BMJ Open

Risk factors for neuroendocrine neoplasms: a protocol for a case-control study based on a record linkage of registry and claims data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-096273
Article Type:	Protocol
Date Submitted by the Author:	07-Nov-2024
Complete List of Authors:	Voigtländer, Sven; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Grundmann, Nina; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Hakimhashemi, Amir; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Meyer, Martin; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Gerlach, Roman; Association of Statutory Health Insurance Physicians of Bavaria Donnachie, Ewan; Bavarian Association of Statutory Health Insurance Physicians Berger, Ursula; Ludwig-Maximilians-University Munich, Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine Tauscher, Martin; Association of Statutory Health Insurance Physicians of Bavaria Müller-Nordhorn, Jacqueline; Bavarian Cancer Registry, Bavarian Health and Food Safety Authority
Keywords:	Gastrointestinal tumours < ONCOLOGY, Respiratory tract tumours < ONCOLOGY, Risk Factors, Case-Control Studies, REGISTRIES

STUDY PROTOCOL

Risk factors for neuroendocrine neoplasms: a protocol for a case-control study based on a record linkage of registry and claims data

Sven Voigtländer,[1] Roman Gerlach,[2] Nina Grundmann,[1] Ewan Donnachie,[2] Ursula Berger,[3] Amir Hakimhashemi,[1] Martin Meyer,[1] Martin Tauscher,[2] Jacqueline Müller-Nordhorn[1]

[1] Bavarian Cancer Registry, Bavarian Health and Food Safety Authority, Schweinauer Hauptstraße 80, 90441 Nuremberg, Germany

[2] Bavarian Association of Statutory Health Insurance Accredited Physicians, Elsenheimerstraße 39, 80687 Munich, Germany

[3] Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine at Ludwig-Maximillians-University Munich, Marchioninistraße 15, 81377 Munich, Germany

Word count (without title page, abstract, statements, and references): 2,547

Correspondence:
Dr. Sven Voigtländer
Bavarian Cancer Registry, Bavarian Health and Food Safety Authority
Schweinauer Hauptstraße 80, 90441 Nuremberg, Germany
Phone: +49-(0)9131-6808-2927
Email: sven.voigtlaender@lgl.bayern.de
ORCID ID: https://orcid.org/0000-0003-3808-9605

ABSTRACT

Background:

Recent studies showed an increase in neuroendocrine neoplasms, especially for the digestive tract. Several risk factors have been suggested to explain this increase including a family history of cancer, tobacco smoking, alcohol consumption, and metabolic disorders including diabetes and obesity. Another risk factor may be depressive disorders, which could increase the risk of neuroendocrine neoplasms either directly or mediated through associated risk behaviours and/or antidepressant medication. Here, we outline the design of our study to identify risk factors for neuroendocrine neoplasms in Germany.

Methods and analysis:

A case-control study of the resident population of Bavaria, the second most populous federal state in Germany, based on a record linkage of data from the Bavarian Cancer Registry and data from the Bavarian Association of Statutory Health Insurance Accredited Physicians. Cases have a diagnosis of a malignant neuroendocrine neoplasm, either of the bronchopulmonary system or the gastroenteropancreatic system, in 2020 or 2021. Controls are sampled from the noncases and matched on sex, birth year (in 5-year intervals), and time of diagnosis (by calendar quarter). Risk factor prevalence of cases and controls is assessed on the basis of assured outpatient diagnoses, i.e. diagnoses documented in at least two out of four consecutive quarters in the eight quarters preceding the diagnosis of a neuroendocrine neoplasm. The analysis uses conditional logistic regression to estimates odds ratios and 95% confidence intervals.

Ethics and dissemination:

This study protocol was approved by the Ethics Committee of the Bavarian State Chamber of Physicians (reference number: 24008). Approval by supervisory authority has been obtained from the Bavarian State Ministry of Health, Care, and Prevention (reference number: G35h-A1080-2023/20-2) and also the Bavarian Data Protection Commissioner stated to have no concerns after presentation of the study protocol (reference number: DSB/7-692/1-275). The results of the case-control study will be presented at national as well as international conferences and be published in the form of scientific articles in peer-reviewed journals.

Registration details:

ClinicalTrials.gov, NCT: 06282016.

Keywords:

Neuroendocrine neoplasms, colorectal cancer, risk factors, case-control study, cancer registry, claims data

ARTICLE SUMMARY

Strengths and limitations of this study

- Cases and controls will be drawn from an unselected source population, i.e. persons, who were insured with the Statutory Health Insurance (85% of all residents) and had at least one outpatient physician contact in 2020 or 2021 within Bavaria
- Cases will comprise all malignant neuroendocrine neoplasms of the gastroenteropancreatic and bronchopulmonary system in the source population in 2020 and 2021 based on high-quality cancer registry data
- The study will measure exposures based on cancer registry data and assured ICD-10 diagnoses and is, therefore, not prone to recall bias
- The study will, different to previous studies, also stratify for neuroendocrine tumours and neuroendocrine carcinomas based on high-quality cancer registry data.
- One limitation is that operationalisation of exposures is limited to exposures, which can be measured based on cancer registry as well claims data

BACKGROUND

Neuroendocrine neoplasms (NEN) are malignancies of neuroendocrine cells.[1] Neuroendocrine cells can be found throughout the body, i.e. where there is epithelium, excluding the central nervous system, bones, and soft tissue. NEN are rare and heterogeneous tumours, which most commonly arise in the gastroenteropancreatic system (GEPS) and the bronchopulmonary system (BPS).[1-3] NEN are categorised according to their differentiation into differentiated neuroendocrine tumours (NET) and poorly differentiated carcinomas (NEC).[4] NET can be further subdivided based on grade while NEC are further categorised into small cell and large cell carcinomas. In addition, there are also mixed neuroendocrine-non-neuroendocrine neoplasms (MiNEN).

Recent studies from the United States (US) and Germany showed an increase in NEN, especially for the digestive tract.[5-9] According to an analysis of data from the US Surveillance, Epidemiology and End Results (SEER) program, age-standardised incidence rates (ASR, US standard population 2000) significantly increased for most gastro-intestinal sites between 1975 and 2008.[7] For instance, the age-adjusted incidence rate for rectal NEN increased from 0.1 in 1975 to 1.1 per 100,000 persons in 2008.[7] The risk for NEN is known to increase with age.[2] Several other risk factors have been suggested by a limited number of studies including a family history of cancer, tobacco smoking, alcohol consumption, and metabolic disorders including diabetes and obesity.[2 10] Apart from this, there may be other risk factors, which have not been confirmed yet, such as depressive disorders, which could increase the risk of neuroendocrine neoplasms either directly or mediated through associated risk behaviours and antidepressant medication. Kenner,[11] for instance, discusses the role of depressive disorders for pancreatic cancer.

The aim of this study is to identify risk factors for neuroendocrine neoplasms based on data of the Bavarian Cancer Registry, Germany, to help understand the increase in neuroendocrine neoplasms.

METHODS

Study design

The study design is a case-control study based on data from the population-based Bavarian Cancer Registry and data from the Bavarian Association of Statutory Health Insurance Accredited Physicians (KVB, German: *Kassenärztliche Vereinigung Bayerns*). The Bavarian Cancer Registry is based on mandatory notifications by physicians and health care providers regarding the diagnosis and treatment of cancer.[12] The KVB collects, on a quarterly basis, diagnosis and treatment data related to its main task, i.e. ensuring and reimbursing

outpatient medical care and psychotherapy for patients with Statutory Health Insurance (GKV, German: *Gesetzliche Krankenversicherung*) in Bavaria. Bavarian Cancer Registry data and KVB data are linked by pseudonymised record linkage following the probability linkage procedure established by the German cancer registries.[13 14] The probability linkage is based on the conversion of KVB identity data into unique tokens, i.e. pseudonymisation, that are compared with the tokens already present in the Bavarian Cancer Registry by probabilistic linkage.[14] After linkage, the pseudonyms are removed resulting in an anonymous dataset.

