

Electronic Supplementary Material: English version of the article
„Selbst berichtete Infektionen in der NAKO Gesundheitsstudie – Einordnung in die gegenwärtige
Forschungslandschaft“

Self-reported infections in the German National Cohort (GNC) in the context of the current research landscape

Max J Hassenstein^{1,2}, Ghazal Aarabi³, Peter Ahnert^{4,5,6}, Heiko Becher⁷, Claus-Werner Franzke⁸, Julia Fricke⁹, Gérard Krause^{1,10,11}, Stephan Glöckner^{1,10}, Cornelia Gottschick¹², André Karch¹³, Yvonne Kemmling¹, Tobias Kerrinnes^{1,10}, Berit Lange^{1,10}, Rafael Mikolajczyk¹², Alexandra Nieters^{14,15}, Jördis J Ott^{1,10}, Wolfgang Ahrens^{16,17}, Klaus Berger¹³, Claudia Meinke-Franze¹⁸, Sylvia Gastell¹⁹, Kathrin Günther¹⁶, Karin Halina Greiser²⁰, Bernd Holleczeck²¹, Johannes Horn¹², Lina Jaeschke²², Annika Jagodzinski^{23,24,25}, Lina Jansen²⁶, Carmen Jochem²⁷, Karl-Heinz Jöckel²⁸, Rudolf Kaaks²⁹, Lilian Krist⁹, Oliver Kuss³⁰, Susan Langer¹², Nicole Legath¹³, Michael Leitzmann²⁷, Wolfgang Lieb³¹, Markus Loeffler^{4,5}, Nina Mangold^{32,33}, Karin B. Michels⁸, Christa Meisinger^{34,35}, Nadia Obi⁷, Tobias Pischon^{22,36,37,38}, Tamara Schikowski³⁹, Sabine Schipf¹⁸, Matthias B. Schulze¹⁹, Andreas Stang²⁸, Sabina Waniek³¹, Kerstin Wirkner^{4,5}, Stefan N. Willich⁹, Stefanie Castell^{1,10}

¹ Department of Epidemiology, Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany

² PhD Programme "Epidemiology" Braunschweig-Hannover, Germany

³ Department of Prosthetic Dentistry, Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf, Germany

⁴ Institute for Medical Informatics, Statistics, and Epidemiology (IMISE), Leipzig University, Leipzig, Germany

⁵ LIFE – Leipzig Research Center for Civilization Diseases, Leipzig University, Leipzig, Germany

⁶ PROGRESS – Associated Member of the German Center for Lung Research (DZL), Giessen, Germany

⁷ Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Germany

⁸ Institute for Prevention and Cancer Epidemiology, Faculty of Medicine and Medical Center, University Medical Center, University of Freiburg, Freiburg, Germany

⁹ Institute for Social Medicine, Epidemiology and Health Economics, Charité - Universitätsmedizin Berlin, Berlin, Germany

¹⁰ German Center for Infection Research (DZIF), Braunschweig, Germany

¹¹ Hannover Medical School (MHH), Hannover, Germany

¹² Institute for Medical Epidemiology, Biometry and Informatics (IMEBI), Medical School of the Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

¹³ Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany

¹⁴ Center for Chronic Immunodeficiency, Medical Center - University of Freiburg, Freiburg, Germany

¹⁵ Center for Biobanking, Medical Center - University of Freiburg, Freiburg, Germany

¹⁶ Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

¹⁷ Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany

¹⁸ Institute for Community Medicine, University Medicine Greifswald, Germany

¹⁹ Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

²⁰ Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany

²¹ Saarland Cancer Registry, Saarbrücken, Germany

²² Molecular Epidemiology Research Group, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

²³ Epidemiological Study Center, University Medical Center Hamburg-Eppendorf (UKE), Hamburg Eppendorf, Germany

²⁴ Department of General and Interventional Cardiology, University Heart and Vascular Center, Hamburg, Germany

²⁵ German Center for Cardiovascular Research (DZHK) Partner Site Hamburg/Lübeck/Kiel, Hamburg, Germany

²⁶ Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

²⁷ Department of Epidemiology and Preventive Medicine, University of Regensburg, Germany

²⁸ Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, Germany

²⁹ Department of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

³⁰ Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany

³¹ Institute of Epidemiology, Kiel University, Kiel, Germany

³² Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany

³³ University Hospital Augsburg (UKA), Augsburg, Germany

³⁴ Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

³⁵ Ludwig-Maximilians-Universität München, Chair of Epidemiology at UNIKA-T Augsburg, Augsburg, Germany

³⁶ Charité – Universitätsmedizin Berlin, Germany

³⁷ German Center for Cardiovascular Research (DZHK), partner site Berlin, Germany

³⁸ MDC/BIH Biobank, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) and Berlin Institute of Health (BIH), Berlin, Germany

³⁹ IUF-Leibniz Institute of Environmental Medicine, Düsseldorf, Germany

Correspondence address

Dr. Stefanie Castell

Helmholtz Centre for Infection Research Department of Epidemiology

Inhoffenstraße 7

38124 Braunschweig

Germany

Stefanie.Castell@helmholtz-hzi.de

Abstract

Background: Infectious diseases continue to play an important role for disease perception, health-economic considerations and public health in Germany. In recent years, infectious diseases have been linked to the development of non-communicable diseases. Analyses of the German National Cohort (GNC) may provide deeper insights into this issue and pave the way for new targeted approaches in disease prevention.

Objectives: The aim is to describe the tools used to assess infectious diseases and to present initial data on infectious disease frequencies. Also, to relate the GNC assessment tools to data collection methods in other studies in Germany.

Methods: As part of the baseline examination, questions regarding infectious diseases were administered using both an interview and a self-administered touchscreen questionnaire. Data from the initial 101,787 GNC participants were analysed.

Results: In the interview, 0.2% (HIV/AIDS) to 8.6% (shingles) of respondents reported ever having a medical diagnosis of shingles, postherpetic neuralgia (in case shingles were reported), hepatitis B/C, HIV/AIDS, and tuberculosis, and sepsis if treated in hospital. In the questionnaire, 12% (cystitis) to 81% (upper respiratory tract infections) of respondents reported having experienced at least one occurrence of upper or lower respiratory tract infections, gastrointestinal infections, cystitis, or fever within the past 12 months.

Outlook: Further cross-sectional analyses of data and tools presented here - for example on determinants of susceptibility to self-reported infections - can be anticipated from the year 2021 onward. Beyond that, more extensive research into infectious disease epidemiology will be possible, particularly once analyses of GNC biological materials have been performed.

Keywords

infectious diseases, infectious disease epidemiology, German National Cohort (GNC), cohort study

Introduction

Infectious diseases continue to play an important role for the subjective disease perception, health economic considerations and public health in Germany, even if they are no longer the leading causes for morbidity and mortality. They contribute considerably to the loss of healthy years of life [1]. Due to their contagiousness and preventability through hygiene measures and vaccination, communicable diseases are important for public health and subject to public debate. In recent years, infectious diseases have been associated with the development of non-communicable diseases in a complex

interplay with the immune system, psyche, microbiome, genetics and lifestyle factors [2, 3]. In addition, the impact of certain infections on carcinogenesis like hepatitis B virus infection (liver cell carcinoma) or human papillomavirus infection (cervical carcinoma) is also known [4]. Data and biosamples of the German National Cohort (GNC) may provide further insights in this regard, and pave the way for more targeted approaches to disease prevention. Especially, the use of several omics technologies in multi-omics analyses of different biosamples will play a role. This approach aims, for example, at the joint analysis of genetic data together with data from comprehensive transcriptome, protein or microbiome analyses. In combination with self-reported infection events, a paradigm shift towards personalised prevention and medicine can be supported.

