Article

Nontuberculous Mycobacterial Disease and Molybdenum in Colorado Watersheds

Ettie M. Lipner [1,2,*], **Joshua French** [3], **Carleton R. Bern** [4], **Katherine Walton-Day** [4],
David Knox [5], **Michael Strong** [1], **D. Rebecca Prevots** [6,†] **and James L. Crooks** [1,2,†]

[1] National Jewish Health, Denver, CO 80206, USA; strongm@njhealth.org (M.S.); crooksj@njhealth.org (J.L.C.)
[2] Department of Epidemiology, Colorado School of Public Health, Aurora, CO 80045, USA
[3] Department of Mathematical and Statistical Sciences, University of Colorado Denver,
    Denver, CO 80204, USA; joshua.french@ucdenver.edu
[4] U.S. Geological Survey, Colorado Water Science Center, Denver, CO 80225, USA; cbern@usgs.gov (C.R.B.);
    kwaltond@usgs.gov (K.W.-D.)
[5] Department of Computer Science, University of Colorado-Boulder, Boulder, CO 80309, USA;
    david.knox@colorado.edu
[6] National Institute of Allergy and Infectious Diseases, National Institutes of Health,
    Bethesda, MD 20814, USA; rprevots@niaid.nih.gov
*   Correspondence: lipnere@njhealth.org
†   D. Rebecca Prevots and James L. Crooks are co-senior authors.

check for
updates

Abstract: Nontuberculous mycobacteria (NTM) are environmental bacteria that may cause chronic lung disease. Environmental factors that favor NTM growth likely increase the risk of NTM exposure within specific environments. We aimed to identify water-quality constituents (Al, As, Cd, Ca, Cu, Fe, Pb, Mg, Mn, Mo, Ni, K, Se, Na, Zn, and pH) 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 constituents associated with disease risk were identified using generalized linear models with Poisson-distributed discrete responses. We observed a highly robust association between molybdenum (Mo) 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 17.0%. We also observed a statistically significant association between calcium (Ca) in the source water and disease risk. The risk of NTM varied by watershed and was associated with watershed-specific water-quality constituents. These findings may inform mitigation strategies to decrease the overall risk of exposure.

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

1. 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 natural environmental reservoirs, primarily soil and water. Environmental conditions related to soil properties, natural source 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]. Although exposure to NTM is extremely common and the NTM 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

Int. J. Environ. Res. Public Health **2020**, 17, 3854

2 of 15

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 given 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 [5,10]. 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 following initial cure. The lack of evidence-based guidance on environmental risk factors is a critical public health data gap for at-risk populations.

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 examine why we observed higher disease risk in these areas. The aim of this study was to assess whether watershed water-quality constituents are associated with increased risk of NTM disease in Colorado. 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 with NTM data from patients residing in the State of Colorado and treated at National Jewish Health (NJH), a leading respiratory hospital in Denver.

2. Methods

2.1. Data Collection

Patient data were obtained from the NJH Electronic Medical Record database. The study population was comprised of all patients with a diagnosis of NTM treated at NJH and who were resident in Colorado (Figure 1) during the study period, from February 2008 through January 2018. Patient address, NTM species, and patient demographic information were extracted from the database. Body site isolation data were not available, therefore NTM disease included pulmonary and extra-pulmonary. This study was approved by the NJH Institutional Review Board (HS-3148). Figure 1 shows the location of the state of Colorado within the continental United States.

Figure 1. Location of Colorado within the continental United States.

2.1.1. NTM Species

NTM species from patient isolates are listed in Supplementary Table S1.

2.1.2. Socio-Demographic Data

Gridded population density datasets of total population as well as age and racial/ethnic categories were obtained for Colorado during 2010 from the Socio-economic Data and Applications Center (SEDAC) by the Center for International Earth Science Information Network (CIESIN), Columbia University [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. Because Colorado has a majority white population, for this analysis we created two categories: "White Alone" versus all other race/ethnic groups, termed "Non-White". Population density, age and racial/ethnic categories per watershed were calculated using ArcGIS 10.2 (ESRI, Boston, MA, USA).

2.1.3. Environmental Exposure Data

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

2.1.4. Water-Quality Data Compilation

We obtained water-quality constituent 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 (EPA) and the National Water Quality Monitoring Council (NWQMC). We accessed the WQP on 8 August 2019 and extracted surface-water data. The cleaned dataset used in this analysis required correcting identified unit errors, incorrect latitude or longitude for specific sampling locations, as well as identifying and excluding particular locations that were not surface-water sampling sites. Appendix S1 details the data cleaning steps that were taken to go from the raw data to the cleaned dataset used in the analysis. The cleaned dataset included 62 water-quality constituents from 1,110,882 total samples collected from 7171 unique sampling sites in Colorado during 1 January 2000 through 31 December 2018. We examined the filtered (dissolved) water- sample fractions (filtered indicates that water was passed through a 0.45 -micrometer filter [19]) (Supplementary Table S2). Table 1 presents the median and standard deviation values of the water-quality constituents obtained from the WQP that were used in our analyses.

2.2. Statistical Analysis

Analysis of data was performed using the R packages, "rgdal" [20], "sp" [21], "arm" [22], and "dplyr" [23]. All surface-water sampling sites were aggregated by watershed using the R package, "sp", and the median value of each water-quality constituent was calculated for each watershed using the R package, "dplyr". The R source code that we created to calculate the watershed medians is available in the Supplementary Materials. Apparent concentration-unit reporting errors were corrected (for example, three orders of magnitude deviations were multiplied by 1000 to align them with the range of the remaining source-specific data). Water-quality constituents were eliminated if data were not available for more than 50 percent of the target watersheds. Following these curation steps, 16 water-quality constituents remained for analysis. We used a natural log transformation of all watershed median variables with a highest to lowest value ratio greater than 3 across watersheds (16 variables). For watersheds with missing data (i.e., watersheds with no data collected for a particular constituent), we imputed the median value using the "imputePCA" function in the "missMDA" R package (version 1.14) [24]. We excluded watersheds in which there were no water samples for any of the water-quality constituents. As a result, we conducted our analyses on 417 out of 575 HUC-10

level watersheds. Drive times between watershed centroids and NJH were calculated using the R "rgeos" package [25]. We categorized watersheds based on whether their centroid center was within a 2.0-h drive to NJH. To account for multiple comparisons, we applied the Bonferroni method to assess statistical significance (*p*-value = 0.007).

