

## Untersuchung regionaler Einflussfaktoren: Mehrebenenanalysen mit Missing Data

Dr. Philipp Jaehn, Institut für Sozialmedizin und Epidemiologie, Medizinische Hochschule Brandenburg Theodor Fontane, Brandenburg an der Havel

### Abstract

Forschung zu regionalen Einflussfaktoren kann einen wichtigen Beitrag für Primärprävention und Versorgung von Krebserkrankungen in Deutschland leisten. Eine systematische Übersicht epidemiologischer Studien aus den USA, die Mehrebenenanalysen zur Untersuchung regionaler Einflussfaktoren unterschiedlicher Outcomes von Krebserkrankungen angewendet haben, zeigt aktuelle methodische und inhaltliche Forschungslücken auf (Zahnd & McLafferty, 2017). So war beispielsweise der Großteil der Studien, die den Einfluss des Zugangs zur Gesundheitsversorgung untersucht haben, limitiert, da lediglich ein Container-Approach Anwendung fand (bspw. Anzahl der Versorgungseinrichtungen pro 100.000 Einwohner\*innen). In Bezug auf soziale Faktoren wurde in den meisten Studien die regionale sozioökonomische Deprivation untersucht. Weitere soziale Faktoren auf regionaler Ebene wie Sozialkapital oder Geschlechterungleichheit wurden deutlich seltener erforscht. Moderne Studien fokussieren des Weiteren explizit auf gesundheitliche Folgen politischer Entscheidungen auf regionaler Ebene. Beispielsweise zeigten Krieger et al., dass ein niedriges Rating einer Region im Rahmen des sogenannten historischen Redlining in den USA verglichen mit einem hohen Rating auch nach Kontrolle für die aktuelle regionale sozioökonomische Lage mit späteren Stadien bei Krebsdiagnose assoziiert ist (Krieger et al., 2020). Durch diesen Ansatz können Ergebnisse erzielt werden, aus denen sich relevante Handlungsempfehlungen ableiten lassen.

Für epidemiologische Studien in Deutschland stehen einige Datenquellen zu regionalen Einflussfaktoren zur Verfügung (Tabelle 1). Die regionale Deprivation kann beispielsweise anhand des German Index of Socioeconomic Deprivation (Kroll et al., 2017) oder durch mehrdimensionale Indizes wie den German Index of Multiple Deprivation abgebildet werden. Letztgenannter schließt neben sozioökonomischen Faktoren auch die Sozialkapital-, Umwelt- und Sicherheitsdeprivation ein (Maier, 2017). Hier ist wichtig zu betonen, dass eine differenzierte Darstellung von Assoziationen für klar abgegrenzte Einzeldimensionen bei der Unterscheidung helfen kann, welche spezifischen Aspekte regionaler Deprivation bedeutsam für ein Gesundheitsoutcome sind. So ist es beispielsweise relevant zu differenzieren, ob unabhängige Assoziationen von Sozialkapitaldeprivation und/oder sozioökonomischer Deprivation vorliegen. Weiterhin besteht ein hohes Potenzial für die Forschung, wenn regionale Faktoren auf eine Art und Weise operationalisiert werden, die über einen Container-Approach hinausgeht (Zahnd & McLafferty, 2017). Letztlich sollte darauf geachtet werden, dass ausreichend Varianz der Exposition vorliegt und dass das Modifiable Areal Unit Problem durch die Auswahl möglichst kleinräumiger regionaler Einheiten minimiert wird (Openshaw, 1984).

<b>Tabelle 1: Sammlung ausgewählter Ressourcen zu regionalen Einflussfaktoren (Link hinterlegt)</b>
<a href="#">Übersicht von Manuela Peters und Hajo Zeeb: Verfügbarkeit offener Daten für die raumbezogene Public-Health-Forschung. GMS German Medical Science 2022, Vol. 20</a>
<a href="#">Github Repository des German Index of Socioeconomic Deprivation</a>
<a href="#">INKAR Datenbank: Umfassende Sammlung sozialer und wirtschaftlicher Faktoren, sowie Faktoren der medizinischen Versorgungsinfrastruktur</a>
<a href="#">Thünen Landatlas: Datenbank des Thünen Instituts zu Faktoren der Siedlungsstruktur, Versorgung und Erreichbarkeit sozialer und medizinischer Einrichtungen nach dem Thünen-Erreichbarkeitsmodell</a>
<a href="#">Regionaler Gender Inequality Index zur Abbildung struktureller Geschlechterungleichheit</a>

