

RESEARCH ARTICLE

Inclusion of deprivation in endemic-epidemic models

Maria Bekker-Nielsen Dunbar^{1,2} , Sverren-Erik Mamelund¹  and Gerardo Chowell³ 

ABSTRACT Deprivation amplification theory suggests that the health effects of individual deprivation are amplified for people who live in areas with greater levels of deprivation. We postulate that health (represented by norovirus incidence) is influenced and amplified by deprivation (a measure that includes socio-economic factors), and believe that this association has been neglected in surveillance models of infectious diseases. We construct a social epidemiological extension of a known surveillance model to evaluate the inclusion of deprivation in surveillance models using the German Index of Socio-economic Deprivation (GISD) in an endemic-epidemic model. We evaluate model types considered in the literature on the basis of Akaike's information criterion. Our results suggest that a social epidemiological endemic-epidemic model with the GISD for enterically transmitted infections does not need to also include time-varying contact matrices as transmission weights.

KEYWORDS Berlin • Deprivation • Endemic-epidemic modelling • Infectious disease surveillance • Norovirus (Norwalk agent) • Social epidemiology

Introduction

Norovirus is a disease of alimentary transmission. It normally causes single-source epidemics, i.e. outbreaks that can be linked to a source of contaminated food. This means that alimentation is a risk factor for norovirus. Alimentation may itself be linked to deprivation. We speculate that in deprived areas, there may be less access to safe food; less education on food handling and food safety or less opportunity to adhere to such recommendations as a result of fewer resources; as well as less access to nutritional options as a result of food deserts (impoverished areas with limited access to nutritional and nourishing food options). Due to these conditions, a greater risk of norovirus exposure is found in areas with greater deprivation. Norovirus outbreaks are initiated by contaminated food (single source epidemic), and the virus causes gastroenteritis (inflammation of the mucous membrane of the digestive system linked to vomiting and diarrhoea). The ongoing transmission occurs via the oral-faecal route, meaning that other transmission routes include contact with fomites or with infected people and their bodily fluids (Mathijs et al., 2012). Norovirus poses significant risks in public places such as schools, hospitals, and nursing homes,

✉ Maria Bekker-Nielsen Dunbar, bl328@uni-heidelberg.de

1 Centre for Research on Pandemics & Society, OsloMet – Oslo Metropolitan University, Oslo, Norway

2 Department of Infectious Disease and Tropical Medicine, Heidelberg University Hospital, Heidelberg, Germany

3 Department of Population Health Sciences, Georgia State University, Atlanta, GA, USA

where close contact and shared spaces facilitate its rapid transmission. In such environments, outbreaks can spread swiftly due to the high density of susceptible individuals and the ease of fomite transmission. Studies have shown that public institutions like hospitals and nursing homes are particularly vulnerable, as the virus can cause severe illness in immunocompromised patients and elderly residents (de Graaf et al., 2016). Schools and childcare institutions are also high-risk settings due to frequent interactions among children, who are often less vigilant about hygiene practices (Gaythorpe et al., 2018b). The importance of understanding intrafamily norovirus transmission has been highlighted in household surveys, which demonstrate how easily the virus can spread in close quarters (Juniastuti et al., 2023). In Japan, for example, long-term surveillance has revealed that norovirus transmission dynamics in public settings remain persistent over time, demonstrating the need for ongoing vigilance in these high-risk environments (Misumi and Nishiura, 2021). Incorporating socio-economic and environmental factors into models of norovirus transmission is therefore crucial, as public institutions in deprived areas may experience even higher transmission rates due to overcrowding and limited resources for infection control.

Recent work in England by Buczkowska et al. (2023) found that an increase in deprivation is linked with an increase in norovirus cases. They suggest an almost 300-fold estimate of the true cases meaning that the norovirus cases reported in the surveillance system are expected to be severely under-ascertained and that the surveillance data “under-represents the true burden of norovirus”. At the time of writing, there is no pre-exposure prophylaxis (vaccine) for norovirus, and norovirus spread is not countered by hand sanitiser. Food-borne diseases pose an increasing risk to public health in Germany (Dietrich et al., 2023). The capital of Germany is Berlin. Berlin is the most culturally diverse city in Germany with a long history of migration and an increasing population. Berlin was divided following the Second World War into East Berlin (Soviet Union occupied; DDR) and West Berlin (American occupied, British occupied, and French occupied regions; BRD) with immigration from Turkey in the west and immigration from Viet Nam and North Korea in the east. We expect Berlin to have differences in levels of deprivation as a result of this historic division.

Deprivation indices measure the lack of advantages in a location. Worse health outcomes are found among residents of disadvantaged areas. Deprivation reflects the inequalities in “exposure to the social determinants of health: (. . .) working conditions, unemployment, access to essential good and services” (Bambra et al., 2021). The deprivation amplification theory states that the health effects of individual deprivation are amplified for people who live in areas that are more deprived (Bambra et al., 2023). Deprivation indices synthesise information gathered on these inequalities of exposure and quantify the poverty and disadvantages of an area numerically. The German Index of Socio-economic Deprivation (GISD) was developed in 2017 and updated in 2022 (Michalski et al., 2022). The GISD is constructed on the basis of three socio-economic domains through nine indicators: three educational, three employment-based, and three social indicators. The GISD takes values from 0 to 1, with a higher value indicating greater deprivation. The GISD is a relative measure, meaning that it is defined relative to the regions that have the best and the worst socio-economic situations. We have chosen to use the continuous deprivation score rather than the categorical deprivation index in our work (Altman and Royston, 2006). Germany is a country with

low social mobility (defined as the ability of a person to attain an improved socio-economic status relative to their parents) (OECD, 2018).

This paper makes several key contributions to the current work on infectious disease modelling. First, it introduces the use of the GISD as a covariate in endemic-epidemic models, offering one of the first studies to explicitly account for socio-economic deprivation in this context. Second, it responds to calls for more region-specific analysis by demonstrating how deprivation influences disease transmission at the district level in Berlin. Third, it advances the application of deprivation amplification theory to infectious disease epidemiology, providing empirical support for the theory. Finally, it innovates by integrating socio-economic factors with time-varying contact matrices, thereby enhancing model accuracy and robustness. Our work addresses various calls for further investigation.

- Michalski et al. (2022) noted that “official notification data for various infectious diseases (. . .) represent further data sources, for which socioeconomic inequalities could be analysed by linking them to regional deprivation measures”, which our work addresses;
- The inclusion of seasonality in the epidemic component in Held et al. (2017) was requested by Meyer and Held (2017), which our work addresses;
- Meyer and Held (2017) also called for the use of time-varying contact matrices, which our work addresses through the inclusion of matrices established using the Arregui et al. (2018) method;
- Bambra et al. (2023) called for health impact assessments to also quantify the impact of health equity, and our work responds to this call through the construction of a social epidemiological endemic-epidemic model; and
- We extend the work of Di Biagio et al. (2022) and Dolcini et al. (2022) who considered only the categorical deprivation index in their endemic-epidemic work – in which deprivation was a proxy for indoor air pollution exposure (and its effects on the coronavirus pandemic of the 2020s) – by considering a continuous deprivation score and focusing on it as its own entity, rather than on its use as a proxy measure for indoor air pollution exposure.

Methods

Material (data)

Deprivation index

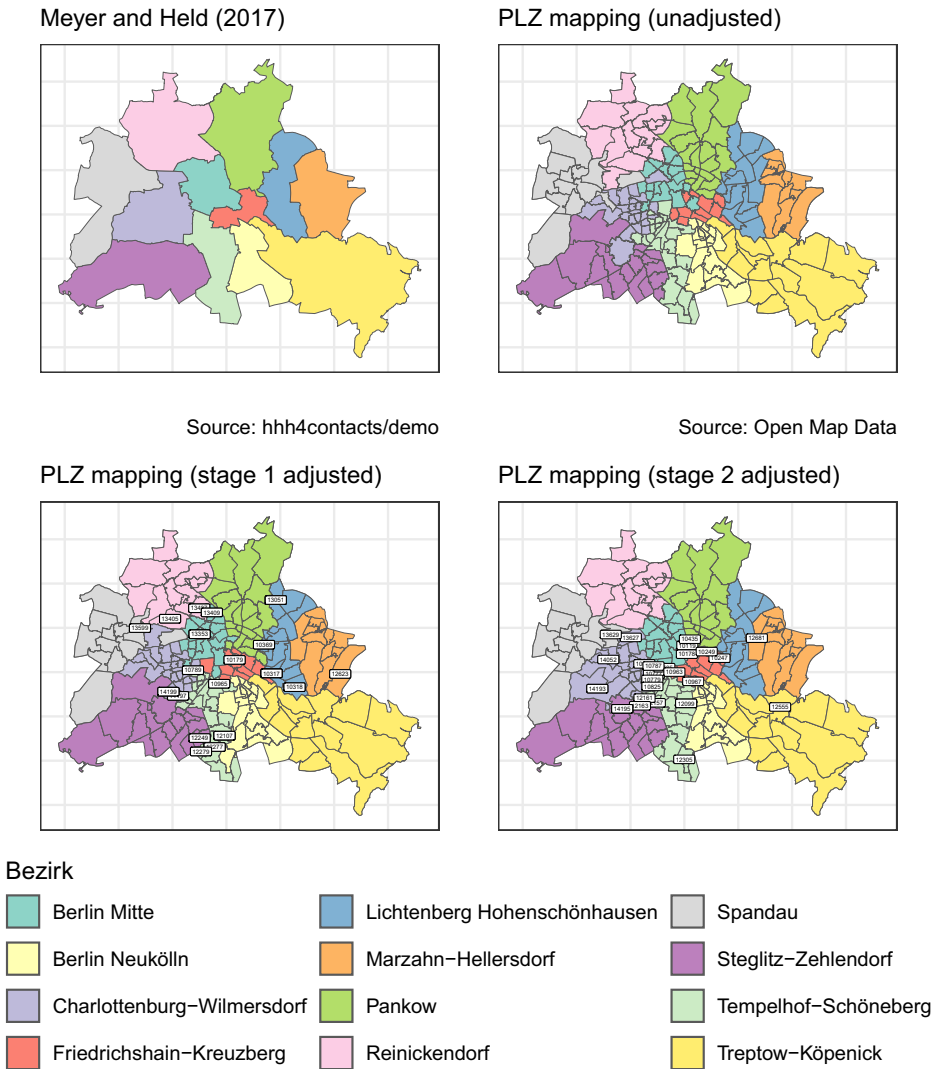
Deprivation indexes are used to capture the multiple dimensions of deprivation in a relative measure. They are used in many countries, including (non-exhaustive list): the United Kingdom, Aotearoa New Zealand, Australia, Canada, South Africa (all English speaking), the United States (English speaking de facto), Germany (relevant for our work and revisited in the discussion), France, Belgium, Switzerland, Italy (all European), and Fiji. We consider the GISD, which is available from Michalski et al. (2022). While others have used the GISD in various work (Table 1), we believe we are the first to use it with a focus on Berlin.

Table 1 Other studies using the GISD

Study	Geography	Coverage
Tetzlaff et al. (2024)	Germany	National
Hoebel et al. (2024)	Germany	National
Reitzle et al. (2024a,b)	Germany	National
Tuncer et al. (2024a,b)	Germany	National
Tetzlaff et al. (2023)	Germany	National
Mercuri et al. (2023)	Four regions	National
Wollschläger et al. (2024)	16 states	National
Rohleder et al. (2022)	401 districts	National
Schultz et al. (2023)	Eight states	Subnational
Yang et al. (2024)	Six states	Subnational
Hanewinkel and Hansen (2024)	13 states	Subnational
Suchert et al. (2023)	13 states	Subnational
Ernst et al. (2023)	One state	Subnational
Moissl et al. (2020)	One city (in three states)	Subnational

In order to use the GISD at the *Bezirk* level (henceforth referred to as “district” as in Meyer and Held (2017) but with the original term retained in reference to data sets used for reproducibility) in Berlin while attaining the same granularity as the surveillance data, we need to link the postcode (PLZ) level GISD (finer granularity) in Berlin to the corresponding district level. As some PLZs are linked to multiple districts we perform a two-stage adjustment (Figure 1). Stage 1 consists of linking the PLZ to the most frequently observed district when tallying and stage 2 consists of a manual adjustment to link the postcode to the district from Meyer and Held (2017). This means that for each PLZ, we investigate whether it is linked with more than one district, and if so which districts it was considered a part of. We then calculate the total number of times the PLZ is linked to each district, and link it to the one with the higher total number. For example, the postcode PLZ 14199 is linked to district Charlottenburg-Wilmersdorf three times (3) and to Steglitz-Zehlendorf once (1), so we link it to Charlottenburg-Wilmersdorf, as this link occurs more frequently. Stage 1 requires the manual adjustments of 19 PLZ codes and ensures that the districts are disjoint units, while stage 2 requires the manual adjustments of 25 PLZ codes and ensures coherence with the established literature. The postcodes that are affected by the adjustments are labelled in Figure 1. For example, PLZ 14195 is considered part of Charlottenburg-Wilmersdorf (the lilac-coloured region), whereas it should be considered part of Steglitz-Zehlendorf (the violet-coloured region) based on visual inspection of the Meyer and Held (2017) plot (top-left panel). For the purposes of completeness and reproducibility, we provide.csv files of the adjustments and an additional plot in the supplementary material (available online at <https://doi.org/10.1553/p-4b4e-mkcd>). This two-stage adjustment provides us with disjoint regions such that when we calculate the GISD score at the district level it is not influenced by neighbouring districts.

