GEUS Bulletin | Geological Survey of Denmark and Greenland

DATA ARTICLE | SHORT

Danish Water Supply Areas and their links to water production facilities: an open-access data set

Jörg Schullehner*[1,2,3]

[1]Geological Survey of Denmark and Greenland, Aarhus, Denmark. [2]Department of Public Health, Environment, Occupation and Health, Aarhus University, Aarhus, Denmark. [3]Danish Big Data Centre for Environment and Health (BERTHA), Aarhus University, Aarhus, Denmark.

Abstract

This data set establishes the missing link between drinking-water quality monitoring data at the water production facility level in the Danish national geodatabase Jupiter and supply areas. Water Supply Areas (WSAs) were collected at municipality level, digitised and linked to the waterworks they are supplied by. Infrastructural changes between 1978 and 2019 were taken into account by allowing WSA polygons to change over time. The number of active WSAs decreased from 3172 in 1978 to 2602 in 2019. The data set consists of longitudinal WSA polygons and a table linking WSAs to the water production facility identification in the Jupiter database, allowing the estimation of current and historical drinking-water quality across Denmark. In combination with the Danish Address Register and the Civil Registration System, this data set allows exposure assessments of drinking-water quality at high spatiotemporal resolution for the entire Danish population. Therefore, this data set is an essential part of studying health effects of drinking-water quality in epidemiological research in Denmark.

Tabular abstract

Geographical coverage	Denmark
Temporal coverage	1978–2019
Subject(s)	Engineering and environmental geology
Data format(s)	Analysed spatial data (shapefile) and linkage table (text file)
Sample collection and analysis	Water Supply Area polygons collected and digitised from various sources, mainly at municipality level. Waterworks and drinking-water quality data from Jupiter
Parameters	Water Supply Area polygons with start and end year of activity, linkage of Water Supply Area polygons (WSAID) to PLANTID (ANLAEGID) in Jupiter.
Related publications	Schullehner & Hansen 2014; Voutchkova *et al*. 2015, 2021; Schullehner *et al*. 2017, 2018, 2019, 2020; Wodschow *et al*. 2018, 2021; Coffman *et al*. 2021; Stayner *et al*. 2021, 2022; Thomsen *et al*. 2021; Thygesen *et al*. 2021; Ebdrup *et al*. 2022; Richter *et al*. 2022; Skaarup *et al*. 2022; Theisen *et al*. 2022.
Potential application(s) for these data	Linkage of water quality registered at production facility level in Jupiter to supply zones, households. Allows estimation of drinking-water quality exposure of the entire Danish population through geocoded residential history and thus linkage to epidemiological studies on health effects of drinking-water quality.

***Correspondence:** *jorg.schullehner@ph.au.dk*
Received: 07 Apr 2022
Accepted: 17 May 2022
Published: 28 June 2022

Keywords: drinking water, monitoring, exposure assessment, epidemiology, Denmark

Abbreviations:

GIS: geographic information system
PFAS: per- and polyfluoroalkyl substances
PIN: personal identification number
PLANTID: water production facility identifier
WSA: Water Supply Area
WSAID: WSA identifier

Edited by: Adam Hambly (Technical University of Denmark)

Reviewed by: Debbie White (British Geological Survey, UK), Martin Rygaard (Technical University of Denmark).

Funding: See page 6

Competing interests: See page 6

Additional files: See page 6

Introduction

Data on drinking-water quality are relevant for a variety of stakeholders, including authorities, interest groups, plumbers, the general public and researchers. In Denmark, drinking-water quality is monitored systematically, and results are registered in a nationwide, publicly accessible geodatabase called Jupiter (Hansen & Pjetursson 2011; GEUS 2022). Jupiter holds drinking water-related information including administrative information on drinking-water production facilities, technical construction and geological description of boreholes, water abstraction volumes, and drinking-water quality samples and their analytical results. Additionally, Jupiter holds information on the location of the water production facilities and their boreholes. However, Jupiter does not hold any information on the areas that water production facilities supply.

This missing link is crucial for a range of tasks, including monitoring of drinking water at population level and estimation of historical drinking-water exposure for use in epidemiological studies on health effects of drinking-water quality.

