eliminated if data did not exist for over
136 50% of watersheds. Sites incorrectly coded as surface water (for example, snow
137 collection sites) were also eliminated. Reporting errors for concentration units were
138 corrected (for example, values that were lower than any others from a particular data-
139 collection entity by three orders of magnitude were multiplied by 1,000 to place them in
140 the range of the other data). Seventeen remaining water-quality characteristics were
141 available for analysis. We log transformed all watershed-median variables with a highest
142 to lowest value ratio greater than 3 across watersheds (17 variables). For watersheds with
143 missing data, we imputed the median value using the "imputePCA" function in the
144 "missMDA" R package (version 1.14). We excluded watersheds where no water samples
145 had been collected. As a result, we conducted our analyses on 412 out of 575 HUC-10
146 level watersheds. Drive time between watershed centroids and NJH were calculated using
147 the R "rgeos" package (23). We categorized watersheds based on whether they were
148 within a 2.0-hour drive to NJH.

149 ***Principal Component Analysis (PCA)***

150 PCA was performed on the HUC-10 level dataset (after data were log transformed and
151 imputed) including 17 metals and nonmetals (Table 1) using the "PCA" function in the R

152 package, “FactoMineR” (version 1.42). We retained the top five principal components for
153 further analysis. These components together explained 78.6% of the variation of the
154 dataset.

155 ***Poisson regression models:***

156 A Poisson regression model was constructed to model disease risk as a function of water-
157 quality characteristics. Our models used the standard log link function and include the log
158 of the population density within each watershed for the state of Colorado as an offset
159 term to account for the differing population densities in each region. NTM case counts
160 were aggregated by watershed. Age, race/ethnicity, and drive-time variables were
161 included in all models to control for potential confounding induced by using a hospitable-
162 based population.

163 *Principal Component Regression*

164 The first Poisson model included the top five principal components from the PCA (Table
165 2). As a sensitivity analysis, we also explored whether including only principal
166 components 1-3 (explaining 67.3% of the variation) changed the results of the regression
167 analysis.

168 *Poisson Regressions with Individual Metals*

169 We identified the individual metals that contributed more than 10% to each significant
170 principal component from the first model (Supplementary table 3). The identified metals
171 were then added as predictor variables for our second Poisson model (Table 3). In our
172 third and fourth Poisson models (Table 4), we constructed separate Poisson regression
173 models for the metals that demonstrated statistical significance from model 2 ($p<0.05$).

174 *Sensitivity Analyses*

We performed two sensitivity analyses of our final results. First, to explore the influence of the drive-time variable on our disease risk estimates, we varied the drive time threshold from 1.5 hours to 2.5 hours. Second, to investigate how our estimates changed when we relaxed our distributional modeling assumption, we used a negative binomial response distribution instead of Poisson.

RESULTS

Our study population comprised 821 patients with NTM disease who had sought treatment at NJH between February 2008 and January 2018 and reside in Colorado. For all NTM patients, the mean age was 64.8 years (± 18.1) and the majority of patients (74.4%) were white. Figure 1 shows the distribution of patients' residential locations across watersheds. Eight hundred and seven patients were available for analysis after accounting for dropped watersheds.

Principal Component Regression

Our findings showed that principal components 1 and 3 were significantly associated with disease risk (Model 1; Table 2). Principal component 1 was associated with a 5.2% increase in disease risk; the highest contributing variables included calcium, magnesium, potassium, sodium, chloride and sulfate (Supplementary table 3). Principal component 3 was associated with an 8.9% increase in disease risk; the highest contributing variables included arsenic, cadmium, manganese, molybdenum, and selenium. In a sensitivity analysis, principal components 1 and 3 remained significant (Supplementary table 4).

The fraction of the population from non-white racial/ethnic groups was a significant protective factor against NTM disease risk. Socio-economic status could be

confounding this association because we did not have up-to-date gridded socio-economic data to include.

Poisson Regression with Individual Metals

From principal component 3, we modeled the risk of NTM disease as a function of the five most highly contributing variables (arsenic, cadmium, manganese, molybdenum, and selenium). We examined the variance-inflation factors and they did not demonstrate collinearity. In a model with all five variables, we observed a significant association between molybdenum and disease risk and a less significant association between arsenic and disease risk, while controlling for the presence of the other metals, drive-time, age and race (Model 2; Table 3). Supplementary table 5 shows the correlation matrix for the five metals tested in model 2 (Table 3). Manganese was significantly correlated with cadmium and selenium. Molybdenum was significantly correlated with arsenic, cadmium, and selenium

From principal component 1, we examined the variance-inflation factors for the six highest contributing variables (calcium, magnesium, potassium, sodium, chloride, sulfate) in a generalized linear model. All six metals were highly collinear (data not shown) and no single metal contributed a substantially higher proportion to the component (Supplementary table 3); these metals are ubiquitous in the environment. Therefore, we chose to include principal component 1 as a single covariate, rather than explore these metals individually. We tested model 2 (Table 3) by including principal component 1 as a covariate to explore how this variable influenced disease risk. Principal component 1 did not demonstrate statistical significance, although molybdenum and arsenic remained significantly associated with disease ($p<0.05$) (Supplementary table 6).

We then modeled the risk of NTM disease as a function of molybdenum and arsenic in separate single-exposure models (Model 3; Table 4). Molybdenum remained significantly associated with disease risk, while arsenic did not. Our results indicate that for every 1-log unit increase in molybdenum concentration in the source water at the HUC-10 watershed level, the risk of NTM disease increases by 16.4% (Model 3; Table 4). After controlling for multiple comparisons using the Bonferroni method (5 models; new p value=0.01), the effect of molybdenum on disease risk remained statistically significant.