Figure 1 PLZ linking to match the GISD given at the PLZ level to the district level



After this PLZ linking process, we are able to determine the weighted average GISD score for the district for each year in the study period (Figure 2). The average GISD score is weighted by the population in the postcode. This population information is provided with the GISD data at the PLZ level, and does not vary over the study period (2011 week 27 to 2015 week 26), unlike the district-level population shown in the population pyramids in Figure 3. There is seemingly no overall pattern to the districts (Table 2). The largest

Figure 2 GISD score by year and population used to calculate the weighted average GISD score at the district level

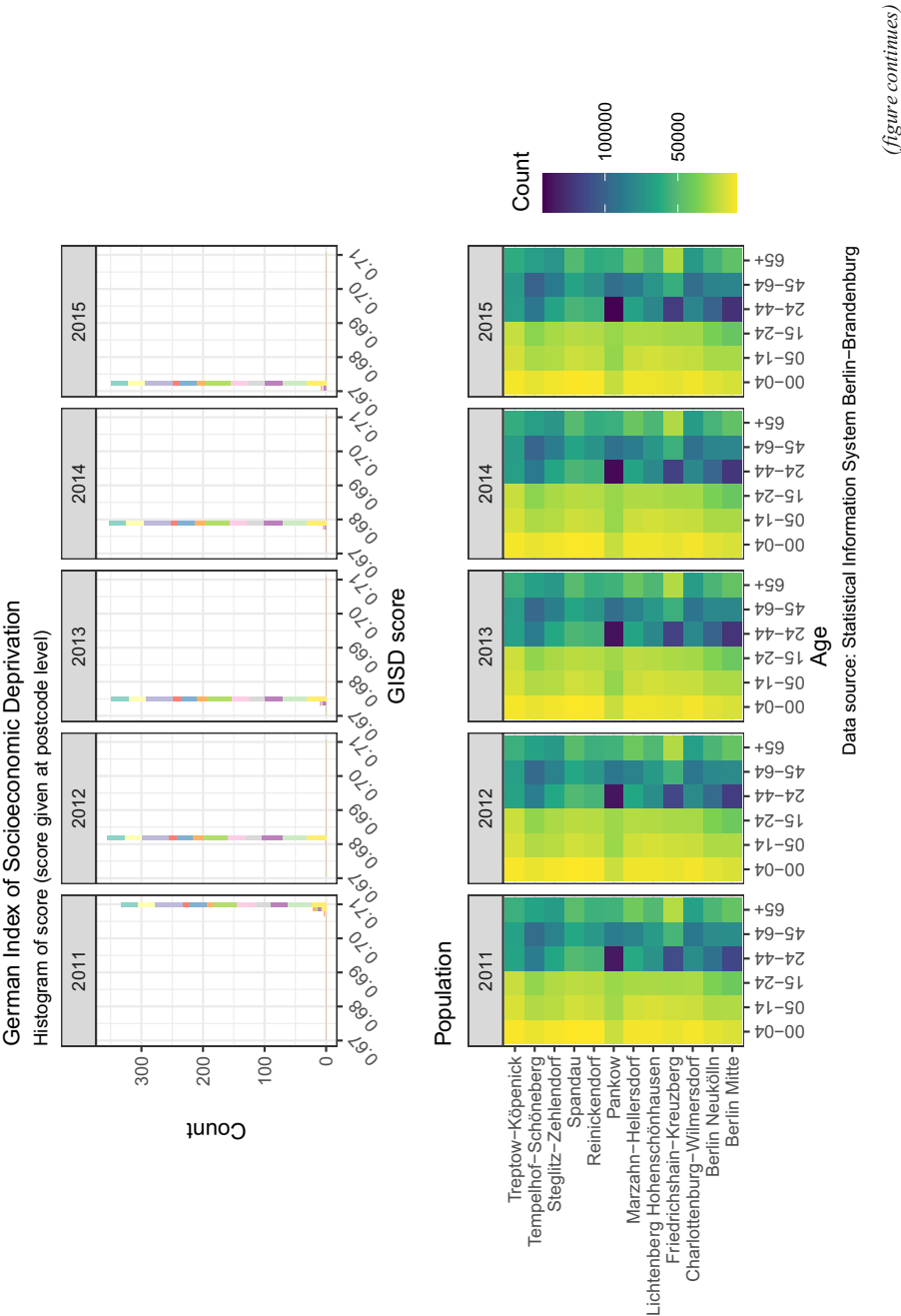


Figure 2 (continued)

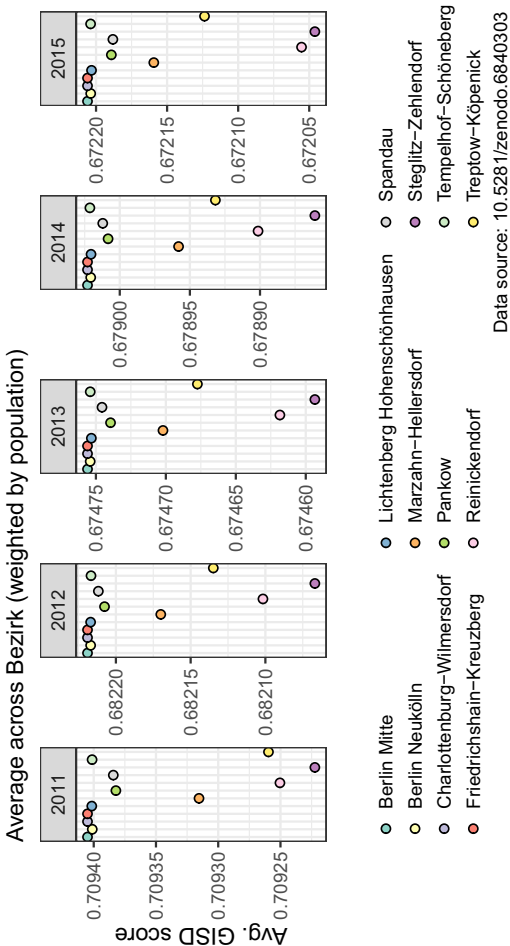


Figure 3 POLYMOD contact matrices updated according to the method by Aregui et al. (2018) using the population data from the Statistical Information System Berlin-Brandenburg. Values of select entries of the contact matrices are shown similarly to those in Meyer and Held (2017): one on the diagonal and one off the diagonal

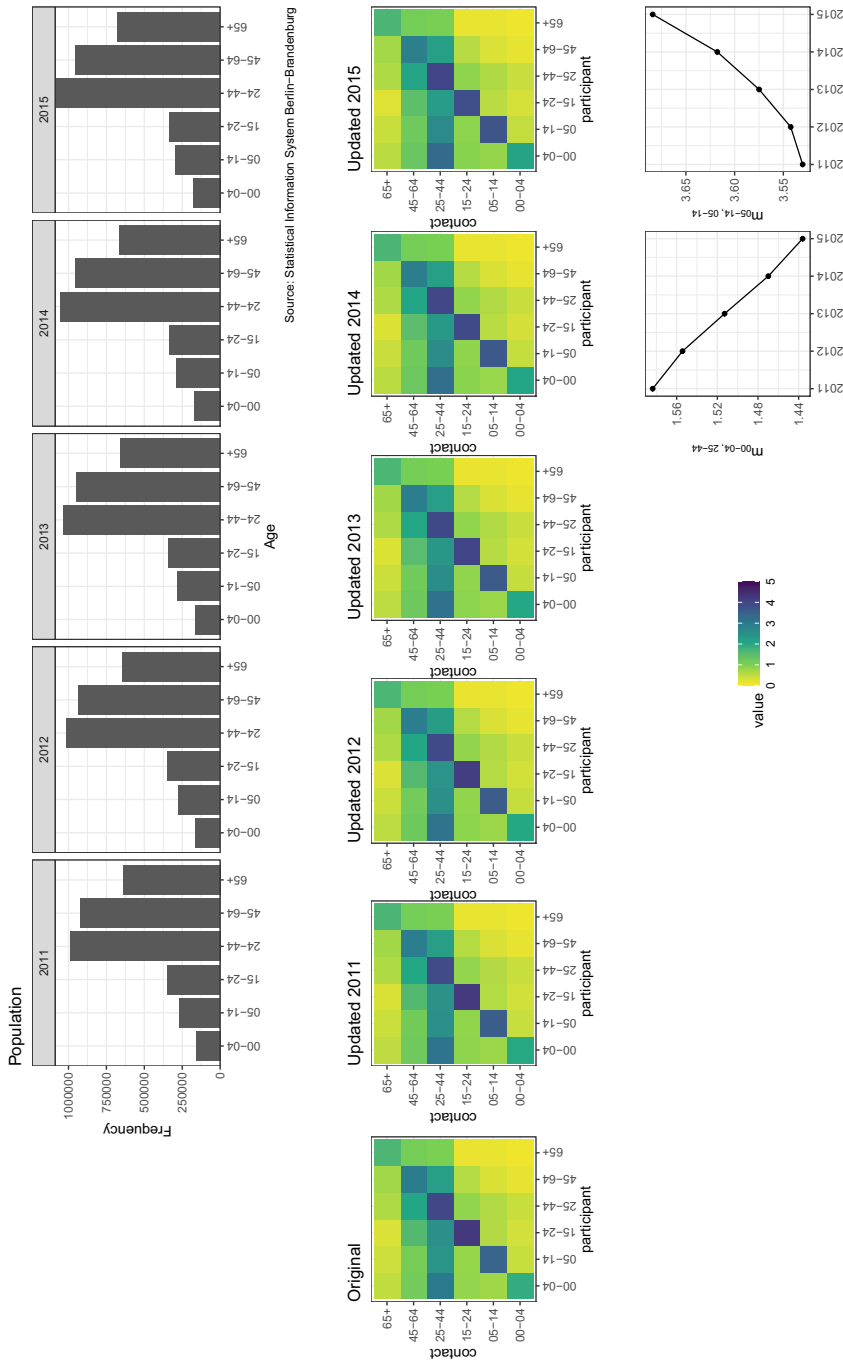


Table 2 Properties of districts. The average GISD across the study period is ranked from 1 (least deprived) to 12 (most deprived) and the population is the average count over the study period. Information on district size is from the data set “Flächennutzung im Land Berlin – Bezirke”

Bezirk	GISD (avg.)	Rank (GISD)	Population (avg.)	Size (km ²)	Density
Berlin Mitte	0.6835218	1	55772.97	39.47	1413.0470
Berlin Neukölln	0.6835193	5	52026.33	44.93	1157.9420
Charlottenburg-Wilmersdorf	0.6835218	2	50396.30	64.72	778.6820
Friedrichshain-Kreuzberg	0.6835218	3	43820.23	20.41	2146.9982
Lichtenberg Hohenschönhausen	0.6835191	6	43697.77	52.12	838.4069
Marzahn-Hellersdorf	0.6834610	9	42018.17	61.78	680.1257
Pankow	0.6835054	8	61685.03	103.06	598.5352
Reinickendorf	0.6833855	11	41116.07	89.31	460.3747
Spandau	0.6835084	7	37037.83	91.87	403.1548
Steglitz-Zehlendorf	0.6833579	12	47604.23	102.56	464.1598
Tempelhof-Schöneberg	0.6835194	4	54340.67	53.03	1024.7156
Treptow-Köpenick	0.6834255	10	40877.73	168.42	242.7131

and the smallest districts are adjacent and the most and the least populous districts have a similar average GISD.