Bei der Forschung zu regionalen Einflussfaktoren sind Krebsregisterdaten vor allem für deskriptive Studien geeignet. Für Forschung, die die Bestimmung von (kausalen) Effekten einer regionalen Größe („kontextueller Effekt“) zum Ziel hat, sind zusätzlich detaillierte Informationen auf individueller Ebene notwendig. Der Grund dafür ist, dass regionale Unterschiede auch durch die Zusammensetzung der Individuen in dieser Region („kompositioneller Effekt“) zustande kommen können. Trotz dieser Limitation sind deskriptive krebsepidemiologische Studien zu regionalen Faktoren von hoher Bedeutung, da sie Informationen darüber liefern können, auf welche Regionen bei der Maßnahmenplanung fokussiert werden sollte.

In der Datenanalyse sind Mehrebenenmodelle eine geeignete Methode, um bei Clustering Effektmaße und Standardfehler korrekt zu schätzen. Außerdem können Mehrebenenmodelle verwendet werden, um Störgrößen auf regionaler Ebene zu berücksichtigen. Bei der Verwendung mehrerer regionaler Faktoren als unabhängige Variablen kann dabei das Problem der Multikollinearität auftreten. Je nach Forschungsfrage ist die Berücksichtigung mehrerer regionaler Faktoren jedoch möglich. So ist beispielsweise die Korrelation von Wahlbeteiligung, German Index of Socioeconomic Deprivation und siedlungsstruktureller Kreistypen (nach Definition des Bundesinstituts für Bau-, Stadt- und Raumforschung) auf Landkreisebene in Deutschland gering und lässt eine Aufnahme aller drei Faktoren in ein Modell zu. Letztlich wies Oakes auf ein weiteres Problem beim Umgang mit Störgrößen auf der regionalen Ebene hin. Oft liegen für bestimmte Kombinationen der Ausprägungen mehrerer regionaler Faktoren keine Beobachtungen vor. Damit entsteht die Problematik, dass Schlussfolgerungen auf der Basis nicht beobachteter Daten gezogen werden könnten (Oakes, 2006).

Eine interessante Ergänzung für die deskriptive Forschung zu regionalen Einflussfaktoren beschrieb Merlo. Nach Merlo sollten sogenannte generelle kontextuelle Effekte berechnet werden, um die Bedeutung des allgemeinen Einflusses der gewählten regionalen Einheit zu bestimmen. Der generelle kontextuelle Effekt kann anhand des Intraklassenkorrelationskoeffizienten (IKK) operationalisiert werden. Ein niedriger IKK weist dabei auf eine geringe Bedeutung des allgemeinen Effekts der regionalen Ebene hin. Dies kann bedeuten, dass bei statistisch signifikanten Assoziationen einzelner regionaler Faktoren mit dem Outcome („spezifische kontextuelle Effekte“) der regionale Effekt bei einem geringen IKK trotzdem als unbedeutend bewertet werden würde (Merlo, 2014). Hohe IKKs zeigen hingegen an, dass die regionale Ebene eine hohe Trennschärfe für die Prädiktion des Outcomes hat und damit geeignet ist, um Regionen zu identifizieren, in denen prioritär interveniert werden sollte.

