Journal of Geochemical Exploration xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/gexplo

Drinking water nitrate estimation at household-level in Danish population-based long-term epidemiologic studies

Jörg Schullehner [a,b,c,*], Nanna Linn Jensen [d], Malene Thygesen [c,e], Birgitte Hansen [a], Torben Sigsgaard [b]

[a] Geological Survey of Denmark and Greenland (GEUS), C.F. Møllers Allé 8, 8000 Aarhus C, Denmark
[b] Department of Public Health, Aarhus University, Bartholins Allé 2, 8000 Aarhus C, Denmark
[c] National Centre for Register-Based Research, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark
[d] Department of Geoscience, Aarhus University, Høegh-Guldbergs Gade 2, 8000 Aarhus C, Denmark
[e] Centre for Integrated Register-based Research at Aarhus University (CIRRAU), Fuglesangs Allé 4, 8210 Aarhus V, Denmark

ARTICLE INFO

Article history:
Received 27 May 2016
Revised 9 March 2017
Accepted 10 March 2017
Available online xxxx

Keywords:
Drinking water
Exposure
Nitrate
Denmark

ABSTRACT

This study presents an approach making drinking water quality data from the Danish national geodatabase Jupiter available for epidemiological studies on long-term health effects of drinking water quality. Drinking water quality was assigned to all Danish households using spatial methods, and the related uncertainty due to sampling density was identified. Nitrate was used as an example contaminant. From 1978, 98% of all Danish households were geocoded with their precise location, yielding a total of 69 million evaluated household-years. Households supplied by a private well (4% of all household-years) were identified; the remaining majority was supplied by public waterworks. Water supply areas were connected to the Jupiter geodatabase to estimate drinking water quality at publicly supplied households. For privately supplied households, the exposure estimate had a substantially higher uncertainty, with 52% of household-years never being sampled, compared with 1% of publicly supplied household-years. For the 18% of all household-years with no nitrate samples, concentrations were estimated by interpolation and an uncertainty score based on the closest nitrate measurement was introduced. While nitrate was used as an example contaminant, the proposed approach is generic and can be used for all monitored drinking water parameters, after preprocessing and validation of the specific data and assumptions.

1. Introduction

Long-term exposure to chemical compounds in drinking water can lead to adverse health effects (Thompson et al., 2007), and a need for well-designed exposure assessments to investigate these has been identified (Villanueva et al., 2014). Amongst others, it was suggested to take advantage of large monitoring databases, carry out large studies with sufficient statistical power and an appropriate exposure contrast, and use environmental and geospatial methods for a thorough exposure assessment (Villanueva et al., 2014). Using the extensive historical registration of drinking water quality measurements in Denmark, we designed an approach addressing these particular challenges.

Denmark is a Scandinavian country in Northern Europe (see Fig. 1) with 5.7 million inhabitants (2015) and a relatively homogenous population, with non-western immigrants and their descendants accounting for 7.2% of the total population in 2014 (Statistics Denmark, 2014). Register-based epidemiological studies of the whole Danish population are conducted widely, given the unique possibilities of linking registers on health and social issues based on a unique personal identification number assigned to all residents of Denmark (Thygesen et al., 2011). Thus, the whole population can be used as a cohort in epidemiological studies (Pedersen et al., 2006). Health-related registers that can be used for research include the Register of Causes of Death, the National Patient Register, the National Health Service Register, the National Prescription Register, the Danish Cancer Registry, the Psychiatric Central Research Register, as well as a number of registers with information on persons with specific diagnoses, such as diabetes, multiple sclerosis and more, see details in Thygesen et al. (2011).

In addition to the Danish registers on health and social issues, the public national geodatabase Jupiter (GEUS, 2016a) includes data on waterworks and drinking water quality for many different parameters; monitored and registered over many decades. Analyses of the drinking water samples that are registered in Jupiter are carried out by certified laboratories (GEUS, 2016b). Therefore, linking the drinking water quality data registered in the Jupiter geodatabase to all Danish households will present the unique possibility of calculating individual-level exposures for all Danish residents via the information on complete residential histories.

Principally, all drinking water in Denmark originates from simple treated groundwater (aeration, sand filtration), and no surface water is used (Danish Nature Agency, 2016). Drinking water quality data are

* Corresponding author at: C.F. Møllers Allé 8, 8000 Aarhus C, Denmark.
E-mail address: jsc@geus.dk (J. Schullehner).

http://dx.doi.org/10.1016/j.gexplo.2017.03.006

Fig. 1. Map of the study area Denmark and its 98 municipalities.

readily uploaded by certified laboratories that analyze the waterworks' samples. The municipal authorities have to approve the uploaded data, before they are accessible by the public. The database itself is hosted at the Geological Survey of Denmark and Greenland (GEUS). The data in Jupiter cannot be used in epidemiological studies without pretreatment and additional information. In Jupiter, the only spatial information available is the location of the waterworks, but not the areas they supply (Schullehner and Hansen, 2014). Furthermore, not all households are connected to the public supply, even if they are located within a water supply area of a public waterworks. These users of private drinking water wells were already shown to have very different drinking water quality compared with the public supplies, with two-thirds violating the drinking water standard for either nitrate, pesticides, or microbiological contamination in four Danish counties (Brüsch et al., 2004; Schullehner and Hansen, 2014).