Deprivation increases the likelihood of worse health outcomes, which our model will reflect. However, if more deprived areas are more isolated or have lower contact rates, this will reduce transmission and delay the onset of epidemics. We expect to see a difference in deprivation levels between the east and the west as Berlin was historically divided into two parts. The Berlin Wall surrounded the western part of Berlin, and was erected on 13 August 1961 with the aim of restricting travel from east to west. The wall fell on 9 November 1989 (the wall stood for 10,315 days). While the wall is no longer present, certain aspects of the city (e.g. its buildings) remain impacted by the historical division to this day.

Social contact

As norovirus has the potential to spread between people, we use the POLYMOD contact matrices to capture contacts between members of different age groups as in Meyer and Held (2017) and Held et al. (2017). We use the density correction from Arregui et al. (2018) to update them to match the demographics of the study period under consideration (2011 week 27 to 2015 week 26). This means we calculate

$$m'_{st} = m_{st} \frac{nn'_t}{n_t n'} \quad (1)$$

where m_{st} denotes the entries of original POLYMOD contact matrix used in previous endemic-epidemic analyses, n_t is the old or unadjusted population in age group t and

Table 3 Population values n' by age group used for updating contact matrices (2014 and 2015 values estimated from redistributing the total population)

	2011	2012	2013	2014	2015
00–04	159143	163051	166182	171814	178152
05–14	267983	272892	279194	286509	295979
15–24	346244	344858	340305	335273	332303
25–44	990766	1010909	1031211	1053302	1077836
45–64	921310	936099	949327	955862	957626
65+	640556	647413	655610	667089	678135
Total	3326002	3375222	3421829	3469849	3520031

$n = \sum_t n_t$, and prime denotes the newer or adjusted demographics. The contact matrix represents the average number of reported contacts, i.e. m_{st} is the average number of contacts age group s has with t . The calculation gives us an updated contact matrix with entries m'_{st} for the average number of contacts age group s has with age group t after demographic changes. The matrices are shown in Figure 3. Contacts are a rough proxy for transmission opportunities.

The population data used to update the matrices comes from the Statistical Information System Berlin-Brandenburg (<https://statistik-statistik-berlin-brandenburg.de>) as this is the source used by Meyer and Held (2017). The data set used is the “Bevölkerungsstand – Berliner Bezirke u. Brandenburger Kreise nach Geschlecht und Jahr” data set. We summarise this population to the age groups considered in the work and use this to inform n_t in (1). The result of incorporating the newer demographics (Table 3) to update contact matrices is given in Figure 3. Note that for legal reasons, the Statistical Information System Berlin-Brandenburg can only provide population counts for Berlin in total (not stratified by district) from 2014 onwards. To obtain the district-level population for 2014 and 2015, we redistribute the total population given according to the fractions from previous years which provides an estimate of the district-level population.

Mixing

Summary measures that seek to describe the pattern have been developed for contact matrices. Most of the summary measures consider the value on or off the diagonal of the matrix. Assortative mixing represents “like mixing with like”, which for contact matrices means that people are more likely to interact with their peer group (in age) and that the greatest values are found on the diagonal. The index of disassortativity (Farrington et al., 2009; Wallinga et al., 2019) considers off-diagonal matrix elements. A greater value of the index of disassortativity means that people mix more with other age groups than with the same age group. The continuous version of the index of disassortativity is presented in Farrington et al. (2009), and Wallinga et al. (2019) outline the discrete version of I_s^2 which we use in our work as we analyse age groups rather than age in years. An additional summary measure

is the Q -statistic (Aral et al., 1999), which measures assortativity. $Q < 1$ indicates disassortative mixing and $Q > 1$ indicates assortative mixing. An alternative assortativity measure that considers more than just the diagonal matrix entries is called r (Newman, 2003). In contrast to Q which considers only the diagonal elements and the size of the contact matrix, r is weighted by the rows and columns of the matrix, providing additional nuance. A third summary measure q (Keeling and Rohani, 2008) considers the ratio of the second-largest to the largest eigenvalue, rather than entries of the matrix (incorrectly listed as λ_1/λ_2 in Bekker-Nielsen Dunbar, 2024).

We calculate the summary measures for the updated contact matrices given in Figure 3. The summary measures indicate that the mixing patterns observed in the study period as a result of demographic updating become more disassortative over time based on their placement on the scale of possible values Q , r , q , and I_s^2 can take (Figure 4). The full suite of summary measure values for our updated contact matrices can be found in Figure 4, which allows for a visual inspection. We observe an increase in the index of disassortativity I over the study period, meaning that the population mixes more heterogeneously over time (as the indices are constructed in comparison to homogeneous mixing), which is also seen in the decrease in assortativity measures Q , r , and q . We observe a decrease in the degree of assortative mixing over time as the values of Q decrease which also indicates more heterogeneous mixing over the study period (Figure 4).

Disease surveillance

The data for laboratory-confirmed norovirus cases comes from the Robert Koch Institute (for more information, see: Meyer and Held, 2017; Held et al., 2017; Bracher and Held, 2022). Our study period is 2011 week 27 to 2015 week 26. These data cover four norovirus seasons, and the number of weeks included for the years 2011 and 2015 is less than that for the other three years (2012, 2013, and 2014) as shown in Figure 5. The second norovirus season (i.e. winter 2012/2013) seems to have fewer cases than the other three based on visual inspection. When plotting the cases by age group and district, as shown in Figure 5, we see that the cases are distributed unevenly across age groups and that district Steglitz-Zehlendorf seems to have the greatest burden. We note that this district has the lowest average GISD score for the study period (Table 2), and is thus the most deprived.

Model

The endemic-epidemic model is a known statistical framework developed for infectious disease surveillance. An overview of the previous endemic-epidemic approaches to norovirus modelling is provided in Table 4. The original model for norovirus incidence in Berlin was presented in Meyer and Held (2017). This model was extended by Held et al. (2017) who included seasonal sine-cosine waves in the endemic component of the model as this extension was highlighted as an option for further work by Meyer and Held (2017). Held et al. (2017) used the data from Meyer and Held (2017) as training data for their predictions.

Figure 4 Values of Q plotted on the scale from purely disassortative mixing (limit: $-1/(k-1)$) to purely assortative mixing (limit: 1) in the style of Garnett et al. (1996) as well as values of r , q , and I_s^2 plotted on a scale from 0 to 1

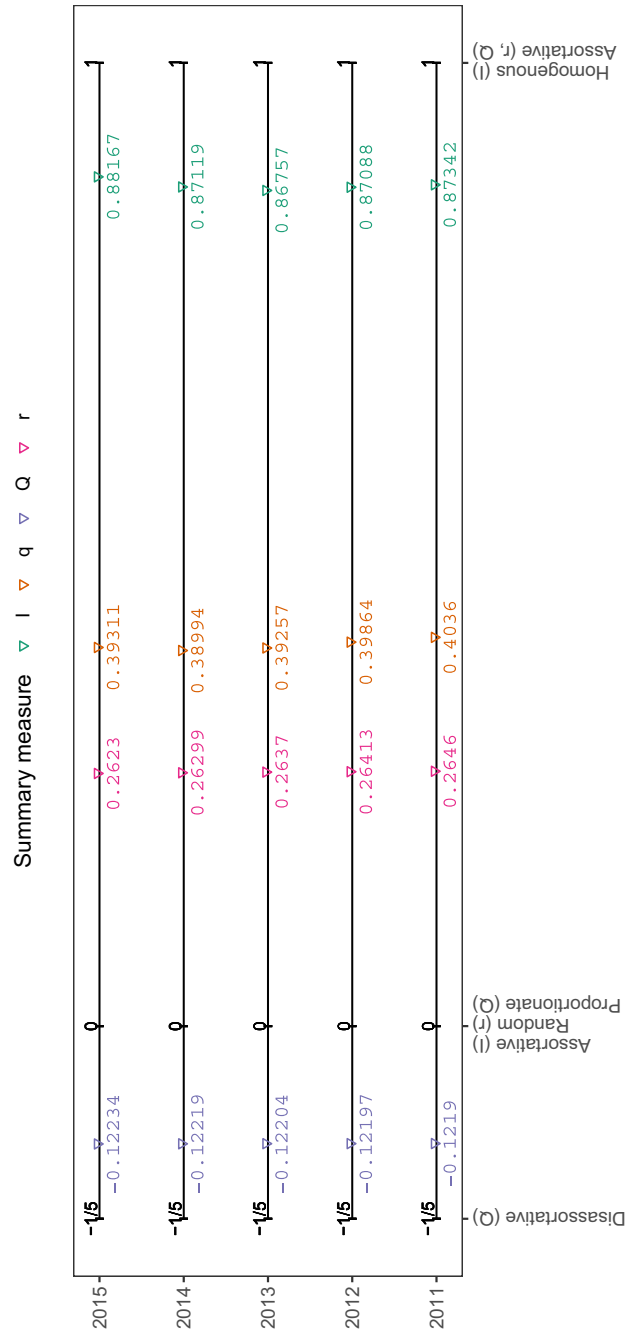
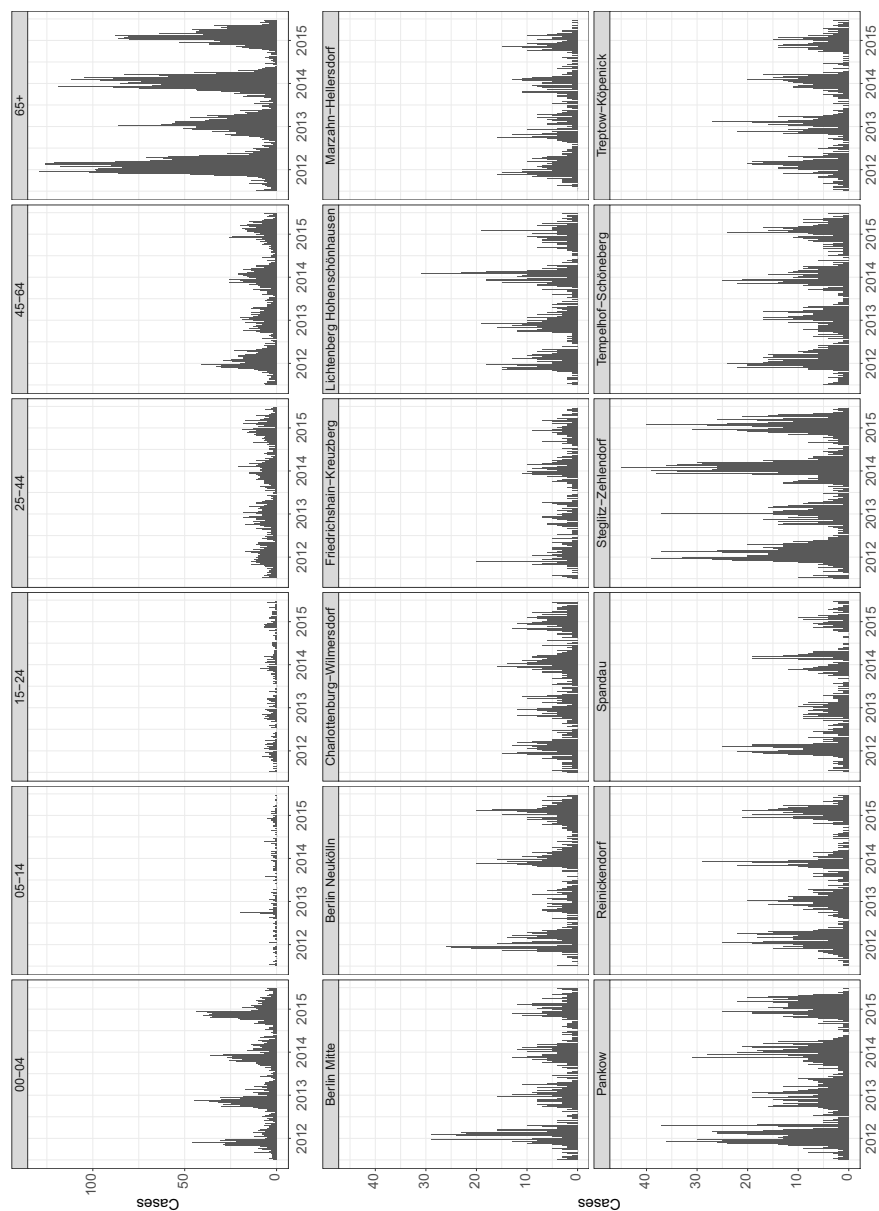


Figure 5 Cases observed in the study period by age group (top row) and districts (bottom two rows)



Data source: hhh4contacts/demo

Table 4 Characteristics of previous norovirus studies

	Meyer and Held (2017)	Held et al. (2017)	Bracher and Held (2022)
Study period w	2011 week 27 to 2015 week 26	2011 week 27 to 2016 week 26	2011 week 1 to 2017 week 52
Age group a	0–4, 5–14, 15–24, 25–44, 45–64, 65+	0–4, 5–14, 15–24, 25–44, 45–64, 65+	—
District b	12 districts	12 districts	12 districts
Model fit evaluation	AIC	AIC	AIC
Verification analysis criterion	—	Prediction (ranked probability score)	Prediction (log score)

Bracher and Held (2022) also explored predictions and evaluated the impact of including a distributional assumption for the lagged observations.