Central assessment tools in the GNC are the computer-assisted interview ('face-to-face') and the self-administered questionnaire that is filled in independently at desktop computers (touchscreen module) by the participants in the study centre [5]. For both tools, the questions on infectious diseases were compiled from 2009 onwards by experts in the area of infections and the immune system. Included in the interview are questions referring to the entire life course and to infectious diseases that may have a chronic progression. These particular questions are asked in a specific way to rule out self-diagnosis. Transient infections are queried in the touchscreen module. Such infections that occur mostly only temporary like acute respiratory infections are also suitable for self-diagnosis. The self-administered questionnaires of the touchscreen module were evaluated in both GNC pretests between 2011 and 2013. The results indicate the suitability of the infection-specific questionnaire to assess the targeted diseases and showed acceptable reliability [6, 7].

The aim of this paper is to describe the GNC tools to assess self-reported infectious diseases. Additionally, first data on disease frequencies are reported, and the assessment tools are put into context with tools used in other studies or in the German notification system, which is based on the Infection Protection Act (German abbreviation: IfSG). In this way, GNC assessment tools are positioned within the existing research landscape.

Methods

The presented data originate from the first 101,791 participants, examined in 18 GNC study centres between 14 March 2014 and 17 March 2017. The GNC is a population-based German cohort study with the aim of including 200,000 participants between the ages of 20 to 69 years. The baseline recruitment was conducted in 2014-2019, while the follow-up period is scheduled for 20-30 years [8]. The questions on infectious diseases were included in the so-called level-1-examination and were intended for all participants. The GNC study protocol has been reviewed by the respective ethics committees and data protection officers. All participants provided signed informed consent before taking part in the study

[4]. The data presented in this publication (age, gender) are not quality-assured. Also, the data on diagnoses are not validated based on medical data. The infectious disease questions considered in both the interview and the self-administered questionnaire, including the respective response options, are presented in Table 1. The interview queries medical diagnoses of infections, or, in the case of sepsis, hospitalisation. In contrast, all responses to the self-administered questionnaire rely on self-diagnosis. In this publication, the interview questions were analysed only concerning the age of the participant at the time he or she visited the study centre. Further analyses, such as the consideration of age at diagnosis, are reserved for future research. Therefore, the age-stratified figures reflect the frequencies at the age at examination. Within the touchscreen module, the response options for the occurrence of the infectious diseases have been aggregated, i.e. the presence of at least one disease/symptom is reported. For both instruments, the response options 'no answer', 'do not know' and missing values have been excluded for the calculation of proportions. These answers are reported separately. Some participants exceeded the maximum recruitment age of 69 years at their first site visit. Therefore, the age group of ≥ 70 years has been specified in age-stratified analyses. Furthermore, four participants have been excluded due to implausible or missing reports on age. For herpes zoster and postherpetic neuralgia, cases have been omitted from the analyses if the age at diagnosis was below 21 because such cases are considered medically implausible.

All analyses have been conducted with R, Version 3.6.0, including the package 'tidyverse'.

Results

In total, 101,787 participants were included in the analysis. The median age site visit was 53 years (interquartile range 45-62 years); 53.6% (54,526/101,787) were female. There were 2,971 (2.9%) participants in the age group of 70 years and older. In the interview, the occurrence of herpes zoster, postherpetic neuralgia, hepatitis B and C, HIV/ AIDS, tuberculosis was queried in terms of a medical diagnosis. For sepsis, hospitalisation was queried. Between 0.2% (HIV/ AIDS) and 8.6% (herpes zoster) of the participants reported the diagnosis of the respective infectious disease during their life span by a medical doctor (Table 2). In the touchscreen module, infection episodes were inquired concerning the past 12 months for lower respiratory tract infections (LRTI), upper respiratory tract infections (URTI), bladder infections, gastrointestinal infections or fever ($>38^{\circ}\text{C}$). For these infections, at least one episode was reported from 12.2% (bladder infection) up to 81.2% (URTI) of responders. (Table 2). Here, URTI were reported most frequently, followed by gastrointestinal infections, fever and lower respiratory tract infections. Figures 1 to 3 present the proportions of participants, stratified by age and sex, which reported the respective disease or symptom within the life course (interview) or within the past 12 months (touchscreen module). Hepatitis C and HIV/ AIDS were the least frequently reported

diseases and are additionally presented in Figure 2 with a modified scale for better presentability. HIV/AIDS were mainly reported by men, and bladder infections mainly by women. Considering data quality, the proportion of data that can be analysed from the interview is very high (97.1% and more, Table 2), while this proportion is lower for the questions of the touchscreen module (90.0-91.2%, Table 2).

Discussion

This publication is the first to present results of the GNC regarding infectious diseases or their symptoms, respectively. Due to the long follow-up-period, the GNC has immense potential for research on risk factors or predictors of infectious diseases and their effects. The storage of biosamples allows future research on biomarkers associated with increased infection risk or on biological mechanisms in the context of infections and non-communicable diseases [8].

Standardised indicators of disease frequencies in terms of prevalences or incidences are not directly to be derived from the presented numbers. The calculation of these numbers is planned for future disease-specific research in order to characterise the GNC accordingly. Here, it should be noted that only survivors report the diseases, therefore, the numbers might be underestimated. Hence, GNC disease frequencies in the framework of existing studies or the German notification system will be discussed in future disease-specific publications. In contrast, this publication serves the purpose to put the questionnaires used into the context of the research landscape and, therefore, to show the potential of the GNC.

The age-specific frequencies of the infectious diseases ascertained in the GNC and presented here can only be compared to a limited degree to notification data, health insurance data or studies with a focus on prevalence or incidence assessment. The reason being that the frequencies presented in Figure 1 and 2 do not consider the age at diagnosis, but the question if the respective disease has ever been diagnosed during the life course. Thus, the figures only show the age at disease assessment on site. Furthermore, it is known that it is a challenge to recall medical events correctly, especially if the events lie further back in time [9, 10]. The diseases and symptoms queried in the touchscreen module (Figure 3) probably provide a more precise representation of the age at diagnosis as the questionnaire queries occurrences within the past 12 months of the time of site visit. In general, validation of the self-reports is reasonable, and can be at least be partially realised by health insurance data in the future. Here, it should be kept in mind that not all symptomatic patients necessarily visit a doctor's practice.

The interview data have excellent quality in terms of valid answers, but the touchscreen module data score considerably less well (Table 2). It cannot be determined if a missing answer equals the absence of the disease, and, thus, results may be biased. The frequencies presented in Table 2 refer to all

participants with a valid answer (without missing values, 'no answer' and 'do not know'). Consequently, the calculated results may be biased. This issue could be addressed in future analyses with appropriate methods, e.g. multiple imputation [11].

In the following, all reported infectious diseases or symptoms are put into the context of the research landscape. Additionally, an outlook on research questions will be given that can be answered with GNC data in the future.