To explore the sensitivity of the results to drive-time, we conducted a sensitivity analysis where the drive time from the center of each watershed to NJH varied from 1.5 hours, to 2.0 hours, to 2.5 hours; molybdenum remained significant in each model. When drive time increased to 2.5 hours, manganese also became statistically significant (Supplementary Table 7). We then tested each significant metal in separate single-exposure models and found that molybdenum and manganese both retained significance (Supplementary Table 8). We then performed a second set of sensitivity analyses switching the Poisson response to a negative binomial response. Using this distribution, we ran separate single-exposure models for each of our final exposures: molybdenum and manganese. We also tested principal component 1 since it was significant in model 1 (Table 2). The estimated coefficients for principal component 1 and molybdenum remained positive and statistically significant (p=0.029 and 0.028, respectively), while manganese was not significant (p=0.575).

DISCUSSION

We found that the presence of molybdenum in the source water is associated with increased risk of NTM disease (Table 3; Supplementary Tables 7 & 8). After removing the non-significant metals from the model, we found that for every one-log unit increase in the molybdenum concentration in the source water, a 16.4% increase in NTM disease risk was observed (Table 4). We created a "drive-time" variable to control for oversampling of patients residing in the Front Range communities, where NTM patients are more likely to be seen at NJH than patients outside of this metropolitan area. By accounting for drive time in our models, we block a non-causal backdoor path in the causal diagram (Figure 4). A sensitivity analysis revealed that while molybdenum was significantly associated with disease risk in each drive-time model (Supplementary Table 7), the model with the furthest drive-time (2.5 hours) also showed that manganese concentrations in the source water were associated with increased disease risk. While prior studies have shown that low manganese concentrations in the soil is associated with increasing disease risk (8, 14), the meaning of our finding in this study is unclear and warrants further research.

The fitted risk estimates across Colorado watersheds based on the regression model with the five principal components (Model 1; Table 2) and the regression model with molybdenum alone (Model 3; Table 4) are shown in Figures 2 & 3, respectively. From our fitted estimates, we observe numerous high-risk watersheds in the mountainous regions to the west of the Continental Divide and along the Front Range to the east of the Continental Divide. Watersheds in the mountainous regions provide the majority of the water supply to highly populated communities in the Front Range (24).

The effect of molybdenum on mycobacteria has been previously described. Several molybdenum enzymes in mycobacteria exert important physiological functions. Tuberculosis and nontuberculous mycobacteria contain many proteins for the import and utilization of molybdenum, including the molybdate transport proteins modA, modB, and modC, and the molybdenum cofactor biosynthesis proteins moaA, moaB, moaC, moaD, and moaE. Some mycobacteria, including *M. tuberculosis* contain additional paralogs of the molybdenum cofactor biosynthesis proteins (25). Molybdenum has been shown to be essential for nitrate assimilation in mycobacteria (26), and is an essential component of many bacterial enzymes involved in carbon, nitrogen, and sulfur metabolism (26). In *M. tuberculosis*, molybdenum cofactor biosynthesis proteins have been suggested to be associated with pathogenesis (27), and with hypoxic persistence (26) potentially contributing to the ability to convert to nitrogen respiration under oxygen-limiting concentrations, as may be encountered in lung granulomas. This literature suggests a physiological connection linking molybdenum and essential metabolism, potentially impacting pathogenesis and persistence of *Mycobacterium tuberculosis*. While this mechanism has not been established for NTM, it offers biological plausibility since NTM and *Mycobacterium tuberculosis* are phylogenetically related organisms (4).

Our study opens many avenues of research to confirm whether molybdenum influences NTM growth in water sources as well as NTM growth in the human host. In a recent Korean study, Oh *et al.* (28) reported that trace element status is associated with mycobacterial lung disease. The authors demonstrated that patients with pulmonary NTM had higher median molybdenum concentrations in their serum (1.70 μg/L) compared with healthy controls (0.96 μg/L) and patients with pulmonary tuberculosis (0.67 μg/L).

Patients and clinicians alike would benefit from knowing whether molybdenum intake from water consumption or even certain dietary profiles (e.g. vitamin supplementation containing molybdenum) increase risk of infection and/or progression of disease. While Oh *et al.* conducted their study in South Korea, it should also be noted that molybdenum is an abundant natural resource in this region and is one of South Korea's main mining products, largely as a by-product of tungsten mining. Molybdenum is also highly abundant in the mountainous regions of Colorado where some of the world's largest producers of molybdenum are located (29).

Molybdenum is mainly used as an alloying agent in the production of steel because of its strength and ability to withstand high temperatures. Small quantities of molybdenum are essential to human, animal and plant life, and it is present in trace quantities in rocks, soil and water, often at concentrations less the 10 µg/L (30). The environmental concentrations of molybdenum can vary widely, and in places where molybdenum is processed, the concentrations in soil and water may increase considerably (31). Molybdenum has "relatively high geochemical mobility – a tendency to enter into solution in water under normal Earth-surface conditions" (31, 32); we hypothesize that perhaps even small amounts of water-soluble molybdenum may act as a metabolic source for NTM in the water supply. Soil moisture is known to influence molybdenum availability: poorly drained wet soils (for example, peat marshes, swampy organic rich soils) tend to accumulate molybdenum to high levels (33). In fact, Falkinham and colleagues have repeatedly shown that peat rich soils and brackish marshes are rich in NTM (4, 34, 35). In addition, molybdenum can form complexes with organic matter, particularly humic and fulvic acids (36). Falkinham and colleagues have also reported

310 that humic and fulvic acids support high numbers of *M. avium* complex (MAC) species
311 (4, 37). We found that the median value of molybdenum across Colorado watersheds was
312 4.3 µg/L, but reached 325 µg/L at one specific watershed.