Setting

The setting is the resident population of Bavaria in the years 2020 and 2021. Bavaria is the second most populous federal state in Germany with about 13 million residents. Though the Bavarian Cancer Registry covers the complete resident population of Bavaria, the source population of the study is limited to persons, who are insured with the GKV and had at least one outpatient physician contact in 2020 or 2021 within Bavaria. Persons who are not insured in the GKV as well as persons without outpatient physician contact in 2020 and 2021 within Bavaria are not included in the KVB data. In Bavaria, about 11.5 million residents, which is 85% of all residents, are insured with the GKV.[15]

Definition and recruitment of cases and controls

Eligible cases are defined as persons with a malignant NEN of the BPS or GEPS diagnosed in 2020 or 2021. Definition of malignant NEN as well as of BPS and GEPS is based on the fifth edition of the World Health Organization's (WHO) Classification of Tumours, also known as the WHO Blue Books,[4 16] and the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10),[17] respectively. Detailed information on the eligible combinations of histology code and tumour site is provided in the Supplement (Supplement 1). The selection of cases is based on the record linkage of Bavarian Cancer Registry and KVB data, i.e. persons with a malignant NEN in 2020 or 2021 according to the Bavarian Cancer Registry data that had a GKV insurance as well as an outpatient physician contact in 2020 or 2021. Eligibility is not limited to age.

Eligible controls are defined as noncases in the same source population as for cases, i.e. the KVB data. Selection of controls is done by random sampling.

Cases will be matched 1:2 with controls on sex, birth year (5-year intervals), and time of diagnosis (by calendar quarter). Two controls will be matched to each case.

Definition of outcomes, exposures, and confounders

Outcomes

Primary outcomes are NEN of the BPS and the GEPS as defined by eligible combinations of histology code and tumour site (see Supplement 1). Secondary outcomes are the two subcategories of NEN, i.e. NET and NEC.

Exposures and potential confounders

The following exposures and potential confounders will be assessed:

Exposures: disease-related exposures

1. Depression or persistent mood disorder
2. Obesity
3. Metabolic syndrome
4. Previous malignant neoplasms
5. Alcohol abuse (only for NEN of the GEPS)
6. Diabetes (only for NEN of the GEPS)
7. Ulcerative colitis (only for NEN of the GEPS)
8. Crohn disease (only for NEN of the GEPS)
9. Allergies (only for NEN of the BPS)
10. Asthma (only for NEN of the BPS)

Exposures: sociodemographic

1. Area deprivation of the residence municipality

Potential confounders: sociodemographic

1. Birth year (5-year intervals)
2. Sex
3. Rurality of the residence district

Potential confounders: time of NEN diagnosis

1. Time of diagnosis (calendar quarter)

Measurement of outcomes, exposures, and confounders

Outcomes are measured using the ICD-10 as well as histological information in the data of the Bavarian Cancer Registry (see Supplement 1). Previous malignant neoplasms, birth year (5-year intervals), sex, and time of diagnosis (by calendar quarter) are also taken from the BKR data. Previous malignant neoplasms includes all neoplasms with ICD-10 codes C00-

C97 (malignant neoplasms) except C44 (other malignant neoplasms of skin).[17] For cases and controls, the prevalence of the disease-related exposures, apart from previous malignant neoplasms, is assessed by specific ICD-10 diagnoses based on the KVB data (see Supplement 2 for further details). Only assured ICD-10 diagnoses will be considered,[18] i.e. diagnoses that were recorded in at least two out of four consecutive quarters in the eight quarters preceding the quarter of the NEN diagnoses for cases as well as controls, which are matched by quarter of NEN diagnosis. Area deprivation of the residence municipality of cases and controls is measured by the Bavarian Index of Multiple Deprivation, which is based on official data.[19] In particular, we use the deprivation quintile of the residence municipality at the time of NEN diagnosis. Rurality of the residence district at the time of NEN diagnosis is based on the dichotomous categorisation of districts into urban area and rural area by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development.[20]

Bias

The case-control study is based on a record linkage of registry data, claims data, and administrative data. These data sources are, unlike survey data, not prone to recall bias. The potential of selection bias is considerably reduced by a small number of matching variables, i.e. birth year, sex, and time of NEN diagnosis, as well as the use of an almost unselected source population for cases and controls, i.e. the KVB data. The KVB data covers all persons with GKV (about 85% of all residents in Bavaria) and at least one outpatient physician contact in Bavaria within the study period of 2020-2021. According to the KVB, 94% of all GKV insured persons had at least one contact with a general practitioner in 2021,[21] so that about 80% of all residents are included in our source population. It is known that the proportion of persons with GKV, who use outpatient care, is higher for women compared to men and lower for younger persons compared to older persons.[18] Regarding potential differential misclassification bias, assessment of disease-related exposures is done in the eight quarters preceding the quarter of the NEN diagnosis to ensure that the prevalence and number of disease-related exposures is not influenced by potentially increased clinical investigation related to the NEN diagnosis.

Confounding is controlled by matching for birth year, sex, and time of NEN diagnosis. Conditional logistic regression will be employed to account for matching of controls to cases. Furthermore, differences in existing infrastructure of outpatient care between urban areas and rural areas may be associated with the prevalence of diagnosis-related exposures. This potential confounding is additionally controlled in sensitivity analyses by adjusting models for the rurality of the residence district.

Study size

Based on the data of the population-based Bavarian Cancer Registry (until 21st June 2024), the number of incident NEN cases in Bavaria, Germany, is 3,935 cases for the study period 2020 and 2021, of which 2,258 were NEN of the BPS and 1,677 were NEN of the GEPS. Taking into account that the KVB data covers 85% of all residents in Bavaria, Germany, 3,345 cases may be expected at maximum after the record linkage of Bavarian Cancer Registry and KVB data. As not all residents with incident NEN in 2020 and 2021 may receive outpatient treatment in Bavaria and as the record linkage may not identify all possible NEN cases in the KVB data, a record linkage for 80% of all cases is probably more realistic. Based on this assumption, about 3,150 cases and 6,300 matched controls would be expected. With more than 3,000 expected cases (about 1,800 BPS cases and 1,340 GEPS cases) and more than 6,000 expected controls, the linked dataset has a considerable size and is the best available approach to exploit outpatient data for a risk factor analysis for NEN. Even for GEPS tumour sites, such as the small intestine and the pancreas, we may expect 360 and 325 cases, respectively.

Analysis

The descriptive analysis, stratified by BPS and GEPS, includes the calculation of frequencies and percentages for categorical variables, the mean and median (with standard deviation (SD) and interquartile range (IQR), respectively) for birth year as well bivariate 2 x 2 tables for the combinations of case-control status and disease-related exposures. In addition, the bivariate analyses are stratified by the matching variables and by the subcategories of the outcome, i.e. NET and NEC. For the stratified analyses, odds ratios (ORs) will be calculated according to Mantel & Haenszel.[22]

Conditional logistic regression models will be estimated separately for BPS and GEPS to obtain estimates of ORs (and 95% confidence intervals (CIs)) for multiple exposures. All models will adjust for birth year (5-year intervals), sex, and time of diagnosis (calendar quarter). After stepwise inclusion of exposure variables, interaction terms will be added to the models to investigate effect modification. All models will also be stratified by the NEN subcategories NET and NEC and all analyses for GEPS will additionally be stratified by tumour site.

In a sensitivity analysis, rurality of the residence district and area deprivation of the residence municipality will be added to the regression models. Rurality of the residence district may be associated with the likelihood of receiving an assured diagnosis of disease-related exposures

as well as outcome measures. Area deprivation of the residence municipality may be linked to patterns of disease-related exposures and, thus, influence the outcome measures.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct or reporting or dissemination plans of this research.

DISCUSSION

This large, population-based case-control study fully exploits the potential of linking cancer registry with outpatient data to investigate risk factors for NEN, which is a rare disease as of now, and to help understand their increase in Germany. Thus, the study will add to previous studies from other countries, of which many were suffering from small study size, had a hospital-based study design, and analysed only selected tumour sites.[2] The study will also investigate exposures, such as depression and metabolic syndrome, which have not been extensively studied so far.[2 10 23] An additional advantage of the study design is that assessment of exposures does not rely on self-reported exposures but is based on assured outpatient diagnoses in the two years preceding the NEN diagnosis, and thus are not subject to recall bias nor subjective perception. With regard to the outcomes, the study will,unlike previous studies, additionally stratify for NET and NEC based on high-quality cancer registry data, allowing for the analysis of potential differences in risk factors between these two entities. The results of this study should provide risk ratios for potential risk factors of NEN and, thus, help to understand the recent NEN increase. The findings of the study may provide valuable insights for government policy on potential preventive measures, while also initiating further research. The study design may also serve as a flagship example of how the linkage of health data of different data sources can yield substantial epidemiological insights, especially in the case of rare diseases.