Tuberculosis

In Germany, every case of tuberculosis (TB) is required by law (IfSG) to be reported to health authorities. The data are published yearly by the Robert Koch-Institute (RKI) in their disease-specific report [12]. The German notification system meets international quality standards [13]. Nevertheless, gaps in notifications are still possible. These might occur in vulnerable populations or only clinically diagnosed forms of TB without microbiologic verification. Further research data on TB are included in health insurance data. Data collections on specific populations who underwent active surveillance either for active or latent TB, e.g. for occupational reasons [14], or because of serial X-ray examinations [15], allow a different type of (historical) comparison [16]. Apart from the mentioned limitations for the GNC self-reports, there is the challenge that latent TB may be falsely reported as tuberculosis. Positive tests (e.g. tuberculin skin test) hint at latent TB, but unlike active tuberculosis, the condition is asymptomatic, and the microbiological detection of *M. tuberculosis complex* remains negative. This may result in an overestimation of the frequencies reported in the GNC. In our view, potential uses for the GNC data especially arise if the diagnosis for active or latent tuberculosis were confirmed, e.g. by information from biomaterials. In this case, crucial issues concerning disease progression after TB and possible consequences (e.g. restriction of lung function) can be investigated. This would exceed the follow-up-period of most of the existing other international cohort studies [17, 18].

Herpes zoster and postherpetic neuralgia

Despite the high disease burden of herpes zoster [19], there is no obligation to report the disease to authorities in Germany because of its low contagiousness. Therefore, studies concerned with the epidemiology of the disease are mostly based on secondary data [20, 21]. These studies were able to detect and quantify previously unknown cardiac and cerebrovascular consequences of a varicella-zoster virus (VZV) reactivation.

For herpes zoster and post-zoster neuralgia, an assessment of the validity of GNC pretest data has been performed. Presumably, earlier herpes zoster episodes were less likely to be reported than more recent episodes [24]. As already mentioned earlier, this embodies an important limitation of the self-reported infectious diseases in the GNC.

From an epidemiological point of view, GNC baseline and follow-up examinations are conducted within an interesting time period, since in Germany, the varicella (chickenpox) vaccination has been recommended by authorities since 2004 [25] and the herpes zoster vaccination since 2019 [26]. In contrast to studies based solely on secondary data, the longitudinal design of the GNC with repeated collection of biosamples and integration of vaccination card data allows research on risk factors and biomarkers for VZV-reactivation and possibly provides insight into molecular causes for vaccination failure. The results could be used for the development of personalised vaccination strategies.

HIV/ AIDS

In Germany, various tools are available to estimate the disease burden of HIV/ AIDS. A case register has been available since 1982, based on voluntary and anonymous reports of patients by physicians [27]. Since 1987, laboratories have been required to anonymously notify positive HIV-tests (currently § 7 par. 3 IfSG). Furthermore, reports on HIV frequencies from routine tests of all blood donations have been available [27]. From these data sources and the cause-of-death statistics of the statistical offices of the federal states as well as the sales data on antiretroviral therapy from pharmacy accounting centres, the RKI estimates incident HIV infections [28]. Additionally, multiple patient registers or cohorts exist, e.g. the HIV-1 seroconverter-study managed by the RKI since 1997 [29] or registries with clinical patient data ([30], www.tp-hiv.de). The 'Translational Platform HIV' (TP-HIV) has an integrated biobank and, therefore, allows molecular-genetic transmission investigations [31, 32]. Due to the low number of cases, it is difficult to answer specific research questions regarding HIV within the GNC; rather, information on HIV-infection may be used as exclusion criteria or confounder for immunological and infectiological GNC analyses.

Hepatitis B and C virus infection

In terms of hepatitis B and C virus infections (HBV/ HBC), the long latency between the infection and clinical disease as well as the often asymptomatic nature of the condition challenge epidemiological research. This becomes apparent in the cause-of-death statistics and the statistics of notifiable infectious diseases. According to updated surveillance case definitions, acute HBV cases are registered independently of stage and clinical presentation [33]. Thereby, asymptomatic cases can be reported, which addresses the underestimation of HBV in notifiable disease records. Additionally, blood donors are tested for both viruses [34]. The screening of pregnant women generates additional information regarding chronic HBV-infection [35] as do studies among high-risk populations [36] and individuals infected with HIV [37]. As for other infections, secondary data from health insurances can be analysed scientifically. However, these are restricted to individuals with insurance and focus on treatment parameters (e.g. [38]). Also, patient registries and studies comprise specific data, e.g. on disease

progression (www.deutsche-leberstiftung.de). Similar to HIV/AIDS, data on hepatitis in the GNC will mainly play a role as exclusion criteria because participants are rarely affected. On the other hand, long-term follow-up data from the GNC can investigate consequences of chronic viral hepatitis infection. Given the existence of serological vaccine markers for HBV, the collection of vaccination data within the GNC provides options for a comparative analysis of self-reported and serologically determined HBV immunity. Also, explorative investigations on proportion and predictors of possible HBV vaccine-non-responders can be conducted.

Infections of the upper respiratory tract

Infections of the upper respiratory tract are among the most frequent infectious diseases [39]. This is reflected in the frequencies obtained from the touchscreen module (Table 2). Moreover, e.g. influenza caused 33,116 disability-adjusted life years (DALYs) in 2005-2007 in Germany [40]. GNC data and if applicable, complemented by health insurance data, can provide insights into susceptibility to respiratory infections. This enables the investigation of a spectrum of immune-mediated phenotypes and, therefore, represents important participant-specific information for longitudinal questions. The construct 'infection susceptibility' can be integrated into these research questions as a global measure considering all information concerning infections and future biosample analysis (e.g. serostatus). Furthermore, the data analysis regarding URTI may contribute to confirm or extend the results of specialised local population studies (AWIS: [3, 41], LIFE-Adult: [42]) or studies focusing only on acute respiratory infections like GrippeWeb [43].

Infections of the lower respiratory tract

In Germany, pneumonia causes over 250,000 hospitalisations each year with a case fatality rate of approximately 13% [44]. Globally, infections of the lower respiratory tract (LRTI) are responsible for approximately 103 million DALYs [45]. With GNC data, existing studies on pneumonia from the clinical perspective ((CAPNETZ [46], PROGRESS [47]) can be expanded by the analysis of molecular factors and lifestyle data, and put on a wider data basis. Furthermore, the longitudinal design allows the assessment of long-term consequences of respiratory infections such as cardiovascular events [48]. Within future publications, the GNC data will be evaluated: It is conceivable, for instance, that respiratory diseases, such as asthma or chronic obstructive pulmonary disease (COPD), are mistaken for LRTI, which would lead to an overestimation of cases. The comprehensive database of the GNC, including information on medication intake, allows for a more differentiated consideration in the future.