Abschließend kann Forschung zu regionalen Faktoren mit Krebsregisterdaten durch fehlende Werte limitiert werden. Hier führten einige kürzlich erschienene Studien aus dem Vereinigten Königreich eine multiple Imputation (MI) durch, um die Unsicherheit durch Missings im Stadium zu berücksichtigen (Barclay et al., 2021; Muller et al., 2020). Die verfügbaren Informationen, um einen Mechanismus der fehlenden Werte zu erklären, sind in Krebsregisterdaten aufgrund weniger unterstützender Variablen jedoch eher gering. Damit ist die Missing At Random (MAR) Annahme bei der Forschung mit Krebsregisterdaten oft nicht plausibel. Eine Studie aus Australien unterstützt diese Überlegung. Hier waren Variablen, die standardmäßig in Krebsregisterdaten vorliegen, unzureichend, um die MAR Annahme zu bestätigen (Luo et al., 2017). Nichtsdestotrotz kommen die Autor\*innen zu dem Schluss, dass MI eine valide Methode für die Krebsregisterforschung darstellen kann, da die mit den fehlenden Werten verbundene Unsicherheit widerspiegelt wird (Luo et al., 2017). Für MI liegt für R letztlich das Paket jomo vor, das die Mehrebenenstruktur eines Datensatzes bei der Imputation berücksichtigen kann (Quartagno et al., 2019).

## Referenzen:

- Barclay, M. E., Abel, G. A., Greenberg, D. C., Rous, B., & Lyratzopoulos, G. (2021). Socio-demographic variation in stage at diagnosis of breast, bladder, colon, endometrial, lung, melanoma, prostate, rectal, renal and ovarian cancer in England and its population impact. *Br J Cancer*, *124*(7), 1320-1329. <https://doi.org/10.1038/s41416-021-01279-z>
- Krieger, N., Wright, E., Chen, J. T., Waterman, P. D., Huntley, E. R., & Arcaya, M. (2020). Cancer Stage at Diagnosis, Historical Redlining, and Current Neighborhood Characteristics: Breast, Cervical, Lung, and Colorectal Cancers, Massachusetts, 2001-2015. *Am J Epidemiol*, *189*(10), 1065-1075. <https://doi.org/10.1093/aje/kwaa045>
- Kroll, L. E., Schumann, M., Hoebel, J., & Lampert, T. (2017). Regionale Unterschiede in der Gesundheit – Entwicklung eines sozioökonomischen Deprivationsindex für Deutschland [in German]. *Journal of Health Monitoring*, *2*(2), 103–120. <https://doi.org/10.17886/rki-gbe-2017-035.2>
- Luo, Q., Egger, S., Yu, X. Q., Smith, D. P., & O'Connell, D. L. (2017). Validity of using multiple imputation for "unknown" stage at diagnosis in population-based cancer registry data. *PLoS One*, *12*(6), e0180033. <https://doi.org/10.1371/journal.pone.0180033>
- Maier, W. (2017). [Indices of Multiple Deprivation for the analysis of regional health disparities in Germany : Experiences from epidemiology and healthcare research]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, *60*(12), 1403-1412. <https://doi.org/10.1007/s00103-017-2646-2>
- Merlo, J. (2014). Invited Commentary: Multilevel Analysis of Individual Heterogeneity—A Fundamental Critique of the Current Probabilistic Risk Factor Epidemiology. *Am J Epidemiol*, *180*(2), 208-212. <https://doi.org/10.1093/aje/kwu108> %J American Journal of Epidemiology
- Muller, P., Woods, L., & Walters, S. (2020). Temporal and geographic changes in stage at diagnosis in England during 2008-2013: A population-based study of colorectal, lung and ovarian cancers. *Cancer Epidemiol*, *67*, 101743. <https://doi.org/10.1016/j.canep.2020.101743>
- Oakes, J. M. (2006). Commentary: advancing neighbourhood-effects research--selection, inferential support, and structural confounding. *Int J Epidemiol*, *35*(3), 643-647. <https://doi.org/10.1093/ije/dyl054>
- Openshaw, S. (1984). *The Modifiable Areal Unit Problem. Concepts and Techniques in Modern Geography*. Geobooks.
- Quartagno, M., Grund, S., & Carpenter, J. R. (2019). jomo: A Flexible Package for Two-level Joint Modelling Multiple Imputation. *The R Journal*, *11*, 205.
- Zahnd, W. E., & McLafferty, S. L. (2017). Contextual effects and cancer outcomes in the United States: a systematic review of characteristics in multilevel analyses. *Ann Epidemiol*, *27*(11), 739-748.e733. <https://doi.org/10.1016/j.annepidem.2017.10.002>