313 Many studies have examined mycobacterial distributions and abundance in
314 different geographic areas. These studies range from examining the presence of NTM in
315 premise plumbing (6, 38-40), in the water distribution systems (41), in the water
316 treatment facilities (42) and in the untreated source water (43) (Figure 5) (44). Studies
317 have demonstrated that NTM exposure and infection occurs in the home (Fig. 5A) (39,
318 45, 46), with household water as a source of exposure. Lande et al. (38), for example,
319 showed genotypic matches between *M. avium* respiratory isolates and isolates from
320 household plumbing. Upstream of premise plumbing, studies have shown the
321 proliferation of NTM in water distribution systems (41) (Fig. 5B). However, the entry
322 point for these organisms into the water distribution system and premise plumbing
323 remains unknown. Further upstream, NTM have been found in water treatment facilities
324 (Fig. 5C). King et al. (42) conducted a survey to obtain information on mycobacteria (as
325 well as other microbial pathogens) in source and treated drinking water collected from
326 drinking water treatment plants (DWTPs) across the United States (Fig. 5C & 5D). *M.*
327 *avium* and *M. intracellulare* were detected in 6 out of 24 source water samples and both
328 samples were detected simultaneously at 4 DWTPs. King et al. also identified 10 out of
329 24 DWTPs that had no mycobacteria detected in source or treated water. The literature
330 indicates that there may in fact be high-risk and low-risk regions with respect to
331 mycobacterial exposure, and that these areas likely correspond to high and low risk areas
332 for disease.

Further research is required to confirm the causal pathway between molybdenum, NTM abundance, and disease prevalence. If molybdenum is a metabolic source for NTM in the environment, it is plausible that the mycobacteria also utilize this trace metal to survive in the host, which may explain higher blood serum concentrations of molybdenum among NTM patients (28). Identification of these factors is critical for the development of prevention strategies to minimize exposure and infection in high-risk regions.

LIMITATIONS

The water-quality data that we extracted from the WQP have their own implicit biases. The sampling locations were not from random or systematically representative locations. Rather they were selected for varying specific purposes. While some sites may have been sampled monthly for years, others were sampled only once for a specific project and some watersheds were not sampled at all. Additionally, data were imputed to some watersheds with missing information. Therefore, we cannot predict how much bias may be influencing the resulting median concentrations at each watershed. Additionally, a lack of data on NTM abundance prevents us from correlating high NTM densities in the source water with high prevalence of disease. There are also limitations inherent to our study population, which has been previously discussed (14). We included a drive-time variable to control for oversampling of patients residing the Front Range communities which are more accessible to NJH. Finally, because we did not have data on body site, we cannot specifically associate these findings with pulmonary disease. While the majority of NTM isolates are pulmonary, these findings may be generalizable to all types

of NTM infection, since environmental exposures likely influence both pulmonary and extrapulmonary NTM infection.

CONCLUSIONS & FUTURE DIRECTIONS

We found that the presence of molybdenum in the source water of Colorado watersheds is consistently associated with increased risk of NTM disease. While this study cannot establish a causal association, numerous factors bolster the validity of our findings. Our results for molybdenum are consistent with reports in the scientific literature. The connection established between molybdenum and *Mycobacterium tuberculosis* offers biological plausibility and elevated blood serum concentrations of molybdenum among NTM patients offer specificity to our finding.

This study opens many new avenues of research for the NTM research community. A water sampling study in Colorado would further support these findings. Future research should also include understanding the dose-response relationship between molybdenum and NTM growth. Importantly, the relationship between molybdenum and NTM growth in the (human) host must also be understood. Answering these questions will not only improve patients' lives, but will also contribute to the development of a prevention plan based on environmental risk factors and substantially decrease the risk of exposure and ultimately disease.

Author Bio: Dr. Ettie Lipner is an Instructor at National Jewish Health and an Adjunct Instructor at the Colorado School of Public Health, in the Department of Epidemiology in Aurora. Dr. Lipner's primary research interests include infectious disease epidemiology and spatial epidemiology.

References

1. Prevots DR, Marras TK. Epidemiology of human pulmonary infection with nontuberculous mycobacteria: a review. Clinics in chest medicine. 2015 Mar;36(1):13-34.
2. Winthrop KL, McNelley E, Kendall B, Marshall-Olson A, Morris C, Cassidy M, et al. Pulmonary nontuberculous mycobacterial disease prevalence and clinical features: an emerging public health disease. American journal of respiratory and critical care medicine. 2010 Oct 1;182(7):977-82.
3. Bottai D, Stinear TP, Supply P, Brosch R. Mycobacterial Pathogenomics and Evolution. Microbiol Spectr. 2014 Feb;2(1):MGM2-0025-2013.
4. Falkinham JO, 3rd. Surrounded by mycobacteria: nontuberculous mycobacteria in the human environment. Journal of applied microbiology. 2009 Aug;107(2):356-67.
5. Strollo SE, Adjemian J, Adjemian MK, Prevots DR. The Burden of Pulmonary Nontuberculous Mycobacterial Disease in the United States. Annals of the American Thoracic Society. 2015 Oct;12(10):1458-64.
6. Gebert MJ, Delgado-Baquerizo M, Oliverio AM, Webster TM, Nichols LM, Honda JR, et al. Ecological Analyses of Mycobacteria in Showerhead Biofilms and Their Relevance to Human Health. MBio. 2018 Oct 30;9(5).
7. Falkinham JO, 3rd, Parker BC, Gruft H. Epidemiology of infection by nontuberculous mycobacteria. I. Geographic distribution in the eastern United States. The American review of respiratory disease. 1980 Jun;121(6):931-7.
8. Adjemian J, Olivier KN, Seitz AE, Falkinham JO, 3rd, Holland SM, Prevots DR. Spatial clusters of nontuberculous mycobacterial lung disease in the United States. American journal of respiratory and critical care medicine. 2012 Sep 15;186(6):553-
9. Adjemian J, Daniel-Wayman S, Ricotta E, Prevots DR. Epidemiology of Nontuberculous Mycobacteriosis. Seminars in respiratory and critical care medicine. 2018 Jun;39(3):325-35.
10. Adjemian J, Olivier KN, Seitz AE, Holland SM, Prevots DR. Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries. American journal of respiratory and critical care medicine. 2012 Apr 15;185(8):881-6.
11. Adjemian J, Olivier KN, Prevots DR. Epidemiology of Pulmonary Nontuberculous Mycobacterial Sputum Positivity in Patients with Cystic Fibrosis in the United States, 2010-2014. Annals of the American Thoracic Society. 2018 Jul;15(7):817-26.
12. Adjemian J, Olivier KN, Prevots DR. Nontuberculous mycobacteria among patients with cystic fibrosis in the United States: screening practices and environmental risk. American journal of respiratory and critical care medicine. 2014 Sep 1;190(5):581-6.
13. Adjemian J, Frankland TB, Daida YG, Honda JR, Olivier KN, Zelazny A, et al. Epidemiology of Nontuberculous Mycobacterial Lung Disease and Tuberculosis, Hawaii, USA. Emerging infectious diseases. 2017 Mar;23(3):439-47.
14. Lipner EM, Knox D, French J, Rudman J, Strong M, Crooks JL. A Geospatial Epidemiologic Analysis of Nontuberculous Mycobacterial Infection: An Ecological Study in Colorado. Annals of the American Thoracic Society. 2017 Jun 08.