Table 1. Median and standard deviation (SD) values of water-quality constituents * obtained from the Water Quality Portal (WQP) used in principal component analysis (PCA). (µg/L = micrograms per liter).

Exposure Characteristics	Median ± SD (µg/L)
Aluminum	18 ± 4371.6
Arsenic	<0.5 ± 49.9
Cadmium	0.1 ± 50.6
Calcium	32,110 ± 70,745.7
Copper	1.6 ± 440.8
Iron	38 ± 26,245.6
Lead	<0.5 ± 326.4
Magnesium	6691 ± 40,822.9
Manganese	22.6 ± 7406.7
Molybdenum	4.3 ± 18.8
Nickel	1.2 ± 37.2
pH	7.93± 0.77
Potassium	1347 ± 6884.6
Selenium	0.06 ± 48.0
Sodium	6100 ± 123,203.3
Zinc	17 ± 5951.9

* The filtered (dissolved) portion of the water sample fractions were used.

2.2.1. Principal Component Analysis (PCA)

PCA was performed on the HUC-10 level dataset (after data were log transformed and imputed) including 16 water-quality constituents using the "PCA" function in the R package, "FactoMineR" (version 1.42) [26]. We retained the top three principal components, which explained 64.6% of the data variability, for further analysis (Supplementary Table S3).

2.2.2. Poisson Regression Models

Five Poisson regression models were constructed to model disease risk as a function of water-quality constituents. Our models used the standard log link function and included the log of the population density within each watershed for the state of Colorado as an offset term to account for the differing population densities in each region. NTM case counts were aggregated by watershed. Age, race/ethnicity, and drive time variables were included in all models to control for potential confounding associated with a hospitable-based population. Relative risk, 95% confidence interval (CI), and a *p*-value are presented for each model variable. ($p < 0.05$). To create Figures 3 & 4, we used the best-fit estimates of the watershed-specific risks from the Poisson models.

Principal Component Regression Model

The first Poisson regression model included the top three principal components from the PCA (Table 2; Model 1). Using the analysis of deviance test, the model containing the top three components demonstrated the best improvement in fit.

Poisson Regression Models with Individual Metals

From Supplementary Table S3, we identified the individual metals that contributed more than 3% to each statistically significant principal component (components 1 and 3). Principal component 2 was not statistically significant and was therefore not further explored in this analysis. The identified

metals from principal components 1 and 3 were added as predictor variables to Poisson Models 2 and 4. We then constructed separate single-exposure regression models (Models 3 and 5) for the metals that demonstrated statistical significance from Models 2 and 4 ($p < 0.05$).

Sensitivity Analyses

We performed two sensitivity analyses of our final results. To investigate how our estimates changed when we relaxed our distributional modeling assumption, we used a negative binomial response distribution instead of Poisson (Supplementary Table S5).

3. 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. Eight hundred and seven patients were available for analysis after accounting for dropped watersheds.

3.1. Principal Component Regression Model

Our findings showed that principal components 1 and 3 were significantly associated with disease risk (Table 2; Model 1).

Table 2. Model 1: Poisson regression model examining principal components and other covariates associated with NTM disease risk. Bolded estimates are statistically significant.

Characteristics	Relative Risk (95% CI) *p*-Value
Age: ≥65 years (%)	0.969 (0.059, 15.319) 0.813
Race: Non-White [a]	**0.118** **(0.046, 0.298)** **1.57×10^{-6}**
Drive Time (>2.0 h to NJH)	**0.634** **(0.485, 0.821)** **1.14×10^{-5}**
Principal Component 1	**1.054** **(1.0007, 1.111)** **0.049**
Principal Component 2	1.035 (0.979, 1.094) 0.223
Principal Component 3	**1.083** **(1.009, 1.161)** **0.026**

[a] Reference group is White Alone.

For every 1- unit increase in source water constituents contributing to principal component 1, there was a 5.4% increase in disease risk (Table 2; Model 1); the highest contributing water-quality constituents included arsenic, calcium, magnesium, molybdenum, potassium, selenium, sodium, and pH (Supplementary Table S3). For every 1- unit increase in source water constituents contributing to principal component 3, there was an 8.3% increase in disease risk; the highest contributing water-quality constituents included aluminum, arsenic, cadmium, manganese, molybdenum, selenium, zinc, and pH.

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 up-to-date gridded socio-economic data were not available for inclusion.

We created a "drive- time" variable to control for oversampling of patients residing in 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 2).
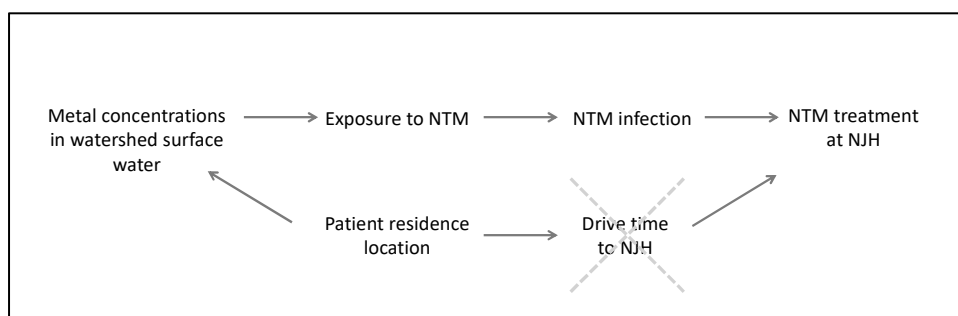

Figure 2. Directed acyclic graph to explore how "drive time" influences disease risk in our study. Controlling for drive time blocks the non-causal backdoor path between metal concentrations in watershed of residence and NTM treatment at NJH.