Gastrointestinal infections

Acute gastrointestinal infections lead to 63.2 million missed working days due to illness [49]. In Germany, gastrointestinal illnesses are among the infectious diseases most frequently reported to the Robert Koch-Institute, which is particularly the case for infections caused by norovirus, rotavirus, campylobacter and salmonella [50]. Mild or moderate gastrointestinal tract infections are often not reported, especially among adults [51]. The results of international studies confirm this [52-55]. Depending on the study, only 20 to 40% of the diseased consulted a doctor; stool samples with pathogen diagnostics were only available for a fraction of these patients. For Germany, only a few epidemiological studies on gastrointestinal infections are available outside of outbreak investigations. In 2009, Wilking et al. surveyed more than 20,000 people once concerning acute gastrointestinal illnesses and their symptoms within the past four weeks [49]. In 2009-2010, a pilot project of the National Reference Network for Enteric Infections (NEI) analysed 185 stool samples of ambulant and hospitalised diarrhoea patients for gastrointestinal pathogens. Only 51% of the samples contained potential enteropathogens. This study aimed to establish a uniform approach for the systematic acquisition of relevant pathogens from stool [56]. This procedure should then be applied in epidemiological studies of NEI. In the context of the GNC, the enquiry of diarrhoeal diseases enables to approximately define the construct 'infection susceptibility' together with other queried infections and put into relation with other diseases addressed (longitudinally) in the GNC. Add-on studies, so called level 3 studies, within the GNC would enable to investigate pathogen-specific research questions by capturing incident events promptly. For instance, this would be feasible within the scope of the 'Integrated DZIF Infection Cohort within the German National Cohort (ZIFCO; www.info-pia.de).

Urinary tract infections

In Germany, there are no comparable studies with similar data collection on urinary tract infections (UTI). Following IfSG, there is no obligation to notify. The former Barmer GEK health insurance determined the UTI prevalence of the insured aged 12 and over for 2013 [57]. However, this study used secondary data which only capture cases in which a doctor has been consulted. The European ARESC-study with study centres in Germany analysed the pathogen spectrum for the assessment of the resistance situation. The goal was to evaluate the recommendations regarding uncomplicated cystitis [58]. Hence, GNC data can fill in a gap. Apart from that, the data can be considered for the construct 'infection susceptibility' (see above section on URTI). Additionally, self-reports on UTI from international studies mostly originate from female study populations [59, 60]. The self-reports from male GNC participants may add to this data to gain a broader insight into the epidemiology of transient infections. Finally, long-term effects of frequent UTI in women can be investigated, e.g. unspecific consequences such as shortened life expectancy.

Fever

Fever is a symptom of infection. It is commonly ascertained in studies as a symptom, e.g. of respiratory tract infection to differentiate between influenza-like illness (ILI) and common cold [61]. This is especially true for studies that complement the notification system by supplementary surveillance such as GrippeWeb [43]. Clinical studies on LRTI or sepsis also collect data on fever and, in some instances, also hypothermia [47]. However, temperature thresholds often differ between studies, which in turn limits cross-cohort evaluations (e.g. LIFE-Adult: $>38.5^{\circ}\text{C}$ in case of infection [42]; PROGRESS: $\geq 38.3^{\circ}\text{C}$ rectally or $\geq 37.8^{\circ}\text{C}$ orally [47]; GNC separate question: $>38^{\circ}\text{C}$). The GNC data can be incorporated into the construct 'infection susceptibility' in future analyses and taken into account in molecular investigations of the immune system.

Sepsis

Sepsis is a common complication of various diseases and especially occurs in hospitals and intensive care. The incidence is approximately 110 cases per 100,000 inhabitants per year and is associated with a high case fatality rate [62]. In sepsis, the initially targeted immune response of the body systemically expands, which can lead to considerable organ damage and death. Therefore, sepsis is an important medical research area. Large cohort studies like the GNC, with information on risk factors, molecular data and health insurance data can assist research on sepsis together with data from clinical cohorts (AlertsNet [63], SepNet study group (www.sepsis-stiftung.eu/sepnet/), PROGRESS (www.capnetz.de/html/progress/project), [47, 64]), population-based cohorts (LIFE-Adult [42]) and additional health insurance data. Here, the GNC also provides data on participants who have overcome sepsis and are healthy for study participation.

In summary, it can be stated that on all infectious diseases queried in the GNC numerous and very diverse research questions can be investigated in the future. Thereby, the GNC complements the existing research landscape in a variety of ways.

Outlook

Based on a cross-sectional design, publications on results from the GNC in the area of infection research can be expected from 2021 onwards, e.g. on determinants of susceptibility to certain self-reported infections. In particular, it will be possible to examine the distribution of infectious diseases by social class. Further research in the field of infectious disease epidemiology will be feasible after GNC biosamples have been analysed.

Separately financed serological determinations of antibodies against the causal pathogen of Lyme disease and analyses of the stool or nasal microbiome up to high-dimensional multi-omics analyses are

also conceivable. Infection-specific subcohorts with timely data and symptomatic biosample collection concerning acute infections as well as the collection of viable blood cells for characterisations of the immune system (www.info-pia.de) require additional funding. These types of investigations have not been conducted within the GNC so far. Future research in the area of infectious disease epidemiology will, therefore, focus on these additional projects and analyses.

Ethics

All described examinations on humans have been conducted with the approval of the responsible ethics committees, following national laws and the Declaration of Helsinki of 1975 (in the current, revised version). All participants provided signed informed consent.

Conflicts of interest

The authors state that there are no conflicts of interest.

Funding

This project was carried out with data from the German National Cohort (www.nako.de). The GNC is funded by the Federal Ministry of Education and Research (BMBF) [project funding reference numbers: 01ER1301A/B/C and 01ER1511D], the federal states and the Helmholtz Association as well as by the participating universities and institutes of the Leibniz Association. Max Hassenstein receives a scholarship from the Life Science Foundation to promote science and research.

Acknowledgements

We want to thank all GNC participants and employees involved, particularly the experts who were involved in the development of questions on infectious diseases.

References

1. Plass D, Vos T, Hornberg C, Scheidt-Nave C, Zeeb H, Krämer A (2014) Trends in Disease Burden in Germany: Results, Implications and Limitations of the Global Burden of Disease Study. *Dtsch Arztebl Int* 111(38):629–638. DOI 10.3238/arztebl.2014.0629
2. Lynch S V, Pedersen O (2016) The Human Intestinal Microbiome in Health and Disease. *N Engl J Med* 375(24):2369–2379. DOI 10.1056/NEJMra1600266
3. Maccioni L, Weber S, Elgizouli M et al. (2018) Obesity and risk of respiratory tract infections: results of an infection-diary based cohort study. *BMC Public Health* 18(1):271. DOI 10.1186/s12889-018-5172-8

4. Moore P S, Chang Y (2010) Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nature Reviews Cancer* 10(12):878–889. DOI 10.1038/nrc2961
5. German National Cohort (GNC) Consortium (2014) The German National Cohort: aims, study design and organization. *Eur J Epidemiol* 29(5):371–382. DOI 10.1007/s10654-014-9890-7
6. Sievers C, Akmatov M K, Kreienbrock L et al. (2014) Evaluation of a questionnaire to assess selected infectious diseases and their risk factors: findings of a multicenter study. *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz* 57(11):1283–1291. DOI 10.1007/s00103-014-2052-y
7. Castell S, Akmatov M K, Obi N et al. (2014) Test-retest reliability of an infectious disease questionnaire and evaluation of self-assessed vulnerability to infections : findings of Pretest 2 of the German National Cohort. *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz* 57(11):1300–1307. DOI 10.1007/s00103-014-2045-x
8. Schipf, S., Schöne, G., Schmidt, B. et al. (2020) Die Basiserhebung der NAKO Gesundheitsstudie: Teilnahme an den Untersuchungsmodulen, Qualitätssicherung und Nutzung von Sekundärdaten. *Bundesgesundheitsbl.* <https://doi.org/10.1007/s00103-020-03093-z>
9. Hess L M, Method M W, Stehman F B, Weathers T D, Gupta P, Schilder J M (2012) Patient Recall of Health Care Events and Time to Diagnose a Suspected Ovarian Cancer. *Clinical Ovarian and Other Gynecologic Cancer* 5(1):17–23. DOI 10.1016/j.cogc.2012.04.001
10. Kjellsson G, Clarke P, Gerdtham U-G (2014) Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *J Health Econ* 35:34–46. DOI 10.1016/j.jhealeco.2014.01.007
11. Sterne J A C, White I R, Carlin J B et al. (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338:b2393. DOI 10.1136/bmj.b2393
12. Robert Koch-Institut (2018) Bericht zur Epidemiologie der Tuberkulose in Deutschland für 2018. Robert Koch-Institut, Berlin. DOI 0.17886/rkipubl-2018-002
13. Hauer B, Fiebig L, Brodhun B, Haas W (2012) Tuberculosis surveillance and control in Germany - An application of the Berlin Declaration Monitoring and Evaluation Framework. *Eur J Microbiol Immunol (Bp)* 2(4):287–291. DOI 10.1556/EuJMI.2.2012.4.7