ETHICS AND DISSEMINATION

This study protocol was approved by the Ethics Committee of the Bavarian State Chamber of Physicians (reference number: 24008). Approval by supervisory authority has been obtained from the Bavarian State Ministry of Health, Care, and Prevention (reference number: G35h-A1080-2023/20-2) and also the Bavarian Data Protection Commissioner stated to have no concerns after presentation of the study protocol (reference number: DSB/7-692/1-275). This study is based on registry and claims data, which are collected on a legal basis without the explicit consent of the patients and which can be used for research purposes by the registry and, under certain conditions, by third parties. Patient consent for a specific study is only required for the use or linkage of plain data, but not for the study protocol presented, which is

based on an anonymized data set that does not contain any personal data. In accordance with point (b) of Article 14(5) of the European Union General Data Protection Regulation, it is not necessary to inform the patients in this case. The study will be conducted in accordance with the Helsinki declaration of the World Medical Association as well as the guidelines and recommendations for ensuring good epidemiological practice.[24]

The data that support the findings of this study will not be publicly accessible because the study partners, i.e. the Bavarian Cancer Registry and the Bavarian Association of Statutory Health Insurance Accredited Physicians, are subject to strict legal regulations regarding the disclosure of data. Upon reasonable request, however, the permissibility of the data provision will be reviewed by the Bavarian Cancer Registry and the Bavarian Association of Statutory Health Insurance Accredited Physicians in accordance with the applicable legal requirements.

The results of the case-control study will be presented at national and international conferences. After final analysis, the results will be published in the form of scientific articles in peer-reviewed journals. In addition, the authors will seek opportunities to share the findings with relevant stakeholders, such as clinicians in cancer centres, and the wider public by using, for instance, newsletters, press releases, and social media platforms.

TRIAL STATUS

The study started 15th October 2024 with the recruitment, i.e. data extraction and record linkage process of registry and claims data. The recruitment should be completed by the end of 2024.

DECLARATIONS

Authors' contributions

Conceptualization: SV, MT, JMN; Methodology: SV, NG, UB, AH, JMN; Writing – Original Draft: SV; Writing – Review & Editing: all authors.

All authors contributed to the study design, critically reviewed the manuscript for important intellectual content, and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests

The authors declare that they have no competing interests.

Patient consent for publication

Not applicable, since no individual person's data was used for this study protocol.

Data availability statement

Not applicable, since no individual person's data was used for this study protocol.

Acknowledgements

We are grateful to the Trust Centre of the Bavarian Cancer Registry (Dr Jana Johne, Mr Stefan Möllenkamp) for substantial support regarding ethical approval and regulatory approval. Moreover, we want to thank the Cancer Registry of North Rhine-Westphalia (Prof Stang) for commenting on an earlier version of the study protocol.

REFERENCES

1. Luley K, Gebauer J, Gebauer N, et al. [Gastroenteropancreatic neuroendocrine neoplasms-Heterogeneity, management and perspectives of treatment and research]. *Internist (Berl)* 2020;61(8):875-90. doi: 10.1007/s00108-020-00832-x
2. Leoncini E, Carioli G, La Vecchia C, et al. Risk factors for neuroendocrine neoplasms: a systematic review and meta-analysis. *Ann Oncol* 2016;27(1):68-81. doi: 10.1093/annonc/mdv505
3. Yao JC, Hassan M, Phan A, et al. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol* 2008;26(18):3063-72. doi: 10.1200/JCO.2007.15.4377
4. WHO Classification of Tumours Editorial Board. Digestive system tumours [Internet]: International Agency for Research on Cancer; 2019 [5th ed.; vol. 1:[Available from: https://tumourclassification.iarc.who.int/chapters/31 accessed 14.06.2024.
5. Montminy EM, Zhou M, Maniscalco L, et al. Contributions of Adenocarcinoma and Carcinoid Tumors to Early-Onset Colorectal Cancer Incidence Rates in the United States. *Ann Intern Med* 2021;174(2):157-66. doi: 10.7326/M20-0068 [published Online First: 2020/12/15]
6. Emrich K, Kraywinkel K. Epidemiology of rectal cancer in Germany. Is the incidence of rectal cancer increasing in young adults? *Der Onkologe* 2020;26:1085-94. doi: 10.1007/s00761-020-00857-9
7. Tsikitis VL, Wertheim BC, Guerrero MA. Trends of incidence and survival of gastrointestinal neuroendocrine tumors in the United States: a seer analysis. *J Cancer* 2012;3:292-302. doi: 10.7150/jca.4502
8. Voigtländer S, Hakimhashemi A, Grundmann N, et al. Trends of colorectal cancer incidence according to age, anatomic site, and histological subgroup in Bavaria: A registry-based study. *Front Oncol* 2022;12:904546. doi: 10.3389/fonc.2022.904546 [published Online First: 20220920]
9. Grundmann N, Voigtländer S, Hakimhashemi A, et al. Site-specific trends in gastroenteropancreatic neuroendocrine neoplasms in Bavaria, Germany. *Cancer Med* 2023;12(19):19949-58. doi: 10.1002/cam4.6510 [published Online First: 20230922]
10. Haugvik SP, Basim Ibrahim I, Hedenström P, et al. Smoking, alcohol and family history of cancer as risk factors for small intestinal neuroendocrine tumors: a systematic review and meta-analysis. *Scand J Gastroenterol* 2017;52(8):797-802. doi: 10.1080/00365521.2017.1310290 [published Online First: 20170410]
11. Kenner BJ. Early Detection of Pancreatic Cancer: The Role of Depression and Anxiety as a Precursor for Disease. *Pancreas* 2018;47(4):363-67. doi: 10.1097/MPA.0000000000001024
12. Katalinic A, Halber M, Meyer M, et al. Population-Based Clinical Cancer Registration in Germany. *Cancers (Basel)* 2023;15(15) doi: 10.3390/cancers15153934 [published Online First: 20230802]
13. Meyer M, Altmann U, Bartholomäus S, et al. Datenzusammenführung, -speicherung und Best-of-Prozess [Data linkage, data storage, and best-of process]. In: Stegmaier C, Hentschel S, Hofstädter F, et al., eds. Das Manual der Krebsregistrierung: GEKID, ADT 2019:81-98.
14. Arndt V, Holleczek B, Kajuter H, et al. Data from Population-based Cancer Registration for Secondary Data Analysis: Methodological Challenges and Perspectives. *Gesundheitswesen* 2020;82(S 01):S62-S71. doi: 10.1055/a-1009-6466
15. Federal Ministry of Health. Mitglieder und Versicherte der gesetzlichen Krankenversicherung (GKV) [Members and insured persons in the Statutory Helth Insurance] 2024 [Available from: https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte accessed 27.05.2024.
16. World Health Organization (WHO), International Agency for Research on Cancer (IARC). WHO Classification of Tumours: Thoracic tumours (5th edition) (Blue Books) 2024 [Available from: https://tumourclassification.iarc.who.int/chapters/35 accessed 27.05.2024.

17. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision 2020 [Available from: https://icd.who.int/browse10/2019/en accessed 14.09.2023.
18. Schubert I, Ihle P, Köster I. Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. [Internal confirmation of diagnoses in routine statutory health insurance data: concept with examples and case definitions]. *Gesundheitswesen* 2010;72(6):316-22. doi: 10.1055/s-0030-1249688 [published Online First: 2010/05/19]
19. Maier W. [Indices of Multiple Deprivation for the analysis of regional health disparities in Germany : Experiences from epidemiology and healthcare research]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2017;60(12):1403-12. doi: 10.1007/s00103-017-2646-2
20. German Federal Institute for Research on Building UAaSD. Städtischer und Ländlicher Raum [Urban and rural areas] 2024 [Available from: https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/kreise/staedtischer-laendlicher-raum/kreistypen.html accessed 27.05.2024.
21. Bavarian Health and Food Safety Authority. General practice: total patients (2021) [Hausärztliche Praxis: Patienten/innen insgesamt (2021)] 2024 [Available from: https://www.lgl.bayern.de/gesundheit/gesundheitsberichterstattung/gesundheitsatlas/ia_report/atlas.html accessed 27.05.2024.
22. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22(4):719-48.
23. Jung YS, Yun KE, Chang Y, et al. Risk factors associated with rectal neuroendocrine tumors: a cross-sectional study. *Cancer Epidemiol Biomarkers Prev* 2014;23(7):1406-13. doi: 10.1158/1055-9965.EPI-14-0132 [published Online First: 20140509]
24. Hoffmann W, Latza U, Baumeister SE, et al. Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology. *Eur J Epidemiol* 2019;34(3):301-17. doi: 10.1007/s10654-019-00500-x [published Online First: 20190304]

