ORIGINAL ARTICLE

An exploration of colorectal cancer incidence rates in North Dakota, USA, via structural equation modeling

Gary G. Schwartz[1] (ID) · Marilyn G. Klug[1] · Bradley C. Rundquist[2]

Abstract

Purpose The state of North Dakota has one of the highest incidence rates for colorectal cancer in the USA. Its high incidence rate, coupled with a large variation in incidence rates among counties within the state, makes North Dakota a "natural laboratory" in which to investigate environmental clues to colorectal cancer. We conducted a hypothesis-generating study to explore potential determinants of colorectal cancer in North Dakota.

Methods We obtained county-specific incidence rates for North Dakota's 53 counties from the statewide cancer registry and corresponding data on county demographic, agricultural, and geophysical features from population-based sources. Candidate demographic/agricultural variables included median household income, population density, colorectal cancer screening rates, average farm size (in acres), and the percent of county fertilized. Geophysical variables included the uranium content of soil, residential radon levels, and source of drinking water (municipal or well water). Statistical analyses were performed via multivariate regression and structural equation modeling.

Results Colorectal cancer incidence rates across North Dakota counties varied 3-fold. The structural equation model identified a significant role for well water use ($p < 0.05$). This finding is consistent with studies that implicate well water in colorectal cancer.

Conclusions Well water contains several agents, e.g., bacteria, disinfection by-products, and nitrates that are potent colorectal carcinogens. Studies of well water use and colorectal cancer risk at the individual level in North Dakota are warranted.

Keywords Colorectal cancer · Epidemiology · Structural equations · North Dakota

Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in the world, with 1.8 million new cases diagnosed annually [1]. The etiology of CRC is incompletely understood. A history of chronic inflammatory bowel disease and of type 2 diabetes are known medical risk factors [2]. Behavioral risk factors for CRC include physical inactivity, obesity, and excessive alcohol consumption. However, many individuals with CRC have no known risk factors. Thus, it is likely that environmental risks for CRC exist that have not been identified [3].

Insights into environmental causes of cancer often come from geographic studies. For example, high rates of mesothelioma on the Eastern seaboard of the USA led to discovery of the association of mesothelioma with shipbuilding and to the identification of asbestos (used in ships' insulation) as the cause [4]. With respect to CRC, there is an unusually high incidence of CRC in the US state of North Dakota (ND). For reasons that are unknown, ND has the second highest CRC rate among white males in the USA and the third highest incidence rate among white females [5] (Fig. 1). We conducted a hypothesis-generating study at the county level in ND in order to explore candidate demographic and geographic determinants of its high CRC incidence rates. We report that county-wide incidence rates for CRC in ND are significantly associated with the use of non-municipal ("well") water.

✉ Gary G. Schwartz
gary.schwartz@und.edu

[1] Department of Population Health, University of North Dakota School of Medicine & Health Sciences, 1301 N. Columbia Road, Grand Forks, ND 58202, USA

[2] Department of Geography, University of North Dakota College of Arts & Sciences, Grand Forks, ND, USA
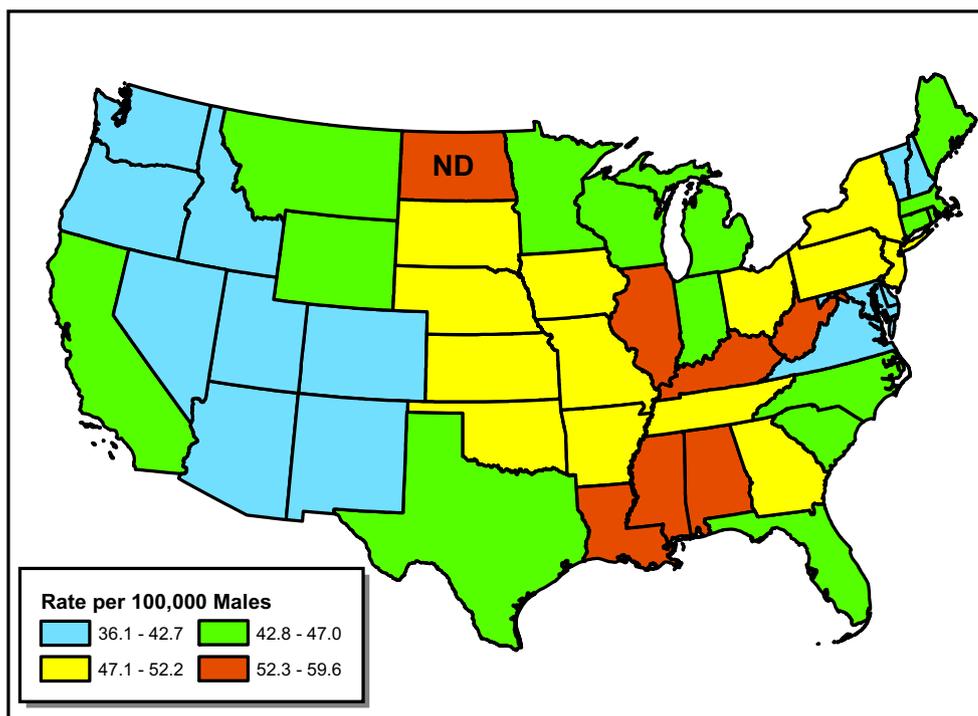
Methods