To overcome these limitations, efforts have been made to establish this missing link by collecting and digitising the Danish Water Supply Areas (WSAs) and linking them to water production facilities in Jupiter (Schullehner & Hansen 2014; Schullehner *et al.* 2017). This data set has already been used in many epidemiological studies (Schullehner *et al.* 2018, 2019, 2020; Coffman *et al*. 2021; Stayner *et al.* 2021, 2022; Thomsen *et al.* 2021; Thygesen *et al*. 2021; Wodschow *et al.* 2021; Ebdrup *et al*. 2022; Richter *et al.* 2022; Theisen *et al.* 2022) and nationwide drinking-water quality mapping and modelling (Wodschow *et al*. 2018; Voutchkova *et al*. 2021, Skaarup *et al*. 2022) and has been continuously developed to include the historical WSAs from 1978 to 2019. Here, this data set is described and published as an open-access data set.

Data collection

Setting

Danish drinking-water supply has several distinctive characteristics: it is based on groundwater, often with limited treatment (typically aeration and sand filtration), without chlorination and decentralised: around 2600 public waterworks supply a population of around 5.8 million (Schullehner & Hansen 2014; Ministry of Environment of Denmark 2022). Groundwater for drinking production is mainly sourced from aquifers in limestone/chalk and Quaternary and Miocene sandy deposits.

Drinking-water quality monitoring is based on the implementation of the EU Drinking Water Directive (European Council 1998) in the Ministerial Order on Water Quality and Control of Water Supply Facilities (Ministry of Environment of Denmark 2021). Historically, the Ministerial Order defined public supplies (in Danish: *almene vandværker*) as serving ten or more households (Schullehner & Hansen 2014). This definition is applied in the presented data set that focuses on public supplies.

The Ministerial Order defines monitoring programmes, including lists of quality parameters to be monitored and the sampling frequency depending on water production volumes. The monitoring programme has changed over time, resulting in data breaks, which need to be considered. According to the Ministerial Order, certified laboratories must take and analyse drinking-water samples, and results of both raw and finished treated drinking-water analyses and annual water production volumes are to be reported to the publicly accessible Jupiter database.

Jupiter currently holds information on more than 370 000 boreholes, more than 40 000 of which are linked to water production facilities as well as around 60 000 water production facilities; the majority of the latter being private wells. Over 700 000 drinking-water samples are registered in Jupiter, with almost 10 million analytical measurements of water quality parameters.

Data sources

Locations of public supplies were initially extracted from Jupiter in October 2013 and have been continuously updated since. Missing coordinates were geocoded using address information (Schullehner & Hansen 2014). Information on WSAs is often available at municipality level. The 98 Danish municipalities are required to devise water supply plans (in Danish: *vandforsyningsplaner*), including which areas of the municipalities are supplied by which public water production facilities. Some municipalities publish online geographic information system (GIS) services with WSA polygons, which were downloaded. The majority of municipalities published maps, which were collected and digitised into polygons in GIS. For larger water supply companies, especially around the larger cities, waterworks and/or municipalities were contacted and supplied either files with supply zone polygons or distribution networks or a description of their supply infrastructure and zones (Schullehner & Hansen 2014; Schullehner *et al.* 2017).

Data processing

In the initial version, used in a number of epidemiological studies (Schullehner *et al.* 2018, 2019, 2020; Coffman *et al*. 2021; Thygesen *et al.* 2021; Wodschow *et al.* 2021; Richter *et al.* 2022; Theisen *et al.* 2022), it was assumed that the WSA polygons did not change over time. The water production facility identifier (PLANTID, in Danish: *ANLAEGID*) was spatially joined to the WSA polygons. In most cases, the water production facility is located within the WSA polygon that it supplies. In other cases, WSAs were created as multipolygons including a small polygon around the plant's address, such that plants are located within the WSA polygons they supply. For the large water supply companies of the major cities, links between PLANTID and the WSA identifier (WSAID) have