3.2. Poisson Regression Models with Individual Metals

We modeled the risk of NTM disease as a function of the variables that contributed at least 3% to principal components 1 and 3 (Models 2–5). This threshold captured 50% of the water-quality constituents that were contributing to each component. We examined the variance-inflation factors for principal component 1 and found that magnesium, calcium, potassium, and sodium were highly correlated. In principal component 3, we detected no collinearity. From principal component 1, we observed a significant association between calcium, magnesium, molybdenum and disease risk, while controlling for the presence of the other water-quality constituents, drive- time, age and race (Table 3; Model 2).

Table 3. Model 2: Poisson regression model examining individual metals from principal component 1 and other covariates associated with NTM disease risk. Bolded estimates are statistically significant.

Characteristics	Relative Risk (95% CI) *p*-Value
Age: ≥65 years (%)	2.16 (0.119, 37.6) 0.599
Race: Non-White [a]	**0.076** **(0.027, 0.211)** **8.3×10^{-7}**
Drive Time (>2.0 h to NJH)	**0.513** **(0.393, 0.663)** **5.1×10^{-7}**
pH	0.810 (0.543, 1.22) 0.306

Table 3. *Cont.*

Characteristics	Relative Risk (95% CI) *p*-Value
Arsenic (1-log unit)	0.946 (0.869, 1.03) 0.082
Calcium (1-log unit)	**1.95 (1.36, 2.77) 0.0003**
Magnesium (1-log unit)	**0.625 (0.475, 0.820) 0.0007**
Molybdenum (1-log unit)	**1.22 (1.07, 1.39) 0.0030**
Potassium (1-log unit)	0.988 (0.785, 1.24) 0.918
Selenium (1-log unit)	0.966 (0.906, 1.03) 0.297
Sodium (1-log unit)	1.04 (0.882, 1.23) 0.637

[a] Reference group is White Alone.

We then modeled the risk of NTM disease as a function of each significant metal from Model 2 in a separate single-exposure model (Table 4; Model 3).

Table 4. Model 3: Poisson regression models examining selected metals from Model 2 and other covariates associated with NTM disease risk in three separate single-exposure models. Bolded estimates are statistically significant.

Characteristics	Relative Risk (95% CI) *p*-Value	Characteristics	Relative Risk (95% CI) *p*-Value	Characteristics	Relative Risk (95% CI) *p*-Value
Age: ≥65 years (%)	1.32 (0.08, 20.5) 0.845	Age: ≥65 years (%)	1.42 (0.092, 21.0) 0.801	Age: ≥65 years (%)	0.86 (0.05, 13.7) 0.916
Race: Non-White[a]	**0.14 (0.06, 0.32) 2.4 × 10⁻⁶**	Race: Non-White [a]	**0.20 (0.09, 0.44) 5.9 × 10⁻⁶**	Race: Non-White m [a]	**0.16 (0.07, 0.34) 3.7 × 10⁻⁶**
Drive Time (>2.0 h to NJH)	**0.52 (0.40, 0.66) 1.7 × 10⁻⁷**	Drive Time (>2.0 h to NJH)	**0.55 (0.43, 0.69) 9.2 × 10⁻⁷**	Drive Time (>2.0 h to NJH)	**0.55 (0.43, 0.69) 7.54 × 10⁻⁷**
Calcium (1-log unit)	**1.19 (1.05, 1.35) 0.0055**	Magnesium (1-log unit)	1.06 (0.96, 1.17) 0.232	Molybdenum (1-log unit)	**1.17 (1.05, 1.29) 0.0041**

[a] Reference group is White Alone.

From principal component 3, 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 water-quality constituents, drive- time, age, and race (Table 5; Model 4).

Int. J. Environ. Res. Public Health **2020**, *17*, 3854

8 of 15

Table 5. Model 4: Poisson regression model examining individual metals from principal component 3 and other covariates associated with NTM disease risk. Bolded estimates are statistically significant.

Characteristics	Relative Risk (95% CI) *p*-Value
Age: ≥65 years (%)	1.17 (0.065, 20.1) 0.915
Race: Non-White [a]	**0.09** **(0.03, 0.26)** **1.1 × 10⁻⁵**
Drive Time (>2.0 h to NJH)	**0.57** **(0.44, 0.73)** **1.3 × 10⁻⁵**
pH	0.99 (0.68, 1.45) 0.964
Aluminum (1-log unit)	1.01 (0.96, 1.07) 0.615
Arsenic (1-log unit)	**0.90** **(0.82, 0.99)** **0.022**
Cadmium (1-log unit)	1.00 (0.96, 1.04) 0.917
Manganese (1-log unit)	1.06 (0.96, 1.18) 0.255
Molybdenum (1-log unit)	**1.23** **(1.06, 1.41)** **0.0049**
Selenium (1-log unit)	1.00 (0.94, 1.06) 0.899
Zinc (1-log unit)	1.02 (0.93, 1.13) 0.653

[a] Reference group is White Alone.

We modeled the risk of NTM disease as a function of each significant metal from Model 4 in a separate single-exposure model (Table 6; Model 5).

Molybdenum and calcium remained significantly associated with disease risk (Table 4; Model 3), while magnesium (Table 4; Model 3) and arsenic did not (Table 6; Model 5). 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 increased by 17% (Table 4; Model 3 and Table 6; Model 5). Our results also indicate that for every 1-log unit increase in calcium concentration in the source water at the HUC-10 watershed level, the risk of NTM disease increased by 19% (Table 4; Model 3). The effect of calcium and molybdenum on disease risk remained statistically significant after controlling for multiple comparisons using the Bonferroni method (7 models; new *p*-value = 0.007). Supplementary Table S4 shows the correlation matrix for the metals tested in Models 2 and 4 (Table 3; Table 5).