SUPPLEMENT

Supplement 1

For cases, eligible combinations of tumour site according to the ICD-10 [1] and histology code according to the WHO Blue Books [2 3] are:

- Gastroenteropancreatic System (GEPS)
  - Tumour sites: ICD-10 codes C15-C25
  - Histological codes based on the WHO Blue Books, categorised according to ICD-10 and excluding histological codes with behaviour codes 0 (benign) and 1 (borderline malignancy):
    - C15, Oesophagus: 8240/3 (NET), 8246/3, 8041/3, 8013/3 (all NEC), 8244/3 (MiNEN)
    - C16, Stomach: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3, 8153/3 (both NET)
      - Additional subtypes: 8242/3, 8156/3 (both NET), 8244/3 (MiNEN)
    - C17, Small intestine: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3, 8153/3, 8156/3 (all NET), 8693/3
    - C18.1, Appendix: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3 (NET)
    - C18-C20 excluding C18.1, Colorectum: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3 (NET)
    - C21, Anus and anal canal: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC)
      - Additional subtypes: 8154/3 (MiNEN)
    - C22, Liver and intrahepatic bile ducts: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN)
    - C23-C24, Gallbladder and other and unspecified parts of the biliary tract: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN)
    - C25, Pancreas:
      - Non-functioning pancreatic NET: 8150/3
      - Functioning pancreatic NET: 8151/3, 8153/3, 8155/3, 8152/3, 8156/3, 8158/3, 8241/3

- Other pancreatic NET[1]: 8240/3, 8249/3
- Pancreatic NEC: 8246/3, 8041/3, 8013/3
- Pancreatic MiNEN: 8154/3

- Bronchopulmonary System (BPS)
  - Tumour site: ICD-10 code C34
  - Histological codes based on the WHO Blue Books, categorised according to ICD-10:
    - C34, Bronchus and lung:
      - NET: 8240/3, 8249/3
      - NEC: 8041/3, 8045/3 (both small cell NEC), 8013/3 (large cell NEC)

[1] The histological codes 8240/3 and 8249/3 are not included in the WHO Blue Books, likely because they cannot be subdivided into non-functioning and functioning tumours. Therefore, both histological codes will be included as NET.

Supplement 2

The prevalence of the disease-related exposures, apart from previous malignant neoplasms, will be measured based on the following ICD-10 [1] diagnoses:

- Depression or persistent mood disorder
  - ICD-10 codes F32 (depressive episode), F33 (recurrent depressive disorder), or F34 (Persistent mood [affective] disorders), each including all four-character categories
- Obesity
  - ICD-10 codes E65 (localised adiposity) or E66 (obesity) including all four-character categories
- Metabolic syndrome (if any three out of the five following diagnoses is present)
  - ICD-10 code E66 (obesity), each including all four-character categories
  - ICD-10 code E78.1 (pure hyperglyceridaemia)
  - ICD-10 code E78.0 (pure hypercholesterolaemia)
  - ICD-10 code I10 (essential (primary) hypertension)
  - ICD-10 codes E10-E14 (diabetes mellitus) or R73 (elevated blood glucose level), each including all four-character categories
- Alcohol abuse (only for NEN of the GEPS)
  - ICD-10 code F10 (mental and behavioural disorders due to use of alcohol), including all four-character categories
- Diabetes (only for NEN of the GEPS)
  - ICD-10 codes E10-E14 (diabetes mellitus), each including all four-character categories
- Ulcerative colitis (only for NEN of the GEPS)
  - ICD-10 code K51 (ulcerative colitis) including all four-character categories
- Crohn disease (only for NEN of the GEPS)
  - ICD-10 code K50 (Chron disease) including all four-character categories
- Allergies (only for NEN of the BPS)
  - ICD-10 J30.1 (allergic rhinitis due to pollen) or J30.4 (allergic rhinitis, unspecified)
- Asthma (only for NEN of the BPS)
  - ICD-10 code J45 (asthma) including all four-character categories

The metabolic syndrome is defined as the presence of any three out of five constituting risk factors,[4] i.e. elevated waist circumference, elevated triglycerides (including drug treatment for elevated triglycerides), reduced high-density lipoprotein cholesterol (HDL-C, including

treatment for reduced HDL-C), elevated blood pressure (including antihypertensive drug treatment), and elevated fasting glucose (including treatment for elevated fasting glucose). As diagnosis data does not include information on the defined cut points, the measurement of the metabolic syndrome is approximated by the presence of any three out of the five following ICD-10 diagnoses: E66 approximates elevated waist circumference, E78.1 approximates elevated triglycerides, E78.0 approximates reduced HDL-C, I10 approximates elevated blood pressure, and E10-E14 or R73 approximates elevated fasting glucose. In case, the measurement of the metabolic syndrome will be limited due to missing four-character categories, such as E78.1, the measurement will be based on the three-character categories E66, E78, I10, and E10-E14.

References

1. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision 2020 [Available from: https://icd.who.int/browse10/2019/en accessed 14.09.2023.
2. WHO Classification of Tumours Editorial Board. Digestive system tumours [Internet]: International Agency for Research on Cancer; 2019 [5th ed.; vol. 1:[Available from: https://tumourclassification.iarc.who.int/chapters/31 accessed 14.06.2024.
3. World Health Organization (WHO), International Agency for Research on Cancer (IARC). WHO Classification of Tumours: Thoracic tumours (5th edition) (Blue Books) 2024 [Available from: https://tumourclassification.iarc.who.int/chapters/35 accessed 27.05.2024.
4. Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;120(16):1640-5. doi: 10.1161/CIRCULATIONAHA.109.192644 [published Online First: 20091005]

BMJ Open

Risk factors for neuroendocrine neoplasms: a protocol for a case-control study based on a record linkage of registry and claims data

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-096273.R1
Article Type:	Protocol
Date Submitted by the Author:	17-Apr-2025
Complete List of Authors:	Voigtländer, Sven; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Gerlach, Roman; Bavarian Association of Statutory Health Insurance Accredited Physicians Grundmann, Nina; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Donnachie, Ewan; Bavarian Association of Statutory Health Insurance Accredited Physicians Berger, Ursula; Ludwig Maximilian University of Munich Faculty of Medicine, Institute for Medical Information Processing, Biometry, and Epidemiology Hakimhashemi, Amir; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Meyer, Martin; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry Tauscher, Martin; Bavarian Association of Statutory Health Insurance Accredited Physicians Müller-Nordhorn, Jacqueline; Bavarian Health and Food Safety Authority, Bavarian Cancer Registry
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Oncology
Keywords:	Gastrointestinal tumours < ONCOLOGY, Respiratory tract tumours < ONCOLOGY, Risk Factors, Case-Control Studies, REGISTRIES

STUDY PROTOCOL

Risk factors for neuroendocrine neoplasms: a protocol for a case-control study based on a record linkage of registry and claims data

Sven Voigtländer[1], Roman Gerlach[2], Nina Grundmann[1], Ewan Donnachie[2], Ursula Berger[3], Amir Hakimhashemi[1], Martin Meyer[1], Martin Tauscher[2], Jacqueline Müller-Nordhorn[1]

[1] Bavarian Cancer Registry, Bavarian Health and Food Safety Authority, Schweinauer Hauptstraße 80, 90441 Nuremberg, Germany

[2] Bavarian Association of Statutory Health Insurance Accredited Physicians, Elsenheimerstraße 39, 80687 Munich, Germany

[3] Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine at Ludwig-Maximillians-University Munich, Marchioninistraße 15, 81377 Munich, Germany

Word count (without title page, abstract, statements, and references): 3,145

Correspondence:
Dr. Sven Voigtländer
Bavarian Cancer Registry, Bavarian Health and Food Safety Authority
Schweinauer Hauptstraße 80, 90441 Nuremberg, Germany
Phone: +49-(0)9131-6808-2927
Email: sven.voigtlaender@lgl.bayern.de
ORCID ID: https://orcid.org/0000-0003-3808-9605

ABSTRACT

Background:

Recent studies showed an increase in neuroendocrine neoplasms, especially for the digestive tract. Several risk factors have been suggested to explain this increase including a family history of cancer, tobacco smoking, alcohol consumption, and metabolic disorders including diabetes and obesity. Another risk factor may be depressive disorders, which could increase the risk of neuroendocrine neoplasms either directly or mediated through associated risk behaviours and/or antidepressant medication. Here, we outline the design of our study to identify risk factors for neuroendocrine neoplasms in Germany.