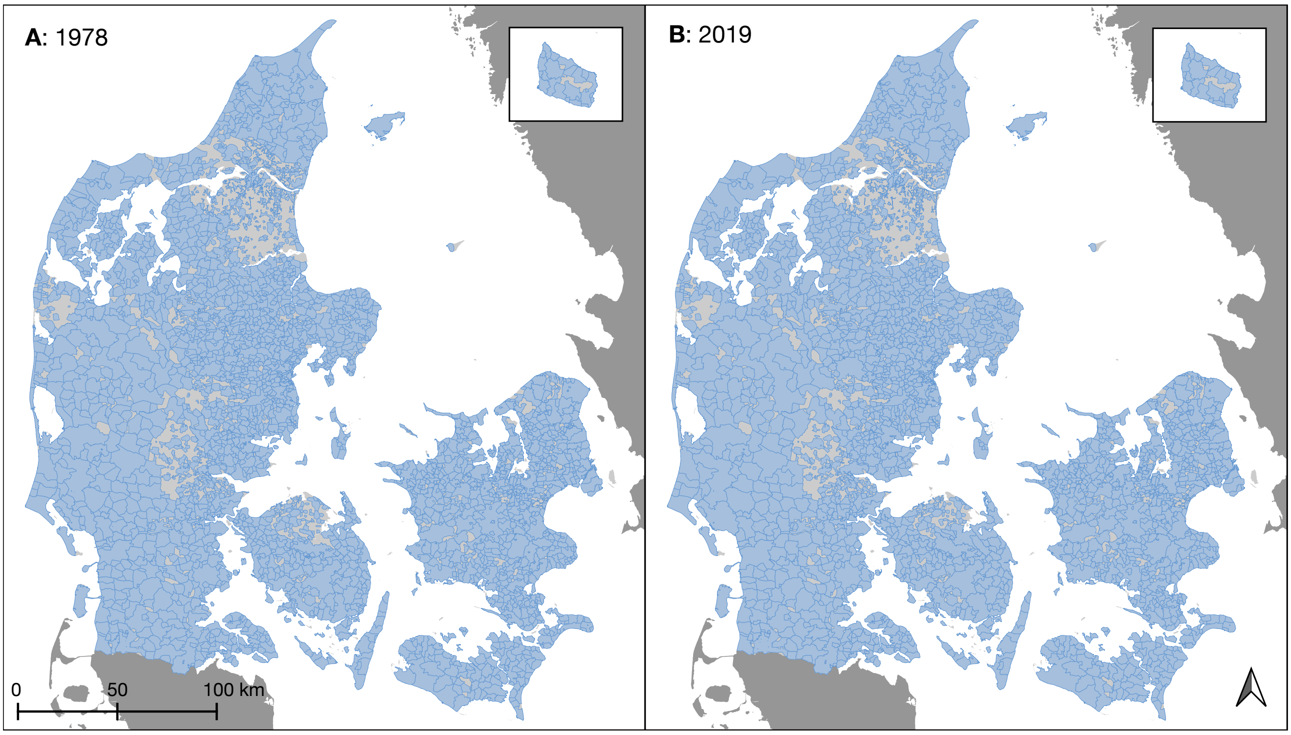

Fig. 1 Extents of the Water Supply Areas (WSAs) in Denmark in (**A**) 1978 and (**B**) 2019. Light grey areas are outside of public water supply zones.

been added manually to the linkage table using information supplied by these companies.

For historical assessments of drinking-water quality, changes in supply infrastructure can be important. Therefore, the initial model was updated and has already been applied in recent epidemiological studies (Stayner *et al*. 2021, 2022; Thomsen *et al.* 2021; Ebdrup *et al.* 2022). In the updated model, the static polygons – corresponding to the infrastructure at the time of data curation – were individually assessed. If more than one water production facility was located within a WSA, the WSA was split into multiple historical WSAs and linked to the corresponding PLANTID. The start and end year of the WSA's activity was estimated by assessing the time series of drinking-water quality samples linked to the respective PLANTID in Jupiter. If a time series of, for example, nitrate measurements at a production facility stopped, it was assumed that this PLANTID closed and was merged with the WSA of the PLANTID with the continuing time series. This results in two polygons with each their WSA being merged into one, with a new WSAID. The start and end year of the active period of the WSAs was updated: the former WSAs' active periods end in the year when the time series stopped, and the merged WSA's start is the following year, resulting in continuity in linkage of supplied households and their water production facilities. The borders of the delineation between WSAs to split were estimated by expert assessment. Map layers of locations of households and private wells were used to guide the decision. In ambiguous cases, additional information was acquired from waterworks' homepages, municipality archives and personal contact with waterworks. As geocoded residential history of the Danish population is available in high quality from 1978 onwards (CIRRAU 2022), and drinking-water quality monitoring achieved good coverage from the 1980s (see *Data description and main features*); the period included in this data set is 1978–2019.