Data and sources

County-specific incidence rates were obtained from the North
Dakota Statewide Cancer Registry (NDSCR) for the time peri-
od 1997–2014 (i.e., from the inception of the registry to the
most current data available). The NDSCR is certified by the
American Association of Central Cancer Registries from which
it has received awards for data quality and completeness
(https://ndcancer.org/aboutmain.html). All rates were age-
adjusted to the 2000 Census. Because of the small numbers of
non-whites, especially in sparsely-populated counties, the anal-
yses were restricted to white, non-Hispanics. In order to prevent
disclosure of potentially identifying information, the NDSCR
suppresses rates for counties with fewer than 10 cases.

We examined candidate environmental and socioeconomic
factors for which there was support in the literature for a role in
CRC. Demographic variables were median household income,
poverty rate, and population density. Geographic factors includ-
ed the average farm size (in acres); the percent of county fertil-
ized; the average amount of nitrate fertilizer used per farm [6];
and the size of the cattle population per capita [7]. Data on the
prevalence of CRC screening practices (whether individuals
have had a fecal occult blood test in the past 2 years or ever
had a colorectal endoscopy) were obtained from national sur-
veys [8]. Geophysical variables were the uranium content of the
soil, average indoor radon levels, and the source of drinking
water (municipal or self-supplied domestic "well water").

Data on household income, poverty rates and population
density were obtained from the US Census [9]. Data on the
percentage of each county that uses self-supplied water in
2005 (non-municipal water from private wells or cisterns,
henceforth "well water") were obtained from the US
Geological Survey [10]. These data were compiled from mul-
tiple sources, including government records for well permits,
state agencies that regulate utility rates, and bills from waste-
water treatment facilities and tax appraisers' offices [11]. Data
on the uranium contest of soil was from the US Department of
Energy [12] and data on indoor radon levels were from the ND
State Radon Survey, as described previously [13].