Methods and analysis:

A case-control study of the resident population of Bavaria, the second most populous federal state in Germany, based on a record linkage of data from the Bavarian Cancer Registry and data from the Bavarian Association of Statutory Health Insurance Accredited Physicians. Cases have a diagnosis of a malignant neuroendocrine neoplasm, either of the bronchopulmonary system or the gastroenteropancreatic system, in the period from 2021 to 2023. Controls are sampled from the noncases and matched on sex, birth year (in 5-year intervals), and time of diagnosis (by calendar quarter). Risk factor prevalence of cases and controls is assessed on the basis of assured outpatient diagnoses, i.e. diagnoses documented in at least two out of four consecutive quarters in the 16 quarters preceding the diagnosis of a neuroendocrine neoplasm. The analysis uses conditional logistic regression to estimates odds ratios and 95% confidence intervals.

Ethics and dissemination:

This study protocol was approved by the Ethics Committee of the Bavarian State Chamber of Physicians (reference number: 24008). Approval by supervisory authority has been obtained from the Bavarian State Ministry of Health, Care, and Prevention (reference number: G35h-A1080-2023/20-2) and also the Bavarian Data Protection Commissioner stated to have no concerns after presentation of the study protocol (reference number: DSB/7-692/1-275). The results of the case-control study will be presented at national as well as international conferences and be published in the form of scientific articles in peer-reviewed journals.

Registration details:

ClinicalTrials.gov, NCT: 06282016.

Keywords:

Neuroendocrine neoplasms, colorectal cancer, risk factors, case-control study, cancer registry, claims data

ARTICLE SUMMARY

Strengths and limitations of this study

- Cases and controls will be drawn from an unselected source population, i.e. persons, who were insured with the Statutory Health Insurance (85% of all residents) and had at least one outpatient physician contact between 2021 and 2023 within Bavaria
- Cases will comprise all malignant neuroendocrine neoplasms of the gastroenteropancreatic and bronchopulmonary system in the source population in the period from 2021 to 2023 based on high-quality cancer registry data
- The study will measure exposures based on cancer registry data and assured ICD-10 diagnoses and is, therefore, not prone to recall bias
- The study will, different to previous studies, also stratify for neuroendocrine tumours and neuroendocrine carcinomas based on high-quality cancer registry data.
- One limitation is that operationalisation of exposures is limited to exposures, which can be measured based on cancer registry as well claims data

BACKGROUND

Neuroendocrine neoplasms (NEN) are malignancies of neuroendocrine cells.[1] Neuroendocrine cells can be found throughout the body, i.e. where there is epithelium, excluding the central nervous system, bones, and soft tissue. NEN are rare and heterogeneous tumours, which most commonly arise in the gastroenteropancreatic system (GEPS) and the bronchopulmonary system (BPS).[1-3] NEN are categorised according to their differentiation into differentiated neuroendocrine tumours (NET) and poorly differentiated carcinomas (NEC).[4] NET can be further subdivided based on grade while NEC are further categorised into small cell and large cell carcinomas. In addition, there are also mixed neuroendocrine-non-neuroendocrine neoplasms (MiNEN).

Recent studies from the United States (US) and Germany showed an increase in NEN, especially for the digestive tract.[5-9] According to an analysis of data from the US Surveillance, Epidemiology and End Results (SEER) program, age-standardised incidence rates (ASR, US standard population 2000) significantly increased for most gastro-intestinal sites between 1975 and 2008.[7] For instance, the age-adjusted incidence rate for rectal NEN increased from 0.1 in 1975 to 1.1 per 100,000 persons in 2008.[7] The risk for NEN is known to increase with age.[2] Several other risk factors have been suggested by a limited number of studies including a family history of cancer, tobacco smoking, alcohol consumption, and metabolic disorders including diabetes and obesity.[2 10] Apart from this, there may be other risk factors, which have not been confirmed yet, such as depressive disorders, which could increase the risk of neuroendocrine neoplasms either directly or mediated through associated risk behaviours and antidepressant medication. Kenner,[11] for instance, discusses the role of depressive disorders for pancreatic cancer. In the past two decades, several studies have explored the relationship between depression and the metabolic syndrome,[12-14] which is a cluster of risk factors including raised blood pressure, dyslipidemia, raised fasting glucose, and central obesity.[15] Prospective cohort studies observed a bidirectional association for depression and the metabolic syndrome.[13] Both diseases, depression and the metabolic syndrome, are linked with insulin resistance and chronic inflammation involving the endocrine and immune systems.[13 14]

The aim of this study is to identify risk factors for neuroendocrine neoplasms based on data of the Bavarian Cancer Registry, Germany, to help understand the increase in neuroendocrine neoplasms.

METHODS

Study design

The study design is a case-control study based on data from the population-based Bavarian Cancer Registry and data from the Bavarian Association of Statutory Health Insurance Accredited Physicians (KVB, German: *Kassenärztliche Vereinigung Bayerns*). The Bavarian Cancer Registry is based on mandatory notifications by physicians and health care providers regarding the diagnosis and treatment of cancer.[16] The KVB collects, on a quarterly basis, diagnosis and treatment data related to its main task, i.e. ensuring and reimbursing outpatient medical care and psychotherapy for patients with Statutory Health Insurance (GKV, German: *Gesetzliche Krankenversicherung*) in Bavaria. As common for claims data, the treatment data of the KVB, however, is limited to fee schedule items including non-specific flat fees, why we will focus on the diagnosis data. Bavarian Cancer Registry data and KVB data are linked by pseudonymised record linkage following the probability linkage procedure established by the German cancer registries.[17 18] The probability linkage is based on the conversion of KVB identity data into unique tokens, i.e. pseudonymisation, that are compared with the tokens already present in the Bavarian Cancer Registry by probabilistic linkage.[18] After linkage, the pseudonyms are removed resulting in an anonymous dataset.

Setting

The setting is the resident population of Bavaria in the period from 2021 to 2023. Bavaria is the second most populous federal state in Germany with about 13 million residents. Though the Bavarian Cancer Registry covers the complete resident population of Bavaria, the source population of the study is limited to persons, who are insured with the GKV and had at least one outpatient physician contact between 2021 and 2023 within Bavaria. Persons who are not insured in the GKV as well as persons without outpatient physician contact between 2021 and 2023 within Bavaria are not included in the KVB data. In Bavaria, about 11.5 million residents, which is 85% of all residents, are insured with the GKV.[19]

Definition and recruitment of cases and controls

Eligible cases are defined as persons with a malignant NEN of the BPS or GEPS diagnosed in the period from 2021 to 2023. Definition of malignant NEN as well as of BPS and GEPS is based on the fifth edition of the World Health Organization's (WHO) Classification of Tumours, also known as the WHO Blue Books,[4 20] and the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10),[21] respectively. Detailed information on the eligible combinations of histology code and tumour site is provided in the Supplement (Supplement 1). The selection of cases is based on the record linkage of Bavarian Cancer Registry and KVB data, i.e. persons with a malignant NEN in the period from 2021 to 2023 according to the Bavarian Cancer Registry data that had a

GKV insurance as well as an outpatient physician contact between 2021 and 2023. Eligibility is not limited to age.

Eligible controls are defined as noncases in the same source population as for cases, i.e. the KVB data. Selection of controls is done by random sampling.

Cases will be matched 1:2 with controls on sex, birth year (5-year intervals), and time of diagnosis (by calendar quarter). Two controls will be matched to each case.

The recruitment of cases and controls is planned to start 15th May 2025 and end 31st August 2025, followed by the data analysis. The recruitment comprises the data extraction at the Bavarian Cancer Registry and the KVB, the record linkage of both datasets and the merging of outcomes, exposures, and confounders.

Definition of outcomes, exposures, and confounders

Outcomes

Primary outcomes are NEN of the BPS and the GEPS as defined by eligible combinations of histology code and tumour site (see Supplement 1). Secondary outcomes are the two subcategories of NEN, i.e. NET and NEC, and the tumour characteristics stage and grade.