15. Chou MP, Clements AC, Thomson RM. A spatial epidemiological analysis of nontuberculous mycobacterial infections in Queensland, Australia. BMC infectious diseases. 2014;14:279.
16. Center for International Earth Science Information Network (CIESIN) CU. U.S. Census Grids (Summary File 1) 2010. In: Center NSDaA, (SEDAC), editors. Palisades NY; 2017.
17. United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), and the Environmental Protection Agency (EPA). The Watershed Boundary Dataset (WBD).
18. US Geological Survey UDoA, National Water Quality Monitoring Council. Water Quality Portal. 2012.
19. Greve A.I. SNE, Van Metre P.C., Wilson J.T. Identification of Water-Quality Trends Using Sediment Cores from Dillon Reservoir, Summit County, Colorado. In: United States of the Interior USGS, editor. Denver, Colorado; 2001.
20. Bivand RS, Keitt T, Rowlingson B. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-4 ed; 2019.
21. Bivand RS, Pebesma E, Gomez-Rubio V. Applied spatial data analysis with R. Springer, NY; 2013.
22. Gelman A & Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models: Cambridge University Press; 2006.
23. Bivand R, Rundel, C., Pebesma, E., Stuetz, R., Hufthammer K.O., Giraaudoux, P., Davis, M., Santilli, S. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). 2019-10-02.
24. Colorado Foundation for Water Education. Citizen's Guide to Where Your Water Comes From. 2005. Available from: https://www.colorado.gov/pacific/sites/default/files/Citizen%27s%20Guide%20to%20Where%20Your%20Water%20Comes%20From.pdf
25. Levillain F, Poquet Y, Mallet L, Mazeres S, Marceau M, Brosch R, et al. Horizontal acquisition of a hypoxia-responsive molybdenum cofactor biosynthesis pathway contributed to Mycobacterium tuberculosis pathoadaptation. PLoS Pathog. 2017 Nov;13(11):e1006752.
26. Williams MJ, Kana BD, Mizrahi V. Functional analysis of molybdopterin biosynthesis in mycobacteria identifies a fused molybdopterin synthase in Mycobacterium tuberculosis. J Bacteriol. 2011 Jan;193(1):98-106.
27. McGuire AM, Weiner B, Park ST, Wapinski I, Raman S, Dolganov G, et al. Comparative analysis of Mycobacterium and related Actinomycetes yields insight into the evolution of Mycobacterium tuberculosis pathogenesis. BMC Genomics. 2012 Mar 28;13:120.
28. Oh J, Shin SH, Choi R, Kim S, Park HD, Kim SY, et al. Assessment of 7 trace elements in serum of patients with nontuberculous mycobacterial lung disease. J Trace Elem Med Biol. 2019 May;53:84-90.
29. Colorado Geological Survey. Molybdenum. Available from: http://coloradogeologicalsurvey.org/mineral-resources/metallic-minerals/moylbdenum/
30. Smedley PL, Kinniburgh, D.G. Molybdenum in natural waters: A review of occurrence, distributionsand controls. Applied Geochemistry. 2017;84:387-432.