We will compare the models from the three studies (Table 4) with the two inclusions provided in this work:

1. The inclusion of the GISD as a model covariate (Figure 2); and
2. The inclusion of the Arregui et al. (2018)-updated contact matrices (Figure 3).

In the endemic-epidemic model for norovirus incidence in Berlin, cases in age group a , in district b , in week w are conditional on the past $w_{\max}$ weeks' surveillance data (i.e. cases) assumed to follow a negative binomial distribution with mean λ and overdispersion ψ

$$Y_{abw} | \mathbf{Y}_{w-1}, \mathbf{Y}_{w-2}, \dots, \mathbf{Y}_{w-w_{\max}} \sim \text{NegBin}(\lambda_{abw}, \psi_a) \quad (2)$$

where $\mathbf{Y}$ denotes cases in all age group and district combinations. The mean is given by

$$\lambda_{abw} = \nu_{abw} n_{ab} / \sum_{a,b} n_{ab} + \phi_{abw} \sum_{w'=1}^{w_{\max}} \sum_{a',b'} u_w c_{a'aw'} r_{b'b} Y_{a'b'w-w'} \quad (3)$$

where $n_{ab} / \sum_{a,b} n_{ab}$ is the population fraction in age group a and district b , ν_{abw} denotes the endemic predictor (described below), ϕ_{abw} denotes the epidemic predictor (described below), u_w denotes the geometric serial interval lag distribution introduced by Bracher and Held (2022), given by

$$u_w = (1 - \kappa)^{w-1} \kappa \quad (4)$$

with a maximum lag of $w_{\max} \equiv 5$ (for norovirus surveillance models cf. Bracher and Held, 2022), $c_{a'aw'}$ is the contact matrix in week w (Figure 3), and $r_{b'b}$ is the connectivity between the districts from Meyer and Held (2017). We estimate κ from the model but have chosen the geometric lag distribution as the serial interval for norovirus is around four days and we therefore place most of the weight on the preceding week's cases. A plot of the estimated distribution is provided in the supplementary material.

The endemic log-linear predictor ν is given by

$$\log(\nu_{abw}) = \alpha^{(\nu)} + \alpha_a^{(\nu)} + \alpha_b^{(\nu)} + \beta x_w + \gamma_a^{(\nu)} \sin\left(\frac{2\pi w}{52}\right) + \delta_a^{(\nu)} \cos\left(\frac{2\pi w}{52}\right) + \xi_b^{(\nu)} z_{bw} \quad (5)$$

with intercept α with reference level age group zero to four years old (00-04) and district Charlottenburg-Wilmersdorf, with age group-specific effects α_a and district-specific effects α_b providing and being interpreted as contrasts to the reference level, effects of holiday season β (since the epidemiological calendar for norovirus includes winter, and behaviour and exposure levels may change during festivities) for the festive indicator x_w , given by

$$x_w = \begin{cases} 1 & w \text{ is week 52 or week 1} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

age group-specific seasonal effects γ_a and δ_a , and ξ_b is the effect of deprivation for the GISD score z_{bw} for district b in week w . The epidemic log-linear predictor ϕ is given by

$$\log(\phi_{abw}) = \alpha^{(\phi)} + \alpha_a^{(\phi)} + \alpha_b^{(\phi)} + \theta \log(n_b) + \gamma_a^{(\phi)} \sin\left(\frac{2\pi w}{52}\right) + \delta_a^{(\phi)} \cos\left(\frac{2\pi w}{52}\right) + \xi_b^{(\phi)} z_{bw} \quad (7)$$

where θ denotes popularity (to reflect the attraction to population centres).

We compare the performance of this model with that of models without z_{bw} (our primary goal), with $c_{a'a}$ in place of $c_{a'aw}$ (our secondary goal), without u_w (for comparison with the addition of weighted lags from Bracher and Held (2022)), without $\gamma_a^{(\phi)}$ and $\delta_a^{(\phi)}$ (for comparison with the addition of seasonality in the epidemic component from Held et al. (2017)), and without θ (for comparison with the addition of a measure of urbanicity from Meyer and Held (2017)). We use Akaike's information criterion (AIC) for model comparison. Models are fit using 2,000 iterations, with a tolerance limit of 0.001, and the Nelder-Mead algorithm is used for fitting the variance.

Results

While the initial results suggested that the default model including every detail had the best fit in terms of AIC, we thought that given the unique historic situation of Berlin (having been split by a physical barrier) the Berlin Wall (Figure 6) might be a stronger determinant of deprivation amplification than the GISD. We considered the inclusion of the historical east and west parts of Berlin in place of the deprivation index to see if this provided a better explanation of the specific differences in Berlin. To investigate this question, we determined which districts were on each side of the wall and categorised this division as a variable. In cases in which the wall was located within the district, we assigned the district to a third category of "both" (meaning both sides), and expected that any conclusions drawn

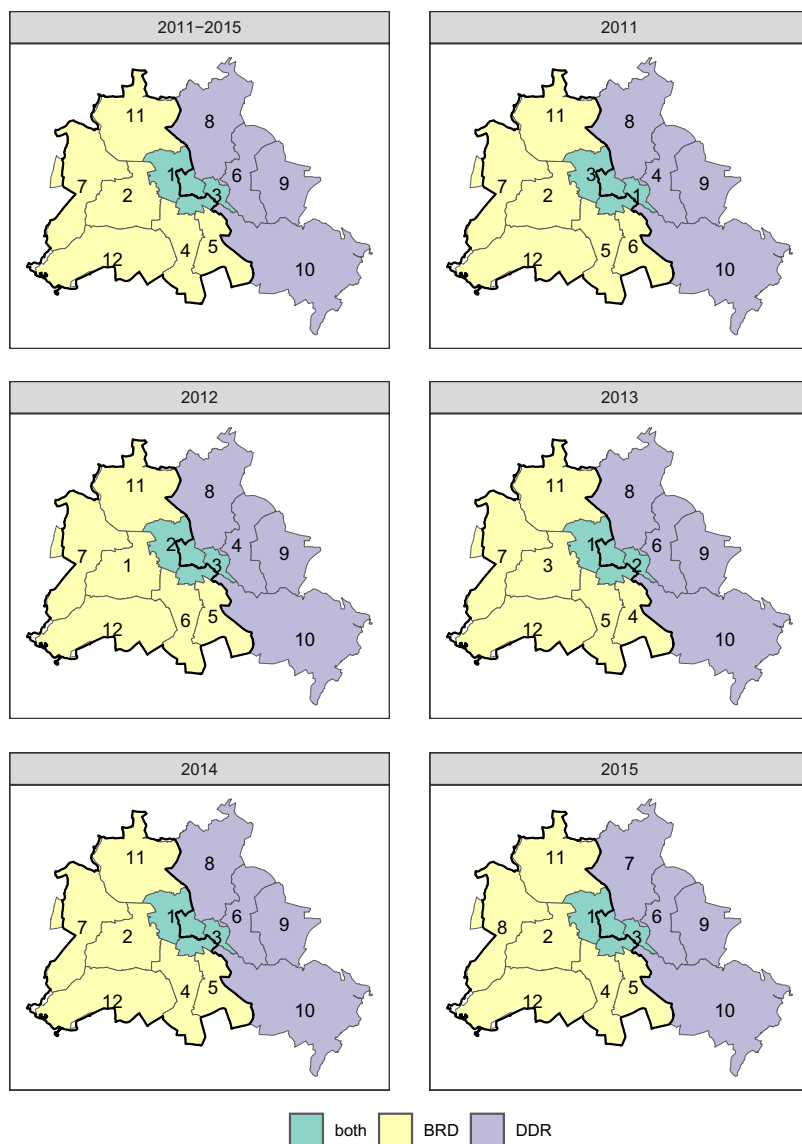
Figure 6 Berlin Wall (marked as a fully drawn line), contemporary Berlin districts boundaries, and ranking of average GISD over the study period (1 is least deprived and 12 is most deprived) as well as by individual year

Ranking of average GISD

1 is least deprived

12 is most deprived

Wall boundaries drawn overlaid



about the effect of the wall would be a bit more complicated for those districts. Rather than the east-west divide we had expected to find, we instead observed a “halo” effect, whereby the innermost districts were less deprived than those surrounding them (the “landlocked” regions in Figure 6 with ranks 1–3). We elected to examine a model with solely an effect of the GISD

$$\log(\nu_{abw}) = \alpha^{(\nu)} + \xi_b^{(\nu)} z_{bw} \quad (8)$$

$$\log(\phi_{abw}) = \alpha^{(\phi)} + \xi_b^{(\phi)} z_{bw} \quad (9)$$

as well as a model with only the historic east and west parts of Berlin (Figure 6), and a model with both for completion.

After this investigation showed that the model with the historical division of the Berlin Wall did not seem to perform better than our initial model, and given that our work is focused on deprivation in general, we considered the model suite in which the GISD appears only in the epidemic component in line with previous work considering such an effect by Di Biagio et al. (2022).

This means that we exclude $\xi^{(\nu)}$ and replace $c_{a'aw}$ with the original $c_{a'a}$ in our full model. All other effects remain as originally intended. The inclusion of time-varying transmission weights (the demographically updated contact matrices $c_{a'aw}$) does not seem necessary according to our results. We then calculated the relative log likelihood L/L_{default} where L denotes the likelihood. We considered calculating the relative likelihood from the AIC as

$$\exp\left(\frac{AIC - AIC_{\text{default}}}{2}\right) = \exp(\kappa) \frac{L}{L_{\text{default}}} \quad (10)$$

but we are not interested in the exponentiated difference in the number of parameters between the two models $\exp(\kappa) = \exp(k - k_{\text{default}})$ which is extremely large.

This is also a pragmatic choice for the situation under investigation given that there seems to be collinearity in the GISD coefficient in the endemic component (Table 5), as the coefficient estimates and associated standard errors are very large and might be inflated. Inclusion of the wall in place of the GISD did not solve the problem of diverging effects for deprivation in the endemic component (in Table 5 the estimates have even greater absolute values when the wall is included in addition to the GISD).