Exposures and potential confounders

The following exposures and potential confounders will be assessed:

Exposures: disease-related exposures

1. Depression or persistent mood disorder
2. Obesity
3. Metabolic syndrome
4. Previous malignant neoplasms
5. Alcohol abuse (only for NEN of the GEPS)
6. Diabetes (only for NEN of the GEPS)
7. Ulcerative colitis (only for NEN of the GEPS)
8. Crohn disease (only for NEN of the GEPS)
9. Allergies (only for NEN of the BPS)
10. Asthma (only for NEN of the BPS)

Exposures: sociodemographic

1. Area deprivation of the residence municipality

Potential confounders: sociodemographic and healthcare-related

1. Birth year (5-year intervals)
2. Sex
3. Rurality of the residence district
4. Healthcare utilisation frequency

Potential confounders: time of NEN diagnosis

1. Time of diagnosis (calendar quarter)

Measurement of outcomes, exposures, and confounders

Primary outcomes as well as their subcategories NET and NEC are measured using the ICD-10 codes and histological information in the data of the Bavarian Cancer Registry (see Supplement 1). Stage, grade, previous malignant neoplasms, birth year (5-year intervals), sex, and time of diagnosis (by calendar quarter) are also taken from the Bavarian Cancer Registry data. Stage and grade are measured according to the TNM classification of the Union of International Cancer Control.[22] Previous malignant neoplasms include all neoplasms with ICD-10 codes C00-C97 (malignant neoplasms) except C44 (other malignant neoplasms of skin).[21] For cases and controls, the prevalence of the disease-related exposures, apart from previous malignant neoplasms, is assessed by specific ICD-10 diagnoses based on the KVB data (see Supplement 2 for further details). Only assured ICD-10 diagnoses will be considered,[23] i.e. diagnoses that were recorded in at least two out of four consecutive quarters in the 16 quarters preceding the quarter of the NEN diagnoses for cases as well as controls, which are matched by quarter of NEN diagnosis. Healthcare utilisation frequency is operationalised based on the use of outpatient medical care by calendar quarter and physician group, as defined by the KVB. Area deprivation of the residence municipality of cases and controls is measured by the Bavarian Index of Multiple Deprivation, which is based on official data.[24] In particular, we use the deprivation quintile of the residence municipality at the time of NEN diagnosis. Rurality of the residence district at the time of NEN diagnosis is based on the dichotomous categorisation of districts into urban area and rural area by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development.[25]

Bias

The case-control study is based on a record linkage of registry data, claims data, and administrative data. These data sources are, unlike survey data, not prone to recall bias. The potential of selection bias is considerably reduced by a small number of matching variables,

i.e. birth year, sex, and time of NEN diagnosis, as well as the use of an almost unselected source population for cases and controls, i.e. the KVB data. The KVB data covers all persons with GKV (about 85% of all residents in Bavaria) and at least one outpatient physician contact in Bavaria within the study period from 2021 to 2023. According to the KVB, 94% of all GKV insured persons had at least one contact with a general practitioner in 2021,[26] so that about 80% of all residents are included in our source population. It is known that the proportion of persons with GKV, who use outpatient care, is higher for women compared to men and lower for younger persons compared to older persons.[23] Regarding potential differential misclassification bias, assessment of disease-related exposures is done in the 16 quarters preceding the quarter of the NEN diagnosis to ensure that the prevalence and number of disease-related exposures is not influenced by potentially increased clinical investigation related to the NEN diagnosis. The potential of reverse causality, i.e. subclinical NEN causing disease-related exposures such as depression and not vice versa, is addressed by sensitivity analyses excluding the eight quarters preceding the quarter of the NEN diagnosis from the assessment of disease-related exposures.

Confounding is controlled by matching for birth year, sex, and time of NEN diagnosis. We will assess the matching process by comparing the distributions of the matching variables between cases and controls graphically and by summary statistics of the distributions. Conditional logistic regression will be employed to account for matching of controls to cases. Furthermore, differences in existing infrastructure of outpatient care between urban areas and rural areas may be associated with the prevalence of diagnosis-related exposures. This potential confounding is additionally controlled in sensitivity analyses by adjusting models for the rurality of the residence district.

Study size

Based on the data of the population-based Bavarian Cancer Registry (until 27th March 2025), the number of incident NEN cases in Bavaria, Germany, is 5,943 cases for the study period from 2021 to 2023, of which 3,274 were NEN of the BPS and 2,669 were NEN of the GEPS. Taking into account that the KVB data covers 85% of all residents in Bavaria, Germany, 5,051 cases may be expected at maximum after the record linkage of Bavarian Cancer Registry and KVB data. As not all residents with incident NEN between 2021 and 2023 may receive outpatient treatment in Bavaria and as the record linkage may not identify all possible NEN cases in the KVB data, a record linkage for 80% of all cases is probably more realistic. Based on this assumption, about 4,750 cases and 9,500 matched controls would be expected. With more than 4,500 expected cases (about 2,620 BPS cases and 2,130 GEPS cases) and more than 9,000 expected controls, the linked dataset has a considerable size

and is the best available approach to exploit outpatient data for a risk factor analysis for NEN. Even for GEPS tumour sites, such as the small intestine and the pancreas, we may expect 590 and 465 cases, respectively.

Analysis

The descriptive analysis, stratified by BPS and GEPS, includes the calculation of frequencies and percentages for categorical variables, the mean and median (with standard deviation (SD) and interquartile range (IQR), respectively) for birth year as well bivariate 2 x 2 tables for the combinations of case-control status and disease-related exposures. This allows to investigate shared exposures of cases and controls. In addition, the bivariate analyses are stratified by the matching variables and by the subcategories of the outcome, i.e. NET and NEC. For the stratified analyses, odds ratios (ORs) will be calculated according to Mantel & Haenszel.[27]

Conditional logistic regression models will be estimated separately for BPS and GEPS to obtain estimates of ORs (and 95% confidence intervals (CIs)) for multiple exposures. All models will adjust for birth year (5-year intervals), sex, and time of diagnosis (calendar quarter). After stepwise inclusion of exposure variables, interaction terms will be added to the models to investigate effect modification, for instance, between depression and the metabolic syndrome. Models will be compared based on Akaike information criterion (AIC) and validated examining their residual plots. The assumption of linearity in the predictors is assessed using additive models with P-splines.[28] All models will also be stratified by the NEN subcategories NET and NEC and all analyses for GEPS will additionally be stratified by tumour site, stage, and grade.

Several sensitivity analyses will be performed. To address the potential of reverse causality, the first sensitivity analysis will measure disease-related exposures based on the fourth and third year preceding the quarter of the NEN diagnosis so that the eight quarters before the NEN diagnosis are excluded. The second sensitivity analysis refers to the study period, which partially coincides with the COVID-19 pandemic that disrupted healthcare utilisation and diagnosis patterns[29 30] and may have led to an underdiagnosis of, both, outcomes as well as disease-related exposures. To control for this potential confounding, we will add healthcare utilisation frequency to the regression models. In a third sensitivity analysis, rurality of the residence district and area deprivation of the residence municipality will be added to the regression models. Rurality of the residence district may be associated with the likelihood of receiving an assured diagnosis of disease-related exposures as well as outcome

measures. Area deprivation of the residence municipality may be linked to patterns of disease-related exposures and, thus, influence the outcome measures.

Missing values may occur in the variables stage and grade of cases. If the number of missing values exceeds an acceptable threshold, multiple imputation (using multiple imputation by chained equation (MICE)) will be applied.[31] All variables, including the matching variables, will be incorporated into the imputation model.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct or reporting or dissemination plans of this research.

DISCUSSION

This large, population-based case-control study fully exploits the potential of linking cancer registry with outpatient data to investigate risk factors for NEN, which is a rare disease as of now, and to help understand their increase in Germany. Thus, the study will add to previous studies from other countries, of which many were suffering from small study size, had a hospital-based study design, analysed only selected tumour sites, and did not include information on stage and grade.[2] The study will also investigate exposures, such as depression and metabolic syndrome, which have not been extensively studied so far.[2 10 32] An additional advantage of the study design is that assessment of exposures does not rely on self-reported exposures but is based on assured outpatient diagnoses in the four years preceding the NEN diagnosis, and thus are not subject to recall bias nor subjective perception. With regard to the outcomes, the study will, unlike previous studies, additionally stratify for NET and NEC based on high-quality cancer registry data, allowing for the analysis of potential differences in risk factors between these two entities.

