# Habilitationsschrift

aus dem Institut für medizinische Informationsverarbeitung,
Biometrie und Epidemiologie (IBE) der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Direktor: Univ. Prof. Dr. Ulrich Mansmann

---

# Spatial and temporal dimensions of health -

# A statistical-epidemiological perspective

---

vorgelegt von

# Dr. oec. publ. Ursula Berger

München

2021

# Contents

# Summary

## Space and time and health

Health has many dimensions, two of which are spatial as well as temporal components and their interplay. The following Habilitation thesis presents a selection of current research that addresses the influence of location and time on health, diseases and health behaviour. This summary puts the twelve contributions of the Habilitation thesis in context and discusses both, impact of the work and statistical and epidemiological aspects, which aim to identify and quantify regional differences and dynamics of health.

In the subsequent summary I refrain from listing further references, as these can be found in detail in the contributions, which are attached.

## The spatial and temporal perspective on health and disease

In epidemiological studies on health or the spread of diseases the spatial and temporal dimension is of utmost importance and must therefore be taken into account in statistical models and analyses in an appropriate manner. Even the simple subdivision of urban and rural areas provides important insights when considering health or the spread of disease, which can be particularly relevant when targeting interventions and investigating their effectiveness. We approached this problem in a particular example, namely the investigation of the global spread of obesity.

> **CONTRIBUTION 1:**
> Chigbu CO, Parhofer KG, Aniebue UU, **Berger U**. (2018).
> Prevalence and sociodemographic determinants of adult obesity: a large representative household survey in a resource-constrained African setting with double burden of undernutrition and overnutrition.
> *J Epidemiol Community Health,* 2018 Aug;72(8):702-707. doi: 10.1136/jech-2018-210573. Epub 2018 Mar 29.
> https://doi.org/10.1055/s-0029-1233487

While obesity has previously mostly affected high-income countries of the world, it is now known to be rising also in resource-constrained countries, and there especially in the urban cities. For many of these countries, apparently due to the overshadowing effect of undernutrition and infectious diseases, there is a lack of obesity research and even more, a lack of reliable and representative population health and surveillance data, which particularly include spatial information, i.e.

including rural as well as urban populations. Such data would allow for reliable and detailed research on the burden of adult obesity and overweight and its socio-demographic determinants. To overcome these deficits, we performed a large, representative study to determine the prevalence of overweight and obesity among adults in Enugu, a state in the southeast of Nigeria (Chigbu et al., 2018). The study is a cross-sectional household population survey of adults aged between 20 and 60 years and took place between 2015 and 2017. Multistage stratified cluster randomised sampling was used, where so-called enumeration areas (EA, small geographical areas defined by the Nigerian National Population Commission for census data collection) formed the clusters. Every consenting household within the selected EA was sampled, and all eligible consenting adults were included. To obtain a representative sample of the adult population, random sampling of the clusters was stratified between urban and rural areas with a sample proportion of 30% to the rural area and 70% to the urban area, i.e. in reverse proportion to the actual population distribution in the state. We further stratified the random selection of the rural EAs by so-called Local Government Areas (LGAs) to ensure the inclusion of different regions of Enugu. The cluster-randomisation of the urban sub-population was additionally stratified by the socio-economic status of the EA, categorised as upper-income class, middle-income class and lower-income class and one university community. Cluster sampling from these urban substrata was done proportional to the population figures of the areas. To estimate the proportion of urban population living in an area of the given category we used the official figures of the 1991 national census and projected them for the year 2015, considering a population growth rate of 3% for Enugu State of Nigeria as obtained from the National Population Commission. This resulted in a distribution of approximately 10% in the upper-class substrata, 40% in the middle-class substrata and 50% in the lower-class substrata, which lead to the randomisation of the corresponding number of EAs from these strata. To estimate the prevalence of overweight and obesity as well as of underweight from the sampled data, the sampling probability has to be taken into account in an appropriate way, where especially for the variance estimation simple weighting of the observed data would not be appropriate, but instead strata-specific variances need to be pooled appropriately. To determine the socio-demographic factors of overweight and obesity, multinomial regression models were fitted, including the information on the stratification. Based on this methodological approach and the large sample including a total of 6628 individuals, we could provide representative and robust population estimates for overweight and obesity. Our results can serve as a source of information for the development of obesity prevention policies in Nigeria. The results of the study revealed that, unlike Western countries, in Nigeria the prevalence of overweight and obesity

increases with higher socio-economic status of the household and of the area of residence. At the same time, a high prevalence for underweight is especially found in rural areas, confirming the double burden of undernutrition and obesity in this country.