Statistical analysis

CRC rates and demographic variables were mapped and ana-
lyzed by county using GIS software (ESRI ArcMap 10.6). We
used multivariate linear regression to identify potential rela-
tionships between the independent variables with age-
adjusted CRC incidence rates. Structural equation models
were then developed using Proc Calis in SAS v 9.4. Briefly,
structural equation modeling is a multivariate statistical anal-
ysis technique that combines factor analysis with multiple
regression analysis. Structural equation models can describe
the interdependence of a set of variables with the goal of
providing a quantitative test of different theoretical models
[14]. Aggregate state data for all counties with available data
were used to estimate covariances (i.e., relationships) between
the independent variables and paths (i.e., prediction
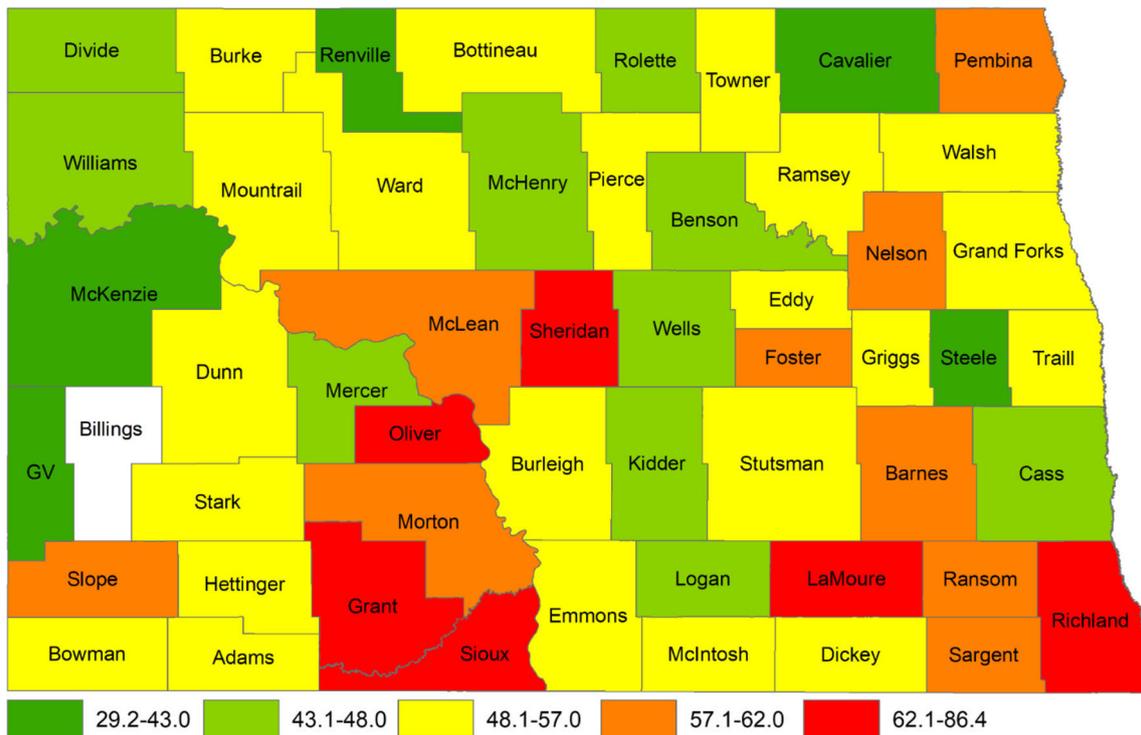
Fig. 2 Colorectal cancer incidence rates (1994–2014), for North Dakota white non-Hispanic males, by county. Data are from the North Dakota Statewide Cancer Registry. The incidence rate was suppressed for Billings County due to the small number of cases

estimations) from the independent variables to the dependent variable. The full model was then reduced by removing co-variances and paths that were not significant.

Results

Incidence data were available for 52 of North Dakota's 53 counties (98%). Data were suppressed for 1 county (Billings) due to small numbers of cases (Fig. 2). Several variables were not normally distributed (nitrate levels, farm size, population density, cattle per capita) and logarithmic or square root transformations were used to normalize these for analyses.

CRC incidence rates varied almost threefold across ND counties, from 29.2 in Renville County to 86.4 per 100,000/year in Oliver County. All of the correlations of the variables with CRC rate were non-significant ($p > 0.05$). When entered into a multiple linear regression, only population density ($p = 0.081$) and well water ($p = 0.067$) significantly predicted CRC rate with p values < 0.10, (using "forward selection" only in the regression model) (Table 1). As none of the variables had a direct relationship with CRC, a path analysis with structural equations was performed in order to determine indirect relationships and correlations.

Population was best modeled as a latent variable or fac-tor represented by income, population density, percent screened, and cattle per person. Farming was also modeled as a latent variable representing cattle per person, nitrate, fertilization, and farmland acres. Farming was correlated with uranium which was also correlated with radon. Farming also was correlated with population size. This in turn was related to well water, which was the only factor found to be directly related to CRC. All causal paths and

Table 1 Multiple linear regression predicting CRC rate with adjusted variables

Variable	B	St. Err	T	p
Intercept	49.56	85.61	0.58	0.566
Nitrate used	−3.63	7.86	−0.46	0.646
% fertilized	0.17	0.17	0.99	0.326
Average farm acres	0.55	18.52	0.03	0.976
Cattle per person	3.36	2.12	1.59	0.121
Median income	0.25	0.41	0.60	0.550
Population density	10.15	6.24	1.79	0.081
Mean radon	0.62	0.77	0.80	0.429
Uranium in soil	−2.76	8.10	−0.34	0.735
Well water	0.68	0.44	1.54	0.131
% screened	−0.21	0.24	−0.86	0.396