469 31. Hem JD. Study and Interpretation of the Chemical Characteristics of Natural
470 Water. In: Survey USG, editor. Third Edition ed. Alexandria, VA; 1985.
471 32. Kaback DS. Transport of molybdenum in mountainous streams, Colorado.
472 Geochimica et Cosmochimica Acta. 1976;40:581-2.
473 33. Kubota J LE, Allaway WH. The effect of soil moisture content upon the uptake
474 of molybdenum, copper, and cobalt by alsike clover. Soil Science Society of America
475 Proceedings 1963;27:679-83.
476 34. Kirschner RA, Jr., Parker BC, Falkinham JO, 3rd. Epidemiology of infection by
477 nontuberculous mycobacteria. Mycobacterium avium, Mycobacterium
478 intracellulare, and Mycobacterium scrofulaceum in acid, brown-water swamps of
479 the southeastern United States and their association with environmental variables.
480 The American review of respiratory disease. 1992 Feb;145(2 Pt 1):271-5.
481 35. Thorel MF, Falkinham JO, 3rd, Moreau RG. Environmental mycobacteria from
482 alpine and subalpine habitats. FEMS Microbiol Ecol. 2004 Sep 1;49(3):343-7.
483 36. Jenne EA. Trace element sorption by sediments and soils—sites and
484 processes. . In: RF G, editor. Molybdenum in the environment. New York, NYU:
485 Marcel Dekker; 1977.
486 37. Falkinham JO, 3rd. Ecology of nontuberculous mycobacteria--where do
487 human infections come from? Seminars in respiratory and critical care medicine.
488 2013 Feb;34(1):95-102.
489 38. Lande L, Alexander DC, Wallace RJ, Jr., Kwait R, Iakhiaeva E, Williams M, et al.
490 Mycobacterium avium in Community and Household Water, Suburban Philadelphia,
491 Pennsylvania, USA, 2010-2012. Emerging infectious diseases. 2019 Mar;25(3):473-
492 81.
493 39. Falkinham JO, 3rd. Nontuberculous mycobacteria from household plumbing
494 of patients with nontuberculous mycobacteria disease. Emerging infectious
495 diseases. 2011 Mar;17(3):419-24.
496 40. Honda JR, Hasan NA, Davidson RM, Williams MD, Epperson LE, Reynolds PR,
497 et al. Environmental Nontuberculous Mycobacteria in the Hawaiian Islands. PLoS
498 Negl Trop Dis. 2016 Oct;10(10):e0005068.
499 41. Falkinham JO, 3rd, Norton CD, LeChevallier MW. Factors influencing numbers
500 of Mycobacterium avium, Mycobacterium intracellulare, and other Mycobacteria in
501 drinking water distribution systems. Applied and environmental microbiology. 2001
502 Mar;67(3):1225-31.
503 42. King DN, Donohue MJ, Vesper SJ, Villegas EN, Ware MW, Vogel ME, et al.
504 Microbial pathogens in source and treated waters from drinking water treatment
505 plants in the United States and implications for human health. Sci Total Environ.
506 2016 Aug 15;562:987-95.
507 43. Falkinham JO, 3rd. Environmental sources of nontuberculous mycobacteria.
508 Clinics in chest medicine. 2015 Mar;36(1):35-41.
509 44. City of Portland Oregon. Portland Water Bureau, Bull Run Treatment
510 Projects.
511 45. Nishiuchi Y, Maekura R, Kitada S, Tamaru A, Taguri T, Kira Y, et al. The
512 recovery of Mycobacterium avium-intracellulare complex (MAC) from the
513 residential bathrooms of patients with pulmonary MAC. Clin Infect Dis. 2007 Aug
514 1;45(3):347-51.

46. Thomson R, Tolson C, Carter R, Coulter C, Huygens F, Hargreaves M. Isolation of nontuberculous mycobacteria (NTM) from household water and shower aerosols in patients with pulmonary disease caused by NTM. J Clin Microbiol. 2013 Sep;51(9):3006-11.

Tables

Table 1. Median and Standard Deviation (SD) values of water quality variables* obtained from the Water Quality Portal (WQP) used in PCA.

Exposure Characteristics	**Median ± SD** (µg/L)
Aluminum	18 ± 4377.4
Arsenic	<0.5 ± 49.9
Cadmium	0.1 ± 47.3
Calcium	31801 ± 70685.9
Chloride	2080 ± 217460.3
Copper	1.6 ± 480.6
Iron	38 ± 26245.6
Lead	<0.5 ± 326.4
Magnesium	6600 ± 40730.3
Manganese	22.6 ± 7406.7
Molybdenum	4.3 ± 18.8
Nickel	1.2 ± 37.2
Potassium	1317 ± 6853.5
Selenium	0.06 ± 48.0
Sodium	5930 ± 123209.6
Sulfate	17000 ± 595293.9
Zinc	17 ± 5951.9

*The filtered portion of the water-sample fractions were used.

529 Table 2.

530 Model 1: Poisson regression model examining principal components and other water-

531 quality characteristics associated with NTM disease risk.

532 Bolded estimates are statistically significant.

Characteristics	Relative Risk 95% CI (p-value)
Age: >= 65 years (%)	0.969 0.059, 15.319 (0.983)
Race: Non-White[a]	**0.118** **0.046, 0.298** **($7.1x10^{-6}$)**
Drive time (>2.0 hours to NJH)	**0.634** **0.485, 0.821** **(0.0007)**
Principal Component 1	**1.052** **1.004, 1.104** **(0.038)**
Principal Component 2	1.036 0.973, 1.103 (0.275)
Principal Component 3	**1.089** **1.009, 1.176** **(0.029)**
Principal Component 4	1.003 0.929, 1.084 (0.940)
Principal Component 5	1.039 0.928, 1.167 (0.512)

533 [a]Reference group is White Alone

534

Table 3.

Model 2: Poisson regression model examining individual metals from principal component 3 associated with NTM disease risk.

Bolded estimates are statistically significant.

Characteristics	Relative Risk 95% CI (p-value)
Age: >= 65 years (%)	0.888 0.052, 14.751 (0.745)
Race: Non-White[a]	**0.097** **0.034, 0.274** **($1.7x10^{-5}$)**
Drive-time (>2.0 hours to NJH)	**0.592** **0.458, 0.756** **($3.9x10^{-5}$)**
Arsenic (1 unit)	**0.915** **0.844, 0.992** **(0.031)**
Cadmium (1 unit)	1.003 0.967, 1.040 (0.886)
Manganese (1 unit)	1.077 0.980, 1.186 (0.129)
Molybdenum (1 unit)	**1.217** **1.062, 1.390** **(0.004)**
Selenium (1 unit)	0.996 0.937, 1.060 (0.908)

[a]Reference group is White Alone

Table 4.

Model 3: Poisson regression models examining significant metals associated with NTM disease risk. Bolded estimates are statistically significant.