To address the diverging coefficients of GISD, we consider instead the final model with log-linear predictors

$$\log(\nu_{abw}) = \alpha^{(\nu)} + \alpha_a^{(\nu)} + \alpha_b^{(\nu)} + \beta x_w + \gamma_a^{(\nu)} \sin\left(\frac{2\pi w}{52}\right) + \delta_a^{(\nu)} \cos\left(\frac{2\pi w}{52}\right) \quad (11)$$

and

$$\log(\phi_{abw}) = \alpha^{(\phi)} + \alpha_a^{(\phi)} + \theta \log(n_b) + \gamma_a^{(\phi)} \sin\left(\frac{2\pi w}{52}\right) + \delta_a^{(\phi)} \cos\left(\frac{2\pi w}{52}\right) + \xi z_{bw} \quad (12)$$

Table 5 Model coefficients for the model with the GISD, the model with the Berlin Wall, and the model with both

	Model with historical division	Model with GISD	Model with both
$-\log(\psi_{00-04})$	0.3469651	0.4305772	0.4439486
$-\log(\psi_{05-14})$	-2.7485215	-2.7425023	-2.7414454
$-\log(\psi_{15-24})$	-1.8214928	-1.8031098	-1.8028735
$-\log(\psi_{25-44})$	-0.9838918	-0.9609265	-0.9600564
$-\log(\psi_{45-64})$	-0.2137250	-0.1545979	-0.1559667
$-\log(\psi_{65+})$	-0.3569037	-0.1168705	-0.1163549
$\alpha^{(\nu)}$	-12.8726431	119.2936438	20.0828292
$\xi^{(\nu)}$ Charlottenburg-Wilmersdorf		-172.7469628	2.4379790
$\xi^{(\nu)}$ Friedrichshain-Kreuzberg		-180.1818905	-30.1377231
$\xi^{(\nu)}$ Lichtenberg Hohenschönhausen		-198.9586816	-329.2994905
$\xi^{(\nu)}$ Marzahn-Hellersdorf		-198.7755820	-330.4153917
$\xi^{(\nu)}$ Berlin Mitte		-198.7279356	-53.9850085
$\xi^{(\nu)}$ Berlin Neukölln		-176.0408282	-26.0247625
$\xi^{(\nu)}$ Pankow		-198.9835305	-329.7287833
$\xi^{(\nu)}$ Reinickendorf		-172.2364939	1.7158731
$\xi^{(\nu)}$ Tempelhof-Schöneberg		-173.5717794	-304.5507942
$\xi^{(\nu)}$ Spandau		-197.6080889	-1.3719600
$\xi^{(\nu)}$ Treptow-Köpenick		-174.8899770	1.8304446
$\xi^{(\nu)}$ Steglitz-Zehlendorf		-173.3813890	0.5121334
Division _{DDR} ^(ν)	10.5374374		187.6287123
Division _{BRD} ^(ν)	14.6853045		-19.1832564
$\alpha^{(\phi)}$	-5.3111100	-8.8650783	-11.3125823
$\xi^{(\phi)}$ Charlottenburg-Wilmersdorf		5.1042249	-3.3444859
$\xi^{(\phi)}$ Friedrichshain-Kreuzberg		4.8011986	8.3657433
$\xi^{(\phi)}$ Lichtenberg Hohenschönhausen		5.6116524	8.4787120
$\xi^{(\phi)}$ Marzahn-Hellersdorf		5.1043175	7.9701659
$\xi^{(\phi)}$ Berlin Mitte		5.5693853	9.1595776
$\xi^{(\phi)}$ Berlin Neukölln		5.2917933	-2.9812313
$\xi^{(\phi)}$ Pankow		6.5715793	9.4397343
$\xi^{(\phi)}$ Reinickendorf		5.4564609	-2.7275413

(table continues)

Table 5 (continued)

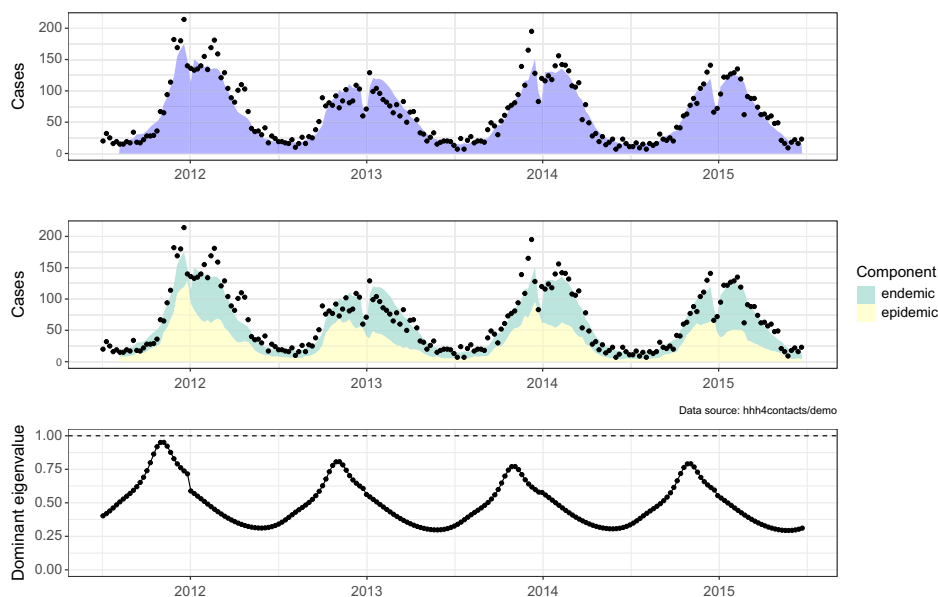
	Model with historical division	Model with GISD	Model with both
$\xi^{(\phi)}$ Tempelhof-Schöneberg		5.8924530	8.7601068
$\xi^{(\phi)}$ Spandau		5.0930752	−3.2297854
$\xi^{(\phi)}$ Treptow-Köpenick		5.5744138	−2.8674592
$\xi^{(\phi)}$ Steglitz-Zehlendorf		6.4468041	−1.8287594
Division _{DDR} ^(ϕ)	0.4519273		0.4932142
Division _{BRD} ^(ϕ)	0.2001090		8.1108283

This model has an AIC of 28,631.64 (other values are given in Table 6 for comparison). The fit of the model is shown in Figure 7 and the model coefficients are given in Table 7. We see that the model shows greater relative influence of the endemic component than usual endemic-epidemic models. This may be indicative of the nature of the disease, which leads to single source outbreaks within a stratum, rather than of person-to-person spread across strata. We recall for the interpretation that the 65+ age group has the most cases followed by the 0–4 age group while the 5–14 and 15–24 age groups have the fewest cases. By district, the most cases seem to be observed in Steglitz-Zehlendorf and Pankow, and the fewest cases are observed in Charlottenburg-Wilmersdorf and Friedrichshain-Kreuzberg (Figure 5). Because the model has log-transformed predictors we can consider the exponentiated values as the multiplicative term of an expected increase when observed cases increase by 1, whereby values close to 1 indicate low-or-no association. See the supplementary material for visualisations of the coefficients on this scale.

Deprivation $\hat{\xi}$ has the greatest effect. We see a negative effect corresponding to a small multiplicative increase for the holiday season $\hat{\beta}$ which is likely responsible for the dips in overall cases observed around the start of a new calendar year (Figure 7). Reductions in

Table 6 AIC of models with only the GISD or the Berlin Wall (above) and AIC of models fitted (below)

	d	AIC		
Model with GISD	32	33337.78		
Model with historical division	12	33772.08		
Model with both	36	33337.29		
	AIC	Δ AIC	AIC/AIC(default)	L/L(default)
Default	28444.58	0.00000	1.000000	1.000000
Without higher order lag u_w	28900.44	455.86025	1.016026	1.016187
Without epidemic seasonality $\gamma^{(\phi)}, \delta^{(\phi)}$	28498.19	53.61689	1.001885	1.002744
Without popularity θ	28464.99	20.41868	1.000718	1.000793

Figure 7 Model fit

norovirus cases in this period are likely due to people eating out less and eating more home-cooked meals, and to people mixing with a smaller group. The estimate for popularity $\hat{\theta}$ suggests more cases would be expected to reach areas with larger populations. The greatest effect of age in the epidemic compartment is $\hat{\alpha}_{65+}^{(\phi)}$. This is the only age effect greater than the intercept in either component (we have provided a plot of the effects relative to the reference level in the supplementary material, which shows that this effect is the only one that is higher than the reference). Besides this effect, the largest effect is found in the 0–4 age group. This suggests an association of infancy and old age with norovirus.

Regarding seasonal effects, re-parameterising the sine-cosine waves in terms of amplitude (variability in size of peak) and phase (variability in timing), we see that the oldest age group (65+) has a different seasonality pattern than the other age groups in both components. This is likely due to how different the older age group's pattern of observed cases is (Figure 5), with its maximum observed counts being more than twice as large as those of other groups. The waves show period patterns that are not flat, indicating that the temporal association with cases is seasonal, which is not surprising when looking at the oscillating peaks and troughs in the observed cases. There seems to be a greater association of seasonality for the 0–14 and 15–24 age groups, with the latter having the highest amplitude overall. The effects of districts are lowest for Friedrichshain-Kreuzberg, Berlin-Mitte, and Berlin-Neukölln, two of which are the most central districts, and could thus be responsible for the “halo” effect.

Table 7 Model coefficients

	Estimate	Std. error		Estimate	Std. error
$\alpha^{(\nu)}$	4.7068638	0.0998811	$\alpha^{(\phi)}$	-4.4699886	1.7640301
$\alpha_{05-14}^{(\nu)}$	-3.0790889	0.5438178	$\alpha_{05-14}^{(\phi)}$	-0.1398778	0.8290772
$\alpha_{15-24}^{(\nu)}$	-2.1445130	0.1121555	$\alpha_{15-24}^{(\phi)}$	-4.0077497	2.1619985
$\alpha_{25-44}^{(\nu)}$	-2.2818494	0.1326340	$\alpha_{25-44}^{(\phi)}$	-2.0785497	0.7420686
$\alpha_{45-64}^{(\nu)}$	-2.1928050	0.1601284	$\alpha_{45-64}^{(\phi)}$	-1.2503831	0.7404071
$\alpha_{65+}^{(\nu)}$	-1.2686618	0.1964524	$\alpha_{65+}^{(\phi)}$	1.0465947	0.7160480
$\alpha_{\text{Berlin Mitte}}^{(\nu)}$	-0.1322536	0.1072115	$\gamma_{00-04}^{(\phi)}$	2.9398249	0.5098491
$\alpha_{\text{Berlin Neukölln}}^{(\nu)}$	-0.2491703	0.1168475	$\gamma_{05-14}^{(\phi)}$	1.1284148	0.4156349
$\alpha_{\text{Friedrichshain-Kreuzberg}}^{(\nu)}$	-0.2360321	0.1134170	$\gamma_{15-24}^{(\phi)}$	3.6153548	1.6997272
$\alpha_{\text{Lichtenberg Hohenschönhausen}}^{(\nu)}$	0.2950535	0.1046013	$\gamma_{25-44}^{(\phi)}$	-0.4946121	0.2569458
$\alpha_{\text{Marzahn-Hellersdorf}}^{(\nu)}$	0.1161014	0.1142798	$\gamma_{45-64}^{(\phi)}$	0.6915391	0.1832723
$\alpha_{\text{Pankow}}^{(\nu)}$	0.5423598	0.0966234	$\gamma_{65+}^{(\phi)}$	0.1992303	0.1214413
$\alpha_{\text{Reinickendorf}}^{(\nu)}$	0.3368040	0.1059892	$\delta_{00-04}^{(\phi)}$	-1.7674017	0.5084232
$\alpha_{\text{Spandau}}^{(\nu)}$	0.1541714	0.1088006	$\delta_{05-14}^{(\phi)}$	1.1037493	0.2939696
$\alpha_{\text{Steglitz-Zehlendorf}}^{(\nu)}$	0.8007635	0.0971840	$\delta_{15-24}^{(\phi)}$	-1.8595850	1.3444790
$\alpha_{\text{Tempelhof-Schöneberg}}^{(\nu)}$	0.3120494	0.0978066	$\delta_{25-44}^{(\phi)}$	-0.0295585	0.1802024
$\alpha_{\text{Treptow-Köpenick}}^{(\nu)}$	0.1848273	0.1113172	$\delta_{45-64}^{(\phi)}$	-0.3434015	0.1836602
$\gamma_{00-04}^{(\nu)}$	0.0871183	0.0954323	$\delta_{65+}^{(\phi)}$	-0.2949957	0.1420252
$\gamma_{05-14}^{(\nu)}$	-0.2713383	0.3212306	ψ_{00-04}	0.2835049	0.0410553
$\gamma_{15-24}^{(\nu)}$	-0.1920094	0.1209598	ψ_{05-14}	1.9680814	0.4971350
$\gamma_{25-44}^{(\nu)}$	0.3925678	0.1209100	ψ_{15-24}	0.3013989	0.1652700
$\gamma_{45-64}^{(\nu)}$	-0.6818247	0.1342511	ψ_{25-44}	0.0426622	0.0435470
$\gamma_{65+}^{(\nu)}$	-0.9775884	0.1939884	ψ_{45-64}	0.1992476	0.0396736
$\delta_{00-04}^{(\nu)}$	-0.6932154	0.0641154	ψ_{65+}	0.5589763	0.0329321
$\delta_{05-14}^{(\nu)}$	-0.9060146	0.5273104	β	-1.1299031	0.2217778
$\delta_{15-24}^{(\nu)}$	-0.6866487	0.0998952	θ	1.3659094	0.1212877
$\delta_{25-44}^{(\nu)}$	-0.5622193	0.0922629	ξ	6.0648150	2.1353590
$\delta_{45-64}^{(\nu)}$	-0.6672792	0.0986491			
$\delta_{65+}^{(\nu)}$	-1.4118994	0.2165542			

Discussion

The main goal of this work was to include deprivation in an endemic-epidemic model. Our motivation is similar to that of Hank's razor which states that any outcome that could be influenced by socio-economic status likely is influenced by socio-economic status. Here we have postulated that health (represented by norovirus incidence) is influenced by deprivation (a measure that includes socio-economic factors) which may additionally act as an amplifier (cf. the deprivation amplification theory). We determined the impact of including the GISD in a known model for norovirus incidence in Berlin that is used for model development in endemic-epidemic surveillance. Including deprivation in the model leads to a lower AIC value (Table 6), indicating its usefulness in capturing variance currently not incorporated in the existing model as an exogenous factor. We could have considered age-standardised norovirus rates, but this would have removed the comparison opportunities with the previous endemic-epidemic modelling work, and we wanted to evaluate the inclusion of deprivation in these model types.