Fig. 3 Structural equation model of risk factors for CRC incidence among counties in ND. Measured variables are shown in rectangles, "latent variables" (i.e., variables that best represent a cluster of measured variables) are shown in oval outline. Uni-directional arrows indicate potential causal pathways; bi-directional arrows indicate a co-varying relationship that is unlikely to be causal. Mn = mean
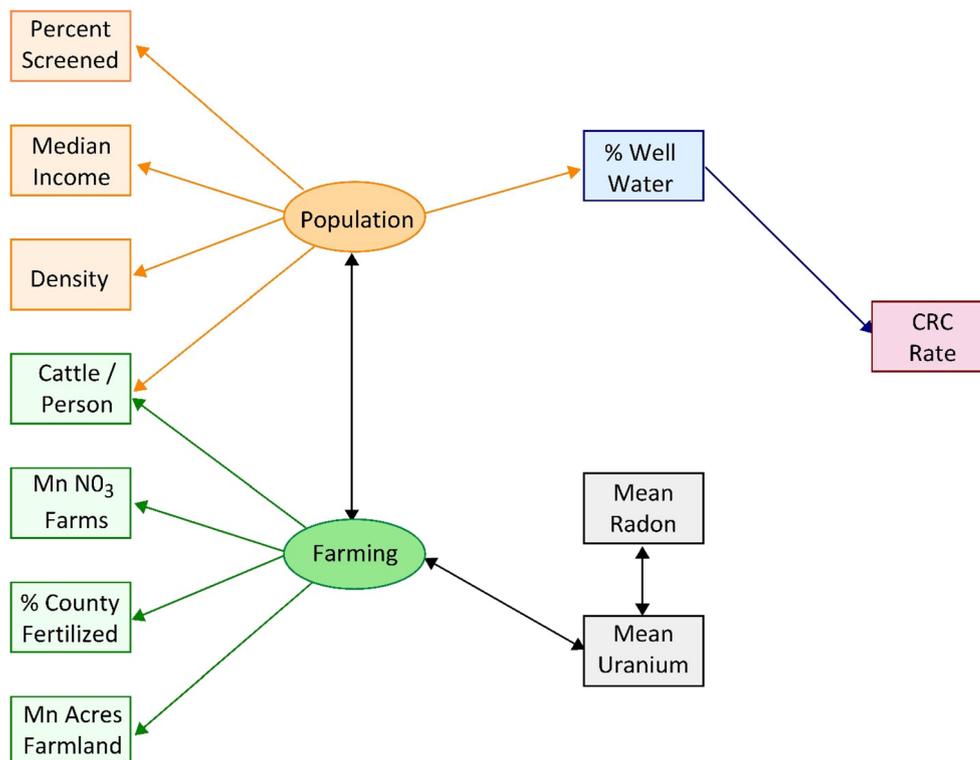
covariances included in the structural equation model were significant at $p < 0.05$ (Fig. 3).

Discussion

ND is the fourth least populous state in the USA and has one of the highest CRC incidence rates of any state. Its rate of 54.4 per 100,000 in males (during 2017–2019) is comparable with some of the highest rates recorded in Europe [15]. Moreover, ND shows an almost three-fold variation in incidence rates among its > 50 counties. Its high incidence rate, plus the large variation in rates within the state, makes it a "natural laboratory" in which to investigate environmental clues to CRC. Our structural equation model identified well water use as a significant predictor of CRC incidence rates. This finding is consistent with numerous investigations in CRC that have implicated well water in CRC risk in the USA and in other parts of the world [16, 17].

ND is a rural state and has a lower prevalence of screening for CRC than the US overall (68% vs. 73%, in 2010) [18]. We considered that the higher incidence rates in some counties might result from low utilization of screening practices. Less rural counties (i.e., those with higher population density) were associated with higher screening rates. However, our analyses did not show a direct association between lower utilization of screening and higher CRC incidence rates.