Another example, with a focus on exploration of rural-urban differences, is our work on the medical needs and care for skin cancer in Bavaria (Tizek et al., 2020). Here, information on the places of residence of the participants at a district level allowed us to analyse regional differences in the prevalence of skin cancer in the study population. In addition, factors related to decisions on consultation in case of a conspicuous skin condition are examined. The data resulted from a cross-sectional, questionnaire based survey run in 2016 during the "*Bavarian Central Agricultural Festival*" (*Bayerisches Zentral-Landwirtschaftsfest*) in Munich. The work focuses on 2483 participants living in different districts of Bavaria. To investigate factors related to people's decision to whether to consult a dermatologist first in the event of a conspicuous skin condition, Generalised Mixed Models were employed. These account for district specific heterogeneity by including district-specific random effects. Besides individual socio-demographic characteristics and medical history, knowledge and health care utilisation related to skin cancer, also information on the type of district were included in the modelling process, differentiating urban areas, the environs and rural areas. The analyses showed that in rural areas the prevalence for skin cancer is higher, but at the same time the utilization of skin cancer screening and dermatological consultation is significantly lower than in urban areas. Here considering the information on the area of residence gives an insight into rural-urban differences, which provides evidence that an improvement in care might be achieved through targeted interventions, particularly in rural areas.

Regional differences may also occur in health-related lifestyles. We explored this question in relation to physical activity and the risk of overweight and obesity in Nigeria in Chigbu et al. (2020) based on the representative sample of the survey described above (Chigbu et al., 2018). Again the estimates on the proportions of the population that are physically active or inactive, respectively, and their standard deviations were obtained by taking the sampling distributions in the different strata into account. In Multinomial Regression Models we evaluated the effect of physical activity on obesity and overweight and assessed predictors of outdoor leisure-time physical exercise by Binary Logistic Regression adjusting for the stratification. The results show significant differences in health-related behaviours between rural and urban populations: overall, individuals from rural areas were more than twice as likely to be physically active as urban residents. However, this physical activity was mainly driven by everyday life activity. Recreational sports and outdoor leisure-time physical exercises, on the other hand, were significantly more prevalent in urban areas, with younger people and people with a higher education and income engaging more in recreational sports. The hindrances to outdoor leisure-time physical exercise differ between urban and rural dwellers: a significantly higher portion of rural dwellers reported lack of time and lack of interest as hindrances to outdoor leisure-time physical exercise than urban dwellers. In contrast, more urban dwellers report fear of criminals, fear of traffic and doing in-door exercises than rural dwellers. However, in both, rural and urban dwellers, the by far largest proportion of respondents (47% of urban residents and 42% of rural residents) had no reason or were not hindered from outdoor leisure-time physical exercise by any factor or were hindered by lack of interest in outdoor leisure-time physical exercise. These findings can have an impact on targeted physical activity policy formulation to battle physical inactivity and with it the pandemic of obesity.

Apparently, a person's health-related lifestyle may change over the course of life, i.e. exhibit a dependency on time. In particular, life-changing circumstances, such as those regarding working life, can have an effect on lifestyle and health behaviour. This question is tackled in Berger et al. (2005), where we examine the impact of retirement on changes in physical activity. The work is based on longitudinal data from *The West of Scotland Twenty-07 Study*. We examined the patterns of physical activity of nearly 700 participants focusing on the oldest cohort, who were aged about 60 years when interviewed in 1991 and were followed up four to five years later. Bayesian hierarchical models for repeated measures with a seasonal component were used for the statistical analyses. Because many of the questions on physical activity referred to recent behaviour, and data collection lasted several months, all models were seasonally adjusted. The seasonal effect was modelled as a second-order random walk, which produces a smooth-effect function over the calendar months. Model parameters were estimated by Markov chain Monte Carlo sampling. The analysis had two parts, each generating a binary outcome variable that indicated whether a person had achieved an active lifestyle, i.e. whether either the 'fitness' or the 'active living' recommendation was met. The estimated seasonal effect confirmed that physical activity was greater during the summer months. We found that higher levels of physical activity associated with a healthier lifestyle (in terms of smoking and diet), and that socio-economic factors played a minor role in determining the level of physical activity. A substantial amount of physical activity occurred at work but was lost by those who had retired. While those who were not working were more physically active at home or at leisure than those in work, the majority of sampled individuals carried out too little physical activity outside work to compensate for the loss of work-related activity after retirement. From our findings, we were able to conclude that special health promotion initiatives that encourage people to become more physically active should be targeted at those who are about to retire.