As Meyer and Held (2017) called for the inclusion of time-varying contact matrices in endemic-epidemic models with contact matrices, we established these matrices using the Arregui et al. (2018) demographic updating method and included them in our modelling approach as a secondary goal. The updated contact matrices were shown visually and summarised using known measures for contact matrices. While norovirus is primarily transmitted through contaminated food, it can also spread via person-to-person contact and contaminated surfaces, which highlights the importance of using contact matrices in the modelling framework. These matrices capture the interactions between different demographic groups, which are crucial for understanding interpersonal transmission in community settings like schools and workplaces. In areas with higher levels of socio-economic deprivation, crowding and limited access to sanitation, person-to-person transmission can be accelerated, making the integration of socio-economic status measures, such as the GISD, essential. By combining contact matrices with socio-economic status data, the model provides a more comprehensive view of norovirus transmission, accounting for both food-related and social transmission pathways, thus enhancing the accuracy of the endemic-epidemic models. We do note, however, that contact matrices are used as a way to inform mixing patterns that rely on contact as a proxy for transmission opportunities and thus only capture direct contact.

Under-ascertainment is a known problem for norovirus data (specific analyses of the data used in this work are given by Bernard et al., 2014; Hofmann et al., 2020). Currently, however, under-ascertained data can only be accounted for in univariate endemic-epidemic models (Bracher and Held, 2021), while the under-ascertainment of norovirus is known to be age-dependent (Gaythorpe et al., 2018a), and is therefore more meaningfully analysed in a multivariate approach. Furthermore, such an approach requires external knowledge of the reporting rate, which adds additional complexity. This means that we have implicitly assumed no under-reporting of data. We have also implicitly assumed that the data quality does not deteriorate or change. However, we know that this is not the case as we noted changes in how population is counted and had to manually redistribute it (Figure 2). A final assumption we have made is that there is no change in the serial interval, but we do not believe that this is an issue for norovirus.

Deprivation indices

We considered in our work the GISD measure of deprivation. According to Michalski et al. (2022): “the use of the index at the level of the municipalities (here: PLZ level, ed.) increases the socioeconomic homogeneity compared to the district (here: bezirk, ed.) level, and there is evidence that health inequalities are underestimated more strongly, the more coarsely the spatial level is resolved”. Consequently, some of the differences within districts may not be evident in our investigation. Deciding what size population strata to consider in epidemiological surveillance involves a trade-off between being able to determine when outbreaks are occurring, protecting personally identifiable information, and having informative denominators and granularity of strata. A PLZ-level analysis would not have been possible as the surveillance data are only available stratified by district.

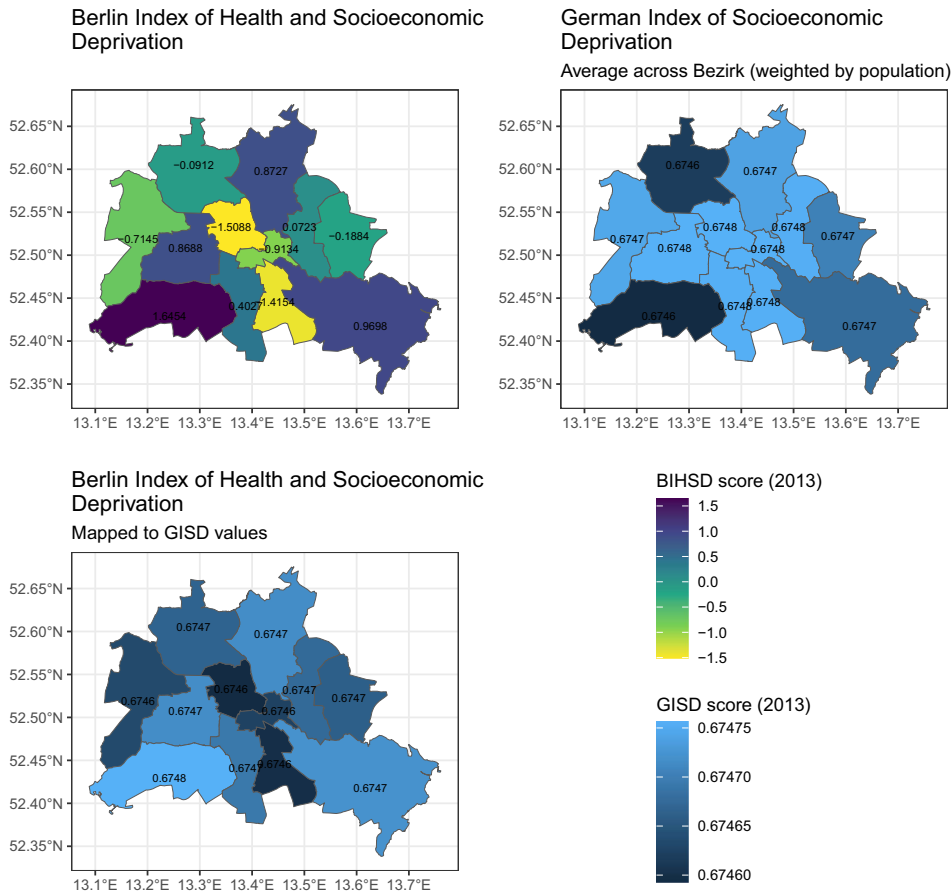
The first index of multiple deprivation for a German setting was created by Maier et al. (2012), who created one for Bavaria. This index was later extended by Maier and Schwettmann (2018) to cover the entire country, and was called the German Index of Multiple Deprivation (GIMD). We considered doing a sensitivity analysis of the GISD data against the GIMD data, but the latter are available at a less fine resolution (Noble et al., 2006; Schuurman et al., 2007; Maier et al., 2012; Hofmeister et al., 2016; Kroll et al., 2017; Schederecker et al., 2019), and thus do not capture differences in Berlin at the district-level. Additional differences between German deprivation indices are outlined in Table 8.

We decided to examine the similarities of our GISD data sub-setting to consider Berlin only using an alternative measure of deprivation, which was developed later specifically with Berlin in mind: the Berlin Index of Health and Social Deprivation (BIHSD) (Zeiher et al., 2022a,b). While these data consider a baseline deprivation level determined for Berlin rather than for all of Germany, they are only available for the years 2013 and 2022, and thus are not maintained at the same temporal granularity as the GISD data. Lest it be mistakenly thought that the level of deprivation is equal across districts (Figure 2) due to our visualisation using the limits for all years rather than each year individually, we showcase the values for 2013 to compare them with the BIHSD in Figure 8. For the sake of completeness, we note that Lakes et al. (2014) also provide a potential alternative to the BIHSD but it does not seem to be available for the years in our study. While we may wish to conclude from the histograms of the GISD score (Figure 2) that the decrease in the GISD during the study period indicates that deprivation has improved, we note that the underlying

Table 8 German country-level deprivation indices

GISD (Michalski et al., 2022)	GIMD (Maier and Schwettmann, 2018)
• GISD is constructed from nine indicators: three educational, three employment-based, and three social indicators (which Michalski et al. (2022) dub “subdimensions”) • GISD is only constructed on the basis of socio-economic considerations • GISD has equal weighting	• GIMD is constructed on the basis of more than just socio-economic considerations and might be more informative since it includes other considerations in its construction • GIMD does not have equal weighting cf. Hofmeister et al. (2016)

Figure 8 Comparison of the GISD and the BIHSD for 2013



comparison is different. The areas with the lowest and highest levels of deprivation by PLZ based on the GISD in our study period unsurprisingly follow the north-south divide found in the original work (Michalski et al., 2022): the highest level of deprivation is found for PLZ 85737 Ismaning in Bavaria for 2011 to 2014 and for PLZ 85774 Unterföhring in Bavaria for 2015, and the lowest level of deprivation is found for PLZ 17353 Strasburg in Mecklenburg-Vorpommern in 2011 and 2012 and for PLZ 39629 Berkau in Saxony-Anhalt in 2013 to 2015. Berlin is located towards the latter two states rather than towards Bavaria. Overall, the GISD scores for Berlin districts seem to be within a similar range (Table 2), and we considered using the rank in place of the scores. However, as they are equidistant values this would introduce artificial differences between districts that are not found in the data (e.g. ranks 1, 2, and 3 have the same value when rounded, which is seen in the table).

It could be interesting to examine an interaction between the historical division of Berlin and the GISD but interaction terms are currently not implemented in the code we used to fit the model. We also think that this is unlikely to improve our model fitting if the reason for the large estimates is singularity (as a result of the GISD scores being very similar for all regions). We note that as the GISD is an index, changes can impact either all parts of the index – e.g. political choices such as austerity policies or economic issues such as recessions – or individual parts of the index. Changes can affect the individual dimensions of: education, employment or income. Furthermore, the index is a relative measure. The GISD is thus sensitive to changes in the underlying socio-economic conditions. Education is a more stable socio-economic measure than income as it stabilises after age 30, whereas changes in income or employment may occur as an effect of health or may cause poor health. It has been suggested that income is the factor that can change the most in a short time period (Galobardes et al., 2006a). Because the GISD is a relative index, an increase in values in the most deprived area (which would be a target for political improvement) should lead to changes in GISD values for other areas, even when there are no changes in these areas. Future investigations should consider studying the performance of individual socio-economic measures and compare them with the results of the composite index (Galobardes et al., 2006b).

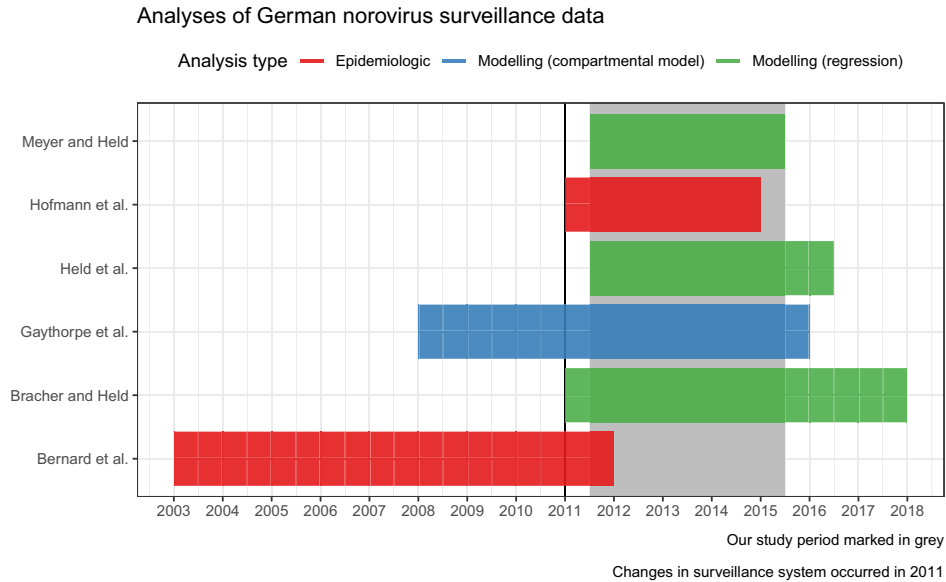
We note the simplicity of our linking approach. For example, when linking the GISD from PLZ to district the PLZ has a 3/4 link with Charlottenburg-Wilmersdorf and a 1/4 link with Steglitz-Zehlendorf, and a weighted approach could be utilised to reflect this. However, we highlight the potential limitations of using the GISD at this level of granularity, and suggest that future work explore the possibility of using more distinct categorical versions of a deprivation index or complementary measures to better capture the relevant socio-economic differences between districts in order to investigate their impact. We contemplated simulating GISD scores to introduce greater variation and investigate this possibility, but determined that the choices required for such an approach would require greater consideration as simulation studies need to be designed well. Thus, we opted to examine a Berlin-specific deprivation index. Future studies could benefit from considering alternative socio-economic measures that would provide more nuanced insights.

Comparison with other models

In addition to the work we build on (Table 4), Gaythorpe et al. (2018a) examined the data using a compartmental MSEIAR (maternal antibodies-susceptible-exposed-infectious-asymptomatic-removed) model (see Figure 9 for an overview). They also used POLYMOD contact data, but assumed a constant population size while we adjusted the contact matrices to reflect demographic changes.