Table 6. Model 5: Poisson regression models examining selected metals from Model 4 and other covariates associated with NTM disease risk in two separate single-exposure models. Bolded estimates are statistically significant.

Characteristics	Relative Risk (95% CI) *p*-Value	Characteristics	Relative Risk (95% CI) *p*-Value
Age: ≥65 years (%)	2.12 (0.14, 30.4) 0.583	Age: ≥65 years (%)	0.86 (0.051, 13.7) 0.916
Race: Non-White [a]	**0.23 (0.11, 0.48) 8.3 × 10⁻⁵**	Race: Non-White [a]	**0.16 (0.07, 0.34) 3.7 × 10⁻⁶**
Drive Time (>2.0 h to NJH)	**0.56 (0.44, 0.70) 1.4 × 10⁻⁶**	Drive Time (>2.0 h to NJH)	**0.55 (0.43, 0.69) 7.54 × 10⁻⁷**
Arsenic (1-log unit)	0.96 (0.89, 1.01) 0.113	Molybdenum (1-log unit)	**1.17 (1.05, 1.29) 0.0041**

[a] Reference group is White Alone.

The estimated risk estimates across Colorado watersheds based on the regression model with the three principal components (Table 2; Model 1), and the regression model with molybdenum alone (Table 4; Model 3 and Table 6; Model 5) are shown in Figure 3; Figure 4, respectively.
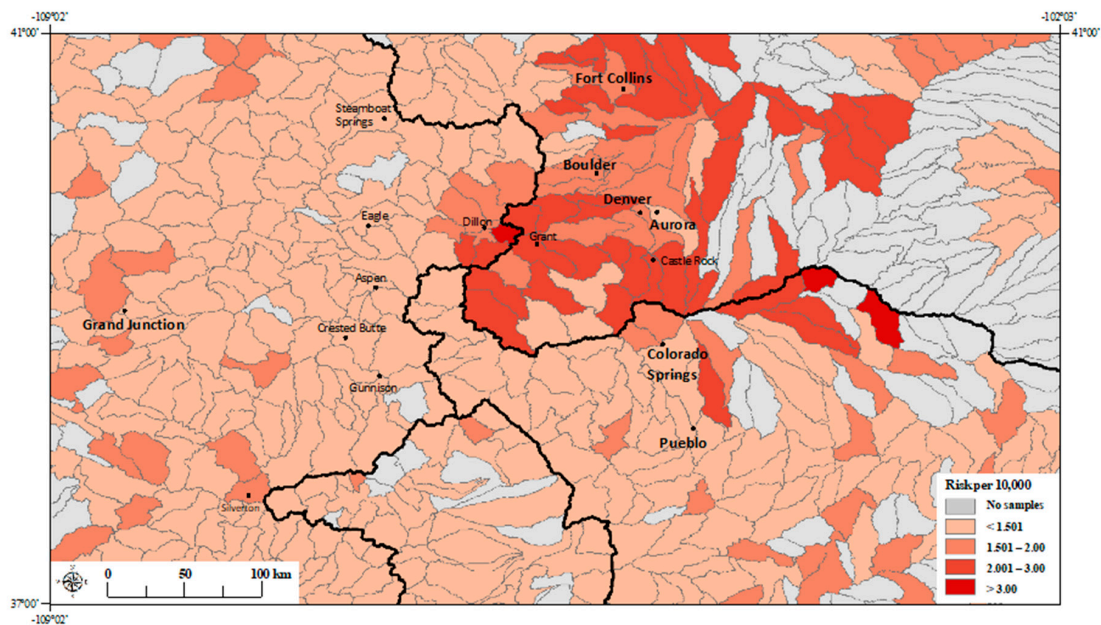

Figure 3. Fitted NTM disease risk estimates per watershed (HUC-10) based on PCA regression model (including principal components 1, 2 and 3) (Model 1; Table 2). *Black lines* represent watershed boundaries of four major watersheds (HUC-2) in Colorado. *Black line* separating the leftmost watershed from the remaining three watersheds represents the Continental Divide. 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.
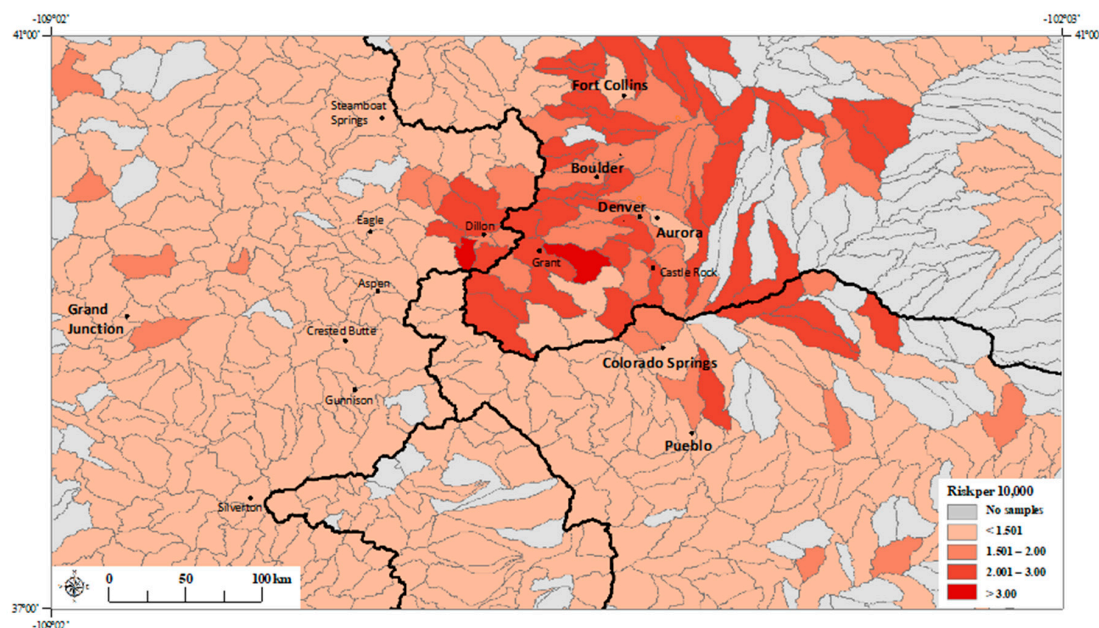
Figure 4. Fitted NTM disease risk estimates per watershed (HUC-10) based on molybdenum regression model (Model 3; Table 4). *Black lines* represent watershed boundaries of four major watersheds (HUC-2) in Colorado. *Black line* separating the leftmost watershed from the remaining three watersheds represents the Continental Divide. 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.