Characteristics	Relative Risk 95% CI (p-value)	**Characteristics**	Relative Risk 95% CI (p-value)
Age: >= 65 years (%)	1.575 0.105, 22.773 (0.741)	Age: >= 65 years (%)	0.628 0.072, 0.348 (0.745)
Race: Non-White[a]	**0.239** **0.115, 0.491** **(0.0001)**	Race: Non-White[a]	**0.159** **0.072, 0.348** **($4.9x10^{-6}$)**
Drive-time (>2.0 hours to NJH)	**0.583** **0.456, 0.735** **($8.9x10^{-6}$)**	Drive-time (>2.0 hours to NJH)	**0.569** **0.445, 0.719** **($4.0x10^{-6}$)**
Arsenic (1 unit)	0.953 0.894, 1.015 (0.138)	Molybdenum (1 unit)	**1.164** **1.049, 1.291** **(0.004)**

[a]Reference group is White Alone

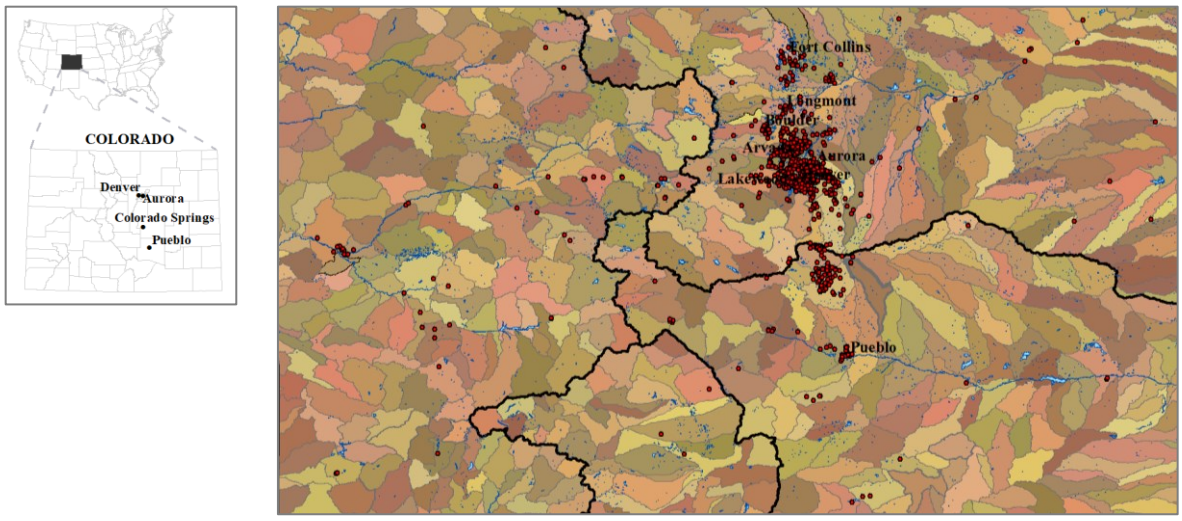


Figure 1. National Jewish Health NTM patient distribution across watersheds in Colorado. *Black lines* represent watershed boundaries of four major watersheds (HUC-2) in Colorado. These four major watersheds are divided into 575 HUC-10 level watersheds (boundaries delineated by gray lines) used in the analyses. *Red dots* indicate patient residence location. *Blue areas* indicate lakes, rivers and reservoirs. The smaller map indicates the location of Colorado within the continental United States.

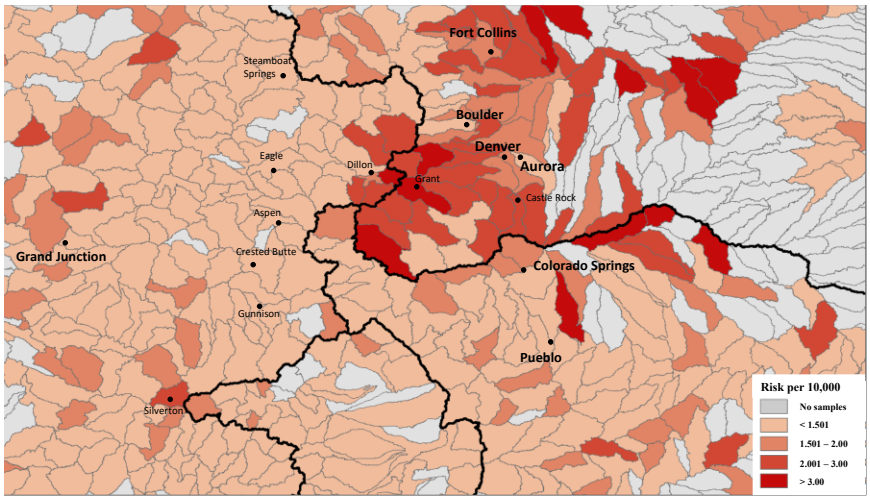


Figure 2. Fitted NTM disease risk estimates per watershed (HUC-10) based on PCA regression model (Model1; Table 2). *Black lines* represent watershed boundaries of four major watersheds (HUC-2) in Colorado. These four major watersheds are divided into 575 HUC-10 level watersheds (boundaries delineated by gray lines). City names are printed in boldface type, town names are printed in smaller font.

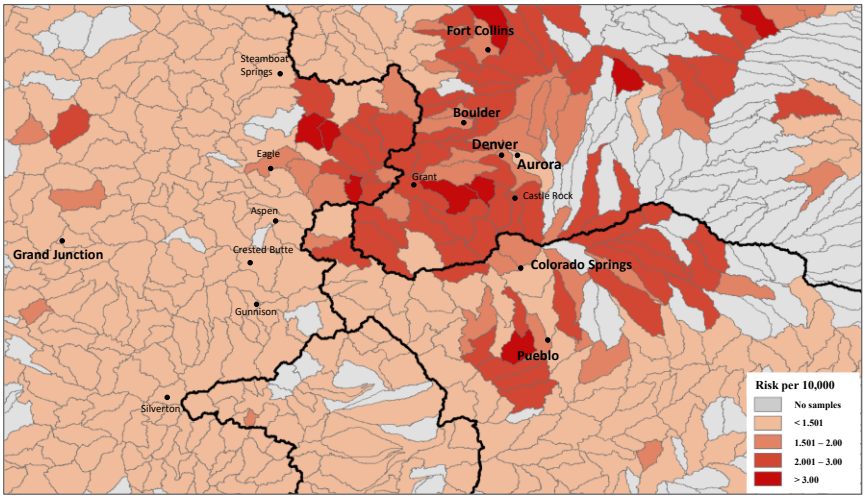
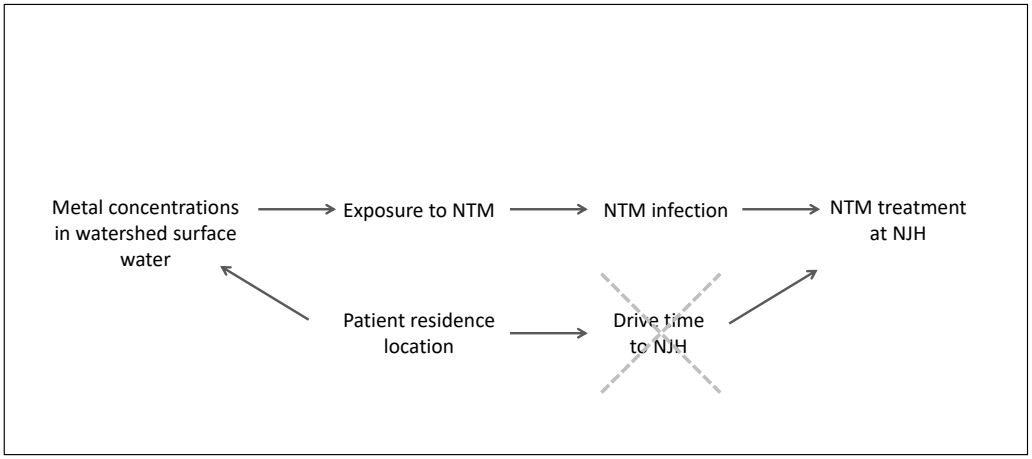