14. Nienhaus A, Schablon A, Preisser A M, Ringshausen F C, Diel R (2014) Tuberculosis in healthcare workers - a narrative review from a German perspective. *J Occup Med Toxicol* 9(1):9. DOI 10.1186/1745-6673-9-9.
15. o. V. (1982) Röntgenreihenuntersuchungen werden eingestellt. *Staatsanzeiger Baden-Württemberg*. Nr. 96 vom 04.12.1982
16. Statistisches Amt des Vereinigten Wirtschaftsgebietes (1949) *Statistische Berichte VIII/3/1-17. Erkrankungen an aktiver Tuberkulose 1949-2.Vj.1952.*
https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEHeft_derivate_00023654/StatBer-8-03-001.pdf Zugriffen: 30.08.2019
17. Fox G J, Nguyen V N, Dinh N S et al. (2019) Post-treatment Mortality Among Patients With Tuberculosis: A Prospective Cohort Study of 10 964 Patients in Vietnam. *Clin Infect Dis* 68(8):1359–1366. DOI 10.1093/cid/ciy665
18. Mkoko P, Naidoo S, Mbanga L C, Nomvete F, Muloiwa R, Dlamini S (2019) Chronic lung disease and a history of tuberculosis (post-tuberculosis lung disease): Clinical features and in-hospital outcomes in a resource-limited setting with a high HIV burden. *S Afr Med J* 109(3):169–173. DOI 10.7196/SAMJ.2019.v109i3.13366.
19. Ultsch B, Köster I, Reinhold T et al. (2013) Epidemiology and cost of herpes zoster and postherpetic neuralgia in Germany. *Eur J Health Econ* 14(6):1015–1026. DOI 10.1007/s10198-012-0452-1
20. Hillebrand K, Bricout H, Schulze-Rath R, Schink T, Garbe E (2015) Incidence of herpes zoster and its complications in Germany, 2005-2009. *J Infect* 70(2):178–186. DOI 10.1016/j.jinf.2014.08.018
21. Ultsch B, Siedler A, Rieck T, Reinhold T, Krause G, Wichmann O (2011) Herpes zoster in Germany: quantifying the burden of disease. *BMC Infect Dis* 11:173. DOI 10.1186/1471-2334-11-173
22. Langan S M, Minassian C, Smeeth L, Thomas S L (2014) Risk of stroke following herpes zoster: a self-controlled case-series study. *Clin Infect Dis* 58(11):1497–1503. DOI 10.1093/cid/ciu098
23. Minassian C, Thomas S L, Smeeth L, Douglas I, Brauer R, Langan S M (2015) Acute Cardiovascular Events after Herpes Zoster: A Self-Controlled Case Series Analysis in Vaccinated and Unvaccinated Older Residents of the United States. *PLoS Med* 12(12):e1001919. DOI 10.1371/journal.pmed.1001919

24. Caputo M, Horn J, Karch A et al. (2019) Herpes zoster incidence in Germany - an indirect validation study for self-reported disease data from pretest studies of the population-based German National Cohort. *BMC Infect Dis* 19(1):99. DOI 10.1186/s12879-019-3691-2
25. Robert Koch-Institut (Hrsg.) (2004) Varizellen (Windpocken). *Epidemiol Bull* 49:421-426
26. Robert Koch-Institut (Hrsg.) (2018) Mitteilung der Ständigen Impfkommision (STIKO) beim RKI: Wissenschaftliche Begründung zur Empfehlung einer Impfung mit dem Herpes zoster-subunit-Totimpfstoff. *Epidemiol Bull* 50:555-562
27. Robert Koch-Institut (2017) HIV/AIDS-Surveillance. AIDS-Fallregister. https://www.rki.de/DE/Content/InfAZ/H/HIVAIDS/Epidemiologie/Surveillance/surveillance_inhalt.html Zugegriffen: 30.08.2019
28. Robert Koch-Institut (Hrsg.) (2017) Schätzung der Zahl der HIV-Neuinfektionen und der Gesamtzahl von Menschen mit HIV in Deutschland. *Epidemiol Bull* 47:533-534
29. Robert Koch-Institut (2017) HIV/AIDS. HIV-1 Serokonverterstudie. <https://www.rki.de/DE/Content/InfAZ/H/HIVAIDS/Studien/SeroKon/SeroKon-tab-gesamt.html> Zugegriffen: 30.08.2019
30. Robert Koch-Institut (2012) Klinische Surveillance der HIV-Erkrankung (ClinSurv). <https://www.rki.de/DE/Content/InfAZ/H/HIVAIDS/Epidemiologie/Surveillance/ClinSurv/ClinSurvHIV.html> Zugegriffen: 30.08.2019
31. Stecher M, Schommers P, Schmidt D et al. (2019) Antiretroviral treatment indications and adherence to the German-Austrian treatment initiation guidelines in the German ClinSurv HIV Cohort between 1999 and 2016. *Infection* 47(2):247–255. DOI 10.1007/s15010-018-1248-8.
32. Stecher M, Hoenigl M, Eis-Hübinger A M et al. (2019) Hotspots of Transmission Driving the Local Human Immunodeficiency Virus Epidemic in the Cologne-Bonn Region, Germany. *Clin Infect Dis* 68(9):1539–1546. DOI 10.1093/cid/ciy744
33. Robert Koch-Institut (Hrsg.) (2015) Auswirkung der Falldefinitionsänderungen auf die Fallzahlen von Hepatitis B und Hepatitis C im Jahr 2015, *Epidemiol Bull* 49:528
34. Offergeld R, Ritter S, Quabeck L, Hamouda O (2010) Infektionsepidemiologische Daten von Blutspendern in Deutschland 2007 (Epidemiological data on infections among blood donors in Germany 2007). *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz* 53(11):1188–1196. DOI 10.1007/s00103-010-1152-6