We conducted a sensitivity analysis switching the Poisson response to a negative binomial response (Supplementary Table S5). Using this distribution, we ran separate single-exposure models for each of our final exposures: calcium and molybdenum. The estimated coefficients for calcium and molybdenum remained positive and statistically significant ($p = 0.028$ and 0.024, respectively).

4. Discussion

We found that the presence of calcium and molybdenum in the source water is associated with an increased risk of NTM disease (Tables 3–6; Models 2–5). After removing the non-significant metals from the model, we found that for every one-log unit increase in the calcium and molybdenum concentrations in the source water, a 19% and 17% increase in NTM disease risk was observed, respectively (Table 4; Model 3). 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 (Figure 3; Figure 4). Watersheds in the mountainous regions provide most of the water supply to highly populated communities in the Front Range [27]. Molybdenum is also highly abundant in the mountainous regions of Colorado [28].

The effect of molybdenum on mycobacteria has been described previously [29–31]. Several molybdenum enzymes in mycobacteria exert important physiological functions. *Mycobacterium tuberculosis* as well as the 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 [29]. Molybdenum has been shown to be essential for nitrate assimilation in mycobacteria [30] and is an essential component of many bacterial enzymes involved in carbon, nitrogen, and sulfur metabolism [30]. In *M. tuberculosis*, molybdenum cofactor biosynthesis proteins have been suggested to be associated with pathogenesis [31] and with hypoxic persistence [30] potentially contributing to the ability to shift nitrogen respiration under

the oxygen-limiting concentrations that may occur in lung granulomas. This literature suggests a physiological connection linking molybdenum and essential metabolism, potentially affecting pathogenesis and persistence of *M. tuberculosis*. Although this mechanism has not been established for NTM, it offers biological plausibility because NTM and *M. tuberculosis* are phylogenetically related organisms [4].

Our study opens many avenues of research to investigate the influence of molybdenum on NTM growth in water sources as well as in the human host. In a recent Korean study, Oh et al. [32] 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 from certain dietary profiles (e.g., vitamin supplementation containing molybdenum) increases the risk of infection and/or progression of disease.

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 than 10 µg/L [33]. The environmental concentrations of molybdenum can vary widely, and in places where molybdenum is processed, the concentrations in soil and water may increase considerably [34]. Molybdenum has "relatively high geochemical mobility—a tendency to enter into solution in water under normal Earth-surface conditions" [34,35]; 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 [36]. Likewise, Falkinham and colleagues have repeatedly shown that peat rich soils and brackish marshes are rich in NTM [4,37,38]. In addition, molybdenum can form complexes with organic matter, particularly humic and fulvic acids [39]. Falkinham and colleagues have also reported that humic and fulvic acids support high numbers of *Mycobacterium avium* complex (MAC) species [4,40]. We found that the median value of molybdenum across Colorado watersheds was 4.3 µg/L (Table 1), but reached 325 µg/L at one specific watershed.

Many studies have examined mycobacterial distributions and abundance in different geographic areas. These studies range from examining the presence of NTM in premise plumbing [6,41–43], in the water distribution systems [44], in the water treatment facilities [45] and in the watershed untreated source water [46] (Figure 5) [47].
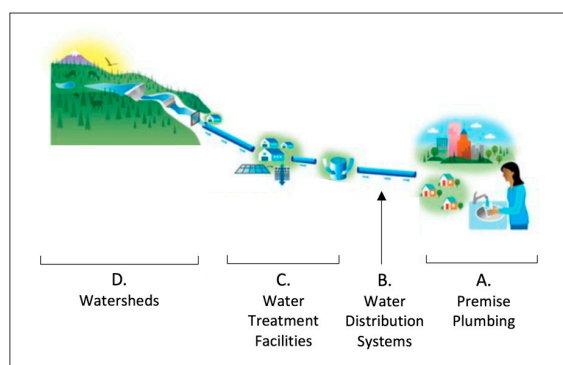

Figure 5. Locations of potential exposure to NTM.

Studies have demonstrated that NTM exposure and infection occurs in the home (Figure 5A) [42,48,49], with household water as a source of exposure. Lande et al. [41], for example, showed genotypic matches between *M. avium* respiratory isolates and isolates from household plumbing. Studies have shown the proliferation of NTM in water distribution systems upstream

from premise plumbing [44] (Figure 5B). However, the entry point for these organisms into the water distribution system and premise plumbing remains unknown. Farther upstream, NTM have been found in water treatment facilities (Figure 5C). King et al. [45] conducted a survey to obtain information on mycobacteria (as well as other microbial pathogens) in source and treated drinking water collected from drinking water treatment plants (DWTPs) across the United States (Figure 5C,D). *M. avium* and *Mycobacterium intracellulare* were detected in 6 out of 24 source water samples and both samples were detected simultaneously at 4 DWTPs. King et al. also identified 10 out of 24 DWTPs that had no mycobacteria detected in source or treated water. The literature indicates that the high-risk and low-risk regions for mycobacterial exposure likely correspond to high and low risk areas for disease.