Figure 3. Fitted NTM disease risk estimates per watershed (HUC-10) based on molybdenum regression model (Model 3; Table 4). *Black lines* represent watershed boundaries of four major watersheds (HUC-2) in Colorado. These four major watersheds are divided into 575 HUC-10 level watersheds (boundaries delineated by gray lines). City names are printed in boldface type, town names are printed in smaller font.



608
609 Figure 4. Directed acyclic graph to explore how “drive-time” influences disease risk in
610 our study. Controlling for drive time blocks the non-causal backdoor path between
611 metal concentrations in watershed of residence and NTM treatment at NJH.

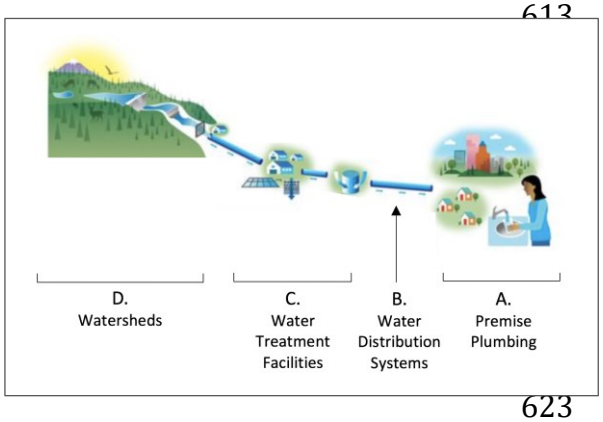


Figure 5. Locations of potential exposure to NTM.

Supplementary Table 1. NTM species from patient isolates.
"M." is an abbreviation for "Mycobacterium".

NTM species	
M. ABSCESSUS	M. MALMOENSE
M. ALVEI	M. MARSEILLENSE
M. AROSIENSE	M. MASSILIENSE
M. ARUPENSE	M. MUCOGENICUM
M. AVIUM	M. NEBRASKENSE
M. AVIUM COMPLEX	M. NEOAURUM
M. BOLLETII	M. NONCHROMOGENICUM
M. BOUCHEDURHONENSE	M. PARAFFINICUM
M. BRISBANENSE	M. PARASCROFULACEUM
M. CELATUM	M. PEREGRINUM
M. CHELONAE	M. PHOCAICUM
M. CHIMAERA	M. PORCINUM
M. COLOMBIENSE	M. RHODESIAE
M. CONCEPTIONENSE	M. SCROFULACEUM
M. FLAVESCENS	M. SENEGALENSE
M. FORTUITUM	M. SEPTICUM
M. GADIUM	M. SHIMOIDEI
M. GENAVENSE	M. SIMIAE
M. GOODII	M. SMEGMATIS
M. GORDONAE	M. SZULGAI
M. INTERJECTUM	M. TIMONENSE
M. INTRACELLULARE	M. TRIPLEX
M. KANSASII	M. VULNERIS
M. KUMAMOTONENSE	M. WOLINSKYI

M. KYORINENSE	M. XENOPI
M. LENTIFLAVUM	M. YONGONENSE
M. MAGERITENSE	

628

629

630 Supplementary Table 2. Water-quality characteristics extracted from the Water Quality
631 Portal (WQP).

Major Metals	**Minor Metals**	**Major Nonmetals**	**Minor Nonmetals**	**Other water-quality characteristics**
Sodium	*Aluminum*	Alkalinity	Antimony	pH
Potassium	Barium	Bicarbonate	*Arsenic*	Specific conductance
Magnesium	Beryllium	Bromide	Boron	Stream flow
Calcium	*Cadmium*	Carbon dioxide	*Selenium*	Total coliform
	Cerium	Carbon		
	Chromium	Carbonate		
	Chromium(III)	*Chloride*		
	Chromium(VI)	Fluoride		
	Cobalt	Hydrogen ion		
	Copper	Inorganic carbon		
	Dysprosium	Oxygen		
	Erbium	Silica		
	Europium	Silicon		
	Gadolinium	*Sulfate*		
	Iron			
	Lanthanum			
	Lead			
	Lithium			
	Manganese			
	Mercury			
	Molybdenum			
	Neodymium			
	Nickel			
	Praseodymium			
	Rhenium			
	Rubidium			
	Samarium			
	Scandium			
	Silver			
	Strontium			
	Thallium			
	Titanium			
	Vanadium			
	Ytterbium			
	Yttrium			
	Zinc			

632 *Italicized variables were included in Principal Component Analysis.