Limitations refer to the observation of specific causal pathways as well as the measurement of disease-related exposures, as the outpatient data does not include information regarding the disease onset and as ICD-10 codes lack diagnostic thresholds. To our knowledge, however, this is the best currently available data in Germany to study such a large and unselected population. The drawback is the lack of detailed information to observe specific causal pathways. Regarding the measurement of disease-related exposures, we will use assured ICD-10 diagnoses, i.e. diagnoses recorded in at least two out of four consecutive quarters, to limit misclassification.

The results of this study should provide risk ratios for potential risk factors of NEN and, thus, help to understand the recent NEN increase. The findings of the study may provide valuable insights for government policy on potential preventive measures, while also initiating further research. The study design may also serve as a flagship example of how the linkage of health data of different data sources can yield substantial epidemiological insights, especially in the case of rare diseases.

ETHICS AND DISSEMINATION

This study protocol was approved by the Ethics Committee of the Bavarian State Chamber of Physicians (reference number: 24008). Approval by supervisory authority has been obtained from the Bavarian State Ministry of Health, Care, and Prevention (reference number: G35h-A1080-2023/20-2) and also the Bavarian Data Protection Commissioner stated to have no concerns after presentation of the study protocol (reference number: DSB/7-692/1-275). This study is based on registry and claims data, which are collected on a legal basis without the explicit consent of the patients and which can be used for research purposes by the registry and, under certain conditions, by third parties. Patient consent for a specific study is only required for the use or linkage of plain data, but not for the study protocol presented, which is based on an anonymized data set that does not contain any personal data. In accordance with point (b) of Article 14(5) of the European Union General Data Protection Regulation, it is not necessary to inform the patients in this case. The study will be conducted in accordance with the Helsinki declaration of the World Medical Association as well as the guidelines and recommendations for ensuring good epidemiological practice.[33]

The data that support the findings of this study will not be publicly accessible because the study partners, i.e. the Bavarian Cancer Registry and the Bavarian Association of Statutory Health Insurance Accredited Physicians, are subject to strict legal regulations regarding the disclosure of data. Upon reasonable request, however, the permissibility of the data provision will be reviewed by the Bavarian Cancer Registry and the Bavarian Association of Statutory Health Insurance Accredited Physicians in accordance with the applicable legal requirements.

The results of the case-control study will be presented at national and international conferences. After final analysis, the results will be published in the form of scientific articles in peer-reviewed journals. In addition, the authors will seek opportunities to share the findings with relevant stakeholders, such as clinicians in cancer centres, and the wider public by using, for instance, newsletters, press releases, and social media platforms.

TRIAL STATUS

The study starts 15th May 2025 with the recruitment, i.e. data extraction and record linkage process of registry and claims data. The recruitment should be completed by 31st August 2025.

DECLARATIONS

Authors' contributions

Conceptualization: SV, MT, JMN; Methodology: SV, NG, UB, AH, JMN; Writing – Original Draft: SV; Writing – Review & Editing: all authors.

All authors contributed to the study design, critically reviewed the manuscript for important intellectual content, and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.SV is the guarantor.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests

The authors declare that they have no competing interests.

Patient consent for publication

Not applicable, since no individual person's data was used for this study protocol.

Data availability statement

Not applicable, since no individual person's data was used for this study protocol.

Acknowledgements

We are grateful to the Trust Centre of the Bavarian Cancer Registry (Dr Jana Johne, Mr Stefan Möllenkamp) for substantial support regarding ethical approval and regulatory approval. Moreover, we want to thank the Cancer Registry of North Rhine-Westphalia (Prof Stang) for commenting on an earlier version of the study protocol.

REFERENCES

1. Luley K, Gebauer J, Gebauer N, et al. [Gastroenteropancreatic neuroendocrine neoplasms-Heterogeneity, management and perspectives of treatment and research]. *Internist (Berl)* 2020;61(8):875-90. doi: 10.1007/s00108-020-00832-x
2. Leoncini E, Carioli G, La Vecchia C, et al. Risk factors for neuroendocrine neoplasms: a systematic review and meta-analysis. *Ann Oncol* 2016;27(1):68-81. doi: 10.1093/annonc/mdv505
3. Yao JC, Hassan M, Phan A, et al. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J Clin Oncol* 2008;26(18):3063-72. doi: 10.1200/JCO.2007.15.4377
4. WHO Classification of Tumours Editorial Board. Digestive system tumours [Internet]: International Agency for Research on Cancer; 2019 [5th ed.; vol. 1:[Available from: https://tumourclassification.iarc.who.int/chapters/31 accessed 14.06.2024.
5. Montminy EM, Zhou M, Maniscalco L, et al. Contributions of Adenocarcinoma and Carcinoid Tumors to Early-Onset Colorectal Cancer Incidence Rates in the United States. *Ann Intern Med* 2021;174(2):157-66. doi: 10.7326/M20-0068 [published Online First: 2020/12/15]
6. Emrich K, Kraywinkel K. Epidemiology of rectal cancer in Germany. Is the incidence of rectal cancer increasing in young adults? *Der Onkologe* 2020;26:1085-94. doi: 10.1007/s00761-020-00857-9
7. Tsikitis VL, Wertheim BC, Guerrero MA. Trends of incidence and survival of gastrointestinal neuroendocrine tumors in the United States: a seer analysis. *J Cancer* 2012;3:292-302. doi: 10.7150/jca.4502
8. Voigtländer S, Hakimhashemi A, Grundmann N, et al. Trends of colorectal cancer incidence according to age, anatomic site, and histological subgroup in Bavaria: A registry-based study. *Front Oncol* 2022;12:904546. doi: 10.3389/fonc.2022.904546 [published Online First: 20220920]
9. Grundmann N, Voigtländer S, Hakimhashemi A, et al. Site-specific trends in gastroenteropancreatic neuroendocrine neoplasms in Bavaria, Germany. *Cancer Med* 2023;12(19):19949-58. doi: 10.1002/cam4.6510 [published Online First: 20230922]
10. Haugvik SP, Basim Ibrahim I, Hedenström P, et al. Smoking, alcohol and family history of cancer as risk factors for small intestinal neuroendocrine tumors: a systematic review and meta-analysis. *Scand J Gastroenterol* 2017;52(8):797-802. doi: 10.1080/00365521.2017.1310290 [published Online First: 20170410]
11. Kenner BJ. Early Detection of Pancreatic Cancer: The Role of Depression and Anxiety as a Precursor for Disease. *Pancreas* 2018;47(4):363-67. doi: 10.1097/MPA.0000000000001024
12. Löwe B, Hochlehnert A, Nikendei C. [Metabolic syndrome and depression]. *Ther Umsch* 2006;63(8):521-7. doi: 10.1024/0040-5930.63.8.521
13. Pan A, Keum N, Okereke OI, et al. Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies. *Diabetes Care* 2012;35(5):1171-80. doi: 10.2337/dc11-2055
14. Qiu W, Cai X, Zheng C, et al. Update on the Relationship Between Depression and Neuroendocrine Metabolism. *Front Neurosci* 2021;15:728810. doi: 10.3389/fnins.2021.728810 [published Online First: 20210831]
15. Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;120(16):1640-5. doi: 10.1161/CIRCULATIONAHA.109.192644 [published Online First: 20091005]
16. Katalinic A, Halber M, Meyer M, et al. Population-Based Clinical Cancer Registration in Germany. *Cancers (Basel)* 2023;15(15) doi: 10.3390/cancers15153934 [published Online First: 20230802]