Software

Data management was done in R (versions 3.2–4.0; R Core Team 2020). GIS tasks were done in ArcMap 10 and QGIS (versions 2.14–3.18; ESRI 2015; QGIS.org 2022).

Data description and main features

The data set presented here consists of the WSA polygons with attributes WSAID (unique key), START and END (start and end year of the WSA's active period) as a shapefile. Additionally, there is a table linking WSAID to the unique key of the water production facilities in Jupiter (PLANTID, in Danish: *ANLAEGID*).

Figure 1 shows the extent of the estimated WSA boundaries in 1978 and 2019. Note that whilst most municipalities divide their entire area into potential WSAs, some (e.g., Aalborg municipality) provided the more precise physical extents, leaving areas without public supply empty. Even though a household lies

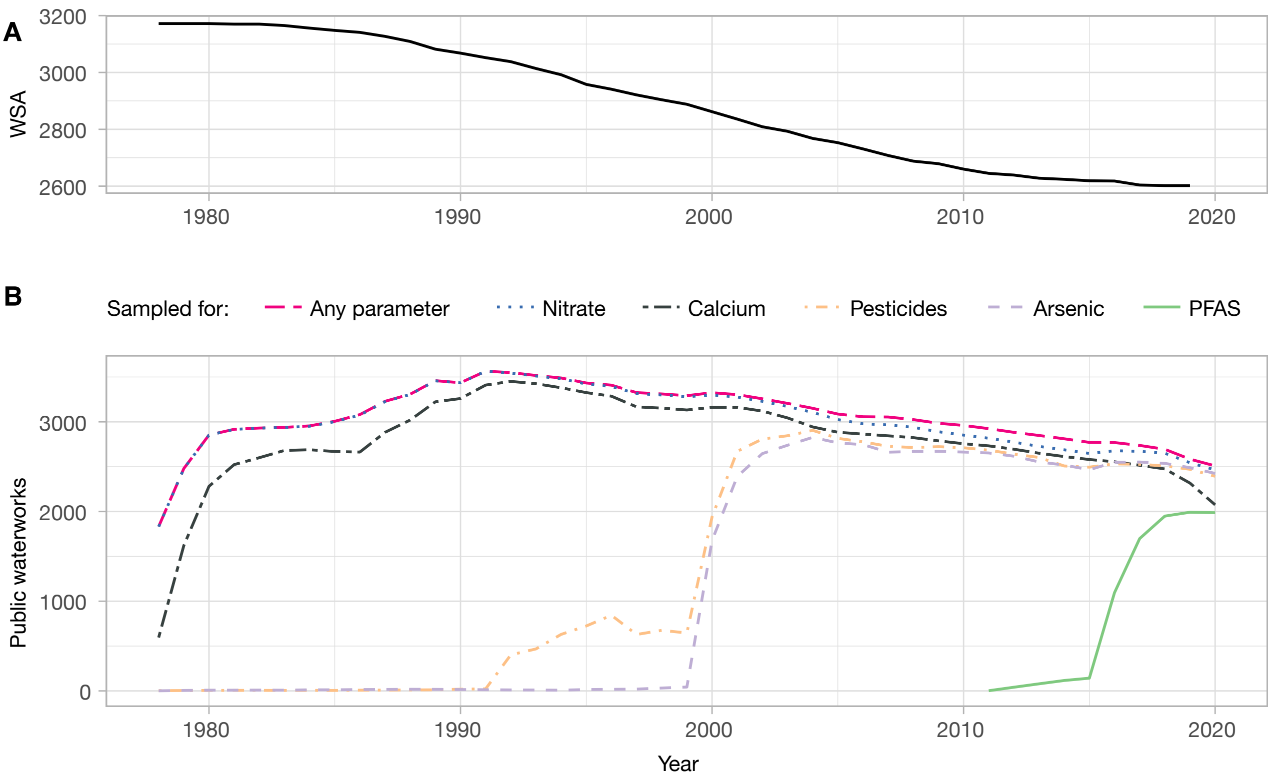

Fig. 2 Annual number of active Water Supply Areas (WSA; **A**) and number of public waterworks with drinking-water samples analysed for different parameters registered in Jupiter (**B**). To account for different sampling frequencies, the annual number is presented here as the rolling count over three years.

within a WSA, it may be supplied by a private well (see later discussion on limitations).

During the study period, a consolidation of the water supply structure can be observed: the number of active WSAs decreased from initially 3172 in 1978 to 2602 in 2019 (Fig. 2A). This corresponds to closures of water production facilities and consequent mergers with neighbouring WSAs. Mergers of WSAs also happen without closures of production facilities, resulting in the merged WSA being supplied by more than one production facility. In 1978, 74% of WSAs were supplied by only a single production facility; however, by 2019, this decreased to 66%.