**The impact of spatial context on health**

Regional living conditions, such as the social and economic infrastructure of a region, the population structure of the neighbourhood, regional employment opportunities or physical environment can have a direct or indirect influence on a person's health and health behaviour in addition to individual factors such as age, gender, education, individual socio-economic status, etc.. However, the underlying mechanisms remain insufficiently understood.

We approached the above question in Voitgländer et al. (2010a) where we investigated regional variation in physical health across Germany and explored whether regional as well as neighbourhood deprivation are negatively associated with physical health, and whether this association can be explained by specific neighbourhood exposures related to physical health. Data for this study were obtained from the German Socio-Economic Panel (SOEP), a large panel survey of private households that provides information on all household members. In the analysis we included individuals aged 18 and above who took part in the survey in 2004. That year's questionnaire included information on perceived health as well as different perceived neighbourhood characteristics. Additional data on the neighbourhood of the households (by microm GmbH) were matched together with regional data at the level of the 439 German districts, provided by the Federal Office of Building and Regional Planning (BBR). We fitted a series of Mixed Linear Models for this *multilevel* data, including step by step regional and neighbourhood proxies of deprivation (e.g. regional unemployment quota, average purchasing power of the street section) as well as specific neighbourhood exposures (e.g. perceived air pollution). Last we differentiated the regions by East and West Germany. We also controlled for individual characteristics including socio-economic status and health behaviour. The distribution of the random effects from the fitted Mixed Linear Models gave additionally insight in the proportion of variation in physical health, which can be attributed to the individual level, to the household and to the regional level. Our analyses revealed how strongly regional as well as neighbourhood deprivation is associated with an individual's health in Germany, even when simultaneously adjusting for individual health determinants. Especially people with lower socio-economic status are more affected by unfavourable

regional living conditions. Therefore the impact of regional and neighbourhood deprivation adds to a further increase of regional health inequalities in Germany.

When considering the spatial distributions of health inequalities, not only the spatial but also the temporal components are of interest, in order obtain insight on how inequalities develop over time. In particular in Germany, the development of regional living conditions are of interest since the reunification. This research focus is taken in Voigtländer et al. (2010b), where we describe the extent of regional disparities in health-related living conditions in Germany and their dynamics between the years 1995 and 2005. Data on the level of the German districts were obtained from the INKAR data source (www.inkar.de), which provides regional data and interactive maps on spatial and urban development. In addition we regarded data of "*Statistik regional*", that is regional data provided by the Federal Statistical Office (Statistisches Bundesamt). We regarded selected indicators of the regional social structure such as unemployment rate, employment rate, disposable income, density of physicians and proportion of females among the 18 – 29 year-olds. Time-wise we compared the years 1995 and 2005 / 2006. The results were stratified for East and West Germany to assess the development of regional disparities within these two large entities, mainly employing descriptive methods regarding the distribution of the selected indicators within districts of East and West Germany and plotting the values in descending order of rank for the two time-points (i.e. disparity plots). For the study period, we found an increase in regional inequality at the district-level as well as an increasing divergence between East and West Germany.