Others have considered comparisons of endemic-epidemic models with univariate autoregressive models, namely autoregressive integrated moving-average models (which focus on the conditional mean of the time series, see Lu and Meyer, 2020, 2022) and integer-valued generalised autoregressive conditional heteroscedasticity models (which focus on the conditional variance of the time series, see Bracher and Held, 2022). We consider now vector autoregressive models. An overview of the models we consider is illustrated

Figure 9 Comparison of study periods considered by Bernard et al. (2014), Meyer and Held (2017), Held et al. (2017), Gaythorpe et al. (2018a), Hofmann et al. (2020), and Bracher and Held (2022)



by a rectangular outline in Figure 10. We compared our model with a VAR(1) model, meaning an autoregression of order 1, which is the formulation given in the illustration. The VAR (1) model has $73 \times 72 = 5,256$ coefficients; each y_{it} in $\mathbf{Y}_t$ has one for each combination of districts (12 total) and age groups (six total) for the previous time period as well as the intercept (72 combinations of districts and age groups and a single intercept). We did not include any additional effects (such as our main effect of interest, deprivation). It provides a good fit to the data (Figure 11), but we do not believe it to be the most useful approach for what we set out to do in this project.

Figure 10 Comparisons with autoregressive (AR) models. The inclusion of moving averages (MA) and conditional heteroskedacity (CH) has been considered by other researchers. The univariate models are on the left while the multivariate models are circled. On the right are formulations of vector autoregressive models (multivariate; VAR) with the option to add exogenous variables

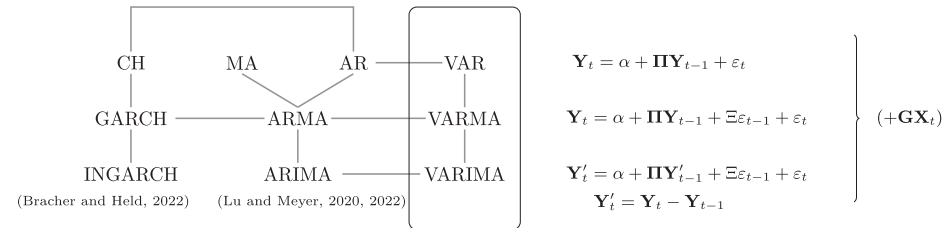
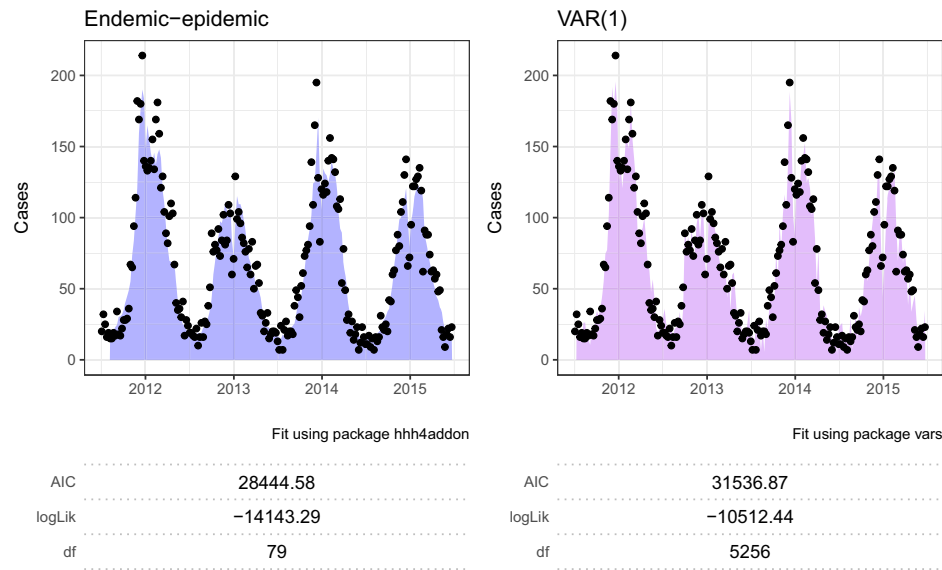


Figure 11 Comparison of our model with a VAR(1) model



Conclusion

In this work we have examined the impact of deprivation within a city. We expected to see more deprived areas in the east (Heblich et al., 2021), not least because of historical division in the setting considered in our work. However, this did not seem to be the case. There are two competing effects in a city that could influence deprivation: cities selectively receive people seeking higher education and employment but they also attract people with lower levels of socio-economic status pursuing opportunity. In a larger analysis, such as a country level analysis, deprivation in rural areas may need to be considered.

In conclusion, while Galobardes et al. (2006b) highlighted that socio-economic indices may not always be essential in every context, our findings suggest that for infectious disease surveillance, particularly for diseases like norovirus, socio-economic variables can play a critical role. Norovirus transmission is influenced not only by biological factors, but also by social determinants, such as deprivation, overcrowding, and access to resources. This underscores the need to integrate socio-economic considerations from the outset of surveillance model development. We propose that a multidisciplinary approach, incorporating both social and biological sciences, is essential to comprehensively capture the dynamics of disease spread, particularly in populations with large socio-economic disparities. Including these elements early in the surveillance process can lead to more accurate models and better-targeted interventions, ultimately improving public health outcomes in vulnerable communities but more work needs to be done to incorporate deprivation indices in surveillance models.

Supplementary material

Available online at <https://doi.org/10.1553/p-4b4e-mkcd>

Supplementary file 1. Codes used in the analysis (codes.zip).

Supplementary file 2. Additional figures (supporting-figs.zip).

Supplementary file 3. Data sets denoting PLZ mapping (plz-mapping.zip).

Supplementary file 4. Model objects (models.zip).



Open science statement

The code used in this analysis is available in the supplementary material. The map data used in our visualisations is available from Open Street Map via <https://suche-postleitzahl.org/downloads> as well as that provided by Meyer and Held (2017).

Contribution

The contributor roles taxonomy is used

MBND: Conceptualisation, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualisation, Writing – original draft, Writing – review & editing

SEM: Funding Acquisition, Writing – original draft, Writing – review & editing

GC: Methodology, Writing – original draft, Writing – review & editing

Disclosure

The Accreditation Council for Continuing Medical Education categories are used

MBND: Salary, Contractual Services (Employment): OsloMet, Contractual Services (Employment): Heidelberg University Hospital

SEM: Grants/Research Support: Centre for Advanced Study at The Norwegian Academy of Science and Letters, Salary, Contractual Services (Employment): OsloMet

GC: Grants/Research Support: Centre for Advanced Study at The Norwegian Academy of Science and Letters, Salary, Contractual Services (Employment): Georgia State University

Acknowledgments

We thank the Centre for Advanced Study at the Norwegian Academy of Science and Letters for their hospitality and support. We would also like to thank the anonymous peer reviewers for providing valuable suggestions for improvement.

ORCID iDs

Maria Bekker-Nielsen Dunbar  <https://orcid.org/0000-0002-7249-3524>

Sverren-Erik Mamelund  <https://orcid.org/0000-0002-3980-3818>

Gerardo Chowell  <https://orcid.org/0000-0003-2194-2251>

References

Altman, D. G., and Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ*, 332(7549), 1080. <https://doi.org/10.1136/bmj.332.7549.1080>

<https://doi.org/10.1553/p-4b4e-mkcd>

- Aral, S. O., Hughes, J. P., Stoner, B., Whittington, W., Handsfield, H. H., Anderson, R. M., and Holmes, K. K. (1999). Sexual mixing patterns in the spread of gonococcal and chlamydial infections. *American Journal of Public Health*, 89(6), 825–833. <https://doi.org/10.2105/AJPH.89.6.825>
- Arregui, S., Aleta, A., Sanz, J., and Moreno, Y. (2018). Projecting social contact matrices to different demographic structures. *PLOS Computational Biology*, 14(12), 1–18. <https://doi.org/10.1371/journal.pcbi.1006638>
- Bambra, C., Lynch, J., Smith, K. E., and Pickett, K. (2021). Pale rider: pandemic inequalities. In *The Unequal Pandemic* (pp. 13–34). Bristol University Press. <https://doi.org/10.51952/9781447361251.ch002>
- Bambra, C., Munford, L., Khavandi, S., and Bennett, N. (2023). *Northern exposure: COVID-19 and regional inequalities in health and wealth*. Bristol University Press <https://doi.org/10.51952/9781447369233>
- Bekker-Nielsen Dunbar, M. (2024). Transmission matrices used in epidemiologic modelling. *Infectious Disease Modelling*, 9(1). <https://doi.org/10.1016/j.idm.2023.11.009>
- Bernard, H., Werber, D., and Höhle, M. (2014). Estimating the under-reporting of norovirus illness in Germany utilizing enhanced awareness of diarrhoea during a large outbreak of Shiga toxin-producing *E. coli* O104:H4 in 2011 – a time series analysis. *BMC Infectious Diseases*, 14(1), 116. <https://doi.org/10.1186/1471-2334-14-116>
- Bracher, J., and Held, L. (2021). A marginal moment matching approach for fitting endemic-epidemic models to underreported disease surveillance counts. *Biometrics*, 77(4), 1202–1214. <https://doi.org/10.1111/biom.13371>
- Bracher, J., and Held, L. (2022). Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *International Journal of Forecasting*, 38(3), 1221–1233. <https://doi.org/10.1016/j.ijforecast.2020.07.002>
- Buczowska, M., Butt, S., Jenkins, C., Hungerford, D., Hawker, J., Verlander, N. Q., O’Connell, A. M., and Byrne, L. (2023). Association between socioeconomic deprivation and incidence of infectious intestinal disease by pathogen and linked transmission route: An ecological analysis in the UK. *Epidemiology & Infection*, 151(e109), 1–29. <https://doi.org/10.1017/S0950268823000869>
- de Graaf, M., van Beek, J., and Koopmans, M. P. G. (2016). Human norovirus transmission and evolution in a changing world. *Nature Reviews Microbiology*, 14, 421–433. <https://doi.org/10.1038/nrmicro.2016.48>
- Di Biagio, K., Baldini, M., Dolcini, J., Serafini, P., Sarti, D., Dorillo, I., Ranzi, A., Settimo, G., Bartolacci, S., Simeoni, T. V., and Prospero, E. (2022). Atmospheric particulate matter effects on SARS-CoV-2 infection and spreading dynamics: A spatio-temporal point process model. *Environmental Research*, 212, 113–117. <https://doi.org/10.1016/j.envres.2022.113617>
- Dietrich, J., Hammerl, J. A., John, A., Kapfenstein, O., Loeffler, C., Karsten Nöckler, K., Rosner, B., Spielmeier, A., Szabo, I., and Richter, M. H. (2023). Impact of climate change on foodborne infections and intoxications. *Journal of Health Monitoring*, 8(S3), 78–92. <https://doi.org/10.25646/11403>
- Dolcini, J., Di Biagio, K., Baldini, M., Serafini, P., Sarti, D., Ranzi, A., Settimo, G., Appignanesi, R., Dorillo, I., Bartolacci, S., Simeoni, T. V., and Prospero, E. (2022). Atmospheric particulate matter effects on Sars-Cov-2 infection and its dynamics spreading: a spatio-temporal point process model. *ISEE Conference Abstracts*, 2022 (1). <https://doi.org/10.1289/isee.2022.P-0647>
- Ernst, V. S., Hoffmann, W., Franze, M., and Stenzel, U. (2023). Association of prevalence rates of infantine social-emotional developmental risks and regional deprivation in Mecklenburg-Western Pomerania: A plea for regional small-scale data. *Research Square*. <https://doi.org/10.21203/rs.3.rs-3426762/v1>
- Farrington, C. P., Whitaker, H. J., Wallinga, J., and Manfredi, P. (2009). Measures of disassortativeness and their application to directly transmitted infections. *Biometrical Journal*, 51(3), 387–407. <https://doi.org/10.1002/bimj.200800160>
- Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., and Davey Smith, G. (2006a). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, 60(1), 7–12. <https://doi.org/10.1136/jech.2004.023531>
- Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., and Davey Smith, G. (2006b). Indicators of socioeconomic position (part 2). *Journal of Epidemiology & Community Health*, 60(2), 95–101. <https://doi.org/10.1136/jech.2004.028092>
- Garnett, G. P., Hughes, J. P., Anderson, R. M., Stoner, B. P., Aral, S. O., Whittington, W. L., Handsfield, H. H., and Holmes, K. K. (1996). Sexual mixing patterns of patients attending sexually transmitted diseases clinics. *Sexually Transmitted Diseases*, 23(3), 248–257. <https://doi.org/10.1097/00007435-199605000-00015>