35. Knorr B, Maul H, Schnitzler P (2008) Prevalence of hepatitis B virus infection among women at reproductive age at a German university hospital. *J Clin Virol* 42(4):422–424. DOI 10.1016/j.jcv.2008.03.009
36. Zimmermann R, Marcus U, Schäffer D et al. (2014) A multicentre sero-behavioural survey for hepatitis B and C, HIV and HTLV among people who inject drugs in Germany using respondent driven sampling. *BMC Public Health* 14:845. DOI 10.1186/1471-2458-14-845.
37. Jansen K, Thamm M, Bock C-T et al. (2015) High Prevalence and High Incidence of Coinfection with Hepatitis B, Hepatitis C, and Syphilis and Low Rate of Effective Vaccination against Hepatitis B in HIV-Positive Men Who Have Sex with Men with Known Date of HIV Seroconversion in Germany. *PLoS ONE* 10(11):e0142515. DOI 10.1371/journal.pone.0142515
38. Westermann C, Wendeler D, Nienhaus A (2018) Hepatitis C in healthcare personnel: secondary data analysis of therapies with direct-acting antiviral agents. *J Occup Med Toxicol* 13:16. DOI 10.1186/s12995-018-0197-6
39. James S L, Abate D, Abate K H et al. (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* 392(10159):1789–1858. DOI 10.1016/S0140-6736(18)32279-7
40. Plass D, Mangen M-J J, Kraemer A et al. (2014) The disease burden of hepatitis B, influenza, measles and salmonellosis in Germany: first results of the burden of communicable diseases in Europe study. *Epidemiol Infect* 142(10):2024–2035. DOI 10.1017/S0950268813003312
41. Nieters A, Blagitko-Dorfs N, Peter H-H, Weber S (2019) Psychophysiological insomnia and respiratory tract infections: results of an infection-diary-based cohort study. *Sleep* 42(8). DOI 10.1093/sleep/zsz098
42. Loeffler M, Engel C, Ahnert P et al. (2015) The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* 15:691. DOI 10.1186/s12889-015-1983-z
43. Robert Koch-Institut (Hrsg.) (2011) Bevölkerungsbasierte Erhebung der Häufigkeit von akuten Atemwegserkrankungen. Erste Ergebnisse aus dem neuen Online-Portal GrippeWeb. 37:343-347
44. Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen (AQUA) (2015) PNEU - Ambulant erworbene Pneumonie. Qualitätsindikatoren. Bundesauswertung

zum Erfassungsjahr 2014.

https://www.sqg.de/downloads/Bundesauswertungen/2014/bu_Gesamt_PNEU_2014.pdf

Zugegriffen: 30.08.2019

45. Troeger C, Forouzanfar M, Rao P C et al. (2017) Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases* 17(11):1133–1161. DOI 10.1016/S1473-3099(17)30396-1
46. Suttorp N, Welte T, Marre R et al. (2016) CAPNETZ. Das Kompetenzzentrum für ambulant erworbene Pneumonie (CAPNETZ. The competence network for community-acquired pneumonia (CAP)). *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz* 59(4):475–481. DOI 10.1007/s00103-016-2318-7
47. Ahnert P, Creutz P, Scholz M et al. (2016) PROGRESS - prospective observational study on hospitalized community acquired pneumonia. *BMC Pulm Med* 16(1):108. DOI 10.1186/s12890-016-0255-8
48. Corrales-Medina V F, Alvarez K N, Weissfeld L A et al. (2015) Association between hospitalization for pneumonia and subsequent risk of cardiovascular disease. *JAMA* 313(3):264–274. DOI 10.1001/jama.2014.18229
49. Wilking H, Spitznagel H, Werber D, Lange C, Jansen A, Stark K (2013) Acute gastrointestinal illness in adults in Germany: a population-based telephone survey. *Epidemiol Infect* 141(11):2365–2375. DOI 10.1017/S0950268813000046
50. Robert Koch-Institut (Hrsg.) (2019) Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2018. Robert Koch-Institut, Berlin, S. 35-39. DOI 10.25646/5978
51. Manthey C, Stallmach A, Lohse A (2018) Gastrointestinale Infektionen – Aktuelles. *Gastroenterologie up2date* 14(02):187–203. DOI 10.1055/s-0043-124882
52. Scallan E, Fitzgerald M, Collins C et al. (2004) Acute gastroenteritis in northern Ireland and the Republic of Ireland: a telephone survey. *Commun Dis Public Health* 7(1):61–67.
53. van Cauteren D, Valk H de, Vaux S, Le Strat Y, Vaillant V (2012) Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiol Infect* 140(4):697–705. DOI 10.1017/S0950268811000999
54. Grilc E, Sočan M (2014) Population based self-reported acute gastrointestinal infection in slovenia: multiplier study. *Zdrav Var* 53(2):125–132. DOI 10.2478/sjph-2014-0013

55. Baumann-Popczyk A, Sadkowska-Todys M, Rogalska J, Stefanoff P (2012) Incidence of self-reported acute gastrointestinal infections in the community in Poland: a population-based study. *Epidemiol Infect* 140(7):1173–1184. DOI 10.1017/S0950268811001853
56. Robert Koch-Institut (Hrsg.) (2011) Gastrointestinale Infektionen. Zur Prävalenz von Enteropathogenen im Einzugsbereich Wernigerode (Pilotprojekt des NEI). *Epidemiol Bull* 20:170-175
57. Glaeske G, Schicktanz C (2015) BARMER GEK Arzneimittelreport 2015. Auswertungsergebnisse der BARMER GEK Arzneimitteldaten aus den Jahren 2013 bis 2014.
58. Wagenlehner F M E, Wagenlehner C, Savov O, Gualco L, Schito G, Naber K G (2010) Klinik und Epidemiologie der unkomplizierten Zystitis bei Frauen. Deutsche Ergebnisse der ARESC-Studie (Clinical aspects and epidemiology of uncomplicated cystitis in women. German results of the ARESC Study). *Urologe A* 49(2):253–261. DOI 10.1007/s00120-009-2145-7
59. Butler C C, Francis N, Thomas-Jones E et al. (2017) Variations in presentation, management, and patient outcomes of urinary tract infection: a prospective four-country primary care observational cohort study. *Br J Gen Pract* 67(665):e830-e841. DOI 10.3399/bjgp17X693641
60. Foxman B, Barlow R, D'Arcy H, Gillespie B, Sobel J D (2000) Urinary Tract Infection: Self-Reported Incidence and Associated Costs. *Annals of Epidemiology* 10(8):509–515. DOI 10.1016/S1047-2797(00)00072-7
61. Fitzner J, Qasmieh S, Mounts A W et al. (2018) Revision of clinical case definitions: influenza-like illness and severe acute respiratory infection. *Bull World Health Organ* 96(2):122–128. DOI 10.2471/BLT.17.194514
62. Hagel S, Brunkhorst F (2011) Sepsis. *Intensivmed* 48(1):57–73. DOI 10.1007/s00390-010-0249-3
63. Karch A, Schmitz R P, Reißner F et al. (2015) Bloodstream infections, antibiotic resistance and the practice of blood culture sampling in Germany: study design of a Thuringia-wide prospective population-based study (AlertsNet). *BMJ Open* 5(12):e009095. DOI 10.1136/bmjopen-2015-009095
64. Ahnert P, Creutz P, Horn K et al. (2019) Sequential organ failure assessment score is an excellent operationalization of disease severity of adult patients with hospitalized community acquired pneumonia - results from the prospective observational PROGRESS study. *Crit Care* 23(1):110. DOI 10.1186/s13054-019-2316-x