One crucial problem in quantifying the effect that spatial living context has on an individual's health is the so-called "selection bias": people do not choose their place of residence at random, and it is possible that people with higher socio-economic status (and thus better health) have simply left disadvantaged regions. Thus, it is unclear to what extent an identified statistical association is an effect of contextual factors or simply due to the composition of the population in a given region. This issue of "context versus composition" with respect to health inequalities has been the subject of many debates among scientists of this field. Through statistical analyses of observational studies, it is difficult to separate the influence of individual effects from the influence of contextual effects. Multilevel studies employing Mixed Regression Models allow adjusting for individual characteristics when exploring the impact of contextual factors on health outcomes. However, even these approaches can only account for a certain number of pre-specified individual characteristics and will never adjust for all relevant individual characteristics, particularly not for unmeasured variables. To rule out self-selection as a possible explanation for regional differences in health would require an experiment in which individuals are randomly assigned to regions of residence and tracked over time. Such an experiment is difficult or even impossible to conduct in a sensible manner.

However, data are available from a comparable situation. In Reiss & Berger et al. (2013), we analyse mortality among resettlers from the former Soviet Union in North Rhine-Westphalia (NRW) using the data of the *AMOR* study, a retrospective cohort study on the mortality of resettlers. The cohort comprises resettlers who moved from former Soviet Union countries to NRW between 1990 and 2001. We took advantage of the unique situation that these resettlers were distributed quasi-randomly and independently of their socio-economic or health status to districts and district-free cities in North Rhine-Westphalia. There is considerable variance among these districts with respect to socio-economic factors at the contextual level mirroring different levels of regional deprivation. Thus, the assigned distribution of the resettlers to the districts can be considered as a "natural experiment", which allows the unbiased investigation of the influence of contextual factors on individual mortality. To investigate the effect of regional deprivation on mortality

among resettlers in NRW, we used a pre-defined categorisation of the districts of NRW according to their deprivation status, which was based on selected indicators of the social structure of the district. Gender-specific standardised mortality ratios (SMRs) were compared between categories and a Cox Model with a non-parametric, non-linear effect for age based on P-splines was employed. If contextual factors had no influence on mortality (as a measure of health), then no mortality differences would be expected due to the quasi-random assignment of place of residence. In fact, however, our quasi-experimental study revealed a clear regional variation in mortality within NRW with respect to regional deprivation. Mortality of resettlers was highest in the "poverty pole" of the Ruhr area and significantly lower in districts categorised as "family zones", "prospering regions and suburban areas" and "heterogeneous districts". Thus, the study could provide clear and important evidence for an independent effect of regional deprivation on mortality.

**The influence of environmental risk factors and of the physical environment on health and health behaviour**

In addition to the influence of the social context of life (such as regional deprivation), characteristics of the physical environment play an important role when studying the spatial component of health inequalities. Area-specific environmental risk factors can have a direct influence on health. Thus, when studying health impairment due to environmental risks, detailed spatial information from geographic information systems can be helpful.

<div style="border:1px solid blue; padding:10px;">

**CONTRIBUTION 8:**
Herrera R, Radon K, von Ehrenstein OS, Cifuentes S, Muñoz DM, **Berger U** (2016).
Proximity to mining industry and respiratory diseases in children in a community in Northern Chile: A cross-sectional study.
*Environmental Health.* 2016 Jun 7;15(1):66.
DOI: https://doi.org/10.3390/ijerph15010039

</div>

This research focus was pursued in Herrera et al. (2016), where we use spatial information on the location of residence to assess the impact of open-pit mining on respiratory diseases in children. Data was collected from children living in a town in the desert area of Northern Chile, where two large open-pit mines are located just at the town boarder. These mines cause air pollution from open-pit mine explosions. Data on the prevalence of respiratory diseases and potential confounders were available from a cross-sectional survey carried out in 2009 among 288 children living in the community. The proximity of the children's place of residence to the two mines was calculated using geographical positioning systems.

We make use of spatial Bayesian modelling to study the impact of air pollution on respiratory diseases in children. In a first step Bayesian Structured Additive Regressions were fitted to estimate a possible non-linear effect of the distance to the mines on the occurrence of respiratory diseases. The results of these analyses were used to calibrate in a subsequent modeling step Bayesian models with a parametric distance function, which were fitted to identify a minimum distance to the point sources beyond which the health impact is considered non statistically significant. To adjust for possible regional variations, which are not captured by distance to the mine or by an individual's characteristics, we additionally included an spatial effect. The analysis allowed us to uncover a substantial increase in respiratory disease risk with increasing proximity to the mines.