Figure 2B shows the number of waterworks with registered drinking-water samples in Jupiter. Before 1990, there are reporting issues depending on the county in which the waterworks were located. After 1990, the number of waterworks with a registered drinking-water sample as an indication of being active decreases from more than 3500 to 2500. Changes in the monitoring programme are also highlighted (Fig. 2B). While major parameters like nitrate and calcium have been monitored consistently from the 1980s onwards, emerging quality issues such as pesticides and per- and polyfluoroalkyl substances (PFAS) were introduced at later stages. The arsenic drinking-water standard in Denmark was lowered to 5 µg/L in 2001 (Ramsay *et al.* 2021), resulting in increased monitoring.

Figure 3 shows how this data set bridges the gap when linking drinking-water quality data from Jupiter to a person's estimated exposure history in an epidemiological study. Whilst Jupiter, this data set, and the Danish Address Register (SDFE 2022) are all publicly accessible data sets, getting access to the administrative and health registers requires permissions, which can be granted to researchers under specific terms complying with data protection legislation (Thygesen *et al.* 2011). Figure 3 is a minimum working example and can be extended thoroughly in both directions. Linking additional information from Jupiter, such as raw water quality and abstraction well geology, and additional data from administrative and health registers, such as socio-economic factors and comorbidities, is easily accomplished using the PLANTID and personal identification number (PIN), respectively.

The strength of this data set is the establishment of the missing link between longitudinal nationwide water quality at production facility level and supply area. In combination with geocoded addresses linked to the Civil Registration System, links to health and administrative databases are possible. This data set takes infrastructural changes into account, by allowing