- Gaythorpe, K. A. M., Trotter, C. L., and Conlan, A. J. K. (2018a). Modelling norovirus transmission and vaccination. *Vaccine*, 36(37), 5565–5571. <https://doi.org/10.1016/j.vaccine.2018.07.053>
- Gaythorpe, K. A. M., Trotter, C. L., Lopman, B., Steele, M., and Conlan, A. J. K. (2018b). Norovirus transmission dynamics: a modelling review. *Epidemiology & Infection*, 146(2), 147–158. <https://doi.org/10.1017/S0950268817002692>
- Hanewinkel, R., and Hansen, J. (2024). Regional socioeconomic deprivation in Germany and nicotine use among children and adolescents. *International Journal of Environmental Health Research*, 34(2), 1023–1033. <https://doi.org/10.1080/09603123.2023.2195155>
- Heblisch, S., Trew, A., and Zylberberg, Y. (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy*, 129(5), 1508–1552. <https://doi.org/10.1086/713101>
- Held, L., Meyer, S., and Bracher, J. (2017). Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture. *Statistics in Medicine*, 36(22), 3443–3460. <https://doi.org/10.1002/sim.7363>
- Hoebel, J., Nowossadeck, E., Michalski, N., Baumert, J., Wachtler, B., and Tetzlaff, F. (2024). Sozioökonomische Deprivation und vorzeitige Sterblichkeit in Deutschland 1998–2021. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 67(5), 528–537. <https://doi.org/10.1007/s00103-024-03862-0>
- Hofmann, F. M., Olawumi, E., Michaelis, M., Hofmann, F., and Stöbel, U. (2020). Challenges in infection epidemiology: On the underreporting of norovirus gastroenteritis cases in Germany. *International Journal of Environmental Research and Public Health*, 17(1). <https://doi.org/10.3390/ijerph17010314>
- Hofmeister, C., Maier, W., Mielck, A., Stahl, L., Breckenkamp, J., and Razum, O. (2016). Regionale Deprivation in Deutschland: Bundesweite Analyse des Zusammenhangs mit Mortalität unter Verwendung des ‘German Index of Multiple Deprivation (GIMD)’. *Das Gesundheitswesen*, 17, 42–48. <https://doi.org/10.1055/s-0034-1390421>
- Juniastuti, Utsumi, T., Yamani, L. N., Dinana, Z., Gunawan, E., Maharani, A. T., Fitria, A. L., Wahyuni, R. M., Soetjipto, Doan, Y. H., Shimizu, H., Ishii, K., Matsui, C., Deng, L., Abe, T., Katayama, K., Lusida, M. I., and Shoji, I. (2023). A household survey of intrafamily norovirus transmission. *Journal of Medical Virology*, 95(10), e29 164. <https://doi.org/10.1002/jmv.29164>
- Keeling, M. J., and Rohani, P. (2008). *Host heterogeneities*, chap. 3. Princeton University Press, <https://doi.org/10.2307/j.ctvcn4gk0.6>
- Kroll, L. E., Schumann, M., Hoebel, J., and Lampert, T. (2017). Regional health differences – developing a socio-economic deprivation index for Germany. *Journal of Health Monitoring*, 2(2), 98–114. <https://doi.org/10.17886/RKI-GBE-2017-048.2>
- Lakes, T., Brückner, M., and Krämer, A. (2014). Development of an environmental justice index to determine socio-economic disparities of noise pollution and green space in residential areas in Berlin. *Journal of Environmental Planning and Management*, 57(4), 538–556. <https://doi.org/10.1080/09640568.2012.755461>
- Lu, J., and Meyer, S. (2020). Forecasting flu activity in the United States: Benchmarking an endemic-epidemic beta model. *International Journal of Environmental Research and Public Health*, 17(4). <https://doi.org/10.3390/ijerph17041381>
- Lu, J., and Meyer, S. (2022). An endemic–epidemic beta model for time series of infectious disease proportions. *Journal of Applied Statistics*, 49(15), 3769–3783. <https://doi.org/10.1080/02664763.2021.1962264>
- Maier, W., Fairburn, J., and Mielck, A. (2012). Regionale Deprivation und Mortalität in Bayern. Entwicklung eines ‘Index Multiplier Deprivation’ auf Gemeindeebene. *Das Gesundheitswesen*, 74(7), 416–425. <https://doi.org/10.1055/s-0031-1280846>
- Maier, W., and Schwettmann, L. (2018). Regionale Deprivation in Deutschland: Der ‘German Index of Multiple Deprivation (GIMD)’. *Public Health Forum*, 26(4), 376–379. <https://doi.org/10.1515/pubhef-2018-0085>
- Mathijs, E., Stals, A., Baert, L., Botteldoorn, N., Denayer, S., Mauroy, A., Scipioni, A., Daube, G., Dierick, K., Herman, L., Van Coillie, E., Uyttendaele, M., and Thiry, E. (2012). A review of known and hypothetical transmission routes for noroviruses. *Food and Environmental Virology*, 4(4), 131–152. <https://doi.org/10.1007/s12560-012-9091-z>
- Mercuri, E., Schmid, L., Poethko-Müller, C., Schlaud, M., Kußmaul, C., Ordóñez-Cruickshank, A., Haller, S., Rexroth, U., Hamouda, O., Schaade, L., Wieler, L. H., Gßwald, A., Rosario, A. S., and RKI-SOEP-2 Study

- Group (2023). *Nationwide population-based infection- and vaccine-induced SARS-CoV-2 seroprevalence in Germany at the end of 2021*. medRxiv. <https://doi.org/10.1101/2023.10.30.23297594>
- Meyer, S., and Held, L. (2017). Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2), 338–351. <https://doi.org/10.1093/biostatistics/kxw051>
- Michalski, N., Reis, M., Tetzlaff, F., Herber, M., Kroll, L. E., Hövener, C., Nowossadeck, E., and Hoebel, J. (2022). German Index of Socioeconomic Deprivation (GISD): Revision, update and application examples. *Journal of Health Monitoring*, 7(S5). <https://doi.org/10.25646/10641>
- Misumi, M., and Nishiura, H. (2021). Long-term dynamics of norovirus transmission in Japan, 2005–2019. *PeerJ*, 9(e11769). <https://doi.org/10.7717/peerj.11769>
- Moissl, A. P., Delgado, G. R., Krämer, B. K., März, W., Kleber, M. E., and Grammer, T. B. (2020). Area-based socioeconomic status and mortality: the Ludwigshafen risk and cardiovascular health study. *Clinical Research in Cardiology*, 109, 103–114. <https://doi.org/10.1007/s00392-019-01494-y>
- Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E*, 67, 026 126. <https://doi.org/10.1103/PhysRevE.67.026126>
- Noble, M., Wright, G., Smith, G., and Dibben, C. (2006). Measuring multiple deprivation at the small-area level. *Environment and Planning A: Economy and Space*, 38(1), 169–185. <https://doi.org/10.1068/a37168>
- OECD (2018). *A broken social elevator? How to promote social mobility*. OECD Publishing. <https://doi.org/10.1787/9789264301085-en>
- Reitzle, L., Heidemann, C., Krause, L., and Hoebel, J. (2024a). Prevalence of gestational diabetes mellitus in Germany: Temporal trend and differences by regional socioeconomic deprivation. *Journal of Health Monitoring*, 9(2), 6. <https://doi.org/10.25646/12086>
- Reitzle, L., Heidemann, C., Krause, L., Hoebel, J., and Scheidt-Nave, C. (2024b). Prävalenz von Gestationsdiabetes mellitus in Deutschland: Trend und Unterschiede nach regionaler sozioökonomischer Deprivation. *Journal of Health Monitoring*. <https://doi.org/10.25646/12009>
- Rohleder, S., Costa, D., and Bozorgmehr, K. (2022). Area-level socioeconomic deprivation, non-national residency, and Covid-19 incidence: A longitudinal spatiotemporal analysis in Germany. *eClinicalMedicine*, 49. <https://doi.org/10.1016/j.eclim.2022.101485>
- Schederecker, F., Kurz, C., Fairburn, J., and Maier, W. (2019). Do alternative weighting approaches for an index of multiple deprivation change the association with mortality? A sensitivity analysis from Germany. *BMJ Open*, 9(8), e028 553. <https://doi.org/10.1136/bmjopen-2018-028553>
- Schultz, A., Schmalefeldt, B., Löhden, I., Tetzlaff, F., and Peters, F. (2023). Faktenblatt: Epidemiologie des Zervixkarzinoms in Deutschland 2009 bis 2019. *Onkologie*, 29, 927–931. <https://doi.org/10.1007/s00761-023-01413-x>
- Schuurman, N., Bell, N., Dunn, J. R., and Oliver, L. (2007). Deprivation indices, population health and geography: An evaluation of the spatial effectiveness of indices at multiple scales. *Journal of Urban Health*, 84(4), 591–603. <https://doi.org/10.1007/s11524-007-9193-3>
- Suchert, V., Hanewinkel, R., Neumann, C., and Hansen, J. (2023). Regional socioeconomic deprivation in Germany and adherence to the 24-h movement guidelines among children and adolescents. *Children*, 10(8), 1392. <https://doi.org/10.3390/children10081392>
- Tetzlaff, F., Nowossadeck, E., Jansen, L., Michalski, N., Barnes, B., Kraywinkel, K., and Hoebel, J. (2023). Widening area-based socioeconomic inequalities in cancer mortality in Germany between 2003 and 2019. *Scientific Reports*, 13(1UR), 17 833. <https://doi.org/10.1038/s41598-023-45254-5>
- Tetzlaff, F., Sauerberg, M., Grigoriev, P., Tetzlaff, J., Mühlichen, M., Baumert, J., Michalski, N., Wengler, A., Nowossadeck, E., and Hoebel, J. (2024). Age-specific and cause-specific mortality contributions to the socioeconomic gap in life expectancy in Germany, 2003–21: an ecological study. *Lancet Public Health*, 9(5), e295–e305. [https://doi.org/10.1016/S2468-2667\(24\)00049-5](https://doi.org/10.1016/S2468-2667(24)00049-5)
- Tuncer, O., Du, Y., Michalski, N., and Reitzle, L. (2024a). Diabetes-related amputations in Germany: analysis of time trend from 2015 to 2022 and differences by area-level socioeconomic deprivation. *Journal of Health Monitoring*, 9(2), 5. <https://doi.org/10.25646/12026>
- Tuncer, O., Du, Y., Michalski, N., and Reitzle, L. (2024b). Diabetesbedingte Amputationen in Deutschland im Trend 2015–2022 und nach sozialräumlicher Lage. *Journal of Health Monitoring*, 9(2), e 12 002.2. <https://doi.org/10.25646/12002.2>

- Wallinga, J., van de Kastelee, J., and Hens, N. (2019). Contact patterns for contagious diseases. In Held, L., Hens, N., O'Neill, Ph., and Wallinga, J. (Eds.) *Handbook of infectious disease data analysis*, (pp. 93–110). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315222912>
- Wollschläger, D., Fückel, S., Blettner, M., and Gianicolo, E. (2024). Übersterblichkeit im Kontext der COVID-19-Pandemie in Deutschland: Methoden, Daten, Muster. *Kardiologie*, 18, 101–108. <https://doi.org/10.1007/s12181-024-00666-z>
- Yang, K., Doege, D., Thong, M. S. Y., Koch-Gallenkamp, L., Weisser, L., Bertram, H., Eberle, A., Holleczek, B., Nennecke, A., Waldmann, A., Zeissig, S. R., Pritzkeleit, R., Jansen, L., Brenner, H., and Arndt, V. (2024). Diabetes mellitus in long-term survivors with colorectal, breast, or prostate cancer: prevalence and prognosis. A population-based study. *Cancer*, 130(7), 1158–1170. <https://doi.org/10.1002/cncr.35133>
- Zeihner, J., Häßler, K., Finger, J., and Hermann, S. (2022a). *Gesundheits- und Sozialstrukturatlas Berlin 2022*. <https://www.berlin.de/sen/gesundheit/service/gesundheitsberichterstattung/gesundheits-und-sozialstruktur/>
- Zeihner, J., Häßler, K., Finger, J. D., Haase, A., and Hermann, S. (2022b). The Berlin index of health and social deprivation: a data based tool to tackle health inequality. *European Journal of Public Health*, 32(Supplement_3), ckac129.031. <https://doi.org/10.1093/eurpub/ckac129.031>