Although we did not find literature that explain the association of calcium in source water on NTM disease risk, we speculate that for this Colorado dataset the association may be explained in part by the correlation of calcium with molybdenum because of environmental factors, rather than any meaningful link between calcium and NTM bacteria or NTM disease prevalence. Watersheds with greater disease risk and dissolved molybdenum concentrations are more prevalent on the eastern plains of Colorado (Figure 4). Compared to much of Colorado, waters on the eastern plains have generally greater concentrations of major constituents, like calcium, because of evapotranspiration in a warmer and drier climate, and more prevalent sedimentary rocks that weather to release calcium [50]. An association of such rocks with molybdenum release remains to be explored and warrants further research.

Additional research is required to confirm the causal pathway between molybdenum, NTM abundance, and disease prevalence. If molybdenum is a metabolic factor for NTM in the environment, it is plausible that the mycobacteria also utilize this trace metal to survive in the host, a possibility that may explain reports of higher blood-serum molybdenum concentrations in NTM patients [32]. Identification of these factors is critical to develop prevention strategies for minimizing exposure and infection in high-risk regions.

5. 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 targeted for varying purposes. Although 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 in Front Range communities that 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 most NTM isolates are from pulmonary sites, these findings may be generalizable to all types of NTM infection because environmental exposures likely influence both pulmonary and extrapulmonary NTM infection.

6. Conclusions

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 *M. 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 could further support these findings, as well as understanding the

dose–response relationship between molybdenum and NTM growth. Importantly, the relationship between molybdenum and NTM growth in the (human) host warrants better understanding. Answering these questions will not only improve patients' lives, but could also contribute to the development of a prevention plan based on environmental risk factors and substantially decrease the risk of exposure and ultimately disease.

Supplementary Materials: The following are available online at http://www.mdpi.com/1660-4601/17/11/3854/s1, Supplementary Table S1. Nontuberculous mycobacteria (NTM) species from patient isolates. Supplementary Table S2. Water-quality constituents extracted from the Water Quality Portal (WQP). Supplementary Table S3. Percent contribution of each metal and nonmetal to each principal component. Supplementary Table S4. Correlation matrix for the most highly contributing water-quality constituents to Principal Components 1 and 3. Supplementary Table S5. Sensitivity analysis examining the significant metals using negative binomial model. Our R source code is provided to calculate the watershed medians for water-quality constituents ("WatershedMedianCalcs.R"). Appendix S1. Data cleaning documentation, GIS files—watershed boundary datasets (HUC-8 and HUC-10).

Author Contributions: Conceptualization: E.M.L.; Data Curation: E.M.L., C.R.B., K.W.-D., D.K. and J.F.; Formal Analysis: E.M.L.; Funding acquisition: E.M.L., J.F., M.S. and J.L.C.; Investigation: E.M.L., K.W.-D., C.R.B., J.L.C., J.F. and D.R.P.; Methodology: E.M.L., K.W.-D., C.R.B., J.L.C., J.F. and D.R.P.; Software and Modeling: E.M.L., J.F. and J.L.C.; Supervision: J.L.C. and D.R.P.; Validation: E.M.L., K.W.-D., C.R.B. and J.F.; Visualization: E.M.L., C.R.B., J.F. and J.L.C.; Writing—Original Draft: E.M.L., M.S. and C.R.B.; Writing—Review and Editing: E.M.L., K.W.-D., J.L.C., D.R.P., J.F., M.S., C.R.B. and D.K. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Prevots, D.R.; Marras, T.K. Epidemiology of human pulmonary infection with nontuberculous mycobacteria: A review. *Clin. Chest Med.* **2015**, *36*, 13–34. [CrossRef]
2. Winthrop, K.L.; McNelley, E.; Kendall, B.; Marshall-Olson, A.; Morris, C.; Cassidy, M.; Saulson, A.; Hedberg, K. Pulmonary Nontuberculous Mycobacterial Disease Prevalence and Clinical Features. *Am. J. Respir. Crit. Care Med.* **2010**, *182*, 977–982. [CrossRef] [PubMed]
3. Bottai, D.; Stinear, T.P.; Supply, P.J.; Brosch, R. Mycobacterial Pathogenomics and Evolution. *Microbiol. Spectr.* **2014**, *2*, 27–47. [CrossRef] [PubMed]
4. Falkinham, J.O., III. Surrounded by mycobacteria: Nontuberculous mycobacteria in the human environment. *J. Appl. Microbiol.* **2009**, *107*, 356–367.
5. Strollo, S.E.; Adjemian, J.; Adjemian, M.K.; Prevots, D.R. The Burden of Pulmonary Nontuberculous Mycobacterial Disease in the United States. *Ann. Am. Thorac. Soc.* **2015**, *12*, 1458–1464. [CrossRef] [PubMed]
6. Gebert, M.J.; Delgado-Baquerizo, M.; Oliverio, A.M.; Webster, T.M.; Nichols, L.M.; Honda, J.R.; Chan, E.D.; Adjemian, J.; Dunn, R.R.; Fierer, N. Ecological Analyses of Mycobacteria in Showerhead Biofilms and Their Relevance to Human Health. *MBio* **2018**, *9*, e01614-18. [CrossRef]
7. Falkinham, J.O., III; Parker, B.C.; Gruft, H. Epidemiology of infection by nontuberculous mycobacteria. I. Geographic distribution in the eastern United States. *Am. Rev. Respir. Dis.* **1980**, *121*, 931–937. [PubMed]
8. Adjemian, J.; Olivier, K.N.; Seitz, A.E.; Falkinham, J.O., III; Holland, S.M.; Prevots, D.R. Spatial clusters of nontuberculous mycobacterial lung disease in the United States. *Am. J. Respir. Crit. Care Med.* **2012**, *186*, 553–558. [CrossRef]
9. Adjemian, J.; Daniel-Wayman, S.; Ricotta, E.; Prevots, D.R. Epidemiology of Nontuberculous Mycobacteriosis. *Semin. Respir. Crit. Care Med.* **2018**, *39*, 325–335. [PubMed]
10. Adjemian, J.; Olivier, K.N.; Seitz, A.E.; Holland, S.M.; Prevots, D.R. Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries. *Am. J. Respir. Crit. Care Med.* **2012**, *185*, 881–886.