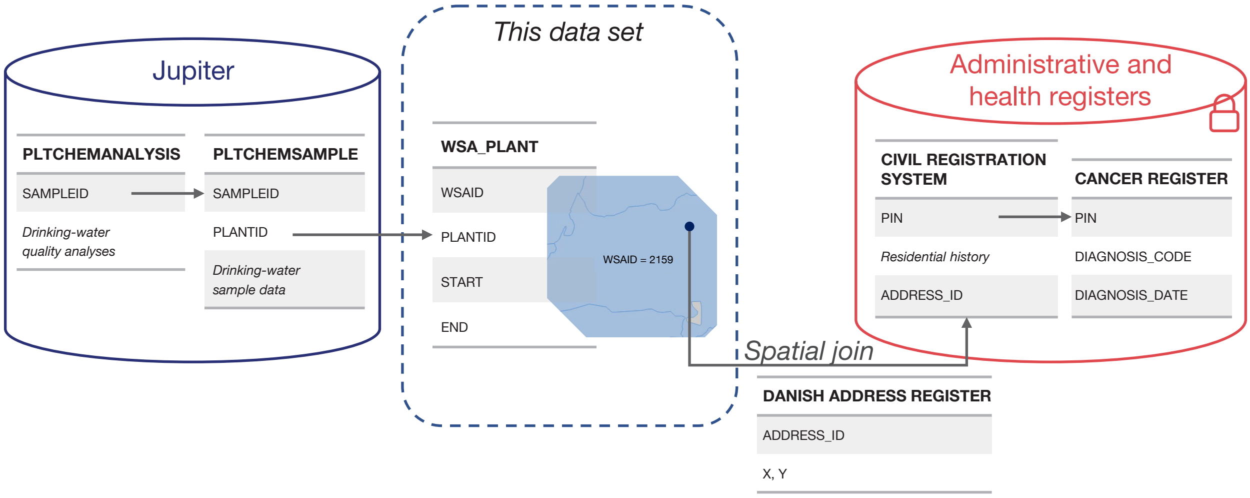

Fig. 3 Minimum working example of using this data set for establishing the linkage between drinking-water quality in Jupiter to a person's estimated exposure history in epidemiological studies. PLTCHEMANALYSIS and PLTCHEMSAMPLE refer to the tables in Jupiter containing the drinking-water plants' chemical analyses and sample data.

WSA polygons and their link to production facilities to change over time.

A major limitation is that this data set is not authoritative, which limits its application for other uses besides research. Boundaries of WSAs can be uncertain, but typically run through areas with very sparse population density, reducing the risk of misclassification. For the WSAs supplied by multiple production facilities, the variation within the WSA may even change during the course of a day depending on production and consumption patterns, but this cannot be estimated with this data set alone. However, it is possible to weigh mean concentrations of water quality parameters by annual production volume registered in Jupiter. When using this data set, it is, therefore, suggested to identify WSAs with large variations for the specific parameter under investigation and examine these more carefully. In this data set, private wells (supplying less than ten households) are not included. However, it is possible to estimate which households are supplied by private wells, for example by proximity analysis (Schullehner *et al.* 2017). It must, however, be noted that private well registration is not complete in Jupiter. Furthermore, it should be noted that raw data in Jupiter often require substantial filtering, clean-up and data management before being useful for nationwide drinking-water quality mapping. The nature of this process depends on many circumstances and should be assessed individually for every water-quality parameter and study period, as, for example, data coverage, analytical methods and detection limits can vary substantially. It is, therefore, not possible here to provide a curated data set for all drinking-water quality parameter data in Jupiter. Research groups who made the effort to curate parameter-specific Jupiter data sets are encouraged to publish these and relevant code along with their studies.

Supported by its wide use in epidemiological studies, it is concluded that this data set is a valuable tool for research purposes. Given its limitations, it is cautioned against using this data set for administrative applications by public authorities. Instead, it is encouraged that the authorities work towards establishing a database, for example, in connection with the Jupiter infrastructure, where waterworks report and update officiary data. For example, by supplying spatial data with their actual supply zones or distribution systems or lists of the supplied addresses.

Data availability and updates

The data set is available at the GEUS Bulletin Dataverse (see *Additional files*), where future updates will be uploaded, under the Creative Commons Licence Attribution 4.0 International (CC BY 4.0). It would be greatly appreciated if users of this data set report any errors or inconsistencies that they encounter.

Acknowledgements

I would like to thank my colleagues Denitza Voutchkova and Birgitte Hansen, Geological Survey of Denmark and Greenland, for feedback and discussions about the construction of this data set, and the reviewers, whose comments improved the quality of this article.

Additional information

Funding statement

This data set has been developed during several years with funding from Innovation Fund Denmark (dNmark research alliance), the Centre for Register-based Research at Aarhus University (CIRRAU), the Aarhus University Research Foundation (grant AUFF-E-2015-FLS-8-61), the United States National Institutes of Health/National Institute of Environmental Health Sciences (grant R01 ES027823-01A1), the Karen Elise Jensens Foundation and the Novo Nordisk Foundation Challenge Programme BERTHA – the Danish Big Data Centre for Environment and Health (grant NNF17OC0027864).

Competing interests

None.

Author contributions

JS is the sole author.