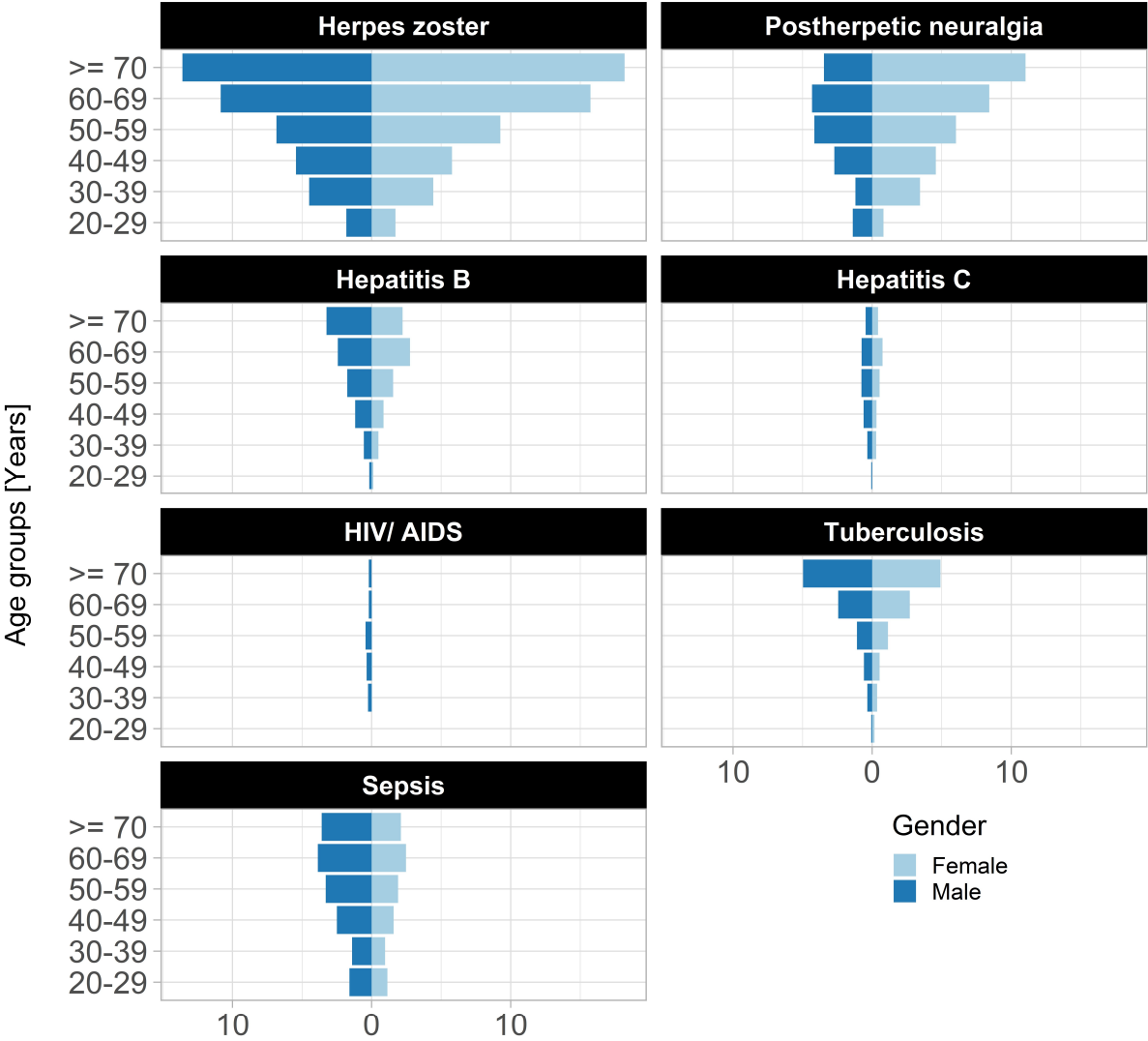
Table 1 Questions and possible answers of GNC survey tools

Question	(Infectious-)Disease	Possible answers
<i>Interview</i>		
Have you ever been medically diagnosed with one of the following infectious diseases?	Tuberculosis	<ul style="list-style-type: none"> • Yes • No • Prefer not to answer • Don't know
	Herpes zoster	
	Postherpetic neuralgia (if herpes zoster was reported)	
	HIV/ AIDS	
	Hepatitis B	
	Hepatitis C	
Have you ever received inpatient treatment in a hospital for blood poisoning, also known as sepsis?		<ul style="list-style-type: none"> • Yes • No • Prefer not to answer • Don't know
<i>Infectious disease questions in the touchscreen-module</i>		
In the past 12 months, how often have you had...	...an infection of the upper respiratory tract (e.g. a cold or infections of the sinus cavities, tonsils, middle ear, throat, voice box)?	<ul style="list-style-type: none"> • Never • 1 time • 2 times • 3- to 4 times • 5- to 6 times • More than 6 times • Prefer not to answer • Don't know
	... an infection of the lower respiratory tract (e.g. bronchitis or lung inflammation)?	
	...a gastrointestinal infection with diarrhoea (diarrhoea is defined as at least three liquid/ mushy bowel movements within 24 hours)?	
	...a bladder infection?	
	...fever (>38°C)?	

Table 2 Frequencies of self-reported infections in the GNC interview und touchscreen

	Reporting of the disease: n (%)	Considered answers: n	Missing values, 'Prefer not to answer', 'Don't know': n
<i>Interview (Self-reports for medical diagnosis in the course of life)</i>			
Tuberculosis	1,423 (1.4)	101,453	334
HIV/ AIDS	156 (0.2)	101,652	135
Hepatitis B	1,615 (1.6)	100,499	1,288
Hepatitis C	557 (0.6)	100,710	1,077
Herpes zoster	8,518 (8.6)	98,793	345
Postherpetic neuralgia	587 (7)	8,418	100
Sepsis as a reason for hospitalisation	2,434 (2.4)	101,366	421
<i>Infectious disease questions in the touchscreen-module (self-reports considering the past 12 months)</i>			
At least one infection of the upper respiratory tract (URTI)	75,380 (81.2)	92,853	8,934
At least one infection of the lower respiratory tract (LRTI)	19,892 (21.7)	91,829	9,958
At least one gastrointestinal infection	35,111 (38.0)	92,293	9,494
At least one bladder infection	11,227 (12.2)	92,232	9,555
At least one occurrence of fever (> 38°C)	23,250 (25.4)	91,609	10,178

Figure 1: Age- and gender-specific frequencies of infectious diseases (%) ever medically diagnosed, from the interview. The age data are aggregated into 10-year groups.



Proportions based on all individuals in the corresponding age and gender groups [%]

Figure 2: Part of figure 1 with modified scale for better representation. Age- and gender-specific frequencies of infections (%) ever medically diagnosed with hepatitis C or HIV/ AIDS, from the interview. The age data are aggregated into 10-year groups.