The work was extended in Herrera et al. (2018), where we quantified the causal attributable risk of living close to the mines on asthma or allergic rhinoconjunctivitis risk burden in these children. We found in the data that prevalences for these diseases were higher in children living within the first quartile of distance to the mines. The idea in this work was to employ targeted maximum likelihood estimation to obtain the causal attributable risk (CAR), which can be used to estimate a possible effect of a hypothetical intervention that increases the distance between place of residence and the mines (e.g., by relocating homes). We used the first quantiles as a threshold for the distance to the mines, corresponding to 1.6 km and 1.9 km, respectively. Children living more than the first quartile away from the mines were considered as the unexposed group. Our results suggest that a hypothetical intervention after which all children lived further away than 1.6 km and 1.9. km to the mines would reduce the prevalence of respiratory disease in the community by around four percentage points. The chosen approach permits to assess the impact of a complex and costly public health strategy beforehand.

The preventive or positive effect of the physical environment can also be the focus of environmental epidemiological studies. For example, a question is to assess and quantify the benefits of parks or green spaces close to home in terms of better health. In Herrera et al. (2018), the relationship between green space and work-related chronic stress in young people was examined based on data of the SOLAR I and SOLAR II studies. Geocoded information of the place of residence was merged with data based on satellite imagery and other spatial information on green space in the residential neighbourhood. We took advantage of the repeated information on the young participants at two time points, 2003 (SOLAR I) and 2007/2008 (SOLAR II). We analysed the data using Generalised Equation Estimation (GEE) for repeated measures, adjusting for socio-demographic characteristics, physical activity and job-related information and psychological measures. The study showed that chronic work-related stress is lower among those who live in a greener neighbourhood, especially among students.

The work of Vogt et al. (2015) investigates (among other things) the benefits of green spaces for the health of older people. The analyses are based on health data from the KORA-Age study from the Augsburg region conducted in 2008/2009 and use geo-information of the residence of the study participants to calculate the distance to nearest green spaces. Mixed Logistic Regression Models were fitted controlling for demographic and socioeconomic factors. Interestingly enough, our work did not show evidence for the relationship between the proximity of the place of residence to green areas and health in this cohort of older people from the Augsburg region.

## Analysis of infectious disease

Infections transmitted from person to person, such as COVID-19, are dynamic events in which both, the spatial and temporal dimensions play an essential role in their spread. For meaningful small-scale monitoring of such infection events, statistical analyses based on complex Generalized Linear Models with different spatio-temporal components are helpful.

In our work (Schneble et al., 2020), we used the COVID-19 registry data for Germany provided by the Robert-Koch-Institute (RKI) and we investigated the development of fatal infections at the district level. We employed Generalized Additive Regression Models, which include both, smooth structured and unstructured spatial-temporal effects to cope with both, the dynamics of the disease and its local spreading. At the spatial level, smooth components are used to elaborate the large-scale distribution of fatal infections, which can be displayed in maps. Unstructured spatio-temporal effects allowed us to identify unusual infection events in isolated districts, such as local outbreaks. Using nowcasting methods, our statistical models allows to predict the number of COVID-19-associated deaths for Germany as well as for individual districts and states as early as the registration date of the infections. Our analyses provide important insights into local infection patterns that goes far beyond the informative value of raw numbers of daily reported, registered cases of infections. Our analyses also allow to get a better picture of future expected deaths from today's registered cases. This in turn can also be used to predict the expected utilization of the regional healthcare system and thus support a district-specific monitoring of the COVID-19 pandemic.

## Statistical-epidemiological methods

The inclusion of spatial and temporal information can provide valuable insight in many epidemiological studies from very different fields, both in describing the distribution of health and disease and their dynamics, and in investigating different variables influencing the development of diseases. In addition to the necessary data bases, a prerequisite is the application of appropriate methodological analysis strategies and, in some cases, complex statistical procedures that adequately account for the spatial and temporal correlation structure of the data. The twelve selected contributions summarised address the issue from both, an epidemiological health science perspective but also with a methodological-statistical focus. The range of applications of the selected contributions is broad in order to underline the flexibility of statistical-epidemiological methods.

# List of Contributions

The present cumulative habilitation is composed of the following twelve contributions. They are listed here in alphabetical order.