Residential radon levels are especially high in ND [13]. Our structural equation model confirmed that radon levels in homes

were significantly associated with uranium measurements in soil. However, neither factor was directly related to CRC rates. Although radon in groundwater has been associated with lung and with stomach cancer, to our knowledge, there is little evidence for an association between radon and CRC [19].

Because our study examined several geographic/demographic variables, it is possible that the association with well water could have arisen by chance. However, statistical concerns about multiple comparisons pertain to hypothesis-testing investigations, not to investigations, such as the present one, whose function is hypothesis-generating. Well water was included as a candidate variable because of its frequent association with CRC in the literature and because common contaminants of well water are known CRC carcinogens.

Our exploratory study has several limitations. First, county-wide data on known risk factors such as diabetes, physical inactivity, and alcohol consumption were not available. Thus, we were unable to control for these variables. Secondly, incidence rates were adjusted for age, but due to small case numbers in many counties, were not adjusted for gender. CRC incidence rates in the USA are approximately 30% higher in males; thus, incidence rates could be biased upwards if a county's population is disproportionately male. ND is predominantly an agricultural state, although it recently has become a major oil producing one. The influx of workers to the ND oil fields, which peaked in 2008, could have influenced incidence rates in some Western counties [20]. However, because the migrants were predominantly young

males, and the mean age at diagnosis for colon cancer in men is 68, this influx is unlikely to have greatly influenced our results. Most importantly, this is an ecologic study in which the units of analysis are individual counties. Thus, it demonstrates that counties with high use of well water are counties with increased risk of CRC and cannot make claims about individuals within those counties.

Conversely, our study has several strengths: the CRC data are population-based, and there was little selection bias, as data were available for 52 of ND's 53 counties. The exposure data of greatest interest, well water, is one that likely was determined with high accuracy, as well water use is documented by multiple records and methods (e.g., water permits, tax records, aerial photography) [10, 11]. Our central finding is consistent with numerous studies that implicate well water in the etiology of CRC [16].

Numerous contaminants of well water have been implicated in colorectal carcinogenesis, including water-borne bacteria, disinfection by-products, and nitrates. Bacteria are believed to influence CRC carcinogenesis via dysbiosis (microbial imbalance) and aberrant gene expression [21]. Treatment of water with chlorinated compounds produces disinfection by-products which themselves are known CRC carcinogens [22]. Nitrate is a common pollutant of drinking water in agricultural regions that derives largely from animal manure and the use of nitrogen-containing fertilizers [23]. Nitrates are converted endogenously to nitrite and the subsequent nitrosation reactions produce highly carcinogenic, N-nitroso compounds [24].

In a large population-based study in Denmark, Schullehner et al. reported a significantly increased risk of CRC among individuals exposed to nitrates in drinking water as low as 3.87 mg/L. This finding is noteworthy because 3.87 mg/L is an order of magnitude lower than the Danish drinking water standard for nitrates, 50 mg/L. [25] The nitrate standard for public drinking water in the USA is 10 mg/L (https://www.wqa.org/Portals/0/Technical/Technical%20Fact%20Sheets/2014_NitrateNitrite.pdf). This standard does not apply to private wells, which commonly exceed it. For example, examination of 218 samples of water from private and municipal wells in ND found nitrate levels > 10 mg/L in 22 samples (22/218 = ~ 10%) [26].

In summary, our structural equation modeling of demographic variables in ND, a state with high and variable incidence rates for CRC, identified well water use as a significant predictor of CRC incidence rates at the county level. This finding is biologically plausible as well water is known to contain potent CRC carcinogens. Studies of well water use and CRC risk in ND at the individual level are warranted.

Compliance with ethical standards

References

1. Macrae FA. Colorectal cancer: epidemiology, risk factors and protective factors. UptoDate. https://www.uptodate.com/contents/colorectal-cancer-epidemiology-risk-factors-and-protective-factors

2. Brenner H, Chen C (2018) The colorectal cancer epidemic: challenges and opportunities for primary, secondary and tertiary prevention. Br J Cancer 119:785–792

3. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, Berry DA (2013) Meta-analyses of colorectal cancer risk factors. Cancer Causes Control 24:1207–1222

4. Tagnon I, Blot WJ, Stroube RB, Day NE, Morris LE, Peace BB, Fraumeni JF Jr (1980) Mesothelioma associated with the shipbuilding industry in coastal Virginia. Cancer Res 40:3875–3879