Additional files

The complete data set is available at *https://doi.org/10.22008/FK2/I5R1SS*

References

CIRRAU. 2022: Data documentation – geocodes. https://cirrau.au.dk/data-resources/data-documentation (accessed April 2022)

Coffman, V.R. *et al.* 2021: Prenatal exposure to nitrate from drinking water and markers of fetal growth restriction: a population-based study of nearly one million Danish-born children. Environmental Health Perspectives **129**(2), 027002. *https://doi.org/10.1289/EHP7331*

Ebdrup, N.H. *et al.* 2022: Nitrate in drinking water and time to pregnancy or medically assisted reproduction in women and men: a nationwide cohort study in the Danish national birth cohort. Clinical Epidemiology **14**, 475–487. *https://doi.org/10.2147/CLEP.S354926*

ESRI. 2015: ArcGIS 10.4. Redlands, CA. Environmental Systems Research Institute, Inc. *https://www.esri.com*

European Council. 1998: Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption. *https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A31998L0083* (accessed May 2022)

GEUS. 2022: National boringsdatabase (Jupiter). Geological Survey of Denmark and Greenland. *https://www.geus.dk/produkter-ydelser-og-faciliteter/data-og-kort/national-boringsdatabase-jupiter* (accessed May 2022)

Hansen, M. & Pjetursson, B. 2011: Free, online Danish shallow geological data. Geological Survey of Denmark and Greenland Bulletin **23**, 53–56. https://doi.org/10.34194/geusb.v23.4842

Ministry of Environment of Denmark. 2021: Bekendtgørelse om vandkvalitet og tilsyn med vandforsyningsanlæg, BEK nr 2361 af 26/11/2021 (Drikkevandsbekendtgørelsen).

Ministry of Environment of Denmark. 2022: Drikkevand. *https://mst.dk/natur-vand/vand-i-hverdagen/drikkevand* (accessed April 2022)

QGIS.org. 2022: QGIS Geographical Information System. QGIS Association. *http://www.qgis.org*

Ramsay, L., Petersen, M.M., Hansen, B., Schullehner, J., van der Wens, P., Voutchkova, D. & Kristiansen, S.M. 2021: Drinking water criteria for arsenic in high-income, low-dose countries: the effect of legislation on public health. Environmental Science & Technology **55**, 3483–3493. *https://doi.org/10.1021/acs.est.0c03974*

R Core Team. 2020: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *https://www.R-project.org/*

Richter F., Kloster, S., Wodschow, K., Hansen, B., Schullehner, J., Kristiansen, S.M., Petersen, M.M., Strandberg-Larsen, K. & Ersbøll, A.K. 2022: Maternal exposure to arsenic in drinking water and risk of congenital heart disease in the offspring. Environment International **160**, 107051. *https://doi.org/10.1016/j.envint.2021.107051*

Schullehner, J. & Hansen, B. 2014: Nitrate exposure from drinking water in Denmark over the last 35 years. Environmental Research Letters **9**, 095001, 9 pp. *https://doi.org/10.1088/1748-9326/9/9/095001*

Schullehner, J., Hansen, B., Thygesen, M., Pedersen, C.B. & Sigsgaard, T. 2018: Nitrate in drinking water and colorectal cancer risk: A nationwide population-based cohort study. International Journal of Cancer **143**, 73–79. *https://doi.org/10.1002/ijc.31306*

Schullehner, J., Jensen, N.L., Thygesen, M., Hansen, B. & Sigsgaard, T. 2017: Drinking water nitrate estimation at household-level in Danish population-based long-term epidemiologic studies. Journal of Geochemical Exploration **183**, 178–186. *https://doi.org/10.1016/j.gexplo.2017.03.006*

Schullehner, J., Paksarian, D., Hansen, B., Thygesen, M., Kristiansen, S.M., Dalsgaard, S., Sigsgaard, T. & Pedersen, C.B. 2019: Lithium in drinking water associated with adverse mental health effects. Schizophrenia Research **210**, 313–315. *https://doi.org/10.1016/j.schres.2019.06.016*

Schullehner, J., Thygesen, M., Kristiansen, S.M., Hansen, B., Pedersen, C.B. & Dalsgaard, S. 2020: Exposure to manganese in drinking water during childhood and association with attention-deficit hyperactivity disorder: a nationwide cohort study. Environmental Health Perspectives **128**(9), 097004. *https://doi.org/10.1289/EHP6391*