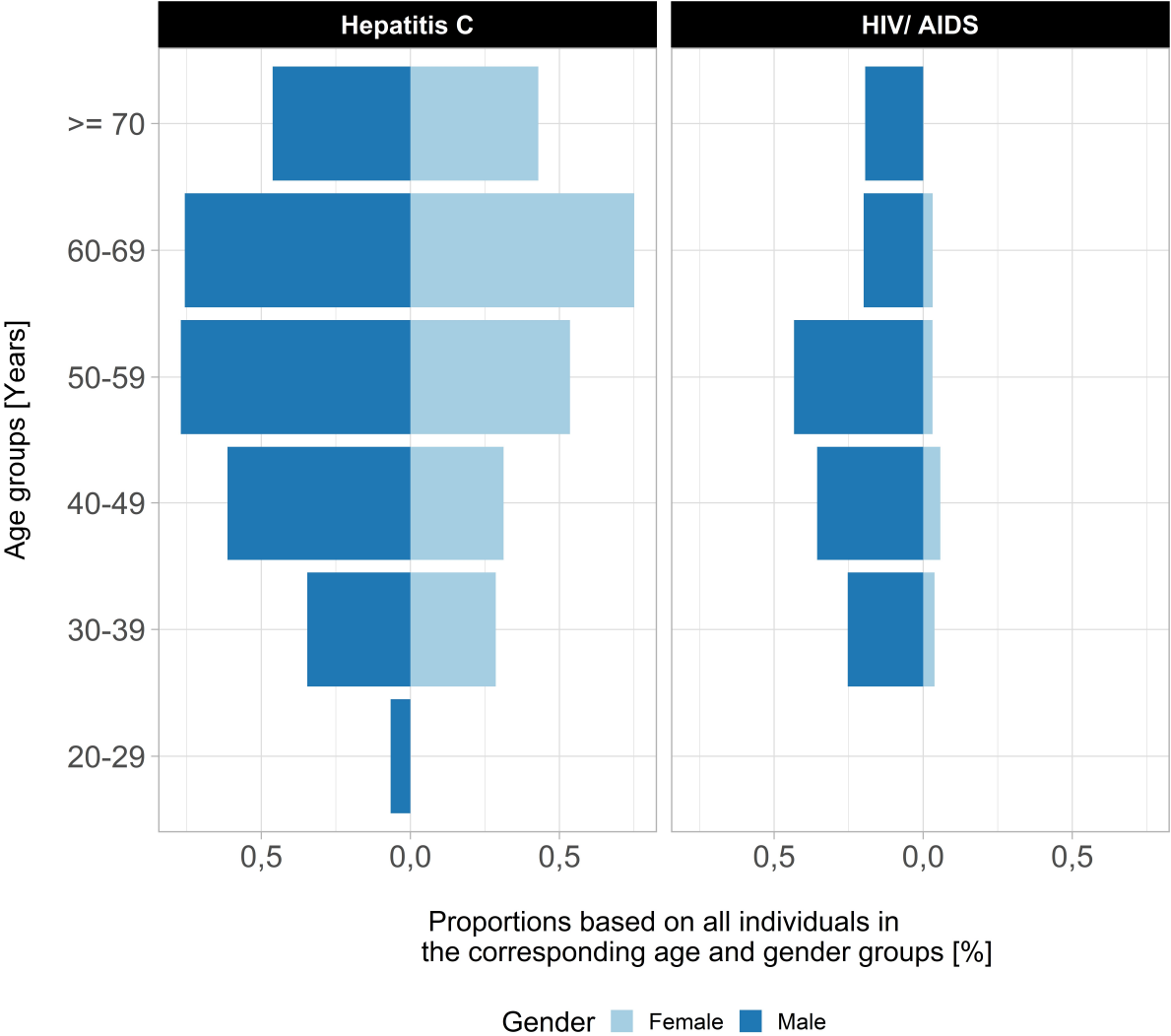
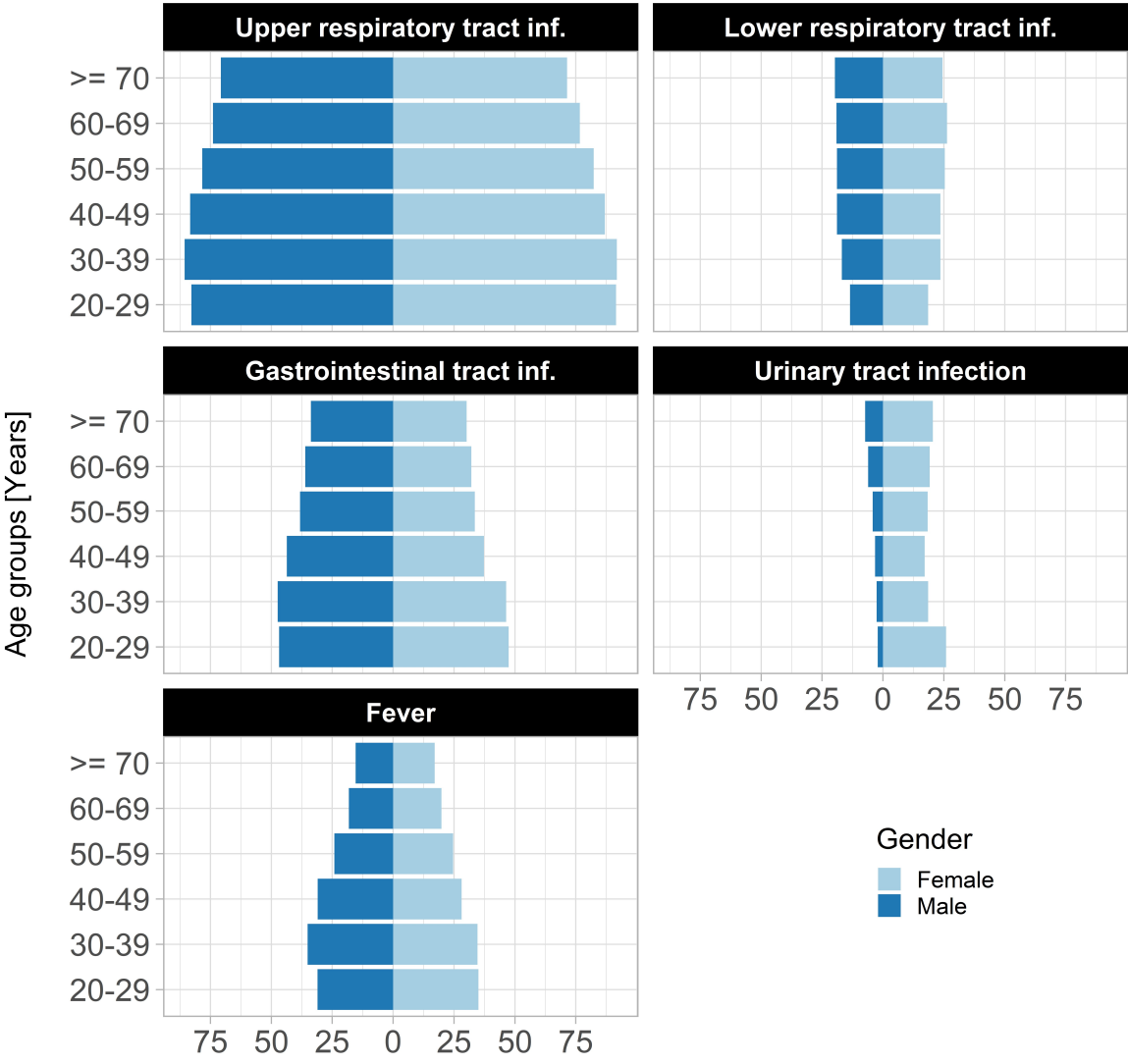


Figure 3: Age- and gender-specific frequencies (%) of the infectious diseases or symptoms within the past 12 months, from the touchscreen module. The age data are aggregated into 10-year groups.



Proportions based on all individuals in the corresponding age and gender groups [%]