5. American Cancer Society. Colorectal cancer facts & figures 2017–2019. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf

6. County-level estimates of nitrogen and phosphorus from commercial fertilizer for the conterminous U.S., 1987–2006. https://water.usgs.gov/GIS/metadata/usgswrd/XML/sir2012-5207_county_fertilizer.xml

7. U.S. Census of Agriculture, 2012. https://www.agcensus.usda.gov/Publications/2012/

8. CINA+Online Cancer in North America. www.cancer-rates.info/naaccr/

9. U.S. Census, 2010. https://www.census.gov/2010census/

10. Kenny JF, Barber NL, Hutson SS, Linsey KS, Lovelace JK, Maupin MA. Estimated use of water in the United States in 2005. Circular 1344, U.S. Department of the Interior, U.S. Geological Survey, Reston, VA; 2009

11. Hutson SS. Guidelines for preparation of state water use estimates for 2005. U.S. Department of the Interior, U.S. Geological Survey. Available at:http://pubs.usgs.gov/tm/2007/tm4e1/pdf/tm4-e1.pdf. Accessed February 10, 2015. [Google Scholar])

12. Aeromagnetic and aeroradiometric data for the conterminous U.S. and Alaska, National Uranium Resource Evaluation (NURE), U.S. Department of Energy. https://pubs.usgs.gov/of/2009/1129/NURE.html

13. Oancea SC, Rundquist BC, Simon I, Swartz S, Zheng Y, Zhou X et al (2017) County level incidence rates of chronic lymphocytic leukemia are associated with residential radon levels. Future Oncol 21:1873–1881

14. Schumaker RE, Lomax RG (2004) A beginner's guide to structural equation modeling, 2nd edn. Lawrence Erlbaum, Mahwah

15. Center MM, Jehmal A, Smith RA, Ward E (2009) Worldwide variations in colorectal cancer. CA Cancer J Clin 59:366–378

16. Ward MH, Jones RR, Brender JD, de Kod TM, Weyer PJ, Nolan BT et al (2018) Drinking water nitrate and human health: an updated review. Int J Environ Res Pub Health 15:1557

17. Gulis G, Campolyova M, Cerhan JR (2002) An ecologic study of nitrite in municipal drinking water and cancer incidence in Trnava District, Slovakia. Environ Res Sect A 88:182–187

18. Perisetti A, Khan H, Vendala R, George NE, Raghavapuram S, Rafiz A et al (2018) Prevalence of non-screening for colorectal cancer: comparing a rural state to the United States. J Nutrition Health Food Sci 6:1–7

19. Messier KP, Serre ML (2017) Lung and stomach cancer associations with groundwater radon in North Carolina, USA. Int J Epidemiol 46:675–685

20. Taylor A. North Dakota's oil boom. The Atlantic, March 13, 2013. https://www.theatlantic.com/photo/2013/03/north-dakotas-oil-boom/100473/. Accessed 18 Sept 2018

21. Zou S, Fang L, Lee M-H (2018) Dysbiosis of gut microbiota in promoting the development of colorectal cancer. Gastroenterol Rep 6:1–12

22. Jones JR, DellaValle CT, Weyer PJ, Robien K, Cantor KO, Krasner VP et al (2019) Ingested nitrate, disinfection by-products, and risk of colon and rectal cancers in the Iowa Women's Health Study. Environ Internat 126:242–251

23. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ (2017) Colorectal carcinoma: a general review and future perspectives in colorectal cancer. Int J Mol Sci 18:197. https://doi.org/10.3390/ijms18010197

24. Nolan BT, Ruddy BC, Hitt EJ, Helsel R (1997) Risk of nitrate in groundwaters of the United States—a national perspective. Environ Sci Technol 21:2229–2236

25. Schullehner J, Hansen B, Thygesen M, Pedersen CB, Sigsgaard T (2018) Nitrate in drinking water and colorectal cancer risk: a nationwide population-based cohort study. Int J Cancer 243:73–79

26. Garklavs G, Nelson R (1986) North Dakota ground water quality. U.S. Geological Survey Open-File Report 87-0744. U.S. Geological Survey, Denver, CO. Available at https://pubs.usgs.gov/of/1987/0744/report.pdf