633 Supplementary Table 3. Percent contribution of each metal and nonmetal to each
634 principal component.

Characteristic	Principal Component 1 (%)	Principal Component 2 (%)	Principal Component 3 (%)	Principal Component 4 (%)	Principal Component 5 (%)
Calcium	12.9	1.19	0.437	0.0186	0.0246
Magnesium	13.6	0.702	0.00925	0.227	0.509
Potassium	12.1	0.228	0.487	2.96	0.0789
Sodium	13.4	0.417	0.131	2.05	0.879
Chloride	12.3	2.25	0.307	0.0139	0.612
Sulfate	13.6	0.596	0.872	0.0937	0.00168
Aluminum	0.0479	18.0	0.827	0.744	0.446
Cadmium	1.26	6.86	10.0	0.426	36.6
Copper	0.650	14.2	2.15	0.0657	2.73
Iron	2.77	8.41	0.789	0.0570	33.9
Lead	0.254	12.3	0.00283	27.6	1.57
Manganese	1.86	2.37	35.9	1.83	15.6
Molybdenum	6.21	0.338	13.7	10.9	0.00264
Nickel	1.71	2.84	2.40	37.9	2.92
Zinc	0.278	13.2	3.82	11.9	0.0569
Arsenic	1.83	11.3	16.2	2.64	0.935
Selenium	5.15	4.76	11.8	0.505	3.01

635
636

637 Supplementary Table 4. Sensitivity analysis for Poisson regression model examining the
638 top 3 principal components. Bolded estimates are statistically significant.
639

Characteristics	Relative Risk 95% CI (p-value)
Age: >= 65 years (%)	1.24 0.075, 19.25 (0.881)
Race: Non-White[a]	**0.136** **0.056, 0.323** **($9.0x10^{-6}$)**
Drive time (>2.0 hours to NJH)	**0.606** **0.468, 0.774** **($8.9x10^{-5}$)**
Principal Component 1	**1.052** **1.004, 1.104** **(0.037)**
Principal Component 2	1.028 0.969, 1.091 (0.355)
Principal Component 3	**1.076** **1.005, 1.152** **(0.035)**

640

641 Supplementary Table 5. Correlation matrix for the five most highly contributing metals to
642 Principal Component 3.

	Arsenic	Cadmium	Manganese	Molybdenum	Selenium
Arsenic	1.000				
Cadmium	0.230	1.000			
Manganese	0.055	0.316	1.000		
Molybdenum	0.397	0.187	0.025	1.000	
Selenium	0.576	0.174	0.123	0.515	1.000

643
644

Supplementary Table 6. Poisson regression model examining principal component 1 and individual metals from principal component 3 associated with NTM disease risk. Bolded estimates are statistically significant ($p<0.05$).

Characteristics	Relative Risk 95% CI (p-value)
Age: >= 65 years (%)	0.836 0.048, 14.147 (0.902)
Race: Non-White*	**0.092** **0.032, 0.263** **($9.9x10^{-5}$)**
Drive-time (>2.0 hours to NJH)	**0.584** **0.451, 0.749** **($3.2x10^{-5}$)**
Principal component 1 (1-log unit)	1.026 0.953, 1.105 (0.499)
Arsenic (1-log unit)	**0.915** **0.844, 0.992** **(0.027)**
Cadmium (1-log unit)	0.999 0.962, 1.038 (0.961)
Manganese (1-log unit)	1.060 0.957, 1.179 (0.273)
Molybdenum (1-log unit)	**1.195** **1.032, 1.379** **(0.016)**
Selenium (1-log unit)	0.989 0.926, 1.056 (0.733)

650 Supplementary Table 7. Sensitivity analysis examining the "Drive-time" variable[a].
651 Bolded estimates are statistically significant (p<0.05).

Drive-time (>1.5 hours to NJH)	Relative Risk 95% CI (p-value)	**Drive-time (>2.0 hours to NJH)**	Relative Risk 95% CI (p-value)	**Drive-time (>2.5 hours to NJH)**	Relative Risk 95% CI (p-value)
Arsenic (1-log unit)	0.924 0.853, 1.001 (0.053)	Arsenic (1-log unit)	**0.916 0.845, 0.993 (0.033)**	Arsenic (1-log unit)	**0.899 0.828, 0.975 (0.011)**
Cadmium (1-log unit)	1.004 0.967, 1.042 (0.848)	Cadmium (1-log unit)	1.001 0.965, 1.039 (0.942)	Cadmium (1-log unit)	1.002 0.966, 1.039 (0.919)
Manganese (1-log unit)	1.086 0.986, 1.199 (0.099)	Manganese (1-log unit)	1.076 0.979, 1.186 (0.134)	Manganese (1-log unit)	**1.103 1.005, 1.214 (0.041)**
Molybdenum (1-log unit)	**1.205 1.049, 1.383 (0.008)**	Molybdenum (1-log unit)	**1.220 1.065, 1.394 (0.004)**	Molybdenum (1-log unit)	**1.183 1.033, 1.351 (0.014)**
Selenium (1-log unit)	0.986 0.927, 1.049 (0.659)	Selenium (1-log unit)	0.996 0.936, 1.060 (0.892)	Selenium (1-log unit)	1.011 0.949, 1.077 (0.739)

652 Each model is also controlled for age and race.
653
654
655
656
657
658
659 Supplementary Table 8. Sensitivity analysis examining the "Drive-time" variable[a].
660 Bolded estimates are statistically significant.

Drive-time (>2.5 hours to NJH)	Relative Risk 95% CI (p-value)	**Drive-time (>2.5 hours to NJH)**	Relative Risk 95% CI (p-value)	**Drive-time (>2.5 hours to NJH)**	Relative Risk 95% CI (p-value)
Arsenic (1-log unit)	0.942 0.883, 1.003 (0.064)	Manganese (1-log unit)	**1.132 1.041, 1.232 (0.004)**	Molybdenum (1-log unit)	**1.138 1.028, 1.261 (0.013)**

661 Each model is also controlled for age and race.
662
663