SDFE. 2022: The Danish Address Register. Agency for Data Supply and Efficiency. *https://eng.sdfe.dk/products-and-services/the-danish-address-register* (accessed May 2022)

Skaarup, C., Wodschow, K., Voutchkova, D.D., Schullehner, J., Raaschou-Nielsen, O., Andersen, H.R., Hansen, B. & Ersbøll, A.K. 2022: Geographical Distribution and pattern of pesticides in Danish drinking water 2002–2018: reducing data complexity. International Journal of Environmental Research and Public Health **19**(2), 823. *https://doi.org/10.3390/ijerph19020823*

Stayner, L.T. *et al*. 2021: Exposure to nitrate from drinking water and the risk of childhood cancer in Denmark. Environment International **155**, 106613. https://doi.org/10.1016/j.envint.2021.106613

Stayner, L.T. *et al*. 2022: Nitrate in drinking water and risk of birth defects: findings from a cohort study of over one million births in Denmark. The Lancet Regional Health – Europe **14**. 100286. *https://doi.org/10.1016/j.lanepe.2021.100286*

Theisen, C.F., Wodschow, K., Hansen, B., Schullehner, J., Gislason, G., Ersbøll, B.K. & Ersbøll, A.K. 2022: Drinking water magnesium and cardiovascular mortality: a cohort study in Denmark, 2005–2016. Environment International **164**, 107277. *https://doi.org/10.1016/j.envint.2022.107277*

Thomsen, A.M.L., Ramlau-Hansen, C.H., Schullehner, J., Ebdrup, N.H., Liew, Z., Coffman, V., Stayner, L., Hansen, B. & Olsen, J. 2021: Prenatal nitrosatable prescription drug intake, drinking water nitrate, and the risk of stillbirth: a register- and population-based cohort of Danish pregnancies, 1997–2017. Environmental Health **20**, 118. *https://doi.org/10.1186/s12940-021-00805-z*

Thygesen, L.C., Daasnes, C., Thaulow, I. & Bronnum-Hansen, H. 2011: Introduction to Danish (nationwide) registers on health and social issues: Structure, access, legislation, and archiving. Scandinavian Journal of Public Health **39**, 12–16. *https://doi.org/10.1177/1403494811399956*

Thygesen, M., Schullehner, J., Hansen, B., Sigsgaard, T., Voutchkova, D.D., Kristiansen, S.M., Pedersen, C.B. & Dalsgaard, S. 2021: Trace elements in drinking water and the incidence of attention-deficit hyperactivity disorder. Journal of Trace Elements in Medicine and Biology **68**, 126828. *https://doi.org/10.1016/j.jtemb.2021.126828*

Voutchkova, D., Schullehner, J., Knudsen, N., Jørgensen, L., Ersbøll, A., Kristiansen, S. & Hansen, B. 2015: Exposure to selected geogenic trace elements (I, Li, and Sr) from drinking water in Denmark. Geosciences **5**, 45–66. *https://doi.org/10.3390/geosciences5010045*

Voutchkova, D.D., Schullehner, J., Skaarup, C., Wodschow, K., Ersbøll, A.K. & Hansen, B. 2021: Estimating pesticides in public drinking water at the household level in Denmark. GEUS Bulletin **47**, 6090. *https://doi.org/10.34194/geusb.v47.6090*

Wodschow, K., Hansen, B., Schullehner, J. & Ersbøll, A.K. 2018: Stability of major geogenic cations in drinking water-an issue of public health importance: a Danish study, 1980-2017. International Journal of Environmental Research and Public Health **15**(9), 1212. *https://doi.org/10.3390/ijerph15061212*

Wodschow, K., Villanueva, C.M., Larsen, M.L., Gislason, G., Schullehner, J., Hansen, B. & Ersbøll, A.K. 2021: Association between magnesium in drinking water and atrial fibrillation incidence: a nationwide population-based cohort study, 2002–2015. Environmental Health **20**, 126. *https://doi.org/10.1186/s12940-021-00813-z*