**Berger U**, Der G, Mutrie N, Hannah MK (2005). The impact of retirement on physical activity. *Aging and Society*, 25, pp. 181-195. DOI: https://doi.org/10.1017/S0144686X04002739

Chigbu CO, **Berger U**, Aniebue U, Parhofer KG (2020). Physical Activity and Outdoor Leisure Time Physical Exercise: A Population Study of Correlates and Hindrances in a Resource-Constrained African Setting. *J Multidiscip Healthc*. 2020;13:1791-1799. DOI: https://doi.org/10.2147/JMDH.S281518

Chigbu CO, Parhofer KG, Aniebue UU, **Berger U**. (2018). Prevalence and sociodemographic determinants of adult obesity: a large representative household survey in a resource-constrained African setting with double burden of undernutrition and overnutrition. *J Epidemiol Community Health,* 2018 Aug;72(8):702-707. Epub 2018 Mar 29. DOI: https://dx.doi.org/10.1136/jech-2018-210573

Herrera R, **Berger U**, von Ehrenstein OS, Díaz I, Huber S, Moraga Muñoz D, Radon K (2018). Estimating the Causal Impact of Proximity to Gold and Copper Mines on Respiratory Diseases in Chilean Children: An Application of Targeted Maximum Likelihood Estimation. *Int. J. Environ. Res. Public Health* 2018, 15, 39. DOI: https://doi.org/10.3390/ijerph15010039

Herrera R, Markevych I, **Berger U**, Genuneit J, Gerlich J, Nowak D, Schlotz W, Vogelberg C, von Mutius E, Weinmayr G, Windstetter D, Weigl M, Heinrich J, Radon K (2018). Greenness and job-related chronic stress in young adults: a prospective cohort study in Germany. *BMJ Open*. 2018 Jun 4;8(6):e021599. DOI: https://doi.org/10.1136/bmjopen-2018-021599

Herrera R, Radon K, von Ehrenstein OS, Cifuentes S, Muñoz DM, **Berger U** (2016). Proximity to mining industry and respiratory diseases in children in a community in Northern Chile: A cross-sectional study. *Environmental Health.* 2016 Jun 7;15(1):66. DOI: https://doi.org/10.1186/s12940-016-0149-5

Reiss K[#], **Berger U[#]**, Winkler V, Voigtländer S, Becher H, Razum O (2013). Assessing the effect of regional deprivation on mortality avoiding compositional bias: a natural experiment. *J Epidemiol Community Health* 2013; 67: 213–218. DOI: https://doi.org/10.1136/jech-2012-201336

Tizek L, Schielein MC, **Berger U**, Seifert F, Biedermann T, Böhner A, Zink A (2020). Regional differences in medical needs and care for skin cancer across Bavaria: confronting the gap. *Eur J Dermatol*, 2020; DOI: https://doi.org/10.1684/ejd.2020.3752

Schneble M, De Nicola G, Kauermann G, **Berger U** (2020). Nowcasting fatal COVID-19 infections on a regional level in Germany. *Biom J.* 2020 Nov 20. DOI: https://doi.org/10.1002/bimj.202000143

Vogt S, Mielck A, **Berger U**, Grill E, Peters A, Döring A, Holle R, Strobl R, Zimmermann A, Linkohr B, Wolf K, Kneißl K, Maier W (2015). Neighborhood and healthy aging in a German city: distances to green space and senior service centers and their associations with physical constitution, disability and health related quality of life. *European Journal of Ageing*, Volume 12, Issue 4, pp.273-283. DOI: https://doi.org/10.1007/s10433-015-0345-0

Voigtländer S, **Berger U**, Razum O (2010a). The impact of regional and neighbourhood deprivation on physical health in Germany: a multilevel study. *BMC Public Health,* 2010 Jul 8;10:403. DOI: https://doi.org/10.1186/1471-2458-10-403

Voigtländer S, **Berger U**, Razum O (2010b). Zunehmende regionale Unterschiede bei den Lebensverhältnissen in Deutschland und ihre Bedeutung für die Erklärung gesundheitlicher Ungleichheit. *Gesundheitswesen*, 2010 May; 72(5), pp.301-8. Epub: 2009 Aug 6. German. DOI: https://doi.org/10.1055/s-0029-1233487