17. Meyer M, Altmann U, Bartholomäus S, et al. Datenzusammenführung, -speicherung und Best-of-Prozess [Data linkage, data storage, and best-of process]. In: Stegmaier C, Hentschel S, Hofstädter F, et al., eds. Das Manual der Krebsregistrierung: GEKID, ADT 2019:81-98.
18. Arndt V, Holleczek B, Kajuter H, et al. Data from Population-based Cancer Registration for Secondary Data Analysis: Methodological Challenges and Perspectives. *Gesundheitswesen* 2020;82(S 01):S62-S71. doi: 10.1055/a-1009-6466
19. Federal Ministry of Health. Mitglieder und Versicherte der gesetzlichen Krankenversicherung (GKV) [Members and insured persons in the Statutory Helth Insurance] 2024 [Available from: https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/zahlen-und-fakten-zur-krankenversicherung/mitglieder-und-versicherte accessed 27.05.2024.
20. World Health Organization (WHO), International Agency for Research on Cancer (IARC). WHO Classification of Tumours: Thoracic tumours (5th edition) (Blue Books) 2024 [Available from: https://tumourclassification.iarc.who.int/chapters/35 accessed 27.05.2024.
21. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision 2020 [Available from: https://icd.who.int/browse10/2019/en accessed 14.09.2023.
22. Brierley JD, Gospodarowicz MK, Wittekind C, editors. *TNM Classification of Malignant Tumours, 8th edition*: Union for International Cancer Control, 2017.
23. Schubert I, Ihle P, Köster I. Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. [Internal confirmation of diagnoses in routine statutory health insurance data: concept with examples and case definitions]. *Gesundheitswesen* 2010;72(6):316-22. doi: 10.1055/s-0030-1249688 [published Online First: 2010/05/19]
24. Maier W. [Indices of Multiple Deprivation for the analysis of regional health disparities in Germany : Experiences from epidemiology and healthcare research]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2017;60(12):1403-12. doi: 10.1007/s00103-017-2646-2
25. German Federal Institute for Research on Building UAaSD. Städtischer und Ländlicher Raum [Urban and rural areas] 2024 [Available from: https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/kreise/staedtischer-laendlicher-raum/kreistypen.html accessed 27.05.2024.
26. Bavarian Health and Food Safety Authority. General practice: total patients (2021) [Hausärztliche Praxis: Patienten/innen insgesamt (2021)] 2024 [Available from: https://www.lgl.bayern.de/gesundheit/gesundheitsberichterstattung/gesundheitsatlas/ia_report/atlas.html accessed 27.05.2024.
27. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22(4):719-48.
28. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996;11(2):89-121, 33.
29. Alom S, Chiu CM, Jha A, et al. The Effects of COVID-19 on Cancer Care Provision: A Systematic Review. *Cancer Control* 2021;28:1073274821997425. doi: 10.1177/1073274821997425
30. Voigtländer S, Hakimhashemi A, Grundmann N, et al. Impact of the COVID-19 pandemic on reported cancer diagnoses in Bavaria, Germany. *J Cancer Res Clin Oncol* 2023 doi: 10.1007/s00432-023-04707-0 [published Online First: 20230324]
31. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30(4):377-99. doi: 10.1002/sim.4067 [published Online First: 20101130]
32. Jung YS, Yun KE, Chang Y, et al. Risk factors associated with rectal neuroendocrine tumors: a cross-sectional study. *Cancer Epidemiol Biomarkers Prev* 2014;23(7):1406-13. doi: 10.1158/1055-9965.EPI-14-0132 [published Online First: 20140509]
33. Hoffmann W, Latza U, Baumeister SE, et al. Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology. *Eur J Epidemiol* 2019;34(3):301-17. doi: 10.1007/s10654-019-00500-x [published Online First: 20190304]

521

522

523

SUPPLEMENT

Supplement 1

For cases, eligible combinations of tumour site according to the ICD-10 [1] and histology code according to the WHO Blue Books [2 3] are:

- Gastroenteropancreatic System (GEPS)
  - Tumour sites: ICD-10 codes C15-C25
  - Histological codes based on the WHO Blue Books, categorised according to ICD-10 and excluding histological codes with behaviour codes 0 (benign) and 1 (borderline malignancy):
    - C15, Oesophagus: 8240/3 (NET), 8246/3, 8041/3, 8013/3 (all NEC), 8244/3 (MiNEN)
    - C16, Stomach: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3, 8153/3 (both NET)
      - Additional subtypes: 8242/3, 8156/3 (both NET), 8244/3 (MiNEN)
    - C17, Small intestine: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3, 8153/3, 8156/3 (all NET), 8693/3
    - C18.1, Appendix: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3 (NET)
    - C18-C20 excluding C18.1, Colorectum: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN), 8241/3 (NET)
    - C21, Anus and anal canal: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC)
      - Additional subtypes: 8154/3 (MiNEN)
    - C22, Liver and intrahepatic bile ducts: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN)
    - C23-C24, Gallbladder and other and unspecified parts of the biliary tract: 8240/3, 8249/3 (both NET), 8246/3, 8041/3, 8013/3 (all NEC), 8154/3 (MiNEN)
    - C25, Pancreas:
      - Non-functioning pancreatic NET: 8150/3
      - Functioning pancreatic NET: 8151/3, 8153/3, 8155/3, 8152/3, 8156/3, 8158/3, 8241/3

- Other pancreatic NET[1]: 8240/3, 8249/3
- Pancreatic NEC: 8246/3, 8041/3, 8013/3
- Pancreatic MiNEN: 8154/3

- Bronchopulmonary System (BPS)
  - Tumour site: ICD-10 code C34
  - Histological codes based on the WHO Blue Books, categorised according to ICD-10:
    - C34, Bronchus and lung:
      - NET: 8240/3, 8249/3
      - NEC: 8041/3, 8045/3 (both small cell NEC), 8013/3 (large cell NEC)

[1] The histological codes 8240/3 and 8249/3 are not included in the WHO Blue Books, likely because they cannot be subdivided into non-functioning and functioning tumours. Therefore, both histological codes will be included as NET.

Supplement 2

The prevalence of the disease-related exposures, apart from previous malignant neoplasms, will be measured based on the following ICD-10 [1] diagnoses:

- Depression or persistent mood disorder
  - ICD-10 codes F32 (depressive episode), F33 (recurrent depressive disorder), or F34 (Persistent mood [affective] disorders), each including all four-character categories
- Obesity
  - ICD-10 codes E65 (localised adiposity) or E66 (obesity) including all four-character categories
- Metabolic syndrome (if any three out of the five following diagnoses is present)
  - ICD-10 code E66 (obesity), each including all four-character categories
  - ICD-10 code E78.1 (pure hyperglyceridaemia)
  - ICD-10 code E78.0 (pure hypercholesterolaemia)
  - ICD-10 code I10 (essential (primary) hypertension)
  - ICD-10 codes E10-E14 (diabetes mellitus) or R73 (elevated blood glucose level), each including all four-character categories
- Alcohol abuse (only for NEN of the GEPS)
  - ICD-10 code F10 (mental and behavioural disorders due to use of alcohol), including all four-character categories
- Diabetes (only for NEN of the GEPS)
  - ICD-10 codes E10-E14 (diabetes mellitus), each including all four-character categories
- Ulcerative colitis (only for NEN of the GEPS)
  - ICD-10 code K51 (ulcerative colitis) including all four-character categories
- Crohn disease (only for NEN of the GEPS)
  - ICD-10 code K50 (Chron disease) including all four-character categories
- Allergies (only for NEN of the BPS)
  - ICD-10 J30.1 (allergic rhinitis due to pollen) or J30.4 (allergic rhinitis, unspecified)
- Asthma (only for NEN of the BPS)
  - ICD-10 code J45 (asthma) including all four-character categories

The metabolic syndrome is defined as the presence of any three out of five constituting risk factors,[4] i.e. elevated waist circumference, elevated triglycerides (including drug treatment for elevated triglycerides), reduced high-density lipoprotein cholesterol (HDL-C, including

treatment for reduced HDL-C), elevated blood pressure (including antihypertensive drug treatment), and elevated fasting glucose (including treatment for elevated fasting glucose). As diagnosis data does not include information on the defined cut points, the measurement of the metabolic syndrome is approximated by the presence of any three out of the five following ICD-10 diagnoses: E66 approximates elevated waist circumference, E78.1 approximates elevated triglycerides, E78.0 approximates reduced HDL-C, I10 approximates elevated blood pressure, and E10-E14 or R73 approximates elevated fasting glucose. In case, the measurement of the metabolic syndrome will be limited due to missing four-character categories, such as E78.1, the measurement will be based on the three-character categories E66, E78, I10, and E10-E14.

References

1. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 10th Revision 2020 [Available from: https://icd.who.int/browse10/2019/en accessed 14.09.2023.
2. WHO Classification of Tumours Editorial Board. Digestive system tumours [Internet]: International Agency for Research on Cancer; 2019 [5th ed.; vol. 1:[Available from: https://tumourclassification.iarc.who.int/chapters/31 accessed 14.06.2024.
3. World Health Organization (WHO), International Agency for Research on Cancer (IARC). WHO Classification of Tumours: Thoracic tumours (5th edition) (Blue Books) 2024 [Available from: https://tumourclassification.iarc.who.int/chapters/35 accessed 27.05.2024.
4. Alberti KG, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;120(16):1640-5. doi: 10.1161/CIRCULATIONAHA.109.192644 [published Online First: 20091005]