Drinking water quality is monitored and registered for a variety of parameters (Ministry of Environment and Food, 2016). Here, nitrate is taken as an example, since nitrate is one of the main parameters that are most frequently measured when a drinking water sample is taken. Additionally, nitrate has been a groundwater and drinking water quality issue in Denmark for many decades (Hansen et al., 2011, 2012; Schullehner and Hansen, 2014). Since Danish drinking water supply is based on groundwater with no chemical disinfection, seasonal variations and changes in nitrate concentrations in the distribution system are negligible, as shown by Schullehner et al. (2017). Possible concentration changes due to nitrification of ammonium present in the drinking water are negligible compared to the nitrate concentration contrast between waterworks. Therefore, nitrate concentrations measured both at the exit of the waterworks, within the distribution system and at the consumer's tap can be considered representative of nitrate concentrations in the whole distribution system (Schullehner et al., 2017).

Studies on the long-term health effects of nitrate in drinking water have yielded equivocal conclusions (Ward et al., 2005), with more recent publications indicating an increased risk of, for example, colorectal and bladder cancer at nitrate levels below the current drinking water standard (Espejo-Herrera et al., 2016; Jones et al., 2016). The presented approach will give epidemiologists the possibility to include nitrate exposure from drinking water in their studies in the Danish setting. However, the presented approach is a blueprint for studies on any monitored Danish drinking water quality parameters and suspected associated health outcomes, which can be conducted in the future.

2. Methods

The data on waterworks in the national geodatabase Jupiter was used as a base dataset and augmented by information from other relevant sources. Waterworks were divided into public supplies (supplying 10 or more households) that were assigned to the water supply areas which they supply (polygons), and private wells (supplying less than 10 households) that were handled as point data. This discrimination between public and private supplies complies largely with a definition by the World Health Organization (WHO, 2011). The public supply dataset has been used for a nationwide study on nitrate in drinking water before (Schullehner and Hansen, 2014), while the private well dataset has been augmented by additional data as described below.

All Danish addresses were classified to be supplied by either a public supply or a private well. As an example, annual nitrate concentrations were assigned to each household. Nitrate concentrations are reported as the whole nitrate-ion. For years with no samples, concentrations were estimated. An uncertainty score was introduced to enable sensitivity analyses. GIS-analyses were carried out in ArcGIS 10.2 (Esri Inc., Redlands, CA, United States) and Python 2.7.6 (Python Software Foundation, DE, United States). Data management and interpolation calculations were carried out in R 3.2.2 (R Core Team, 2015), including the packages data.table (Dowle et al., 2014) and *zoo* (Zeileis and Grothendieck, 2005). The study is approved by the Danish Data Protection Agency. The following subsections describe the approach in detail.

2.1. Geocoding of residences

Full residential history of all Danish residents since 1968 is available in the Civil Registration System (CRS). The Centre for Integrated Register-based Research at Aarhus University (CIRRAU) supplied a database covering all Danish household addresses from the establishment of the CRS in 1968 until July 2013. The database had one record for each address, at which at least one individual was registered in the CRS in Denmark since 1968, corresponding to a total of 52,485,728 residences (i.e. a person living in a household for a period of time before moving to another household or emigration/death/disappearance). For each household, the database had information about the exact date on which an individual moved to and from the address. For most of the households, information about the full location (municipality, road, house number and, if relevant, door) was available.

The geocoding of the households was performed by linking the address with information from the Register on Official Standard Addresses and Coordinates, which consists of all official addresses in Denmark and their geographical coordinates. Of the 52,485,728 residences in the database, 99.2% were successfully geocoded. The quality of the geocoded coordinates varied depending on the extent of the missing information on the address. 96.2% of the households could be matched by municipality, road, house number and side door, or by municipality, road and house number, which were the matches considered of the highest quality. For 3.0% of the households, it was not possible to allocate exact geographical coordinates, since part of the household address was missing, in most cases the house number. Therefore, alternative methods were used in order to allocate geographical coordinates to these addresses, e.g. allocating a geographical coordinate, which was constructed as an average of the geographical coordinates on that particular road. For the remaining 0.78% of the households, it was not possible to allocate any geographical coordinates. This was the case for household addresses that were not an official household address, consisted of an abandoned road, or included road codes that had been renamed, which made it impossible to link to the Register on Official Standard Addresses and Coordinates.

Please cite this article as: Schullehner, J., et al., Drinking water nitrate estimation at household-level in Danish population-based long-term epidemiologic studies, J. Geochem. Explor. (2017), http://dx.doi.org/10.1016/j.gexplo.2017.03.006