11. Adjemian, J.; Olivier, K.N.; Prevots, D.R. Epidemiology of Pulmonary Nontuberculous Mycobacterial Sputum Positivity in Patients with Cystic Fibrosis in the United States, 2010–2014. *Ann. Am. Thorac. Soc.* **2018**, *15*, 817–826. [CrossRef]

12. Adjemian, J.; Olivier, K.N.; Prevots, D.R. Nontuberculous Mycobacteria among Patients with Cystic Fibrosis in the United States. Screening Practices and Environmental Risk. *Am. J. Respir. Crit. Care Med.* **2014**, *190*, 581–586. [CrossRef] [PubMed]

13. Adjemian, J.; Frankland, T.B.; Daida, Y.G.; Honda, J.R.; Olivier, K.N.; Zelazny, A.; Honda, S.; Prevots, D.R. Epidemiology of Nontuberculous Mycobacterial Lung Disease and Tuberculosis, Hawaii, USA. *Emerg. Infect. Dis.* **2017**, *23*, 439–447. [CrossRef]

14. Lipner, E.; Knox, D.; French, J.; Rudman, J.; Strong, M.; Crooks, J.L. A Geospatial Epidemiologic Analysis of Nontuberculous Mycobacterial Infection: An Ecological Study in Colorado. *Ann. Am. Thorac. Soc.* **2017**, *14*, 1523–1532. [CrossRef]

15. Chou, M.P.; Clements, A.C.; Thomson, R.M. A spatial epidemiological analysis of nontuberculous mycobacterial infections in Queensland, Australia. *BMC Infect. Dis.* **2014**, *14*, 279. [CrossRef] [PubMed]

16. Center for International Earth Science Information Network (CIESIN) Columbia University. *U.S. Census Grids (Summary File 1) 2010*; Center NSDaA, Ed.; SEDAC: Washington, DC, USA, 2017.

17. U.S. Department of Agriculture, Natural Resources Conservation Service (USDA-NRCS) and the Environmental Protection Agency (EPA). The Watershed Boundary Dataset (WBD). Available online: https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/water/watersheds/ (accessed on 12 June 2015).

18. US Geological Survey UDoA, National Water Quality Monitoring Council: Water Quality Portal. 2012. Available online: https://www.waterqualitydata.us/portal/ (accessed on 8 August 2019).

19. Greve, A.I.; Van Metre, P.C.; Wilson, J.T. *Identification of Water-Quality Trends Using Sediment Cores from Dillon Reservoir, Summit County, Colorado*; United States of the Interior USGS: Denver, CO, USA, 2001.

20. Bivand, R.; Keitt, T.; Rowlingson, B. Rdal: Bindings for the 'Geospatial' Data Abstraction Library. R Package Version 1.4-8. 2019. Available online: https://CRAN.R-project.org/package=rgdal (accessed on 25 April 2020).

21. Roger, S.; Bivand, E.P.; Gomez-Rubio, C. *Applied Spatial Data Analysis with R*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2013.

22. Gelman, A.; Su, Y.-S. Data Analysis Using Regression and Multilevel/Hierarchical Models: R package version 1.10-1. 2018. Available online: https://CRAN.R-project.org/package=arm (accessed on 25 April 2020).

23. Wickham, H.; Francois, R.; Henry, L.; Muller, K. Dplyr: A Grammar of Data Manipulation. R package version 0.8.3. 2019. Available online: https://CRAN.R-project.org/package=dplyr (accessed on 25 April 2020).

24. Josse, J.; Husson, F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *J. Stat. Softw.* **2016**, *70*. [CrossRef]

25. Bivand, R.; Rundel, C.; Pebesma, E.; Stuetz, R.; Hufthammer, K.O.; Giraaudoux, P.; Davis, M.; Santilli, S. *Interface to Geometry Engine-Open Source ('GEOS')*; Rgeos: Washington, DC, USA, 2019.

26. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*. [CrossRef]

27. Colorado Foundation for Water Education: Citizen's Guide to Where Your Water Comes From. Denver, CO. 2005. Available online: https://www.colorado.gov/pacific/sites/default/files/Citizen%27s%20Guide%20to%20Where%20Your%20Water%20Comes%20From.pdf (accessed on 25 April 2020).

28. Molybdenum. Available online: http://coloradogeologicalsurvey.org/mineral-resources/metallic-minerals/moylbdenum/ (accessed on 25 April 2020).

29. Levillain, F.; Poquet, Y.; Mallet, L.; Mazeres, S.; Marceau, M.; Brosch, R.; Bange, F.-C.; Supply, P.J.; Magalon, A.; Neyrolles, O. Horizontal acquisition of a hypoxia-responsive molybdenum cofactor biosynthesis pathway contributed to Mycobacterium tuberculosis pathoadaptation. *PLOS Pathog.* **2017**, *13*, e1006752. [CrossRef]

30. Williams, M.J.; Kana, B.D.; Mizrahi, V. Functional Analysis of Molybdopterin Biosynthesis in Mycobacteria Identifies a Fused Molybdopterin Synthase in Mycobacterium tuberculosis. *J. Bacteriol.* **2010**, *193*, 98–106. [CrossRef]

31. McGuire, A.M.; Weiner, B.; Park, S.T.; Wapinski, I.; Raman, S.; Dolganov, G.; Peterson, M.; Riley, R.; Zucker, J.; Abeel, T.; et al. Comparative analysis of Mycobacterium and related Actinomycetes yields insight into the evolution of Mycobacterium tuberculosis pathogenesis. *BMC Genom.* **2012**, *13*, 120. [CrossRef]

32. Oh, J.; Shin, S.H.; Choi, R.; Kim, S.; Park, H.-D.; Kim, S.-Y.; Han, S.A.; Koh, W.-J.; Lee, S.-Y. Assessment of 7 trace elements in serum of patients with nontuberculous mycobacterial lung disease. *J. Trace Elem. Med. Boil.* **2019**, *53*, 84–90. [CrossRef] [PubMed]

33. Smedley, P.L.; Kinniburgh, D.G. Molybdenum in natural waters: A review of occurrence, distributionsand controls. *Appl. Geochem.* **2017**, *84*, 387–432. [CrossRef]

34. Hem, J.D. *Study and Interpretation of the Chemical Characteristics of Natural Water*, 3rd ed.; Survey USG, Ed.; US Geological Survey: Alexandria, VA, USA, 1985.

35. Kaback, D.S. Transport of molybdenum in mountainous streams, Colorado. *Geochim. Cosmochim. Acta* **1976**, *40*, 581–582. [CrossRef]

36. Kubota, J.; Lemon, E.R.; Allaway, W.H. The Effect of Soil Moisture Content Upon the Uptake of Molybdenum, Copper, and Cobalt by Alsike Clover. *Soil Sci. Soc. Am. J.* **1963**, *27*, 679–683. [CrossRef]

37. Kirschner, R.A.; Parker, B.C.; Falkinham, J.O. Epidemiology of Infection by Nontuberculous Mycobacteria: Mycobacterium avium, Mycobacterium intracellulare, and Mycobacterium scrofulaceum in Acid, Brown-Water Swamps of the Southeastern United States and Their Association with Environmental Variables. *Am. Rev. Respir. Dis.* **1992**, *145*, 271–275. [CrossRef]

38. Thorel, M.F.; Falkinham, J.O., III. Moreau RG: Environmental mycobacteria from alpine and subalpine habitats. *FEMS Microbiol. Ecol.* **2004**, *49*, 343–347. [CrossRef] [PubMed]

39. Jenne, E.A. Trace element sorption by sediments and soils—sites and processes. In *Molybdenum in the Environment*; Marcel Dekker: New York, NY, USA, 1977.

40. Falkinham, J.O., III. Ecology of nontuberculous mycobacteria–where do human infections come from? *Semin. Respir. Crit. Care Med.* **2013**, *34*, 95–102. [CrossRef]

41. Lande, L.; Alexander, D.C.; Wallace, R.J.; Kwait, R.; Iakhiaeva, E.; Williams, M.; Cameron, A.D.S.; Olshefsky, S.; Devon, R.; Vasireddy, R.; et al. Mycobacterium avium in Community and Household Water, Suburban Philadelphia, Pennsylvania, USA, 2010–2012. *Emerg. Infect. Dis.* **2019**, *25*, 473–481. [CrossRef]

42. Falkinham, J.O., III. Nontuberculous mycobacteria from household plumbing of patients with nontuberculous mycobacteria disease. *Emerg. Infect. Dis.* **2011**, *17*, 419–424. [CrossRef]

43. Honda, J.R.; Hasan, N.A.; Davidson, R.M.; Williams, M.D.; Epperson, L.E.; Reynolds, P.R.; Smith, T.; Iakhiaeva, E.; Bankowski, M.J.; Wallace, R.J., Jr.; et al. Environmental Nontuberculous Mycobacteria in the Hawaiian Islands. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0005068. [CrossRef]

44. Falkinham, J.O.; Norton, C.D.; Lechevallier, M.W. Factors Influencing Numbers of Mycobacterium avium, Mycobacterium intracellulare, and Other Mycobacteria in Drinking Water Distribution Systems. *Appl. Environ. Microbiol.* **2001**, *67*, 1225–1231. [CrossRef] [PubMed]

45. King, D.N.; Donohue, M.J.; Vesper, S.J.; Villegas, E.N.; Ware, M.W.; Vogel, M.; Furlong, E.F.; Kolpin, D.W.; Glassmeyer, S.T.; Pfaller, S.L. Microbial pathogens in source and treated waters from drinking water treatment plants in the United States and implications for human health. *Sci. Total. Environ.* **2016**, *562*, 987–995. [PubMed]

46. Falkinham, J.O., III. Environmental sources of nontuberculous mycobacteria. *Clin. Chest Med.* **2015**, *36*, 35–41. [CrossRef] [PubMed]

47. *Bull Run Water Treatment*. 2019. Water Quality Report. West Slope Water District. Available online: http://www.wswd.org/document_center/Water%20Quality/2019%20Water%20Quality%20Report.pdf (accessed on 25 April 2020).

48. Nishiuchi, Y.; Maekura, R.; Kitada, S.; Tamaru, A.; Taguri, T.; Kira, Y.; Hiraga, T.; Hirotani, A.; Yoshimura, K.; Miki, M.; et al. The Recovery of Mycobacterium avium-intracellulare Complex (MAC) from the Residential Bathrooms of Patients with Pulmonary MAC. *Clin. Infect. Dis.* **2007**, *45*, 347–351. [CrossRef] [PubMed]

49. 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**, *51*, 3006–3011. [PubMed]

50. Ivahnenko, T.; Ortiz, R.F.; Stogner, R. Characterization of streamflow, water quality, and instantaneous dissolved solids, selenium, and uranium loads in selected reaches of the Arkansas River, southeastern Colorado, 2009–2010. *Sci. Investig. Rep.* **2013**, 5234.